medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

A review of behavioural rating scales for the assessment of Attention Deficit/Hyperactivity Disorder among adults in research.

Authorship

Marius Grandjean[1,2], Shachar Hochman[1], Raja Mukherjee[3], Roi Cohen Kadosh[1]

[1]School of Psychology, University of Surrey, Guildford GU2 7XH, United Kingdom

[2]Psychological Sciences Research Institute, UCLouvain, Louvain-la-Neuve, Belgium

[3]Adult Neurodevelopmental Service, Horizon House, Epsom, KT17 4QJ, United Kingdom

Correspondence: Dr Shachar Hochman (s.hochman@surrey.ac.uk)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Objective: This review aims to provide researchers with a contemporary and comprehensive understanding of the current state of behavioural rating scales used in evaluating adult ADHD for research purposes. The objective is to offer guidance that enables researchers to make informed decisions when selecting the most suitable scale for their studies. Moreover, our intention was to map and compare these scales, with a specific focus on detecting feigned or invalid symptom presentation—an aspect notably overlooked in prior reviews.

Method: We reviewed the most recent literature on behavioural rating scales for adult ADHD assessment. We evaluated the scales and compared them based on their psychometric properties and the range of symptoms that they assessed.

Results: The Conners' Adult ADHD Rating Scales (CAARS), Mind Excessively Wandering Scale (MEWS), and Wender Utah Rating Scale (WURS) have emerged as the most accurate measures for assessing adult ADHD. Moreover, there is an increasing emphasis on the development of assessment tools, either integrated within existing scales or as independent measures, to evaluate feigning or invalid symptom presentation. In that regard, stand-alone measures have demonstrated greater effectiveness compared to embedded measures, with the ADHD Symptom Infrequency Scale (ASIS) being identified as the most accurate scale for the detection of feigning.

Conclusion: Based on this review, we provide recommendations for the behavioural rating scales with the most accurate measurement of relevant variables in research-related settings.

Keywords

Attention deficit hyperactivity disorder; behavioural rating scale; questionnaire; psychometrics; adult; assessment.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder exhibiting clinical heterogeneity, marked by the presence of a sustained tendency towards inattention and distractibility, as well as hyperactivity and impulsivity, causing interference in functioning or development (American Psychiatric Association, 2022). In many cases, ADHD continues to affect individuals into adulthood, and can be associated with an increased risk of developing other mental health disorders, as well as negative outcomes such as educational underachievement, challenges with employment and interpersonal relationships, and potential involvement with criminal activities (Sayal et al., 2018). The phenomenon of adult ADHD can be differentiated into two distinct subtypes, namely persistent and symptomatic adult ADHD. The former subtype is characterised by the enduring presence of ADHD symptoms over an extended period, spanning from childhood through adulthood. Conversely, the latter subtype refers to the presence of clinically significant symptoms of ADHD that cause impairment in daily functioning without a specific childhood-onset (Song et al., 2021). Recent research has reported the global prevalence rates of persistent and symptomatic adult ADHD to be 2.58% and 6.76% respectively, corresponding to 139.84 million and 366.33 million of affected adults worldwide (Song et al., 2021).

When assessing adult ADHD, various methods are used, including interviews, continuous performance tests (CPTs), and behavioural rating scales. CPTs are computer-based neuropsychological tests that require the individual to respond to a series of stimuli, measuring specific cognitive functions that are impaired in ADHD. Meanwhile, behavioural rating scales are subjective measures completed by either the individual or an informant. These scales usually assess the frequency and severity of ADHD-related behaviours. While clinical interviews have often been regarded as the gold standard for ADHD assessment (Murphy & Gordon, 1998; Ramsay, 2015), a recent comprehensive review by Marshall et al. (2021) found that a combination of interviews, CPTs and behavioural rating scales result in the highest rate of diagnostic accuracy. Unfortunately, such an extensive intervention may not always be feasible (Kessler et al., 2005), and it is resource intensive. In research settings, behavioural rating scales and/or CPTs are often used as the primary outcome of studies on new treatment for ADHD (Allenby et al., 2018; Alyagon et al., 2020; Berger et al., 2021; McGough et al., 2019; Nahum et al., 2023; Paz et al., 2018; Weaver et al., 2012). While several

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

studies have examined the effectiveness of CPTs (Baggio et al., 2020; Epstein et al., 2003; Ogundele et al., 2011; Riccio & Reynolds, 2001; Vaughn et al., 2011), only a few have assessed the usefulness of behavioural rating scales (Harrison & Edwards, 2023; Marshall et al., 2021; Taylor et al., 2011). Given their relatively low cost and quick administration, rating scales are often favoured as the main outcome measure in studies focused on new treatments for ADHD (Alyagon et al., 2020; McGough et al., 2019; Paz et al., 2018; Weaver et al., 2012). However, there is a relative lack of a clear understanding of the quality and adequacy of these rating scales. This prevents researchers to make informed decisions when selecting which scale to use and for what purpose.

