The Ability of Self-Report Methods to Accurately Diagnose Attention Deficit Hyperactivity Disorder: A Systematic Review

Allyson G. Harrison1,2 and Melanie J. Edwards1

Abstract
Objective: To identify and analyze all studies validating rating scales or interview-based screeners commonly used to evaluate ADHD in adults. Method: A systematic literature search identified all studies providing diagnostic accuracy statistics, including sensitivity and specificity, supplemented by relevant articles or test manuals referenced in reviewed manuscripts. Results: Only 20 published studies or manuals provided data regarding sensitivity and specificity when tasked with differentiating those with and without ADHD. While all screening measures have excellent ability to correctly classify non-ADHD individuals (with negative predictive values exceeding 96%), false positive rates were high. At best, positive predictive values in clinical samples reached 61%, but most fell below 20%. Conclusion: Clinicians cannot rely on scales alone to diagnose ADHD and must undertake more rigorous evaluation of clients who screen positive. Furthermore, relevant classification statistics must be included in publications to help clinicians make statistically defensible decisions. Otherwise, clinicians risk inappropriately diagnosing ADHD. (J. of Att. Dis. XXXX; XX(X) XX-XX)

Keywords
ADHD, adult, assessment, screening, review

Attention deficit hyperactivity disorder (ADHD) is a common neurodevelopmental disorder that affects between 5 and 15% of school-aged children (American Psychiatric Association [APA], 2013). Research suggests that up to half of those diagnosed with ADHD in childhood no longer meet diagnostic criteria for this condition as adults (Caye et al., 2016), which may explain the lower base rate in adults, previously reported as between 2.5% and 4.5% (APA, 2013; Kessler et al., 2006; Simon et al., 2009; Song et al., 2021). As a neurodevelopmental disorder, diagnosis requires that a large number of impairing symptoms must have been present starting before age 12 (APA, 2013), meaning that the disorder does not first present in adolescence or adulthood.

Recent studies (e.g., Ahmad et al., 2019; Caye et al., 2016; Sibley, Rohde et al., 2018) have demonstrated that previously non-symptomatic adults who first endorse experiencing symptoms of ADHD in adolescence or young adulthood were most often experiencing these symptoms for reasons other than ADHD such as normal fluctuations in cognitive abilities, a comorbid disorder, or the cognitive effects of substance use. These studies underscore the fact that symptoms associated with many other common psychological conditions that affect adolescents and adults can mimic those of ADHD, and that clinicians must investigate and rule out such causes before making a diagnosis of ADHD in these populations.

It is therefore difficult to understand the recent trend whereby an increasing number of adults are seeking out and subsequently being given a first-time diagnosis of ADHD (Chung et al., 2019; P. Marshall et al., 2021; Oehrlein et al., 2016; Olsson et al., 2014; Sasayama et al., 2022). According to the Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (DSM-5; APA, 2013), diagnosis of ADHD requires that the clinician undertake a comprehensive evaluation of current and historical symptoms, document the...
functional impairment arising from the symptoms, establish chronicity, and rule out other possible causes prior to making this diagnosis. In childhood, this process is often fairly easy (Sibley, 2021). Indeed, a clinician can usually canvass parents and teachers to obtain confirmation of a sufficient number of inattentive and/or hyperactive symptoms in various environments and can typically obtain both educational and medical records to confirm both symptoms and impairment. Furthermore, the few conditions that can mimic the symptoms of ADHD in childhood (e.g., oppositional defiant disorder, substance use disorder, metabolic disorders, mood and anxiety disorders) are easily identified and ruled out (Sibley et al., 2018).

By contrast, diagnosis of ADHD in those over age 18 is more difficult and complex (Kolar et al., 2008; Sibley, 2021), especially in those seeking a first-time diagnosis (Ahmad et al., 2019; Sibley et al., 2018; Sibley, Rohde et al., 2018). Here, it is often more difficult to obtain childhood educational and medical records, ensure that collateral sources who know the person well provide input about both childhood and current symptoms, and rule out other common conditions that mimic the symptoms of ADHD (Ahmad et al., 2019; Caye et al., 2017; Sibley, 2021; Sibley et al., 2018; Sibley, Rohde et al., 2018; Weis et al., 2019). Adult retrospective recall of childhood symptoms is unreliable (Breda et al., 2020; Mannuzza et al., 2002; Miller et al., 2010), making it difficult to determine, with a high degree of confidence, whether an adult met the diagnostic criteria for ADHD in childhood based simply on self-report.

Unfortunately, it seems that many clinicians rely mainly on self-reported symptoms (expressed in semi-structured interviews or on self-report questionnaires) when diagnosing young adults with ADHD. For instance, research has found that the majority of diagnostic reports submitted by young adults seeking academic accommodations at postsecondary schools or on medical licensing exams failed to ensure that all five DSM diagnostic criteria were met before rendering the diagnosis (e.g., Joy et al., 2010; Nelson et al., 2019; Weis et al., 2019). The majority of these submitted reports conferred a diagnosis of ADHD based primarily or exclusively on current self-reported symptoms, with most failing to obtain collateral reports, confirm childhood onset, establish functional impairment, or rule out other potential causes for the reported symptoms.

These trends are worrisome. We know that young adults without ADHD often report experiencing symptoms of ADHD (Harrison, 2004; Harrison et al., 2013; J. A. Suhr & Johnson, 2022) especially when they experience high levels of stress, depression, and/or anxiety (Harrison et al., 2013; Lewandowski et al., 2008; J. A. Suhr & Johnson, 2022), meaning that symptom report alone is not sufficient to confirm this diagnosis. We also know that when clinicians rely on self-reported symptoms alone it increases the false positive rate of diagnosis (Faraone et al., 2003). Indeed, both Gathje et al. (2008) and Gordon et al. (2006) showed that the number of individuals diagnosed with ADHD is dramatically (40%–70%) higher when using symptom report alone relative to when other DSM criteria such as functional impairment are considered prior to diagnosis.

In recent years, there have been a number of contributory issues that might increase levels of stress, anxiety, and depression symptoms in the general population and lead to the experience of ADHD-like symptoms. For instance, the recent COVID −19 pandemic has caused many teens and young adults to report increased levels of stress, anxiety, and depression (Statistics Canada, 2021; World Health Organization, 2022) resulting in increased reports of problems such as difficulty concentrating, disrupted sleep patterns, and increased concerns about academic performance (Rashid & Di Genova, 2020; Son et al., 2020). These young adults may seek to find a cause for their current symptoms not understanding that inattention and concentration difficulties are common to many psychiatric disorders (e.g., mental health disorders, substance use disorders; Ahmad et al., 2019; Sibley et al., 2018; Sibley, Rohde et al., 2018). We also know, for example, that misinformation on social media platforms such as TikTok may be responsible for more young adults now believing they may have ADHD (Pugle, 2022; Yeung et al., 2022), and that provision of such inaccurate information leads previously non-symptomatic students to now report experiencing higher levels of ADHD symptoms (Privitera et al., 2015). Pressures to achieve academically have also been blamed for the rise in students who exaggerate or fabricate symptoms in an effort to obtain academic accommodations or stimulant medication (e.g., Benson et al., 2015; Johnson & Suhr, 2021; P. Marshall et al., 2021). However, these trends should not result in a significant increase in adult-aged diagnoses if diagnosticians carefully apply all of the DSM-5 diagnostic criteria when conducting an evaluation (Sibley et al., 2018; Sibley, Rohde et al., 2018).

