

# Tools for the Diagnosis of ADHD in Children and Adolescents: A Systematic Review

Bradley S. Peterson, MD,<sup>a,b</sup> Joey Trampush, PhD,<sup>b</sup> Morah Brown, MPH,<sup>c</sup> Margaret Maglione, MPP,<sup>c</sup> Maria Bolshakova, PhD,<sup>c</sup> Mary Rozelle, MPH,<sup>c</sup> Jeremy Miles, PhD,<sup>c</sup> Sheila Pakdaman, PhD,<sup>c</sup> Sachi Yagyu, MS,<sup>c</sup> Aneesa Motala, BA,<sup>c</sup> Susanne Hempel, PhD<sup>c</sup>

**CONTEXT:** Correct diagnosis is essential for the appropriate clinical management of attention-deficit/hyperactivity disorder (ADHD) in children and adolescents.

**OBJECTIVE:** This systematic review provides an overview of the available diagnostic tools.

**DATA SOURCES:** We identified diagnostic accuracy studies in 12 databases published from 1980 through June 2023.

**STUDY SELECTION:** Any ADHD tool evaluation for the diagnosis of ADHD, requiring a reference standard of a clinical diagnosis by a mental health specialist.

**DATA EXTRACTION:** Data were abstracted and critically appraised by 1 reviewer and checked by a methodologist. Strength of evidence and applicability assessments followed Evidence-based Practice Center standards.

**RESULTS:** In total, 231 studies met eligibility criteria. Studies evaluated parental ratings, teacher ratings, youth self-reports, clinician tools, neuropsychological tests, biospecimen, EEG, and neuroimaging. Multiple tools showed promising diagnostic performance, but estimates varied considerably across studies, with a generally low strength of evidence. Performance depended on whether ADHD youth were being differentiated from neurotypically developing children or from clinically referred children.

**LIMITATIONS:** Studies used different components of available tools and did not report sufficient data for meta-analytic models.

**CONCLUSIONS:** A valid and reliable diagnosis of ADHD requires the judgment of a clinician who is experienced in the evaluation of youth with and without ADHD, along with the aid of standardized rating scales and input from multiple informants across multiple settings, including parents, teachers, and youth themselves.



<sup>a</sup>Institute for the Developing Mind, Children's Hospital Los Angeles, Los Angeles, California; <sup>b</sup>Department of Psychiatry, Keck School of Medicine at the University of Southern California, Los Angeles, California; and <sup>c</sup>Southern California Evidence Review Center, Los Angeles, California

Drs Peterson and Hempel conceptualized and designed the study, collected data, conducted the analyses, drafted the initial manuscript, and critically reviewed and revised the manuscript; Dr Trampush conducted the critical appraisal; Ms Brown, Ms Maglione, Drs Bolshakova and Pakdaman, and Ms Rozelle screened citations and abstracted the data; Dr Miles conducted the analyses; Ms Yagyu designed and executed the search strategy; Ms Motala served as data manager; and all authors provided critical input for the manuscript, approved the final manuscript as submitted, and agree to be accountable for all aspects of the work.

This trial has been registered at PROSPERO (identifier CRD42022312656).

**To cite:** Peterson BS, Trampush J, Brown M, et al. Tools for the Diagnosis of ADHD in Children and Adolescents: A Systematic Review. *Pediatrics*. 2024;153(4):e2024065854

Attention-deficit/hyperactivity disorder (ADHD) is one of the most prevalent neurodevelopmental conditions in youth. Its prevalence has remained constant at ~5.3% worldwide over the years, and diagnostic criteria have remained constant when based on rigorous diagnostic procedures.<sup>1</sup> Clinical diagnoses, however, have increased steadily over time,<sup>2</sup> and currently, ~10% of US children receive an ADHD diagnosis.<sup>3</sup> Higher rates of clinical compared with research-based diagnoses are because of an increasing clinician recognition of youth who have ADHD symptoms that are functionally impairing but do not fully meet formal diagnostic criteria.<sup>4</sup> The higher diagnostic rates over time in clinical samples also results from youth receiving a diagnosis incorrectly. Some youth, for example, are misdiagnosed as having ADHD when they have symptoms of other disorders that overlap with ADHD symptoms, such as difficulty concentrating, which occurs in many other conditions.<sup>5</sup> Moreover, ADHD is more than twice as likely to be diagnosed in boys than in girls,<sup>3</sup> in lower-income families,<sup>6</sup> and in white compared with nonwhite youth<sup>7</sup>; differences that derive at least in part from diagnostic and cultural biases.<sup>8–11</sup>

Improving clinical diagnostic accuracy is essential to ensure that youth who truly have ADHD benefit from receiving treatment without delay. Similarly, youth who do not have ADHD should not be diagnosed since an incorrect diagnosis risks exposing them to unbeneficial treatments.<sup>12,13</sup> Clinician judgement alone, however, especially by nonspecialist clinicians, is poor in diagnosing ADHD<sup>14</sup> compared with expert, research-grade diagnoses made by mental health clinicians.<sup>15</sup> Accurately diagnosing ADHD is difficult because diagnoses are often made using subjective clinical impressions, and putative diagnostic tools have a confusing, diverse, and poorly described evidence base that is not widely accessible. The availability of valid diagnostic tools would especially help to reduce misdiagnoses from cultural biases and symptom overlap with ADHD.<sup>12,16–19</sup>

This review summarizes evidence for the performance of tools for children and adolescents with ADHD. We did not restrict to a set of known diagnostic tools but instead explored the range of available diagnostic tools, including machine-learning assisted and virtual reality-based tools. The review aimed to assess how diagnostic performance varies by clinical setting and patient characteristics.

## METHODS

The review aims were developed in consultation with the Agency for Healthcare Research and Quality (AHRQ), the Patient-Centered Outcomes Research Institute, the topic nominator American Academy of Pediatrics, key informants, a technical expert panel (TEP), and public input. The TEP reviewed the protocol and advised on key outcomes. Subgroup analyses and key outcomes were prespecified. The review is registered in PROSPERO (CRD42022312656)

and the protocol is available on the AHRQ Web site as part of a larger evidence report on ADHD. The systematic review followed Methods of the (AHRQ) Evidence-based Practice Center Program.<sup>20</sup>

## Selection Criteria

Population: age <18 years.

Interventions: any ADHD tool for the diagnosis of ADHD.

Comparators: diagnosis by a mental health specialist, such as a psychologist, psychiatrist, or other provider, who often used published scales or semistructured diagnostic interviews to ensure a reliable DSM-based diagnosis of ADHD.

Key outcomes: diagnostic accuracy (eg, sensitivity, specificity, area under the curve).

Setting: any.

Study design: diagnostic accuracy studies.

Other: English language, published from 1980 to June 2023.

## Search Strategy

We searched PubMed, Embase, PsycINFO, ERIC, and ClinicalTrials.gov. We identified reviews for reference-mining through PubMed, Cochrane Database of Systematic Reviews, Campbell Collaboration, What Works in Education, PROSPERO, ECRI Guidelines Trust, G-I-N, and ClinicalKey. The peer reviewed strategy is in the Supplemental Appendix. All citations were screened by trained literature reviewers supported by machine learning (Fig 1). Two independent reviewers assessed full text studies for eligibility. The TEP reviewed studies to ensure all were captured. Publications reporting on the same participants were consolidated into 1 record.