To enhance the reader's understanding of the commonly used behavioural rating scales used in the assessment of adult ADHD, we will provide a brief and non-exhaustive overview of the most frequently used measures for assessing adult ADHD. These rating scales are either available publicly at no cost or commercially. Among the publicly available scales, there is the Wender Utah Rating Scale (WURS) which is a retrospective self-reported measure consisting of 61 items that assess experienced childhood symptoms of ADHD in adults. A shorter version of the WURS known as WURS-25 is also available (Ward et al., 1993). Another widely used measure is the Adult ADHD Self-Report Scales (ASRS), which assesses the 18 symptoms of ADHD outlined in the DSM-4 (Adler et al., 2006). In addition to the free measures previously mentioned, there are also commercially available measures. To begin, Barkley Adult ADHD Rating Scale (BAARS-IV) has various subscales for assessing current and childhood symptoms based on DSM-5 criteria. Each scale consists of 27 items and can be administered either through self-report or by an observer (Barkley, 2011). Then, there is the Brown Attention-Deficit Disorder Scales (BADDS), which is a self-reported questionnaire consisting of 40 to 50 items divided into six subscales. It assesses symptoms associated with ADHD and executive function impairments (Brown, 1996), as such impairments has been suggested to be one of the underlying mechanisms of ADHD (Barkley, 1997). Lastly, the Conners' Adult ADHD Rating Scales (CAARS, long version) has 66 items that assess the presence and severity of ADHD symptoms through statements about daily activities and behavioural tendencies. Similarly, to the BAARS-IV, the CAARS also has both self- and observer-reported versions that comprise identical items.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Interestingly, The CAARS also includes an Inconsistency Index, which is a measure for careless or random responding (Conners et al., 1999). In addition to the original Inconsistency Index, researchers have developed two additional indexes for the CAARS aiming to support the psychometric value of the CAARS; The first is the CAARS Infrequency Index (CII), which consists of items from the CAARS that are rarely endorsed by individuals with or without ADHD. The CII is useful for identifying overreporting of symptoms (J. A. Suhr et al., 2011). The second Index Is the Exaggeration Index (EI), which includes items from the Dissociative Experiences Scale (DES) that are integrated into the CAARS. The EI measures the extent to which an individual reports exaggerated or extreme symptoms of ADHD (Harrison & Armstrong, 2016).

In previous reviews, the CAARS was argumentatively recommended as the most useful measure due to its psychometric properties (Taylor et al., 2011), its coverage of a large constellation of symptoms and its ability to identify invalid symptom presentation (Marshall et al., 2021), which is becoming an increasingly concerning issue. Over the last decade, studies have shown that behavioural rating scales can be easily falsified (Jachimowicz & Geiselman, 2004; Lee Booksh et al., 2010; Quinn, 2003). This issue is compounded by the fact that some individuals may feign ADHD symptoms for a variety of reasons, such as to provide a socially acceptable excuse for their difficulties (J. Suhr & Wei, 2013) or to obtain benefits associated with ADHD medication (Hinshaw & Scheffler, 2014). This is particularly concerning because the non-specific and subjective nature of ADHD symptoms makes it easier for individuals to feign symptoms during formal evaluations, which is especially prevalent among adults, particularly college students. Additionally, the diagnostic criteria for adult ADHD, as emphasised by the DSM-5, rely heavily on subjective symptoms rather than cognitive or functional deficits. Therefore, when using behavioural rating scales to assess ADHD symptoms, it is crucial to consider the possibility of feigned results. The primary objective of this paper is to scrutinize the latest advancements concerning the utilisation of behavioural rating scales in the evaluation of ADHD symptoms. More specifically, this review aims to provide researchers with an updated and comprehensive understanding of the current state of behavioural rating scales utilized in the assessment of adult ADHD, to aid them in making informed decisions when selecting an appropriate scale for their research (for a comprehensive analysis of the diagnostic efficacy of behavioural rating scales for clinical utility, see Harrison & Edwards, 2023).