Despite this, the reported increase in adult-aged diagnosis of ADHD, along with evidence that these diagnoses are often made using self-report alone, suggests that many clinicians may be unaware of the problems inherent in using self-report screening measures diagnostically, namely that these usually have a high false positive rate (see Gilbert et al., 2001; Trevethan, 2017). Previous studies demonstrate that clinicians frequently ignore or misunderstand the predictive validity of a positive score on a screening test (Labarge et al., 2003; Morgan et al., 2021). Indeed, clinicians consistently and significantly overestimate the probability of disease/disorder both before and after obtaining test results, which may contribute to overdiagnosis of disorders (Labarge et al., 2003; Morgan et al., 2021). In the case of ADHD, clinicians may incorrectly believe that self-report measures or interviews have a higher level of diagnostic accuracy than is supported by the research, and may not
understand that base rate of the disorder influences the interpretation of obtained scores. Screening tests are not designed to diagnose but rather to identify individuals whose symptoms require more careful evaluation. Because screening tests are often used to identify uncommon disorders (e.g., ones with a low base rate) the cut scores suggested for use on these tests are designed to err on the side of caution, overidentifying many more people than truly have the condition. By contrast, because these screening tests are overly sensitive they rarely miss those who are symptomatic (Gilbert et al., 2001). Similar to previous studies (e.g., Labarge et al., 2003; Morgan et al., 2021), most clinicians diagnosing ADHD in adults may not understand the actual probability of a true positive diagnosis based on a positive screening test score, leading to overdiagnosis.

A Brief Refresher on Sensitivity, Specificity, Positive, and Negative Predictive Values

Given studies showing that many clinicians fail to understand the predictive statistics that inform screening test results, a brief refresher seems in order. Interested readers may also consult any of the good review articles that provide a more comprehensive discussion of these terms (e.g., Gilbert et al., 2001; Lange & Lippa, 2017; Trevethan, 2017).

All tests function on probabilities; a screening test provides the user with a score that is felt to maximize the probability that a true positive case will not be missed while ensuring that very few individuals with a negative score are really symptomatic. Sensitivity is the actual percentage of true positives; how many known positive cases the test detects. In essence, it answers the question, “I already know that my client has the illness in question. What is the chance that this test will show that my client has it?” Specificity, by contrast, is the actual percentage of true negatives; how many known negative cases are correctly classified as such using this test. In essence, it answers the question, “I already know that my client does not have the illness in question. What is the chance that this test shows my client does not have it?”

While these are useful metrics to know about a test, they are usually employed to determine whether a new test works as well as the gold standard method of diagnosis (Lange & Lippa, 2017; Trevethan, 2017). Because sensitivity and specificity are determined by comparing known diagnoses with obtained test scores, they are not influenced by the base rate of the condition.

However, knowing the sensitivity and specificity of a given test does not help a clinician interpret data from a screening test given to an individual client. When evaluating a client in one’s office, the clinician does not already know what the true answer is (e.g., they don’t know for certain whether the client has the illness or not), and so they rely on the test scores to help decide whether a client’s symptoms are consistent with a particular diagnosis. To obtain this type of clinical information, one must instead know the positive predictive value (PPV) and negative predictive value (NPV) of a given test; these predictive values are influenced heavily by the base rate of the disorder or illness within a specified population (Labarge et al., 2003).

The PPV answers the question, “my client just tested positive on this test. What is the chance that my client truly has this illness?” The NPV, by contrast, answers the question, “my client just tested negative on this test. What is the chance that my client does not have this illness?” As one can see, these are clinically relevant questions asked by most evaluators completing diagnostic evaluations. To understand how base rate affects PPV and NPV it may be instructive to use a clinical example.

Assume that you have 60 adults whom you know have ADHD (based on gold standard diagnostic procedures). You administer a new ADHD self-report measure to these adults as well as to 60 adults whom you know do not have ADHD. The new test performs as shown in Table 1. As may be seen, the new test correctly identifies 90% of your ADHD sample as having ADHD and 72% of your non-ADHD group as not having ADHD. Hence, sensitivity is 90% and specificity is 72%. Note, too, that these scores would not change depending on how common ADHD is in your sample, because these metrics simply say how often the test correctly identifies persons whose status is already known.

However, it is easy for a test to identify people correctly when half of them have the condition in question. In this example, when half of the people in the sample have ADHD then the PPV is 76.3% and the NPV is 87.8%. In reality, however, ADHD occurs in only about 5% of the adult population (e.g., Kessler et al., 2006). In order to evaluate how the new test functions clinically (when the true diagnosis is not known), we would need to evaluate how the new test performs in a population in which only 5% of people have the condition (rather than 50% as was the case in Table 1). Using the 90% specificity and 72% sensitivity values obtained when testing against the gold standard, we can calculate the PPV and NPV of this new test when the base rate of ADHD is 5%. Table 2 presents the resulting identification rates that would occur if we used this test to determine who did or did not have ADHD in a population of 1,000 people, where only 5% actually have the condition of interest.

Here, out of 1,000 people only 50 truly have ADHD (e.g., 5%) and 950 do not. However, the clinician does not know who has the condition and who does not, and so we use our new test to make this determination. Table 2 shows how our new test performs in this scenario. With a known sensitivity of 90% (e.g., I already know you do have ADHD, and 90/100 times the test gets it right) the new ADHD test
Table 1. Performance of New ADHD Self-Report Test Compared With Gold Standard.

<table>
<thead>
<tr>
<th>Actual diagnosis/reality</th>
<th>Test Says Not ADHD</th>
<th>Tests Says ADHD</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not ADHD</td>
<td>43</td>
<td>17</td>
<td>60</td>
</tr>
<tr>
<td>ADHD</td>
<td>6</td>
<td>54</td>
<td>60</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>71</td>
<td>120</td>
</tr>
</tbody>
</table>

Table 2. Ability of New Test to Correctly Identify ADHD When Base Rate is 5%.

<table>
<thead>
<tr>
<th>Actual diagnosis/reality</th>
<th>Test Says Not ADHD</th>
<th>Test Says ADHD</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not ADHD</td>
<td>684</td>
<td>266</td>
<td>950</td>
</tr>
<tr>
<td>ADHD</td>
<td>5</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>689</td>
<td>311</td>
<td>1,000</td>
</tr>
</tbody>
</table>

will correctly identify 45/50 individuals as having ADHD. However, applying specificity of 72% to these data (e.g., I already know that you don’t have ADHD, and for the 950 people without ADHD the test gets it right 72% of the time), we can see that the new test also falsely identifies 266 of the normal† (not ADHD) adults as having ADHD. In other words, for every 311 people the test identifies as ADHD, it is wrong 266 times. Hence, when the base rate of a condition is low, the false positive rate of the screening test will be high. Here, at a base rate of 5%, the false positive rate of the test is 86% (266 of the 311 adults are incorrectly identified by the test as ADHD), whereas the false negative rate is less than 1% (see Trevethan, 2017 for a detailed description of how to complete these types of calculations). This analysis shows how the base rate of a condition influences the diagnostic accuracy of any screening test, and why PPV and NPV are important statistics to know clinically when interpreting the results of any screening test.