## Data Extraction

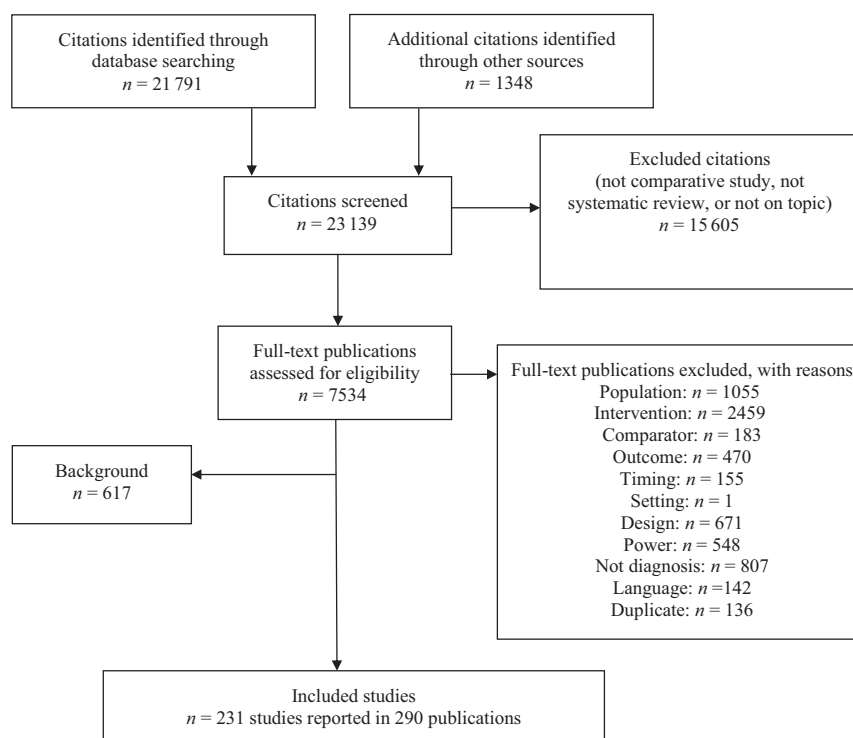
The data abstraction form included extensive guidance to aid reproducibility and standardization in recording study details, results, risk of bias, and applicability. One reviewer abstracted data and a methodologist checked accuracy and completeness. Data are publicly available in the Systematic Review Data Repository.

## Risk of Bias and Applicability

We assessed characteristics pertaining to patient selection, index test, reference standard, flow and timing that may have introduced bias, and evaluated applicability of study results, such as whether the test, its conduct, or interpretation differed from how the test is used in clinical practice.<sup>21,22</sup>

## Data Synthesis and Analysis

We differentiated parent, teacher, and youth self-report ratings; tools for clinicians; neuropsychological tests; biospecimens; EEG; and neuroimaging. We organized analyses



**FIGURE 1**  
Literature flow diagram.

according to prespecified outcome measures. A narrative overview summarized the range of diagnostic performance for key outcomes. Because lack of reported detail in many individual studies hindered use of meta-analytic models, we created summary figures to document the diagnostic performance reported in each study. We used meta-regressions across studies to assess the effects of age, comorbidities, racial and ethnic composition, and diagnostic setting (differentiating primary care, specialty care, school settings, mixed settings, and not reported) on diagnostic performance. One researcher with experience in use of specified standardized criteria<sup>23</sup> initially assessed the overall strength of evidence (SoE) (see Supplemental Appendix) for each study, then discussed it with the study team to communicate our confidence in each finding.

## RESULTS

We screened 23 139 citations and 7534 publications retrieved as full text against the eligibility criteria. In total, 231 studies reported in 290 publications met the eligibility criteria (see Fig 1).

Methodological quality of the studies varied. Selection bias was likely in two-thirds of studies; several were determined to be problematic in terms of reported study flow and timing of assessments (eg, not stating whether diagnosis was known before the results of the index test);

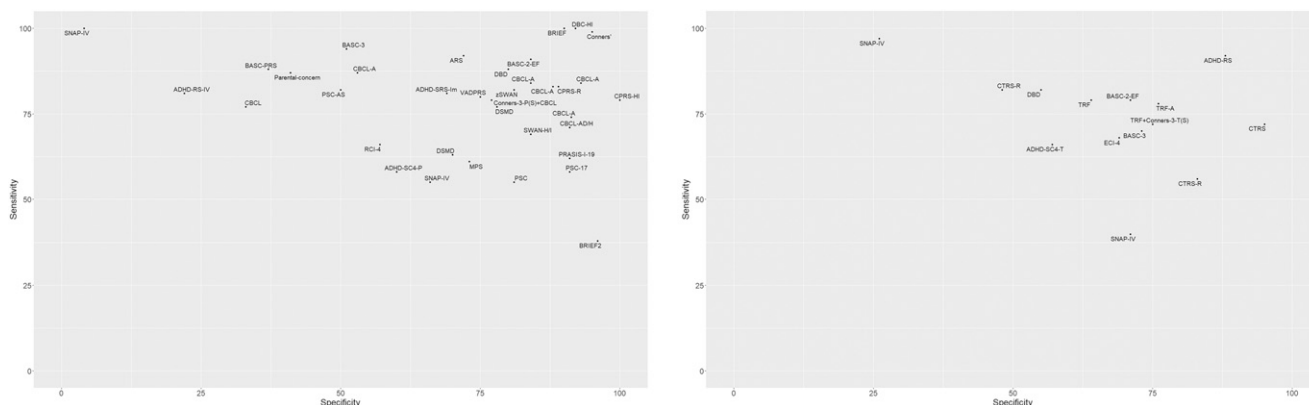
and several lacked details on diagnosticians or diagnostic procedures (Supplemental Fig 1). Applicability concerns limited the generalizability of findings (Supplemental Fig 2), usually because youth with comorbidities were excluded. Many different tools were assessed within the broader categories (eg, within neuropsychological tests), and even when reporting on the same diagnostic tool, studies often used different components of the tool (eg, different subscales of rating scales), or they combined components in a variety of ways (eg, across different neuropsychological test parameters).

The evidence table (Supplemental Table 10, Supplemental Appendix) shows each study's finding. The following highlights key findings across studies.

### Parent Ratings

Fifty-nine studies used parent ratings to diagnose ADHD (Fig 2). The most frequently evaluated tool was the CBCL (Child Behavior Checklist), alone or in combination with other tools, often using different score cutoffs for diagnosis, and evaluating different subscales (most frequently the attention deficit/hyperactivity problems subscale). Sensitivities ranged from 38% (corresponding specificity = 96%) to 100% (specificity = 4% to 92%).<sup>24,25</sup>

Area under the curve (AUC) for receiver operator characteristic curves ranged widely from 0.55 to 0.95 but 3 CBCL studies reported AUCs of 0.83 to 0.84.<sup>26–28</sup> Few



**FIGURE 2**  
Diagnostic performance parent and teacher ratings. For a complete list of scales see Supplemental Appendix.

studies reported measurement of reliability. SoE was downgraded for study limitation (lack of detailed reporting), imprecision (large performance variability), and inconsistent findings (Supplemental Table 1).

### Teacher Ratings

Twenty-three studies used teacher ratings to diagnose ADHD (Fig 2). No 2 studies reported on rater agreement, internal consistency, or test-retest reliability for the same teacher rating scale. The highest sensitivity was 97% (specificity = 26%).<sup>25</sup> The Teacher Report Form, alone or in combination with Conners teacher rating scales, yielded sensitivities of 72% to 79%<sup>29</sup> and specificities of 64% to 76%.<sup>30,32</sup> reported AUCs ranged from 0.65 to 0.84.<sup>32</sup> SoE was downgraded to low for imprecision (large performance variability) and inconsistency (results for specific tools not replicated), see Supplemental Table 2.