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

The traditional approach to compare rating scales used for ADHD assessment is by examining their psychometric properties. However, there is no agreed-upon gold standard regarding which statistics are more important or how to determine when one scale is better than another. In the present review, we chose to focus on evaluating the sensitivity, specificity, PPV, and NPV of the rating scales. Consequently, before delving further into this review, we briefly describe the fundamental statistics that we used. To begin, *sensitivity*, also referred to as the true positive rate, reflects the likelihood that a test or measurement accurately identifies the presence of a specific condition. To elucidate, an ADHD rating scale with a sensitivity of 0.60 would correctly identify ADHD in 60% of cases. Moving on, *specificity*, known as the true negative rate, signifies the probability of a test or measurement accurately recognizing the absence of a particular condition. For instance, an ADHD rating scale boasting a specificity of 0.30 would accurately identify the absence of ADHD in 30% of cases. Transitioning to *positive predictive value* (PPV), which gauges the probability that an individual possesses a condition if the test or measurement identifies them as having it. For example, an ADHD rating scale with a PPV of 0.80 signifies that if the test indicates ADHD presence, there's an 80% likelihood the individual indeed has ADHD. In contrast, *negative predictive value* (NPV) assesses the probability that an individual lacks a given condition if the test or measurement designates them as not having it. To provide an example, an ADHD rating scale with an NPV of 0.70 would imply a 70% probability of an individual not having ADHD if the test indicates so (Ivnik et al., 2001).

In simpler terms, *sensitivity* measures how well a test can correctly find positive cases. On the other hand, PPV looks at how likely it is that a positive test result truly means the person has the condition. This way, for example, it is possible to have a test that does not catch all positive cases (low *sensitivity*), but when it does yield a positive result, it is usually correct (high PPV). In contrast, *specificity* measures how well the test can correctly identify those without the condition (negative cases). The NPV, meanwhile, gauges how likely it is that a negative test result truly means the person doesn't have the condition. Some authors argued that although sensitivity and specificity are crucial in determining true positive and true negative, they fail to account for the confidence level a test score can provide in either confirming or excluding a diagnosis or condition. Likewise, PPV and NPV are useful in clinical decision-making because they are influenced by the base rate (BR), which is the prevalence of the condition in the

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

population but do not indicate the frequency with which the test identifies the targeted behaviour. Hence, it is strongly recommended to interpret all four values concurrently and use qualitative descriptors to interpret their clinical usefulness (Lange & Lippa, 2017) (see Table 1).

Table 1. Recommended qualitative descriptors for sensitivity, specificity, PPV and NPV values as proposed by Lange & Lippa (2017).

Value as a percentage	Qualitative descriptor
<10	Very low
10-24	Low
25-39	Low-moderate
40-59	Moderate
60-74	Moderate-high
75-89	High
90-100	Very-high

Note. PPV = positive predictive value, NPV = negative predictive value.

Although behavioural rating scales should not be relied upon solely to clinically diagnose adult ADHD (Harrison & Edwards, 2023), they are often regarded as a valuable tool in research settings. This is particularly notable when the objective is to measure changes in symptoms during the implementation of new treatments, prioritizing symptom assessment over establishing a conclusive diagnosis. This situation often prompts researchers to deliberate on the most appropriate scale that aligns with their specific research objectives. This paper sets out to elucidate the current status of these scales, offering in-depth insights into their accuracy, the symptoms they effectively capture, and their distinct characteristics. Through practical recommendations, this paper aims to empower researchers to make informed decisions when selecting the most suitable scale for their research endeavours.

Materials and Methods

We conducted a comprehensive search in the PubMed database between 2019 and February 2023, using the search terms “ADHD" AND "attention deficit hyperactivity disorder" AND "assessment or testing or evaluation” AND "adult" AND "diagnosis" (see Figure 1). We specifically chose these terms to match those used in the most recent review of adult ADHD assessment (Marshall et al., 2021) which examined the literature published between 1998 and 2019. Therefore, we selected 2019 as our starting point to provide an updated view of the

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

state of behavioural rating scales for adult ADHD assessment. Electronic publications were the only publications considered for inclusion in the review. In the initial search, a total of 1264 abstracts of journal articles were identified, from which 73 were considered potentially relevant to adult ADHD assessment. After a further consideration of the abstracts, the full texts of 51 journal articles were reviewed, in addition to their bibliographies and citations, which led to the identification and review of an additional 5 articles. The final phase of the literature search aimed at identifying articles that met the inclusion criteria culminating in a final sample of 11 articles (see Table 2). A detailed description of the reviewed papers can be found in Appendix 1.

Figure 1. Flow diagram illustrating the literature search process and article selection.