Purpose of the Present Study

Two previous systematic reviews of ADHD self-report assessment measures (P. Marshall et al., 2021; Taylor et al., 2011) identified that the majority of studies on which ADHD ratings scales are based were of poor quality, had low statistical power, and failed to report sufficient details to conduct a meta-analysis. Furthermore, most ratings scales, while having adequate sensitivity, had poor specificity (i.e., a high false positive rate). Neither review, however, reported on the actual positive and negative predictive values (PPV and NPV) of these rating scales.

Therefore, the purpose of the present study was to systematically review the literature to identify and analyze all studies validating ADHD rating scales or interview-based screeners commonly used to identify or screen for ADHD in adults. We set out to describe the psychometric properties of these scales and to provide clinicians with information about the PPV and NPV of each scale both when differentiating between normal adults and those with ADHD and also when attempting to differentiate individuals with ADHD from those with other clinical conditions or concerns.

Method

A systematic literature search was undertaken using the MEDLINE, PsycARTICLES and PsycINFO databases from 1998 through June 2022. The terms used in the search are included in Appendix A.

The initial electronic database search identified 1,812 abstracts of journal articles and book chapter titles after duplicates were removed. These were all reviewed by the first author. The 400 abstracts that appeared potentially relevant to the assessment of adult ADHD using self-report were then retrieved and read. After this review was completed, journal articles whose abstracts suggested they were relevant were reviewed. The bibliographies and citations of these journal articles were also scrutinized for potentially relevant articles. As a result, the full texts of an additional seven relevant articles and/or ADHD screening test manuals were obtained and reviewed. The final phase of this literature search focused more narrowly on identifying articles or published tests that met the inclusion criteria (see Figure 1; Table 3). Here, it was essential that the study either report the sensitivity and specificity of the tests evaluated or that the data provided allowed for manual calculation of these statistics.

Articles or tests not meeting inclusion criteria most commonly failed to report diagnostic classification statistics associated with test itself (e.g., sensitivity and specificity statistics not provided in the test manual/study) or with the tests and measures utilized in the study (e.g., how many were correctly or incorrectly classified). This included the technical manuals/normative studies of a number of
commercially available tests such as the Clinical Assessment of Attention Deficit-Adult (CAT-A; Bracken & Boatwright, 2005), the Copeland Symptom Checklist for Adult Attention Deficit Disorders (Copeland, 1989), the Connors Adult ADHD Diagnostic Interview for DSM-IV (CAADID; Epstein et al., 2001), and the Barkley Adult ADHD Self-Report Forms-IV (BAARS-IV; Barkley, 2011). However, research articles evaluating these tests and reporting on resulting diagnostic classification statistics were included.

Results

Method of ADHD Verification

Many of the reviewed studies investigating the performance of various ADHD measures failed to initially confirm the ADHD diagnosis using a comprehensive evaluation or a “gold standard.” Most often, ADHD status of participants was confirmed/made using another screening measure, typically one with a high false positive rate; an unvalidated semi-structured interview; or else the specific method of diagnosis was not provided.

For instance, in the Dunlop et al. (2018) study the diagnosis of ADHD was given retrospectively to individuals with an existing diagnosis of major depressive disorder (but no prior ADHD diagnosis) by means of a semi-structured screening interview. Similarly, ADHD was diagnosed retrospectively in a group of adult patients undergoing treatment for drug and alcohol addiction in the Luty et al. (2009) study by means of an interview. In the Dvorsky et al. (2016) study an ADHD diagnosis was confirmed using the CAADID interview (which, as noted above, lacks...
Table 3. Inclusion and Exclusion Criteria.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>• In peer reviewed journals or published test manual</td>
<td>• Publication not in English</td>
</tr>
<tr>
<td>• Participants aged 18 or older</td>
<td>• Sensitivity and specificity scores not provided or calculable</td>
</tr>
<tr>
<td>• Group study investigating interviews, behavior rating scales, and/or neuropsychological tests for screening or identification of ADHD</td>
<td>• Scales not specific to ADHD symptoms</td>
</tr>
<tr>
<td>• ADHD rating scales commercially available or in public domain</td>
<td></td>
</tr>
<tr>
<td>• Comparison groups: adults diagnosed with ADHD vs. control participants</td>
<td></td>
</tr>
<tr>
<td>and/or participants with psychiatric disorders or clinical complaints</td>
<td></td>
</tr>
<tr>
<td>• Results provide diagnostic accuracy statistics, at minimum sensitivity and specificity</td>
<td></td>
</tr>
</tbody>
</table>

information regarding diagnostic sensitivity or specificity). A similar method of diagnosis (e.g., use of a semi-structured interview) was used in the Brown (1996), Erhardt et al. (1999), Pettersson et al. (2018), Ustun et al. (2017), and van de Glind et al. (2013) studies. In the Hines et al. (2012) study a randomly selected group of patients presenting at eight different primary care medical practices for a routine appointment (e.g., not attending due to suspected ADHD) were administered the ASRS-v1.1 6-item screening questionnaire (Kessler et al., 2005). Those who scored 4 or more out of 6 on this questionnaire were assumed to have ADHD, and those with lower scores were assigned to the control sample. In the Ward et al. (1993) study, ADHD status was confirmed using the Utah criteria for ADHD, which requires only self-reporting of childhood and current symptoms. Sometimes (e.g., Hines et al., 2012; Pettersson et al., 2018; Solanto et al., 2004; Van Voorhees et al., 2011) the rating scale being evaluated had also been used to inform diagnostic status.

In most other studies, the actual method of ADHD diagnosis for participants was not provided, with most (e.g., Brevik et al., 2020) saying it was a “well validated” group or a group who simply self-identified as ADHD (e.g., J. Suhr et al., 2009). In one study (Kessler et al., 2005) the composition of the groups, final numbers per group, and method of identification were opaque. Nowhere do the authors of the Kessler et al. study actually identify the final number of persons who were or were not considered to have ADHD, and the method by which diagnosis was given is not explained operationally. Notably, in none of the 20 studies reviewed were any performance or symptom validity measures utilized in the assessment or diagnosis phase or when evaluating self-reported symptoms.

Diagnostic Accuracy of Screening Measures

Tables 4 and 5 provide details regarding the classification performance of the various screening measures. While all of the studies (overtly or not) included the specificity and sensitivity of the measure in question, none provided relevant PPV and NPV metrics according to expected base rate of ADHD. Hence, we have provided these in both tables. We chose to provide data for base rates of both 5% (aligning with the higher estimated base rate of adult ADHD in the general population) and 10% (given previous suggestions that the prevalence of adult ADHD in general medical practices is as much as twice the population prevalence (see Kessler et al., 2005)).