### Youth Self-Reports

Six studies used youth self-reports to diagnose ADHD. No 2 studies used the same instrument. Sensitivities ranged from 53% (specificity = 98%) to 86% (specificity = 70%).<sup>35</sup> AUCs ranged from 0.56 to 0.85.<sup>36</sup> We downgraded SoE for domain inconsistency (only 1 study reported on a given tool and outcome), see Supplemental Table 3.

### Combined Rating Scales

Thirteen studies assessed diagnostic performance of ratings combined across informants, often using machine learning for variable selection. Only 1 study compared performance of combined data to performance from single informants, finding negligible improvement (AUC youth = 0.71; parent = 0.85; combined = 0.86).<sup>37</sup> Other studies reported on limited outcome measures and used ad hoc methods to combine information from multiple informants. The best AUC was reported by a machine learning supported study combining parent and teacher ratings (AUC = 0.98).<sup>38</sup>

### Additional Clinician Tools

Twenty-four studies assessed additional tools, such as interview guides, that can be used by clinicians to aid diagnosis of ADHD. Sensitivities varied, ranging from 67% (specificity = 65%) to 98% (specificity = 100%); specificities ranged from 36% (sensitivity = 89%) to 100% (sensitivity = 98%).<sup>39</sup> Some of the tools measured activity levels objectively using an actometer or commercially available activity tracker, either alone or as part of a diagnostic test battery. Reported performance was variable (sensitivity range 25% to 100%,<sup>40</sup> specificity range 66% to 100%,<sup>40</sup> AUCs range 0.75–0.9996<sup>41</sup>). SoE was downgraded for imprecision (large performance variability) and inconsistency (outcomes and results not replicated), see Supplemental Table 4.

### Neuropsychological Tests

Seventy-four studies used measures from various neuropsychological tests, including continuous performance tests (CPTs). Four of these included 3- and 4-year-old children.<sup>42–44</sup> A large majority used a CPT, which assessed omission errors (reflecting inattention), commission errors (impulsivity), and reaction time SD (response time variability). Studies varied in use of traditional visual CPTs, such as the Test of Variables of Attention, more novel, multifaceted “hybrid” CPT paradigms, and virtual reality CPTs built upon environments designed to emulate real-world classroom distractibility. Studies used idiosyncratic combinations of individual cognitive measures to achieve the best performance, though many reported on CPT attention and impulsivity measures.

Sensitivity for all neuropsychological tests ranged from 22% (specificity = 96%) to 100% (specificity = 100%)<sup>45</sup> (Fig 3), though the latter study reported performance for unique composite measures without replication. Specificities ranged from 22% (sensitivity = 91%)<sup>46</sup> to 100% (sensitivity = 100% to 75%).<sup>45,47</sup> AUCs ranged from

0.59 to 0.93.<sup>48</sup> Sensitivity for all CPT studies ranged from 22% (specificity = 96) to 100% (specificity = 75%).<sup>49</sup> Specificities for CPTs ranged from 22% (sensitivity = 91%) to 100% (sensitivity = 89%)<sup>47</sup> (Fig 3). AUCs ranged from 0.59 to 0.93.<sup>50,51</sup> SoE was deemed low for imprecise studies (large performance variability), see Supplemental Table 5.

## Biospecimen

Seven studies assessed blood or urine biomarkers to diagnose ADHD. These measured erythropoietin or erythropoietin receptor, membrane potential ratio, micro RNA levels, or urine metabolites. Sensitivities ranged from 56% (specificity = 95%) to 100% (specificity = 100% for erythropoietin and erythropoietin receptors levels).<sup>52</sup> Specificities ranged from 25% (sensitivity = 79%) to 100% (sensitivity = 100%).<sup>52</sup> AUCs ranged from 0.68 to 1.00.<sup>52</sup> Little information was provided on reliability of markers or their combinations. SoE was downgraded for inconsistent and imprecise studies (Supplemental Table 6).

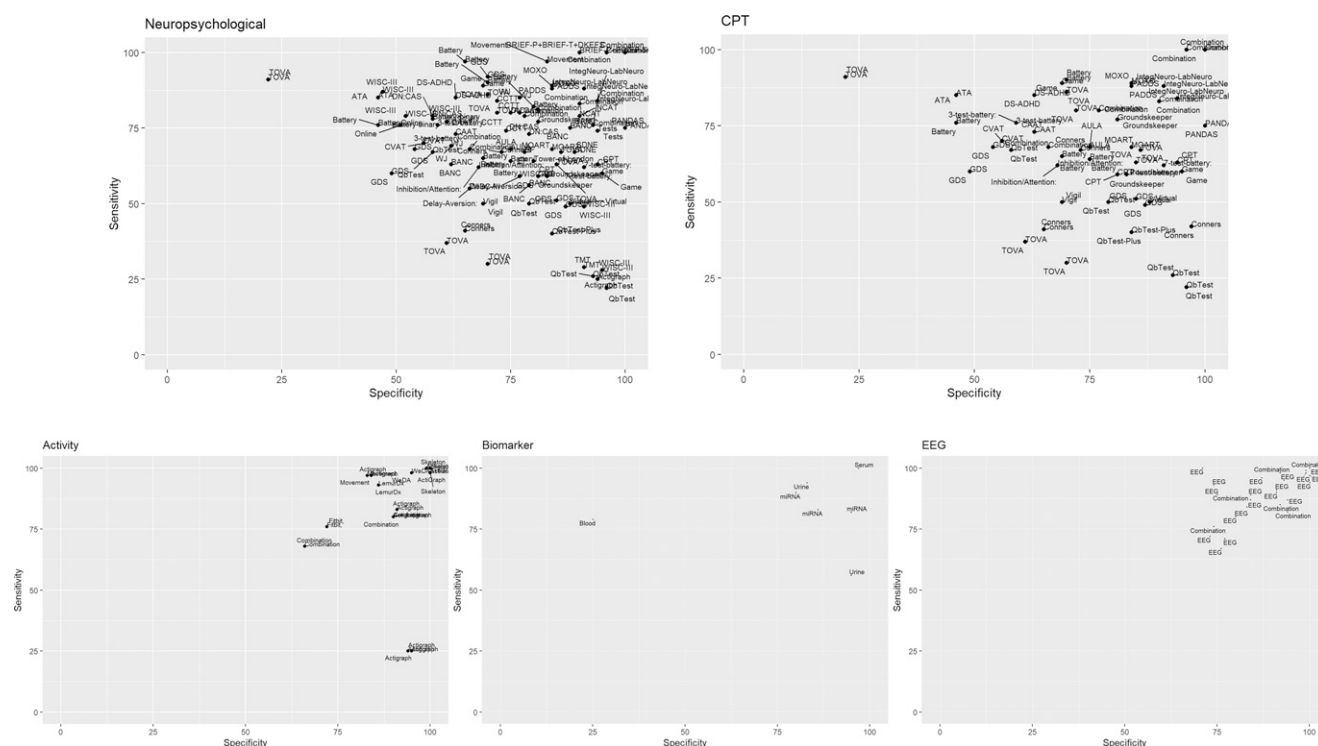
## EEG

Forty-five studies used EEG markers to diagnose ADHD. EEG signals were obtained in a variety of patient states, even during neuropsychological test performance. Two-thirds used machine learning algorithms to select classification parameters. Several combined EEG with demographic variables or rating scales. Sensitivity ranged widely

from 46% to 100% (corresponding specificities 74 and 71%).<sup>53,54</sup> One study that combined EEG with demographics data supported by machine learning reported perfect sensitivity and specificity.<sup>54</sup> Specificity was also variable and ranged from 38% (sensitivity = 95%) to 100% (specificities = 71% or 100%).<sup>53–56</sup> Reported AUCs ranged from 0.63 to 1.0.<sup>57,58</sup> SoE was downgraded for study imprecision (large performance variability) and limitations (diagnostic approaches poorly described), see Supplemental Table 7.