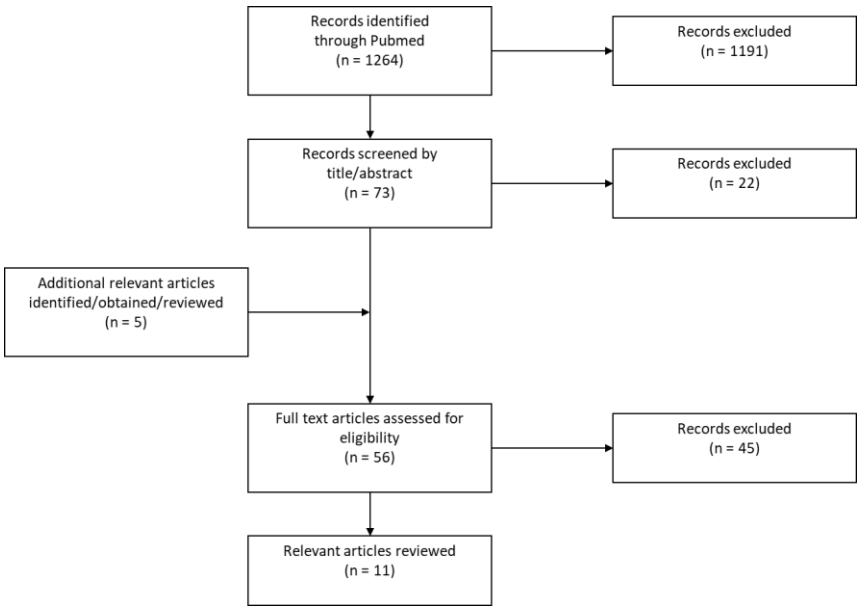


Table 2. Inclusion criteria

- Published in English peer reviewed journals. - ADHD is diagnosed by widely used clinical or research semi-structured or structured interview. - Group study investigation behaviour rating scales for diagnosis, screening, or identification of ADHD.	- Neuropsychological tests are standardized and have normative data. - Results provide diagnostic accuracy statistics, at minimum sensitivity and specificity. - Participants are adults, 18 years old or older.

Results

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Behavioural rating scales for the screening of ADHD

Two studies focused on distinguishing adults with ADHD from those without the condition, while another two studies aimed to differentiate adults with ADHD from individuals diagnosed with other psychiatric disorders.

The first study used the short version of the WURS (WURS-25) and the ASRS. They found that the WURS-25 had a high level of sensitivity 90% and specificity 88% and the ASRS scale demonstrated a high level of sensitivity of 80% and specificity of 88% in distinguishing between adults who were seeking assessment for ADHD and healthy controls. However there were no mention of PPV nor NPV (Brevik et al., 2020). The second study compared the full version of the WURS with the WURS-25 and found that both versions had high sensitivity and specificity values when distinguishing between control and ADHD groups. Specifically, the full WURS had a sensitivity of 95%, a specificity of 93%, a PPV of 94%, and a NPV of 84%, while the WURS-25 had a sensitivity of 91%, a specificity of 94%, a PPV of 88%, and an NPV of 91%. When differentiating between ADHD and other psychiatric disorders, the full WURS had very high specificity of 94% and NPV of 91% and high sensitivity of 84% and PPV of 88%. In contrast, the WURS-25 demonstrated high specificity of 85% and NPV of 79% and moderate-to-high sensitivity of 62% and PPV of 73% (Gift et al., 2021). Then, Mowlem and colleagues (2019) developed the Mind Excessively Wandering Scale (MEWS) to assess the presence of mind wandering in individuals ADHD. The MEWS defines mind wandering as *"periods in time in which attention switches from a current task to unrelated thoughts and feelings*" (Smallwood & Schooler, 2015). The MEWS is a self-report measure consisting of 12 or 15 items, which are rated on a 4-point Likert-type scale ranging from 0 (indicating "not at all or rarely") to 3 (indicating "nearly all of the time or constantly"). In their study, the MEWS was found to have a high level of sensitivity of 90% and specificity of 90% in the diagnosis of ADHD using sample of patients with ADHD and healthy controls. The authors reported no PPV nor NPV (Mowlem et al., 2019). Finally, a study was focused on the main scale of the CAARS and the t-scores, which are standardised scores that compares the individual score with the reference group. They found that utilizing a t-score threshold of > 65 for the identification of ADHD was associated with a moderate-to-high sensitivity of 64%, a high specificity of 86%, a moderate PPV of 51%, and a moderate-to-high NPV of 71%. In contrast, employing a t-score threshold

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

of > 70 yielded a low sensitivity of 14%, a very-high specificity of 92%, a moderate PPV of 47%, and a moderate-to-high NPV of 68% (Harrison et al., 2019).