Differentiating ADHD From Normal/Non-treatment Seeking Adults

Table 4 presents the results from the 12 evaluations that compared individuals said to have ADHD with non-ADHD individuals. Classification results were, in all but one study, compared with individuals said to be normal, non-ADHD, or adults attending a medical practice for routine complaints other than possible ADHD. Only the Kessler et al. (2005) study was opaque regarding the control group composition (see Table 4 for sample descriptions).

Sensitivity is the true positive value of a test. The higher the score, the fewer false negative results. Table 4 shows that, for about half of the tests reviewed, individuals already known to have ADHD are accurately classified relative to normal individuals. Indeed, nine tests reviewed had a sensitivity of over 90%, whereas 11 screening tests fell below 90% when differentiating non-symptomatic individuals from those said to have ADHD, depending on cut score employed for identification. The lowest sensitivities when differentiating between normal and ADHD individuals were the WURS-25 (J. Suhr et al., 2009) and the ASRS 18 items (Kessler et al., 2005), meaning that a large proportion of those who truly had ADHD were not correctly identified in these studies.

Specificity is a test’s ability to correctly identify those without the disease (the true negatives). A highly specific test means that there are few false positive results. Depending on the consequences of incorrect identification, specificity of 90% or higher is often recommended in order to ensure that the false positive rate is low (e.g., Schroeder et al., 2021).
<table>
<thead>
<tr>
<th>Test</th>
<th>Reference</th>
<th>Sample</th>
<th># Items/ Scale Used</th>
<th>Cut Score Used</th>
<th>Sen (%)</th>
<th>Spe (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>Estimated rate of ADHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASRS</td>
<td>Brevik et al. (2020)</td>
<td>646 clinically diagnosed ADHD patients (34 years) vs. 908 normal controls (29 years) (18–40 years)</td>
<td>18 (A + B)</td>
<td>Total score $\geq 35$</td>
<td>80</td>
<td>88</td>
<td>26</td>
<td>99</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total score $\geq 21$</td>
<td>95</td>
<td>45</td>
<td>8</td>
<td>99</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total score $\geq 16$</td>
<td>98</td>
<td>22</td>
<td>6</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kessler et al. 2005</td>
<td>Exact composition of groups not reported. Data derived from 154 participants (18 to 44 years) taken from US National Comorbidity Survey Replication (Kessler et al., 2006). Interviewees were divided into four groups, but no information reported regarding who did or did not receive confirmation of ADHD diagnosis.</td>
<td>6 items (Part A)</td>
<td>$\geq 4/6$ questions endorsing sometimes/ often/very often for questions 1 to 3; and endorsing often/very often for questions 4 to 6</td>
<td>69</td>
<td>99.5</td>
<td>88</td>
<td>98</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\geq 9/18$ items endorsing sometimes/ often/very often for questions 1 to 3, 9, 12, 16, &amp; 18; and endorsing often/ very often for remaining questions</td>
<td>56</td>
<td>98</td>
<td>60</td>
<td>98</td>
<td>10%</td>
</tr>
<tr>
<td>ASRS for DSM-5</td>
<td>Ustun et al. (2017)</td>
<td>193 adults undergoing ADHD evaluation (33 years) and 107 adult controls (age not specified)</td>
<td>6 items (Part A)</td>
<td>Total score $\geq 14$</td>
<td>92</td>
<td>74</td>
<td>16</td>
<td>99</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total score $\geq 17$</td>
<td>76</td>
<td>93</td>
<td>37</td>
<td>99</td>
<td>10%</td>
</tr>
<tr>
<td>BAARS-IV</td>
<td>Dvorsky et al. (2016)</td>
<td>57 college students diagnosed with ADHD using CAADDI interview (20 years, 58% male), 27 college students not diagnosed with ADHD (21 years, 41% male).</td>
<td>9 items</td>
<td>Total of 3 symptoms endorsed as often or very often on the 9-item current Inattentive subscale</td>
<td>89</td>
<td>30</td>
<td>6</td>
<td>98</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total of 5 symptoms endorsed as often or very often on the 9 item childhood inattentive subscale</td>
<td>78</td>
<td>39</td>
<td>8</td>
<td>98</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>Graves (2020)</td>
<td>156 assessment-seeking adults diagnosed with ADHD (27 years) vs. 22 non-clinical controls (30 years).</td>
<td>Subscale Score</td>
<td>Inattention cut score $1.06$ ADHD</td>
<td>79</td>
<td>59</td>
<td>9</td>
<td>98</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>Brown (1996)</td>
<td>143 non-clinical controls, 142 DSM-III diagnosed high IQ ADHD adults (18-40 + y).</td>
<td></td>
<td>Impulsivity cut score 1.13 ADHD</td>
<td>78</td>
<td>55</td>
<td>8</td>
<td>98</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hyperactivity cut score 0.45 ADHD</td>
<td>95</td>
<td>36</td>
<td>7</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$T &gt; = 50$ on Total score</td>
<td>96</td>
<td>89</td>
<td>31</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>CAARS-S:S</td>
<td>Hines et al. (2012)</td>
<td>200 adults (18–65-years) with no current adult ADHD dx vs. 30 randomly chosen negative patients presenting at 8 primary care practices. 25 ADHD dx made using ASRS-A and confirmed using CAARS-self short form.</td>
<td>ADHD Index subscale</td>
<td>$T \geq 65$</td>
<td>92</td>
<td>69</td>
<td>14</td>
<td>99</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>82</td>
<td>87</td>
<td>25</td>
<td>99</td>
<td>10%</td>
</tr>
<tr>
<td>CAARS-S:L</td>
<td>Erhardt et al. (1999)</td>
<td>39 ADHD adults, 16 females 36 years, males 37 years compared to 39 normal adults matched to ADHD sample.</td>
<td>ADHD Index subscale</td>
<td>$T \geq 65$</td>
<td>82</td>
<td>87</td>
<td>25</td>
<td>99</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>82</td>
<td>87</td>
<td>25</td>
<td>99</td>
<td>10%</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Test</th>
<th>Reference</th>
<th>Sample</th>
<th># Items/Scale Used</th>
<th>Cut Score Used</th>
<th>Sen (%)</th>
<th>Spe (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WURS-25</td>
<td>Brevik et al. (2020)</td>
<td>646 clinically diagnosed ADHD patients (34 years) vs. 908 normal controls (29 years).</td>
<td>25</td>
<td>≥ 46</td>
<td>95</td>
<td>75</td>
<td>17</td>
<td>100</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Gift et al. (2021)</td>
<td>137 treatment-seeking adults diagnosed with ADHD (31 years) vs. 120 normal adults with no reported psychiatric or treatment history (35 years).</td>
<td>25</td>
<td></td>
<td>91</td>
<td>92</td>
<td>37</td>
<td>99</td>
<td>56</td>
<td>99</td>
</tr>
<tr>
<td>WURS-25</td>
<td>McCann et al. (2000)</td>
<td>141 adults referred to ADHD clinic, 68 (34 years, 72.1% male) classified as ADHD, 73 (38 years, 66% male) classified as not ADHD.</td>
<td>25</td>
<td>≥ 46</td>
<td>72</td>
<td>58</td>
<td>8</td>
<td>98</td>
<td>16</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>J. Suhr et al. (2009)</td>
<td>1166 individuals with no history of ADHD or current psychiatric problems (64% female), and 104 individuals with existing ADHD dx (34% female in ADHD only, 61% female in ADHD comorbid group). None were seeking assessments. All groups 19 years, range 18 to 25.</td>
<td>37</td>
<td>98</td>
<td>49</td>
<td>97</td>
<td>67</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ward et al. (1993)</td>
<td>81 adult outpatients with ADHD dx (met Utah criteria for ADHD: self-report of childhood and current symptoms), vs. 100 &quot;normal&quot; adults. No information provided on ages or sex distribution.</td>
<td>96</td>
<td>96</td>
<td>56</td>
<td>100</td>
<td>72</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Sen = Sensitivity; Spe = Specificity; PPV = positive predictive value; NPV = negative predictive value; y = years old; dx = diagnosis; dx'd = diagnosed; ASRS = Adult ADHD Self-report Scale; BAARS-IV = Barkley Adult ADHD Rating Scale-IV; CAADID = Conners Adult ADHD Diagnostic Interview for DSM IV; BADDS = Brown Attention-Deficit Disorder Scale; CAARS-S:S = Conners Adult ADHD Rating Scale Self-Report—Short Form; CAARS-S:L = Conners Adult ADHD Rating Scale Self-Report— Long Form; WURS 25 = Wender Utah Rating Scale 25.
Table 5. Predictive Values at Two Base Rates of ADHD Comparing Individuals With ADHD and Treatment Seeking/Clinical Samples.