## Neuroimaging

Nineteen studies used neuroimaging for diagnosis. One public data set (ADHD-200) produced several analyses. All but 2 used MRI: some functional MRI (fMRI), some structural, and some in combination, with or without magnetic resonance spectroscopy (2 used near-infrared spectroscopy). Most employed machine learning to detect markers that optimized diagnostic classifications. Some combined imaging measures with demographic or other clinical data in the prediction model. Sensitivities ranged from 42% (specificity = 95%) to 99% (specificity = 100%) using resting state fMRI and a complex machine learning algorithm<sup>56</sup> to differentiate ADHD from neurotypical youth. Specificities ranged from 55% (sensitivity = 95%) to 100%<sup>56</sup> using resting state fMRI data. AUCs ranged from 0.58 to over 0.99,<sup>57</sup> SoE was downgraded for imprecision (large performance variability) and study



**FIGURE 3**

Diagnostic performance neuropsychological tests, CPTs, activity monitors, biospecimen, EEG.



limitations (diagnostic models are often not well described, and the number and type of predictor variables entering the model were unclear). Studies generally did not validate diagnostic algorithms or assess performance measures in an independent sample (Supplemental Table 8).

### Variation in Diagnostic Accuracy With Clinical Setting or Patient Subgroup

Regression analyses indicated that setting was associated with both sensitivity ( $P = .03$ ) and accuracy ( $P = .006$ ) but not specificity ( $P = .68$ ) or AUC ( $P = .28$ ), with sensitivities lowest in primary care (Fig 4). Sensitivity, specificity, and accuracy were also lower when differentiating youth with ADHD from a clinical sample than from typically developing youth (sensitivity  $P = .04$ , specificity  $P < .001$ , AUC  $P < .001$ ) (Fig 4), suggesting that clinical population is a source of heterogeneity in diagnostic performance. Findings should be interpreted with caution, however, as they were not obtained in meta-analytic models and, consequently, do not take into account study size or quality.

Supplemental Figs 3–5 in the Supplemental Appendix document effects by age and gender. We did not detect statistically significant associations of age with sensitivity ( $P = .54$ ) or specificity ( $P = .37$ ), or associations of the

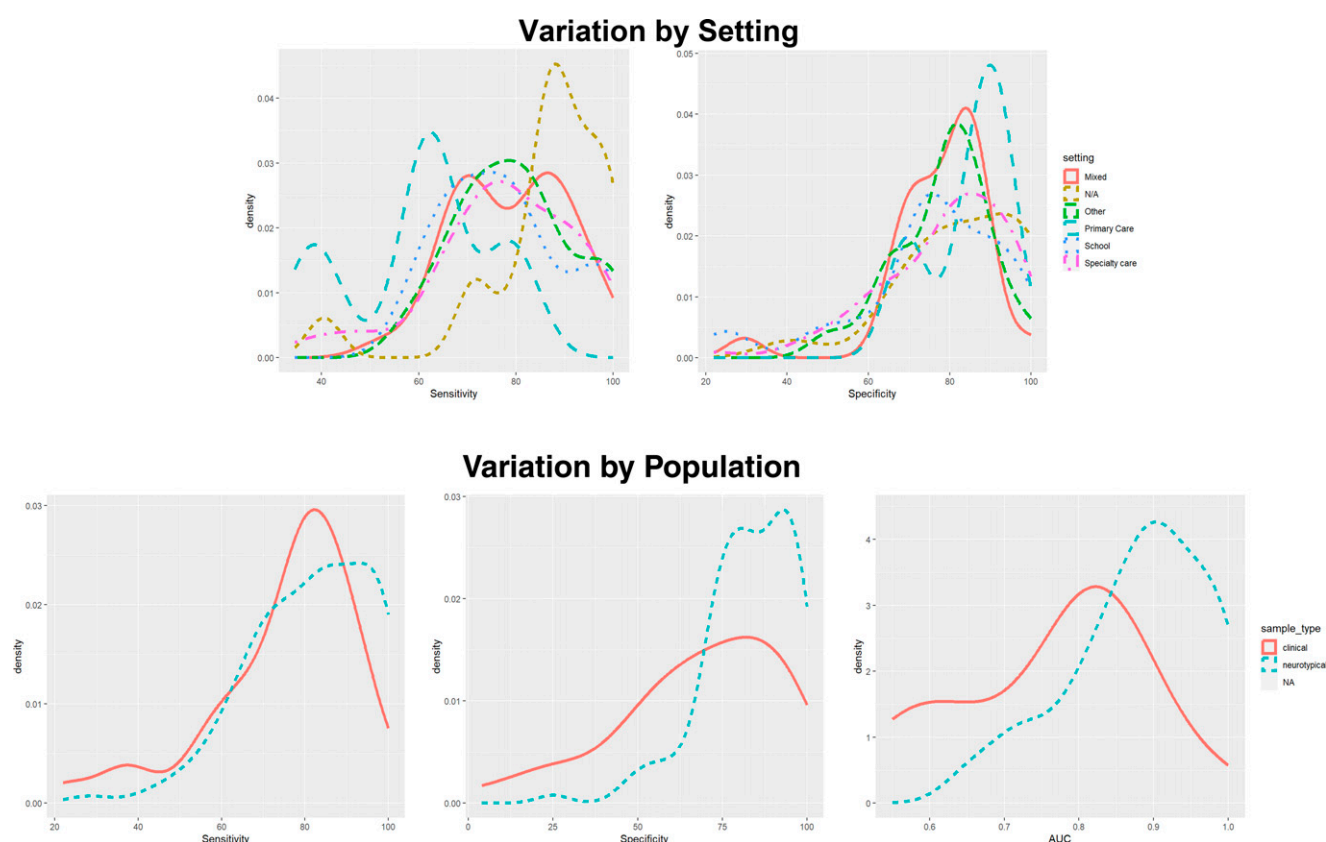
proportion of girls with sensitivity ( $P = .63$ ), specificity ( $P = .80$ ), accuracy ( $P = .34$ ), or AUC ( $P = .90$ ).

## DISCUSSION

We identified a large number of publications reporting on ADHD diagnostic tools. To our knowledge, no prior review of ADHD diagnostic tools has been as comprehensive in the range of tools, outcomes, participant ages, and publication years. Despite the large number of studies, we deemed the strength of evidence for the reported performance measures across all categories of diagnostic tools to be low because of large performance variability across studies and various limitations within and across studies.