The aforementioned studies presented varying outcomes in comparison to prior research conducted beyond the scope of the reviewed timeframe. For instance, Brevik et al., (2020) found higher specificity (88%) for the ASRS than previously reported by other studies (ranging from 27% to 68%) (Dunlop et al., 2018; Pettersson et al., 2018; Söderström et al., 2014; van de Glind et al., 2013). Similarly, Gift et al., (2021) found a higher specificity for the WURS (93-94%) compared to prior findings (ranging from 57% to 70%) (Luty et al., 2009; McCann et al., 2000). Finally, although the specificity of the CAARS generally remained high across studies, the sensitivity found by Harrison et al. (2019) (64%) was lower than previously reported (97%) when the diagnosis is based on self and observer reports (Luty et al., 2009). These discrepancies between the studies may then be attributed to factors such as differences in sample size, addition of informant's feedback, the proportion of ADHD subtypes, or whether there was a clinical group. The latter is of particular issue because clinical settings typically involve patients with other psychiatric conditions, such as depression and anxiety, that are known to share comorbidities with ADHD (Anker et al., 2018; Kessler et al., 2006). As a result, there is a higher risk of these patients being erroneously diagnosed.

Practical recommendations

Our findings yielded the following results, when differentiating between healthy controls and patients with ADHD, the WURS exhibited very-high sensitivity along with the WURS-25 and the MEWS, whereas ASRS had high sensitivity, and the CAARS had moderate sensitivity. Moreover, the WURS and the MEWS showed very-high specificity, while the WURS-25, ASRS and CAARS demonstrated high specificity. The WURS also displayed very-high PPV with a high NPV, whereas the CAARS exhibited moderate PPV with a moderate-to-high. Similarly, when distinguishing between clinical patients and patients with ADHD, the WURS demonstrated very-high sensitivity, and the WURS-25 displayed moderate-to-high sensitivity. Both the WURS and WURS-25 exhibited high specificity. The WURS had a high PPV and a very-high NPV, while the WURS-25 displayed moderate-to-high PPV and high NPV.

Overall, research to date suggest that the CAARS, the MEWS and the WURS are the measures with the best statistical properties for the assessment of adult ADHD (see appendix 2).

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Behavioural Rating scales for the detection of feign ADHD symptoms

A total of seven studies explored indices designed to assess the validity of ADHD symptoms and the potential for feigning results.

To begin, two studies were focused on the validity indexes of the CAARS, which assess for the detection of feigning and non-credible symptom presentation. Becke et al. (2021) created the ADHD credibility index (ACI), a 12-items scale which was then compared with the CII. For the detection of feigning, the ACI had a very high specificity of 98%, a low-to-moderate sensitivity of 30%, and PPV and NPV ranging from 69% to 95% and 58% to 92%, respectively, varying between moderate and very high. Meanwhile, the CII had a very high specificity of 95%, a moderate sensitivity of 46%, and PPV and NPV ranging from 51% to 90% and 63% to 94%, respectively, also varying between moderate and very high (Becke et al., 2021). The authors also created the CII-ACI-Compound Index, a combination of items from both validity indexes. The CII-ACI-Compound Index had a high specificity of 87% to 92%, a moderate sensitivity of 41% to 50%, and PPV ranging from 15% to 87% and NPV ranging from 60% to 97%, varying between low-to-high and moderate-to-high and very high (Becke et al., 2022). Instead of utilising embedded validity indices to detect feigning of ADHD symptoms, Courrégé and colleagues (2019) developed a self-administered measure known as the ADHD Symptom Infrequency Scale (ASIS). Comprising 52 true/false items, the ASIS is subdivided into two scales: the ADHD scale, which includes 19 items designed to align with DSM-5 diagnostic criteria, and the Infrequency scale, which includes 33 items intended to be endorsed more frequently by individuals simulating ADHD than those with a genuine diagnosis. Using a sample composed of patients with diagnosed ADHD and instructed simulators, the infrequency scale of the ASIS demonstrated a high sensitivity of 79% to 86%, specificity of 89%, PPV of 71% to 79%, and a very high NPV of 92% to 93% in the detection of feigning (Courrégé et al., 2019). Another study, which included a clinical control group, reported that the infrequency scale of the ASIS had a very high specificity of 90%, moderate-to-high sensitivity of 71%, moderate-to-high PPV of 65%, and a very high NPV of 93% for detecting feigning (Skeel et al., 2022). Another recent stand-alone measure, the Multidimensional ADHD Rating Scale (MARS), was developed by Potts and colleagues (2021). The MARS includes three categories of items: 18 symptom items, 22 impairment items, and 4 symptom-validity items. Additionally, three "catch" items are incorporated into the assessment to measure effort and attention during the