<table>
<thead>
<tr>
<th>Test</th>
<th>Reference</th>
<th>Sample</th>
<th># Items/ Scale Used</th>
<th>Cut Score Used</th>
<th>Sens (%)</th>
<th>Spec (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASRS</td>
<td>Dunlop et al. (2018)</td>
<td>40 adult patients with major depressive disorder (49 years), 55 healthy adult control subjects (44 years). Five of the MDD group also met DSM-IV criteria for ADHD based on scores from a screening interview.</td>
<td>6 items (Part A)</td>
<td>Total score ≥14</td>
<td>60</td>
<td>69</td>
<td>9</td>
<td>97</td>
<td>18</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Pettersson et al. (2018)</td>
<td>108 patients referred to outpatient clinics in Sweden. 60 (28 years) met criteria for ADHD (method of diagnosis not described). 48 assessment-seeking patient controls not diagnosed ADHD (33 years).</td>
<td></td>
<td></td>
<td>92</td>
<td>27</td>
<td>6</td>
<td>98</td>
<td>12</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Söderström et al. (2014)</td>
<td>41 adults clinically diagnosed with ADHD (32 years) and 20 adults clinically diagnosed without ADHD (30 years).</td>
<td></td>
<td></td>
<td>90</td>
<td>35</td>
<td>7</td>
<td>99</td>
<td>13</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>van de Gindt et al. (2013)</td>
<td>1138 adults (842 male) seeking substance abuse treatment; 13% met ADHD diagnostic criteria based on semi-structured interview (all patients 18–65 years).</td>
<td></td>
<td></td>
<td>84</td>
<td>66</td>
<td>12</td>
<td>99</td>
<td>22</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>BAARS-IV Graves (2020)</td>
<td>156 assessment-seeking adults diagnosed with ADHD (27 years) vs 49 adults diagnosed with Mood or Anxiety Disorders (30 years).</td>
<td>Subscale Score</td>
<td>Inattention cut score 1.17</td>
<td>68</td>
<td>55</td>
<td>7</td>
<td>97</td>
<td>14</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Impulsivity cut score 1.38</td>
<td>67</td>
<td>65</td>
<td>9</td>
<td>97</td>
<td>18</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hyperactivity cut score 1.5</td>
<td>40</td>
<td>80</td>
<td>10</td>
<td>96</td>
<td>18</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>BADDS Solanto et al. (2004)</td>
<td>70 adults diagnosed with ADHD (35 years) and 33 adults diagnosed with depressive or anxiety disorders (44 years).</td>
<td>Subscale Score</td>
<td>Memory Complaints Subtest</td>
<td>92</td>
<td>33</td>
<td>7</td>
<td>99</td>
<td>13</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSM-IV ADHD Symptom Total T ≥ 65</td>
<td>64</td>
<td>86</td>
<td>19</td>
<td>98</td>
<td>34</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>CAARS-SL Harrison et al. (2019)</td>
<td>249 adults clinically diagnosed with ADHD (21 years) and 507 adults clinically assessed but not diagnosed with ADHD (22 years).</td>
<td>Subscale Score</td>
<td>Memory Complaints Subtest</td>
<td>92</td>
<td>33</td>
<td>7</td>
<td>99</td>
<td>13</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSM-IV ADHD Symptom Total T ≥ 65</td>
<td>64</td>
<td>70</td>
<td>10</td>
<td>97</td>
<td>19</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Luty et al. (2009)</td>
<td>107 adult patients undergoing substance abuse treatment; 37 diagnosed retrospectively with ADHD and 59 not diagnosed with ADHD (38 years).</td>
<td>Subscale Score</td>
<td>Inattention cut score 1.17</td>
<td>68</td>
<td>55</td>
<td>7</td>
<td>97</td>
<td>14</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Impulsivity cut score 1.38</td>
<td>67</td>
<td>65</td>
<td>9</td>
<td>97</td>
<td>18</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hyperactivity cut score 1.5</td>
<td>40</td>
<td>80</td>
<td>10</td>
<td>96</td>
<td>18</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>CAARS-SL Van Voorhees et al. (2011)</td>
<td>269 adults (32 years) referred for ADHD evaluation. 184 diagnosed with ADHD, 85 diagnosed with other (mood, anxiety) or no diagnosis. No information about group make-up provided.</td>
<td>Subscale Score</td>
<td>Memory Complaints Subtest</td>
<td>92</td>
<td>33</td>
<td>7</td>
<td>99</td>
<td>13</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSM-IV ADHD Symptom Total T ≥ 65</td>
<td>64</td>
<td>86</td>
<td>19</td>
<td>98</td>
<td>34</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>CSS Söderström et al. (2014)</td>
<td>41 adults clinically diagnosed with ADHD (32 years) and 20 adults clinically diagnosed without ADHD (30 years).</td>
<td>Subscale Score</td>
<td>Memory Complaints Subtest</td>
<td>92</td>
<td>33</td>
<td>7</td>
<td>99</td>
<td>13</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSM-IV ADHD Symptom Total T ≥ 65</td>
<td>64</td>
<td>70</td>
<td>10</td>
<td>97</td>
<td>19</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>DIVA 2.0 Pettersson et al. (2018)</td>
<td>108 patients referred to outpatient clinics in Sweden. 60 (28 years) met criteria for ADHD (method of diagnosis not described). 48 assessment-seeking patient controls not diagnosed ADHD (33 years).</td>
<td>Subscale Score</td>
<td>Memory Complaints Subtest</td>
<td>92</td>
<td>33</td>
<td>7</td>
<td>99</td>
<td>13</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSM-IV ADHD Symptom Total T ≥ 65</td>
<td>64</td>
<td>86</td>
<td>19</td>
<td>98</td>
<td>34</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>WURS-25 Gift et al. (2021)</td>
<td>137 adults with ADHD (31 years) vs 228 adults with Generalized Anxiety Disorder or Major Depressive Disorder (38 years).</td>
<td>Subscale Score</td>
<td>Memory Complaints Subtest</td>
<td>92</td>
<td>33</td>
<td>7</td>
<td>99</td>
<td>13</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSM-IV ADHD Symptom Total T ≥ 65</td>
<td>64</td>
<td>86</td>
<td>19</td>
<td>98</td>
<td>34</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>WURS-25 Luty et al. (2009)</td>
<td>107 adult patients undergoing substance abuse treatment, 42 diagnosed retrospectively with ADHD, 65 not diagnosed with ADHD (38 years).</td>
<td>Subscale Score</td>
<td>Memory Complaints Subtest</td>
<td>92</td>
<td>33</td>
<td>7</td>
<td>99</td>
<td>13</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSM-IV ADHD Symptom Total T ≥ 65</td>
<td>64</td>
<td>86</td>
<td>19</td>
<td>98</td>
<td>34</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>J Suhr et al. (2009)</td>
<td>181 psychiatric controls (78% female), and 104 people with existing ADHD (34% female in ADHD only, 61% female in ADHD comorbid group). None were seeking assessments. All groups 19 years, range 18 to 25 years.</td>
<td>Subscale Score</td>
<td>Memory Complaints Subtest</td>
<td>92</td>
<td>33</td>
<td>7</td>
<td>99</td>
<td>13</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSM-IV ADHD Symptom Total T ≥ 65</td>
<td>64</td>
<td>86</td>
<td>19</td>
<td>98</td>
<td>34</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Ward et al. (1993)</td>
<td>81 adult outpatients with ADHD diagnosis (met Utah criteria for ADHD a self-report of childhood and current symptoms, vs. 70 adults with unipolar depression. No demographic information provided. Study used scores from only 25/61 WURS items that best differentiated between ADHD and normal groups.</td>
<td>Subscale Score</td>
<td>Memory Complaints Subtest</td>
<td>92</td>
<td>33</td>
<td>7</td>
<td>99</td>
<td>13</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSM-IV ADHD Symptom Total T ≥ 65</td>
<td>64</td>
<td>86</td>
<td>19</td>
<td>98</td>
<td>34</td>
<td>96</td>
</tr>
</tbody>
</table>