### Measures for Diagnostic Performance

We required that studies report diagnoses when using the tool compared with diagnoses made by expert mental health clinicians. Studies most commonly reported sensitivity (true-positive rate) and specificity (true-negative rate) when a study-specific diagnostic threshold was applied to measures from the tool being assessed. Sensitivity and specificity depend critically on that study-specific threshold, and their values are inherently a trade-off,



**FIGURE 4**  
Diagnostic performance by setting and population.

such that varying the threshold to increase either sensitivity or specificity reduces the other. Interpreting diagnostic performance in terms of sensitivity and specificity, and comparing those performance measures across studies, is therefore challenging. Consequently, researchers more recently often report performance for sensitivity and specificity in terms of receiver operating characteristics (ROC) curves, a plot of sensitivity versus specificity across the entire range of possible diagnostic thresholds. The area under this ROC curve (AUC) provides an overall, single index of performance that ranges from 0.5 (indicating that the tool provides no information above chance for classification) to 1.0 (indicating a perfect test that can correctly classify all participants as having ADHD and all non-ADHD participants as not having it). AUC values of 90 to 100 are commonly classified as excellent performance; 80 to 90 as good; 70 to 80 as fair; 60 to 70 as poor; and 50 to 60 failed performance.

### Available Tools

Most research is available on parental ratings. Overall, AUCs for parent rating scales ranged widely from “poor”<sup>58</sup> to “excellent.”<sup>59</sup> Analyses restricted to the CBCL, the most commonly evaluated scale, yielded more consistent “good” AUCs for differentiating youth with ADHD from others in clinical samples, but the number of studies contributing data were small. Internal consistency for rating scale items was generally high across most rating scales. Test-retest reliability was good, though only 2 studies reported it. One study reported moderate rater agreement between mothers and fathers for inattention, hyperactivity, and impulsivity symptoms. Few studies included youth under 7 years of age.

AUCs for teacher rating scales ranged from “failed”<sup>33</sup> to “good.”<sup>34</sup> Internal consistency for scale items was generally high. Teacher ratings demonstrated very low rater agreement with corresponding parent scales, suggesting either a problem with the instruments or a large variability in symptom presentation with environmental context (home or school).

Though data were limited, self-reports from youth seemed to perform less well than corresponding parent and teacher reports, with AUCs ranging from “failed” for CBCL or ASEBA when distinguishing ADHD from other patients<sup>33</sup> to “good” for the SWAN in distinguishing ADHD from neurotypical controls.<sup>36,37</sup>

Studies evaluating neuropsychological tests yielded AUCs ranging from “poor”<sup>60,61</sup> to “excellent.”<sup>50</sup> Many used idiosyncratic combinations of cognitive measures, which complicates interpretation of the results across studies. Nevertheless, extracting specific, comparable measures of inattention and impulsivity from CPTs yielded diagnostic performance ranging from “poor” to “excellent” in differentiating ADHD youth from neurotypical controls and “fair” in

differentiating ADHD youth from other patients.<sup>42,60,62</sup> No studies provided an independent replication of diagnosis using the same measure.

Blood biomarkers yielded AUCs ranging from “poor” (serum miRNAs)<sup>63</sup> to “excellent” (erythropoietin and erythropoietin receptors levels)<sup>52</sup> in differentiating ADHD from neurotypical youth. None have been independently replicated, and test-retest reliability was not reported. Most EEG studies used machine learning for diagnostic classification. AUCs ranged from “poor”<sup>64</sup> to “excellent” when differentiating ADHD youth from neurotypical controls.<sup>65</sup> Diagnostic performance was not prospectively replicated in any independent samples.

Most neuroimaging studies relied on machine learning to develop diagnostic algorithms. AUCs ranged from “poor”<sup>66</sup> to “excellent” for distinguishing ADHD youth from neurotypically developing controls.<sup>57</sup> Most studies used pre-existing data sets or repositories to retrospectively discriminate youths with ADHD from neurotypical controls, not from other clinical populations and not prospectively, and none assessed test-retest reliability or the independent reproducibility of findings. Reporting of final mathematical models or algorithms for diagnosis was limited. Activity monitors have the advantage of providing inexpensive, objective, easily obtained, and quantified measures that can potentially be widely disseminated and scaled.

Studies of combined approaches, such as integrating diagnostic tools with clinician impressions, were limited. One study reported increased sensitivity and specificity when an initial clinician diagnosis combined EEG indicators (the reference standard was a consensus diagnosis from a panel of ADHD experts).<sup>67</sup> These findings were not independently replicated, however, and no test-retest reliability was reported.

### Importance of the Comparator Sample

Many studies aimed to distinguish ADHD youth from neurotypical controls, which is a distinction of limited clinical relevance. In clinically referred youth, most parents, teachers, and clinicians are reasonably confident that something is wrong, even if they are unsure whether the cause of their concern is ADHD. To be informed by a tool that the child is not typically developing is not particularly helpful. Moreover, we cannot know whether diagnostic performance for tools that discriminate ADHD youth only from neurotypical controls is determined by the presence of ADHD or by the presence of any other characteristics that accompany clinical “caseness,” such as the presence of comorbid illnesses or symptoms shared or easily confused with those of other conditions, or the effects of chronic stress or current or past treatment. The clinically more relevant and difficult question is, therefore, how well the tool distinguishes youth with ADHD from those who have other emotional and behavioral

problems. Consistent with these conceptual considerations that argue for assessing diagnostic performance in differentiating youth with ADHD from those with other clinical conditions, we found significant evidence that, across all studies, sensitivity, specificity, and AUC were all lower when differentiating youth with ADHD from a clinical sample than when differentiating them from neurotypical youth. These findings also suggest that the comparison population was a significant source of heterogeneity in diagnostic performance.

### Clinical Implications

Despite the large number of studies on diagnostic tools, a valid and reliable diagnosis of ADHD ultimately still requires the judgement of a clinician who is experienced in the evaluation of youth with and without ADHD, along with the aid of standardized rating scales and input from multiple informants across multiple settings, including parents, teachers, and youth themselves. Diagnostic tools perform best when the clinical question is whether a youth has ADHD or is healthy and typically developing, rather than when the clinical question is whether a youth has ADHD or another mental health or behavioral problem. Diagnostic tools yield more false-positive and false-negative diagnoses of ADHD when differentiating youth with ADHD from youth with another mental health problem than when differentiating them from neurotypically developing youth.

Scores for rating scales tended to correlate poorly across raters, and ADHD symptoms in the same child varied across settings, indicating that no single informant in a single setting is a gold-standard for diagnosis. Therefore, diagnosis using rating scales will likely benefit from a more complete representation of symptom expression across multiple informants (parents, school personnel, clinicians, and youth) across more than 1 setting (home, school, and clinic) to inform clinical judgement when making a diagnosis, thus, consistent with current guidelines.<sup>68–70</sup> Unfortunately, methods for combining scores across raters and settings that improve diagnosis compared with scores from single raters have not been developed or prospectively replicated.

Despite the widespread use of neuropsychological testing to “diagnose” youth with ADHD, often at considerable expense, indirect comparisons of AUCs suggest that performance of neuropsychological test measures in diagnosing ADHD is comparable to the diagnostic performance of ADHD rating scales from a single informant. Moreover, the diagnostic accuracy of parent rating scales is typically better than neuropsychological test measures in head-to-head comparisons.<sup>44,71</sup> Furthermore, the overall SoE for estimates of diagnostic performance with neuropsychological testing is low. Use of neuropsychological test measures of executive functioning, such as the CPT, may help inform a clinical diagnosis, but they are not definitive either in ruling in or ruling out a diagnosis of ADHD. The sole use of CPTs and other neuropsychological

tests to diagnose ADHD, therefore, cannot be recommended. We note that this conclusion regarding diagnostic value is not relevant to any other clinical utility that testing may have.