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

administration. The items are rated on a scale ranging from 0 to 8, where the symptom scale ranges from 0 ("never") to 8 ("very often"), and the invalid scale ranges from 0 ("not at all") to 8 ("severe"). For the detection of ADHD, the MARS yielded a high sensitivity of 86% to 92%, a moderate-to-high specificity of 58% to 67%, and PPV ranging from 60% to 65% and high NPV of 86% to 92%. For the detection of feigning, the infrequency scale of the MARS showed a very high specificity of 92%, a moderate-to-high sensitivity of 65%, very high PPV of 92%, and moderate-to-high NPV of 65% (Potts et al., 2021). In another study using a similar sample, the infrequency scale of the MARS and found a specificity of 88%, a moderate-to-high sensitivity of 62%, and moderate-to-high PPV of 63-88% and NPV of 69-87% for the detection of feigning (Potts et al., 2022). Finally, a recent study by Harrison et al. (2022) examined the validity of two new indicators proposed by Aita et al. (2018) for identifying feigned ADHD using the Personality Assessment Inventory (PAI). The PAI is a self-report personality measure consisting of 344 items divided into four scales in which respondents are required to rate each item on a four-point Likert scale ranging from 1 (indicating "false") to 4 (indicating "very true") (Morey, 1991). The study utilized a sample of patients with genuine ADHD and healthy controls who were asked to either respond truthfully or feign ADHD. The results revealed that the Item-FAA (Feigned Adult ADHD index using PAI items) had a high specificity of 78%, a moderate sensitivity of 45%, a low PPV of 19%, and a very high NPV of 92%, while the Scale-FAA (Feigned Adult ADHD index using PAI Scales) had a high specificity of 79%, a low-to-moderate sensitivity of 36%, a low PPV of 17%, and a very high NPV of 91% for the detection of feigning (Harrison et al., 2022).

It is noteworthy that the CAARS stands out among behavioural rating scales in that it possesses validity indicators capable of detecting feigned symptoms. In addition to its initial infrequency index, the CII (J. A. Suhr et al., 2011) and the EI (Harrison & Armstrong, 2016), a new index, the ACI, was recently developed (Becke et al., 2021). Although the ACI had very-high sensitivity, its specificity was found to be low-to-moderate. To improve its overall classification accuracy, the authors of the study developed the CII-ACI-Compound, which showed a very high sensitivity but only a moderate specificity (Becke et al., 2022). Although these indexes show promise, further research is necessary to improve their accuracy. Additionally, they are based on different theoretical foundations and may identify various subgroups of examinees as non-credible (Becke et al., 2021). Conversely, as opposed to measures embedded within

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

existing rating scales, some authors have created stand-alone measures that assess ADHD directly while accounting for the existence of feigned symptoms. In spite of the fact that the MARS has shown promise in previous studies (Potts et al., 2021, 2022), its efficacy has not yet been tested on clinical groups, which may have led to inflated results, as explained earlier. On the contrary, the ASIS has exhibited good psychometric properties in its initial validation study (Courrégé et al., 2019) and these findings have already been replicated with a sample of patients reporting symptoms of depression and anxiety (Skeel et al., 2022).

Practical recommendations

In recent years, there has been a growing emphasis on developing assessment tools aimed at detecting feigned symptoms. This focus arises from the recognition of the inherent subjectivity and lack of specificity in the symptoms evaluated by behavioural rating scales. Current research indicates that stand-alone measures like the ASIS and the MARS exhibit greater accuracy compared to embedded measures. Although these stand-alone measures hold considerable promise, it's crucial to acknowledge that they are still in the process of refinement and development. Given this evolving landscape, we suggest utilizing the CAARS in conjunction with either the CII or the EI. This recommendation is made with the understanding that further research is necessary to optimize these emerging indices and augment their efficacy.