Note. Sen = Sensitivity; Spe = Specificity; PPV = positive predictive value; NPV = negative predictive value; y = years; dx = diagnosis; dx’ed = diagnosed; ASRS = Adult ADHD Self-report Scale; BAARS IV = Barkley Adult ADHD Rating Scale-IV; BADDS = Brown Attention-Deficit Disorder Scale; CAARS = Conners Adult ADHD Rating Scale; CCS = Barkley Current Symptoms Scale-self report; SD = standard deviation; DIVA 2.0 = Diagnostic Interview for ADHD Adults; WURS 25 = Wender Utah Rating Scale 25.
(Kessler et al., 2005) to a low of 22% (Brevik et al., 2020), with most falling in the mid-range of 40-60% (see Table 4); only six studies found a specificity of 90% or better, meaning that many known normal individuals were falsely identified as having ADHD using these tests.

Of greater interest was the variation in PPV scores when the assumed base rate of ADHD is either 5% or 10%. Here, PPV ranges between a low of 6% (ASRS 18 using a cut score of ≥ 16 and a base rate of 5%; Brevik et al., 2020; BAARS-IV when the base rate is 5%; Brevik et al., 2016) to a high of 88-94% using ASRS-part A and a 5-10% base rate (Kessler et al., 2005). As may be seen in Table 4, however, a positive score in any of these studies typically had, at best, chance ability to correctly identify those with true ADHD compared with normal adults. By contrast, all screening tests had excellent ability to correctly classify non-ADHD individuals, meaning that there is a very small chance that someone with a score below published cut-offs really has ADHD.

**Differentiating ADHD From Other Clinical Samples**

Table 5 provides the classification statistics for the 13 studies where individuals said to have ADHD were compared with treatment seeking or clinical samples. The make-up of the clinical samples differed; some were seeking an assessment for ADHD but did not receive a clinical diagnosis, whereas other studies compared individuals with presumed ADHD to those with mental health or other psychiatric conditions (e.g., anxiety disorders, major depressive disorders, substance use disorders). None of the comparator groups were said to be “symptom-free.”

Sensitivity and specificity scores were lower in this sample (see Table 5). Here, sensitivity ranged from 97% (Luty et al., 2009) to 37% (J. Suhr et al., 2009); only six studies found a sensitivity of 90% or greater. Regarding specificity, no test achieved a specificity score above 90%; six were at or above 80% and the lowest two were at 27%.

In almost all cases the self-report screening tests had extremely good NPV when differentiating between ADHD individuals and a clinical sample. At either estimated base rate, a negative score on these measures very rarely misses true cases of ADHD, even in those with comorbid conditions. Exceptions were the ability of the CAARS and the WURS-25 to differentiate substance abuse treatment participants diagnosed retrospectively with ADHD from those who did not screen positive for ADHD (Luty et al., 2009).

The positive predictive value of a screening test score in these clinical samples, by contrast, had only weak ability to correctly classify true cases of ADHD. When tasked with differentiating true ADHD from psychiatric or assessment-seeking populations, the tests with the highest correct classification accuracy at 10% or 5% base rates were: the CAARS (60% and 57% chance that a substance abuse client also had ADHD given a high score; Luty et al., 2009); and the WURS (61% and 59%; Luty et al., 2009). No other studies found that an ADHD screening test/interview had a better than chance ability to correctly identify true ADHD when compared with clinical samples. Indeed, the second-best positive prediction scores were found for the CAARS at a 10% base rate (a high score has a 34% chance of accurate classification; Harrison et al., 2019) and the WURS-25 at the same base rate (33%; Ward et al., 1993). Most had less than a 10% chance of accurate diagnosis given a positive test score (see Table 5).