No independent replication studies have been conducted to validate EEG, neuroimaging, or biospecimen to diagnose ADHD, and no clinical effectiveness studies have been conducted using these tools to diagnose ADHD in the real world. Thus, these tools do not seem remotely close to being ready for clinical application to aid diagnosis, despite US Food and Drug Administration approval of 1 EEG measure as a purported diagnostic aid.<sup>67,72</sup>

### Future Research

All studies of diagnostic tools should report data in more detail (ie, clearly report false-positive and -negative rates, the diagnostic thresholds used, and any data manipulation undertaken to achieve the result) to support meta-analytic methods. Studies should include ROC analyses to support comparisons of test performance across studies that are independent of the diagnostic threshold applied to measures from the tool. They should also include assessment of test-retest reliability to help discern whether variability in measures and test performance is a function of setting or of measurement variability over time. Future studies should address the influence of co-occurring disorders on diagnostic performance and how well the tools distinguish youth with ADHD from youth with other emotional and behavioral problems, not simply from healthy controls. More studies should compare the diagnostic accuracy of different test modalities, head-to-head. Independent, prospective replication of performance measures of diagnostic tools in real-world settings is essential before US Food and Drug Administration approval and before recommendations for widespread clinical use.

Research is needed to identify consensus algorithms that combine rating scale data from multiple informants to improve the clinical diagnosis of ADHD, which at present is often unguided, ad hoc, and suboptimal. Diagnostic studies using EEG, neuroimaging, and neuropsychological tests should report precise operational definitions and measurements of the variable(s) used for diagnosis, any diagnostic algorithm employed, the selected statistical cut-offs, and the number of false-positives and false-negatives the diagnostic tool yields to support future efforts at synthetic analyses.

### CONCLUSIONS

Objective, quantitative neuropsychological test measures of executive functioning correlate only weakly with the clinical symptoms that define ADHD.<sup>73</sup> Thus, many youth with ADHD have normal executive functioning profiles on neuropsychological testing, and many who have impaired executive functioning on testing do not have ADHD.<sup>74</sup> Future research is needed to understand how test measures of executive functioning and the real-world



functional problems that define ADHD map on to one another and how that mapping can be improved.

One of the most important potential uses of systematic reviews and meta-analyses in improving the clinical diagnosis of ADHD and treatment planning would be identification of effect modifiers for the performance of diagnostic tools: determining, for example, whether tools perform better in patients who are younger or older, in ethnic minorities, or those experiencing material hardship, or who have a comorbid illness or specific ADHD presentation. Future studies of ADHD should more systematically address the modifier effects of these patient characteristics. They should make available in public repositories the raw, individual-level data and the algorithms or computer code that will aid future efforts at replication, synthesis, and new discovery for diagnostic tools across data sets and studies.

Finally, no studies meeting our inclusion criteria assessed the consequences of being misdiagnosed or labeled as either having or not having ADHD, the diagnosis of ADHD specifically in preschool-aged children, or the potential adverse consequences of youth being incorrectly diagnosed with or without ADHD. This work is urgently needed.

### ACKNOWLEDGMENTS

We thank Cynthia Ramirez, Erin Tokutomi, Jennifer Rivera, Coleman Schaefer, Jerusalem Belay, Anne Onyekwuluje, and Mario Gastelum for help with data acquisition. We thank Kymika Okechukwu, Lauren Pilcher, Joanna King, and Robyn

Wheatley from the American Academy of Pediatrics (AAP), Jennie Dalton and Paula Eguino Medina from PCORI, Christine Chang and Kim Wittenberg from AHRQ, and Mary Butler from the Minnesota Evidence-based Practice Center. We thank Glendy Burnett, Eugenia Chan, MD, MPH, Matthew J. Gormley, PhD, Laurence Greenhill, MD, Joseph Hagan, Jr, MD, Cecil Reynolds, PhD, Le'Ann Solmonson, PhD, LPC-S, CSC, and Peter Ziemkowski, MD, FAAFP who served as key informants. We thank Angelika Claussen, PhD, Alysa Doyle, PhD, Tiffany Farchione, MD, Matthew J. Gormley, PhD, Laurence Greenhill, MD, Jeffrey M. Halperin, PhD, Marisa Perez-Martin, MS, LMFT, Russell Schachar, MD, Le'Ann Solmonson, PhD, LPC-S, CSC, and James Swanson, PhD who served as a technical expert panel. Finally, we thank Joel Nigg, PhD, and Peter S. Jensen, MD for their peer review of the data.

### ABBREVIATIONS

ADHD: attention-deficit/hyperactivity disorder  
AUC: area under the curve  
CBCL: Child Behavior Checklist  
CPT: continuous performance test  
fMRI: functional magnetic resonance imaging  
ROC: receiver operating characteristics  
SoE: strength of evidence  
TEP: technical expert panel

**DOI:** <https://doi.org/10.1542/peds.2024-065854>

Accepted for publication Jan 26, 2024

Address correspondence to Susanne Hempel, PhD, Southern California Evidence Review Center, University of Southern California, 1845 N. Soto St, Los Angeles, CA 90032. E-mail: [susanne.hempel@med.usc.edu](mailto:susanne.hempel@med.usc.edu)

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2024 by the American Academy of Pediatrics

**FUNDING:** The work is based on research conducted by the Southern California Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract 75Q80120D00009). The Patient-Centered Outcomes Research Institute (PCORI) funded the research (PCORI Publication No. 2023-SR-03). The findings and conclusions in this manuscript are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ or PCORI, its Board of Governors, or Methodology Committee. Therefore, no statement in this report should be construed as an official position of PCORI, AHRQ or of the US Department of Health and Human Services.

**CONFLICT OF INTEREST DISCLOSURES:** The authors have indicated they have no conflicts of interest to disclose.

**COMPANION PAPER:** A companion to this article can be found online at [www.pediatrics.org/cgi/doi/10.1542/peds.2024-065787](http://www.pediatrics.org/cgi/doi/10.1542/peds.2024-065787).

**Data sharing statement:** Data are available in SRDRPlus.

### REFERENCES

1. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am J Psychiatry*. 2007;164(6):942–948
2. Center for Disease Control and Prevention. Data and statistics about ADHD. Available at: <https://www.cdc.gov/ncbddd/adhd/data.html>. Accessed August 23, 2021
3. Danielson ML, Bitsko RH, Ghandour RM, Holbrook JR, Kogan MD, Blumberg SJ. Prevalence of parent-reported ADHD diagnosis and associated treatment among U.S. children and adolescents, 2016. *J Clin Child Adolesc Psychol*. 2018;47(2):199–212
4. Hong SB, Dwyer D, Kim JW, et al. Subthreshold attention-deficit/hyperactivity disorder is associated with functional impairments