Discussion

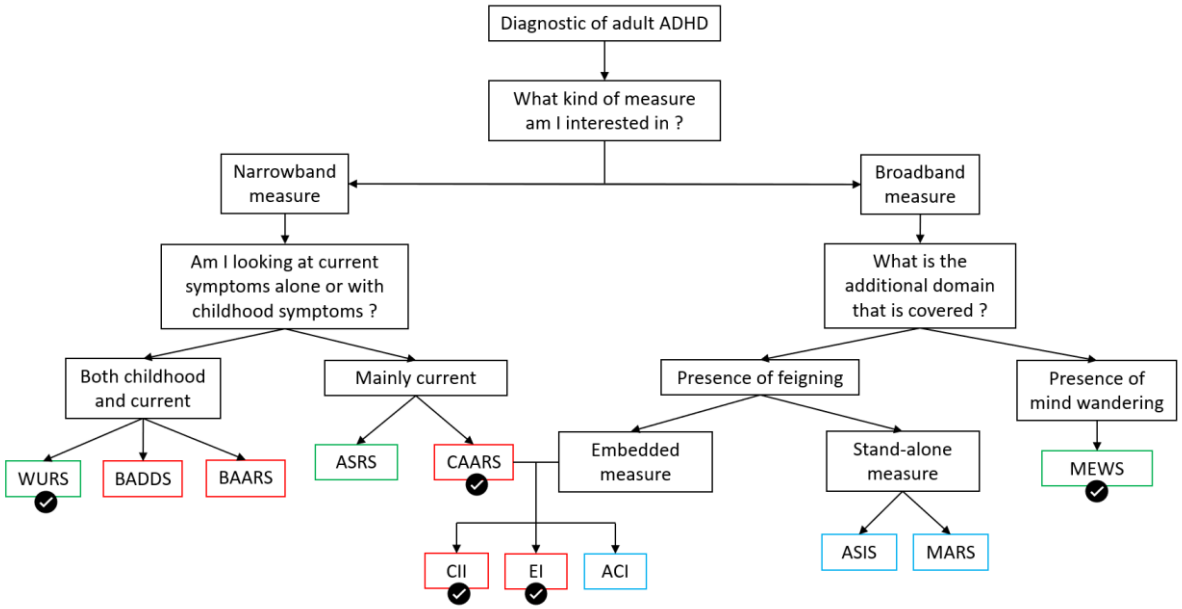
Recommendations for research settings

The examination of the existing literature on the use of behavioural rating scales for adult ADHD assessment revealed that they come in various forms and can evaluate different ranges of symptoms (see Figure 2). For example, the WURS primarily assesses childhood symptoms, whereas the CAARS assesses current symptoms and can be used to monitor symptom changes over time. Therefore, we emphasize that the selection of a behavioural rating scale should be based on the primary variable of interest or research objectives. Nonetheless, for a measure that is suitable across a wide range of contexts, we endorse the findings of previous reviews (Marshall et al., 2021; Taylor et al., 2011) and strongly recommend the utilisation of the CAARS. It should be completed collaboratively by both the patient and a pertinent informant, along with either the CII or the EI.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

We propose this scale for the following compelling reasons. Firstly, the literature has identified the CAARS as one of the most accurate measures available. Secondly, the CAARS is a comprehensive tool that assesses a broad range of ADHD symptoms beyond those outlined in the DSM-5 as opposed to the WURS or the ASRS. Thirdly, the CAARS is strengthened as a standalone instrument by its inclusion of validity indicators, which not only enhance its overall utility but also assist in identifying cases where symptoms may have been inaccurately presented. Finally, as previously mentioned, the CAARS is particularly advantageous for studies assessing the efficacy of new ADHD treatments due to its ability to evaluate the current presence and severity of ADHD symptoms. Unlike the WURS, which focuses on historical symptoms, the CAARS can monitor changes in symptoms over time, allowing for a comprehensive assessment of treatment effectiveness. Although the ASIS and MARS are more effective at detecting invalid symptom presentation as well as feigning, they are still underdevelopment and requires independent validation of their effectiveness by other researchers. Accordingly, the validity indicators of the CAARS remain the most optimal choice thus far.

Figure 2. Tree-graph representing behavioural rating scales used in the assessment of adult ADHD.



Note. The measures are highlighted in green if they are publicly available, in red if they are commercially available and in blue if they are still under development. Note that the CII and EI are in red because they rely on the utilisation of the CAARS. The validation mark emphasizes the rating scales that we recommend based on their accuracy. Narrowband measures refer to scales that focus solely on evaluating specific symptoms associated with ADHD. Broadband measures are assessing a wider range of behaviours. WURS = Wender Utah Rating Scale. ASRS = ADHD self-report scale. CAARS =

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Conner's Adult ADHD Rating Scales. CII = CAARS Infrequency Index. ACI = ADHD Credibility Index. ASIS = ADHD Symptom Infrequency Scale. BAARS-IV = Barkley Adult ADHD Rating Scale-Fourth Edition. EI = Exaggeration Index. MEWS = Mind Excessively Wandering Scale. MARS = Multidimensional ADHD Rating Scale.