**Cut Scores Used for Identification**

Across studies there was also inconsistency regarding the cut scores used when identifying those with ADHD. For instance, on the ASRS some researchers used four out of six items endorsed on part A (requiring endorsement of sometimes or more for questions 1–3 and endorsing often or more for questions 4–6), whereas others used total score cut-offs for either Part A or for the total score on the entire questionnaire. For the BAARS-IV, Dvorsky et al. (2016) used total number of symptoms endorsed as often or very often in childhood and adulthood (but based these on best cut scores calculated for the sample rather than those recommended in the manual) whereas Graves created cut scores for each of the three subscales of the BAARS-IV using optimized sensitivity and specificity thresholds for her sample. For the BAADS, Brown used a T-score of 50 or more on the total score as evidence of ADHD even though 50 represents an average score. Solanto et al. (2004), by contrast, used only the Memory Complaints Scale of the BAADS to identify those with ADHD, again requiring only a T-score of 50 or more. On the WURS, some researchers choose scores of 46 or more as the cut-off whereas two (Ward et al., 1993 & Luty et al., 2009) used scores of 36 or greater. Finally, while most researchers used an ADHD Index score of 65 or greater as the cut-off for the CAARS, Luty et al. (2009) used the total raw score (a score that could range from 0 to 198) and set the cut score as 91; no subscale scores were employed or reported. This is despite the fact that raw score interpretation on the CAARS differs depending on age and sex and that standardized T-scores must be calculated to correctly interpret the meaning of any obtained raw scores.

**Discussion**

This study set out to systematically review the literature to identify and analyze all studies validating ADHD rating scales or interview-based screeners commonly used to identify or screen for ADHD in adults. We set out to describe the diagnostic accuracy statistics of these scales and to provide clinicians with information about the PPV and NPV of each
scale both when differentiating between normal adults and those with ADHD and, of more clinical relevance, when attempting to differentiate individuals with ADHD from those with other clinical conditions or concerns.

It was noteworthy that only about half (nine) of the studies/manuals reviewed actually provided PPV and NPV data for the screening measure being evaluated. For those that did, they almost always reported only PPV and NPV based on the base rate in their current evaluation, typically at a 50% prevalence rate or higher. Given that PPV and NPV are the metrics that clinicians require in order to accurately interpret test scores, and that these are highly influenced by base rate of the condition, the lack of such information being provided to clinicians is worrisome. We know that clinicians often misinterpret screening test scores because they don’t know how to interpret correctly the predictive value of a positive screening test score (e.g., Morgan et al., 2021), and so it is essential that screening measures provide clinicians with such information, ideally for a variety of appropriate base rates.

Results of the present literature review show that, for almost all ADHD screening measures that provide information on sensitivity and specificity (be they self-report or semi-structured interview), these ADHD screening measures have excellent ability to identify non-ADHD individuals correctly; a negative score is almost always correct. By contrast, when using these measures in a clinical setting where one does not know the correct diagnosis a priori, such screening measures have a weak ability to correctly identify those with ADHD when compared to either non-symptomatic individuals or to non-ADHD individuals with conditions that mimic symptoms of ADHD. The latter group of patients are the ones with whom clinicians usually interact; clinicians are rarely asked to determine if a non-symptomatic, otherwise normal individual has ADHD. The results of this review show that, in clinical situations, ADHD screening measures typically have less than chance ability to accurately differentiate those with true ADHD from those with other disorders that also produce symptoms that mimic ADHD. In other words, clinicians who rely mainly or exclusively on these screening measures to diagnose ADHD in adults will overidentify far more people who do not have ADHD than accurately diagnose this condition.

Overidentification is what good screening tests should do—their primary function is to identify all possible individuals who might have the condition of interest, but because they err on the side of caution they falsely identify far more normal people than those who truly have the condition. Clinicians may not appreciate that screening measures overidentify potential cases as being positive, using such scores instead as indicators of diagnostic classification (e.g., Morgan et al., 2021; Thombs et al., 2018). Indeed, even in many of these validation studies the method of classifying participants as having ADHD was based exclusively on self-report measures with a high false positive rate. Clinicians need to be aware that the smaller the base rate of the condition, the higher the false positive rate on screening measures. Unfortunately, as shown by this study and supported in previous research (e.g., P. Marshall et al., 2021; J. A. Suhr & Johnson, 2022; Taylor et al., 2011; Thombs et al., 2018), most self-report checklists or semi-structured interviews that measure symptoms of ADHD, functional impairment, and/or mental health conditions are screening measures with a priori high false positive rates (see Thombs et al., 2018 for a detailed explanation). Screening measures are meant only to alert the clinician to the fact that something might be wrong with the client, not to make diagnostic judgments. Because these measures have such a high false positive rate, clinicians cannot conclude that something is wrong or even confirm the cause of the client’s concerns (e.g., see also Harrison & Sparks, 2022; Morgan et al., 2021), only that further in-depth investigation is warranted.

Research shows that, especially for young adults seeking a first-time evaluation for ADHD, such further investigation must formally evaluate and rule out other, more likely causes for the current symptoms (e.g., Sibley, 2021). Given that there are many other sources of cognitive dysfunction that might lead to false positive symptom endorsement in semi-structured interviews or on self-report questionnaires, it is therefore essential for screening measures to provide information on their ability to discriminate between those with ADHD and those who have conditions with symptoms that mimic ADHD. This should include individuals with depression, anxiety, addiction disorders, and other psychiatric conditions. Without such information, the possibility of false positive identification becomes high.

The need for such clinical validation is shown most clearly by the performance of the ASRS. The initial study on which this screening measure was based is not clear regarding exactly who comprised the ADHD group; however, the authors compared obtained scores only with individuals who endorsed no ADHD symptoms in either childhood or currently. Against this otherwise normal group of individuals, the ASRS part A (six items) had an 88% hit rate when the base rate of ADHD is 5%, and 94% when the base rate is 10%. However, in subsequent studies where this screening measure was tasked with discriminating those with ADHD from other clinical disorders, it failed to demonstrate that high scores are diagnostic; in all subsequent validation studies using clinical comparison groups, a high score on the ASRS had, at best, only a 22% chance of accurately identifying those with true ADHD. This echoes the findings of Chamberlain et al. (2021), who estimated that the ASRS has only about an 11.5% chance of accurately diagnosing ADHD in a large sample of non-treatment seeking young adults in two different countries. These authors caution clinicians that a high score on the ASRS is not synonymous with a diagnosis of ADHD, and that many other
conditions lead to false-positive diagnoses in young adults. The discrepancy in positive predictive value between initial development and practical application of a screening test demonstrates why it is vital for such screening measures to be independently validated against clinical samples.

Clinically, differentiating between ADHD and other, more common conditions that mimic symptoms of ADHD is also a vital step in accurate diagnosis of adults. Many authors have provided advice regarding how to make this differential diagnosis in previously undiagnosed adults (P. Marshall et al., 2021; Ramsey, 2015; Sibley, 2021). This includes establishing chronicity of symptoms, ensuring that others who know the person well verify the presence of the symptoms, determining that the symptoms cause substantial impairment in a variety of major life areas, and most importantly, ruling out other possible causes for the reported symptoms. Such history taking involves interviewing parents or other collateral informants who know the client well, reviewing elementary school report cards or other childhood medical records, and obtaining a detailed timeline of onset, remission, and recurrence of both ADHD and the comorbid symptoms of a patient as well as other major life disruptions or general stressors (e.g., COVID-19) that may influence symptom expression. Sibley (2021) also notes that substance use disorders and other addictive behaviors (e.g., dependence on electronic devices) often mimic symptoms of ADHD and thus it is important to determine whether ADHD symptoms preceded or followed onset of any such conditions. She also recommends always including a measure of adult psychopathology (preferably one that includes symptom validity scales) in all ADHD screenings.