- across domains: a comprehensive analysis in a large-scale community study. *Eur Child Adolesc Psychiatry*. 2014;23(8):627–636
5. MedicalNewsToday. What to know about ADHD misdiagnosis. Available at: <https://www.medicalnewstoday.com/articles/325595>. Accessed May 5, 2023
  6. Charach A, Dashti B, Carson P, et al. *Attention Deficit Hyperactivity Disorder: Effectiveness of Treatment in At-Risk Preschoolers; Long-Term Effectiveness in All Ages; and Variability in Prevalence, Diagnosis, and Treatment*. Agency for Healthcare Research and Quality; 2011
  7. Shi Y, Hunter Guevara LR, Dykhoff HJ, et al. Racial disparities in diagnosis of attention-deficit/hyperactivity disorder in a US national birth cohort. *JAMA Netw Open*. 2021;4(3):e210321
  8. Hinshaw SP, Nguyen PT, O'Grady SM, Rosenthal EA. Annual research review: attention-deficit/hyperactivity disorder in girls and women: underrepresentation, longitudinal processes, and key directions. *J Child Psychol Psychiatry*. 2022;63(4):484–496
  9. Greven C, Richards JS, Buitelaar JK. Sex differences in ADHD. In: Banaschewski T, Coghill D, Zuddas A, eds. *Oxford Textbook of Attention Deficit Hyperactivity Disorder*. Oxford University Press; 2018:154–160
  10. Morgan PL, Hillemeier MM, Farkas G, Maczuga S. Racial/ethnic disparities in ADHD diagnosis by kindergarten entry. *J Child Psychol Psychiatry*. 2014;55(8):905–913
  11. Fadus MC, Ginsburg KR, Sobowale K, et al. Unconscious bias and the diagnosis of disruptive behavior disorders and ADHD in African American and Hispanic youth. *Acad Psychiatry*. 2020;44(1):95–102
  12. Ford-Jones PC. Misdiagnosis of attention deficit hyperactivity disorder: 'normal behaviour' and relative maturity. *Paediatr Child Health*. 2015;20(4):200–202
  13. Scituito MJ, Eisenberg M. Evaluating the evidence for and against the overdiagnosis of ADHD. *J Atten Disord*. 2007;11(2):106–113
  14. Chan E, Hopkins MR, Perrin JM, Herreras C, Homer CJ. Diagnostic practices for attention deficit hyperactivity disorder: a national survey of primary care physicians. *Ambul Pediatr*. 2005;5(4):201–208
  15. 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. *J Consult Clin Psychol*. 2014;82(6):1151–1162
  16. DosReis S, Barksdale CL, Sherman A, Maloney K, Charach A. Stigmatizing experiences of parents of children with a new diagnosis of ADHD. *Psychiatr Serv*. 2010;61(8):811–816
  17. Cook J, Knight E, Hume I, Qureshi A. The self-esteem of adults diagnosed with attention-deficit/hyperactivity disorder (ADHD): a systematic review of the literature. *Atten Defic Hyperact Disord*. 2014;6(4):249–268
  18. Lebowitz MS. Stigmatization of ADHD: a developmental review. *J Atten Disord*. 2016;20(3):199–205
  19. Wiener J, Malone M, Varma A, et al. Children's perceptions of their ADHD symptoms: positive illusions, attributions, and stigma. *Can J Sch Psychol*. 2012;27(3):217–242
  20. Agency for Healthcare Research and Quality. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Agency for Healthcare Research and Quality; 2008
  21. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898
  22. University of Bristol. QUADAS-2. Available at: <https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/>. Accessed February 10, 2021
  23. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. *Content Last Reviewed March 2021. Effective Health Care Program*. Agency for Healthcare Research and Quality; 2021
  24. Gargaro BA, May T, Tonge BJ, Sheppard DM, Bradshaw JL, Rinehart NJ. Using the DBC-P Hyperactivity Index to screen for ADHD in young people with autism and ADHD: a pilot study. *Res Autism Spectr Disord*. 2014;8(9):1008–1015
  25. Hall CL, Guo B, Valentine AZ, et al. The validity of the SNAP-IV in children displaying ADHD symptoms. *Assessment*. 2020;27(6):1258–1271
  26. Elkins RM, Carpenter AL, Pincus DB, Comer JS. Inattention symptoms and the diagnosis of comorbid attention-deficit/hyperactivity disorder among youth with generalized anxiety disorder. *J Anxiety Disord*. 2014;28(8):754–760
  27. Hong N, Comer JS. High-end specificity of the attention-deficit/hyperactivity problems scale of the child behavior checklist for ages 1.5-5 in a sample of young children with disruptive behavior disorders. *Child Psychiatry Hum Dev*. 2019;50(2):222–229
  28. Rishel CW, Greeno C, Marcus SC, Shear MK, Anderson C. Use of the child behavior checklist as a diagnostic screening tool in community mental health. *Res Soc Work Pract*. 2005;15(3):195–203
  29. Tripp G, Schaugency 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
  30. Edwards MC, Sigel BA. Estimates of the utility of Child Behavior Checklist/Teacher Report Form Attention Problems scale in the diagnosis of ADHD in children referred to a specialty clinic. *J Psychopathol Behav Assess*. 2015;37(1):50–59
  31. Gomez R, Vance A, Watson S, Stavropoulos V. ROC analyses of relevant Conners 3-Short Forms, CBCL, and TRF scales for screening ADHD and ODD. *Assessment*. 2021;28(1):73–85
  32. Power TJ, Doherty BJ, Panichelli-Mindel SM, et al. The predictive validity of parent and teacher reports of ADHD symptoms. *J Psychopathol Behav Assess*. 1998;20(1):57–81
  33. 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. *J Am Acad Child Adolesc Psychiatry*. 2017;56(5):401–409
  34. Karr JE, Kibby MY, Jagger-Rickels AC, Garcia-Barrera MA. Sensitivity and specificity of an executive function screener at identifying children with ADHD and reading disability. *J Atten Disord*. 2021;25(1):134–140
  35. Bergeron L, Smolla N, Berthiaume C, et al. Reliability, validity, and clinical utility of the Dominic Interactive for Adolescents-Revised (A DSM-5-based self-report screen for mental disorders, borderline personality traits, and suicidality). *Can J Psychiatry*. 2017;62(3):211–222
  36. Burton CL, Wright L, Shan J, et al. SWAN scale for ADHD trait-based genetic research: a validity and polygenic risk study. *J Child Psychol Psychiatry*. 2019;60(9):988–997
  37. Gibbons RD, Kupfer DJ, Frank E, et al. Computerized adaptive tests for rapid and accurate assessment of psychopathology