This investigation highlighted a few significant problems with research undertaken to develop and/or validate ADHD screening questionnaires/interviews. First, in clinical studies of various screening measures reviewed in the present survey, ADHD status was often determined by self-report alone, typically employing screening measures/interviews with high false positive rates to retrospectively classify individuals with other disorders as also having ADHD. This clearly bypasses all of the other DSM-5 criteria required to make this diagnosis and ignores research showing that false positive diagnosis is likely when ignoring criteria such as functional impairment (e.g., Gathje et al., 2008; Gordon et al., 2006). As such, it is not possible to know whether the individuals studied would have really met diagnostic criteria for ADHD, thus eroding the credibility of the data obtained in their study.

A major weakness of self-report (in interview or when using rating scales) is that it reflects a subjective impression of behavior rather than being an objective and impartial evaluation (P. Marshall et al., 2021). Validity of self-report is thus a significant issue, especially for a diagnosis that relies so heavily on self-reported symptoms and history. Bias and exaggeration in self-reporting, whatever the cause (see Harrison et al., 2021 for a review of various reasons), can interfere with accurate evaluation of ADHD in adults, and noncredible reporting is estimated to occur in up to half of ADHD evaluations (P. S. Marshall et al., 2016; J. A. Suhr et al., 2020; Sullivan et al., 2007).

A myriad of studies over the past decade have shown how easily non-ADHD individuals can feign or exaggerate symptoms of ADHD in a manner that allowed for diagnosis (e.g., P. S. Marshall et al., 2016; Musso et al., 2016; Tucha et al., 2015). Furthermore, studies also review the secondary gain potentials that exist for young adults to feign ADHD in order to obtain access to stimulant medication (e.g., Cook et al., 2021; Weyandt et al., 2016), academic accommodations (e.g., Harrison et al., 2021), or access to disability funding grants or payments (e.g., Gordon et al., 2015; Harrison, 2017). Even if an individual is not deliberately exaggerating symptoms, they may inadvertently start to believe they have ADHD due to inaccurate information provided to them (e.g., Privitera et al., 2015; J. A. Suhr, 2016). For instance, recent reports of factitious disorder and illness identity created via social media (e.g., Haltigan et al., 2023; Harness & Getzen, 2022) show that social media platforms may act as a vehicle of transmission for the social contagion of self-diagnosed mental health conditions, particularly in stressed or vulnerable young women.

Unfortunately, studies show that clinical judgment alone is insufficient to detect such noncredible presentation during clinical evaluations (Dandachi-Fitzgerald et al., 2017; Guilmette, 2013), meaning that clinicians must rely on objective symptom and performance validity measures to evaluate response credibility (Sherman et al., 2020). Despite this, recent research shows that most assessors do not formally evaluate the credibility of self-reported symptoms (Musso & Gouvier, 2014; Nelson et al., 2019); notably, none of the present studies reviewed undertook any objective evaluation of self-report credibility. In order to be certain that screening questionnaires accurately discriminate those with ADHD from other conditions, it is therefore essential that the ADHD subjects in any study be ones for whom a gold-standard assessment occurred, including evaluating symptom credibility and ruling out other possible causes for the symptoms (see Sibley, 2021, for a summary of empirically-informed guidelines for first-time adult ADHD diagnosis). Clinicians must also include objective measures of self-report and performance credibility in any assessment of ADHD in those over age 18 (J. A. Suhr & Berry, 2017).

**Conclusions**

Initial screening for conditions such as ADHD is a vital and cost-effective first step in determining who may require a more comprehensive evaluation; however, results from this review show that screening tests are much better at ruling out ADHD as opposed to confirming the diagnosis. Clinicians who use self-report screening tests or who
administer semi-structured interviews need to be aware that a positive screening outcome, especially in a clinical setting, has an extremely high false positive rate and a low positive predictive value. This means that clinicians must undertake a rigorous evaluation of clients with positive screening scores, including objective reviews of past history, obtaining opinions from knowledgeable collateral sources, evaluating whether symptoms have caused substantial impairment both historically and currently, and most importantly, ruling out the causal influence of many other, higher base rate disorders such as anxiety, depression, addictions, or symptom overreporting. Furthermore, those who develop ADHD screening measures have a responsibility to evaluate how well these measures predict actual ADHD when compared with a sample of assessment-seeking clients and provide data regarding the positive and negative predictive values of their tests at expected population base rates. Without this validation, clinicians run the risk of inappropriately diagnosing and treating clients for ADHD.

**Appendix**

**Search Terms**

Attention deficit hyperactivity disorder  
ADHD  
Attention deficit disorder without Hyperactivity  
Attention deficit disorder with Hyperactivity  
Attention deficit disorder  
Self report  
Self disclosure  
Psychiatric diagnosis  
Instruments  
Screening  
Screening tools  
Screening tests  
Screening scale  
Diagnostic tool  
Diagnostic scale  
Diagnosis  
Testing  
Assessment  
Evaluation  
Questionnaires  
Validation  
Psychometric properties  
Psychometrics  
Specificity  
Sensitivity  
Adult

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding to support this research was provided by the Ontario Ministry of Colleges and Universities. The authors received no funding for authorship and/or publication of this article.

**ORCID iD**

Allyson G. Harrison [https://orcid.org/0000-0002-0426-2011]

**Note**

1. In this manuscript, the use of the term “normal” refers to non-treatment seeking individuals who were not being actively treated for ADHD or another diagnosed psychological condition.

**References**


Labarge, A. S., McCaffrey, R. J., & Brown, T. A. (2003). Neuropsychologists’ abilities to determine the predictive value of diagnostic tests. *Archives of Clinical Neuropsychology*, 18(2), 165–175. citation2003/03/01 00:00Labarge AS


Rashid, T., & Di Genova, L. (2020). Campus mental health in times of COVID-19 pandemic: Data-informed challenges and
opportunities. *Campus Mental Health: Community of Practice (CoP)*. Canadian Association of Colleges and University Student Services. file:///C/Users/RARCUS~1/AppData/Local/Temp/Campus-MH-COVID-19-Rashid-DiGenova.pdf


van de Glind, G., van Den Brink, W., Koeter, M. W., Carpentier, P. J., van Emmerik-van Oortmerssen, K., Kaye, S., Skutle,


**Author Biographies**

**Allyson G. Harrison** is the former clinical director of the Regional Assessment and Resource Centre at Queen’s University, one of two centers mandated to support postsecondary students through the provision of assessments, applied research, and professional development for psychological practitioners. She also holds an appointment as an assistant professor in the Department of Clinical Psychology at Queen’s University. Her research is dedicated to issues of differential diagnosis in LD and ADHD.

**Dr. Melanie J. Edwards** is a clinical psychologist working at the Regional Assessment and Resource Centre at Queen’s University.