- dimensions in youth. *J Am Acad Child Adolesc Psychiatry*. 2020;59(11):1264–1273
38. Longridge R, Norman S, Henley W, Newlove Delgado T, Ford T. Investigating the agreement between the clinician and research diagnosis of attention deficit hyperactivity disorder and how it changes over time; a clinical cohort study. *Child and adolescent mental health*. 2019;24(2):133–141
39. Amado-Caballero P, Casaseca-de-la-Higuera P, Alberola-Lopez S, et al. Objective ADHD diagnosis using convolutional neural networks over daily-life activity records. *IEEE J Biomed Health Inform*. 2020;24(9):2690–2700
40. Lee W, Lee D, Lee S, Jun K, Kim MS. Deep-learning-based ADHD classification using children's skeleton data acquired through the ADHD screening game. *Sensors (Basel)*. 2022;23(1):246
41. Kam HJ, Shin YM, Cho SM, Kim SY, Kim KW, Park RW. Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity. *Appl Clin Inform*. 2010;1(4):377–393
42. Breaux RP, Griffith SF, Harvey EA. Preschool neuropsychological measures as predictors of later attention deficit hyperactivity disorder. *J Abnorm Child Psychol*. 2016;44(8):1455–1471
43. Hall CL, Selby K, Guo B, Valentine AZ, Walker GM, Hollis C. Innovations in practice: an objective measure of attention, impulsivity and activity reduces time to confirm attention deficit/hyperactivity disorder diagnosis in children - a completed audit cycle. *Child Adolesc Ment Health*. 2016;21(3):175–178
44. Öztekin I, Finlayson MA, Graziano PA, Dick AS. Is there any incremental benefit to conducting neuroimaging and neurocognitive assessments in the diagnosis of ADHD in young children? A machine learning investigation. *Dev Cogn Neurosci*. 2021;49:100966
45. Bledsoe JC, Xiao C, Chaovalitwongse A, et al. Diagnostic classification of ADHD versus control: support vector machine classification using brief neuropsychological assessment. *J Atten Disord*. 2020;24(11):1547–1556
46. Zelnik N, Bennett-Back O, Miari W, Goetz HR, Fattal-Valevski A. Is the test of variables of attention reliable for the diagnosis of attention-deficit hyperactivity disorder (ADHD)? *J Child Neurol*. 2012;27(6):703–707
47. Mwamba HM, Fourie PR, den Heever DV. PANDAS: paediatric attention-deficit/hyperactivity disorder application software. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:1444–1447
48. Li F, Zheng Y, Smith SD, et al. A preliminary study of movement intensity during a Go/No-Go task and its association with ADHD outcomes and symptom severity. *Child Adolesc Psychiatry Ment Health*. 2016;10(1):47
49. Zulueta A, Díaz-Orueta U, Crespo-Eguilaz N, Torrano F. Virtual reality-based assessment and rating scales in ADHD diagnosis. *Psicol Educ*. 2019;25(1):13–22
50. Berger I, Slobodin O, Cassuto H. Usefulness and validity of continuous performance tests in the diagnosis of attention-deficit hyperactivity disorder children. *Arch Clin Neuropsychol*. 2017;32(1):81–93
51. Johansson V, Norén Selinus E, Kuja-Halkola R, et al. The quantified behavioral test failed to differentiate ADHD in adolescents with neurodevelopmental problems. *J Atten Disord*. 2021;25(3):312–321
52. Gungor M, Kurutas EB, Oner E, et al. Diagnostic performance of erythropoietin and erythropoietin receptors levels in children with attention deficit hyperactivity disorder. *Clin Psychopharmacol Neurosci*. 2021;19(3):530–536
53. Markovska-Simoska S, Pop-Jordanova N. Quantitative EEG in children and adults with attention deficit hyperactivity disorder: comparison of absolute and relative power spectra and theta/beta ratio. *Clin EEG Neurosci*. 2017;48(1):20–32
54. Beriha SS. Computer aided diagnosis system to distinguish ADHD from similar behavioral disorders. *Biomed Pharmacol J*. 2018;11(2):1135–1141
55. Quintana H, Snyder SM, Purnell W, Aponte C, Sita J. Comparison of a standard psychiatric evaluation to rating scales and EEG in the differential diagnosis of attention-deficit/hyperactivity disorder. *Psychiatry Res*. 2007;152(2-3):211–222
56. Chen Y, Tang Y, Wang C, Liu X, Zhao L, Wang Z. ADHD classification by dual subspace learning using resting-state functional connectivity. *Artif Intell Med*. 2020;103:101786
57. Ekhlasi A, Nasrabadi AM, Mohammadi M. Analysis of EEG brain connectivity of children with ADHD using graph theory and directional information transfer. *Biomed Tech (Berl)*. 2022;68(2):133–146
58. Jahanshahloo HR, Shamsi M, Ghasemi E, Kouhi A. Automated and ERP-based diagnosis of attention-deficit hyperactivity disorder in children. *J Med Signals Sens*. 2017;7(1):26–32
59. Tang Y, Sun J, Wang C, et al. ADHD classification using auto-encoding neural network and binary hypothesis testing. *Artif Intell Med*. 2022;123:102209
60. Jacobson LA, Pritchard AE, Koriakin TA, Jones KE, Mahone EM. Initial examination of the BRIEF2 in clinically referred children with and without ADHD symptoms. *J Atten Disord*. 2020;24(12):1775–1784
61. Yeh SC, Lin SY, Wu EH, et al. A virtual-reality system integrated with neuro-behavior sensing for attention-deficit/hyperactivity disorder intelligent assessment. *IEEE Trans Neural Syst Rehabil Eng*. 2020;28(9):1899–1907
62. Hult N, Kadesjö J, Kadesjö B, Gillberg C, Billstedt E. ADHD and the QbTest: diagnostic validity of QbTest. *J Atten Disord*. 2018;22(11):1074–1080
63. Faraone SV, Newcorn JH, Antshel KM, Adler L, Roots K, Heller M. The groundskeeper gaming platform as a diagnostic tool for attention-deficit/hyperactivity disorder: sensitivity, specificity, and relation to other measures. *J Child Adolesc Psychopharmacol*. 2016;26(8):672–685
64. Williams LM, Hermens DF, Thein T, et al. Using brain-based cognitive measures to support clinical decisions in ADHD. *Pediatr Neurol*. 2010;42(2):118–126
65. Zadehbagheri F, Hosseini E, Bagheri-Hosseiniabadi Z, Rekabdarkolaei HM, Sadeghi I. Profiling of miRNAs in serum of children with attention-deficit hyperactivity disorder shows significant alterations. *J Psychiatr Res*. 2019;109:185–192
66. Chow JC, Ouyang CS, Chiang CT, et al. Novel method using Hjorth mobility analysis for diagnosing attention-deficit hyperactivity disorder in girls. *Brain Dev*. 2019;41(4):334–340

67. Marcano JL, Bell MA, Beex AAL. Classification of ADHD and non-ADHD subjects using a universal background model. *Biomed Signal Process Control*. 2018;39:204–212
68. Zhou X, Lin Q, Gui Y, Wang Z, Liu M, Lu H. Multimodal MR images-based diagnosis of early adolescent attention-deficit/hyperactivity disorder using multiple kernel learning. *Front Neurosci*. 2021;15:710133
69. Snyder SM, Rugino TA, Hornig M, Stein MA. Integration of an EEG biomarker with a clinician's ADHD evaluation. *Brain Behav*. 2015;5(4):e00330
70. Pliszka S; AACAP Work Group on Quality Issues. Practice parameter for the assessment and treatment of children and adolescents with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*. 2007;46(7):894–921
71. Barbaresi WJ, Campbell L, Diekroger EA, et al. Society for developmental and behavioral pediatrics clinical practice guideline for the assessment and treatment of children and adolescents with complex attention-deficit/hyperactivity disorder. *J Dev Behav Pediatr*. 2020;41(Suppl 2S):S35–S57
72. Wolraich ML, Hagan JF Jr, Allan C, et al; Subcommittee on Children and Adolescents with Attention-Deficit/Hyperactive Disorder. Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Pediatrics*. 2019;144(4):e20192528
73. Davidson F, Cherry K, Corkum P. Validating the Behavior Rating Inventory of Executive Functioning for children with ADHD and their typically developing peers. *Appl Neuropsychol Child*. 2016;5(2):127–137
74. Snyder SM. *Systems and Methods to Identify a Subgroup of ADHD at Higher Risk for Complicating Conditions*. US Patent and Trademark Office; 2010
75. Cedergrén K, Östlund S, Åsberg Johnels J, Billstedt E, Johnson M. Monitoring medication response in ADHD: what can continuous performance tests tell us? *Eur Arch Psychiatry Clin Neurosci*. 2022;272(2):291–299
76. Hall CL, Valentine AZ, Groom MJ, et al. The clinical utility of the continuous performance test and objective measures of activity for diagnosing and monitoring ADHD in children: a systematic review. *Eur Child Adolesc Psychiatry*. 2016;25(7):677–699