Strength and limitations

When interpreting the findings of this review, it is important to consider both its strengths and limitations. One potential limitation is that only the Medline database was searched, and the search only retrieved electronic articles. As a result, there's a possibility of publication bias due to the lack of other sources of information. Another constraint of this paper lies in the instability observed in the base rates of PPV and NPV among the studies reviewed. The variations in these base rates made it challenging to establish a meaningful direct comparison between them and might have led to overestimation of their accuracy. While this paper aimed to update and expand the results of Marshall et al. (2021), we decided not to control for the presence of a clinical group. Therefore, some of the experiments reviewed may not accurately reflect the use of rating scales in a clinical environment, where patients may present other clinical diagnoses that are comorbid with ADHD. Moreover, our decision to apply inclusion criteria that were more flexible than those of Marshall et al. (2021) is a distinct strength of our approach. This choice not only enabled a comprehensive review of a greater number of papers, but also aligned with the research-oriented nature of our focus. Certainly, we recognize that researchers frequently give precedence to evaluating symptoms rather than exclusively concentrating on achieving a diagnosis, as typically seen in clinical contexts. Another strength is our focus on scales detecting feigned or invalid symptom presentation which were often overlooked in prior reviews.

Conclusion

The objective of this paper was to provide updated insights for researchers by examining recent developments in the use of behavioural rating scales to assess adult ADHD for research purposes. The growing body of research on these scales has raised concerns about their susceptibility to feigning, highlighting the need to consider the objectives of the ADHD assessment when choosing rating scales. While standalone measures may be more appropriate for research settings, combining different measures can enhance diagnostic accuracy in clinical settings. To enhance the practical applicability of feigning measures, further research should involve clinical groups to validate their effectiveness in settings that

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

closely reflects the clinical reality. Future studies should also systematically report sensitivity and specificity along with PPV and NPV to make to facilitate their comparison with existing measures.

Funding details.

The work was supported by the MRC Impact Acceleration Grant to Roi Cohen Kadosh.

Disclosure statement.

R. Mukherjee has delivered presentations for several pharmaceutical companies specializing in ADHD, whereby the funds generated were specifically allocated to the neurodevelopmental teams, with no direct financial compensation received by the author. The remaining authors declare no conflicts of interest.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

References

Adler, L. A., Spencer, T., Faraone, S. V., Kessler, R. C., Howes, M. J., Biederman, J., & Secnik, K. (2006). Validity of Pilot Adult ADHD Self- Report Scale (ASRS) to Rate Adult ADHD Symptoms. *Annals of Clinical Psychiatry*, *18*(3), 145-148. https://doi.org/10.3109/10401230600801077

Allenby, C., Falcone, M., Bernardo, L., Wileyto, E. P., Rostain, A., Ramsay, J. R., Lerman, C., & Loughead, J. (2018). Transcranial direct current brain stimulation decreases impulsivity in ADHD. *Brain Stimulation*, *11*(5), 974-981. https://doi.org/10.1016/j.brs.2018.04.016

Alyagon, U., Shahar, H., Hadar, A., Barnea-Ygael, N., Lazarovits, A., Shalev, H., & Zangen, A. (2020). Alleviation of ADHD symptoms by non-invasive right prefrontal stimulation is correlated with EEG activity. *NeuroImage : Clinical*, *26*, 102206. https://doi.org/10.1016/j.nicl.2020.102206

American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders*. DSM Library. https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425787

Anker, E., Bendiksen, B., & Heir, T. (2018). Comorbid psychiatric disorders in a clinical sample of adults with ADHD, and associations with education, work and social characteristics : A cross-sectional study. *BMJ Open*, *8*(3), e019700. https://doi.org/10.1136/bmjopen-2017-019700

Baggio, S., Hasler, R., Giacomini, V., El-Masri, H., Weibel, S., Perroud, N., & Deiber, M.-P. (2020). Does the Continuous Performance Test Predict ADHD Symptoms Severity and ADHD Presentation in Adults? *Journal of Attention Disorders*, *24*(6), 840-848. https://doi.org/10.1177/1087054718822060

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions : Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*(1), 65-94. https://doi.org/10.1037/0033-2909.121.1.65

Barkley, R. A. (2011). *Barkley Adult ADHD Rating Scale-IV (BAARS-IV)* (p. x, 150). The Guilford Press.

Becke, M., Tucha, L., Weisbrod, M., Aschenbrenner, S., Tucha, O., & Fuermaier, A. B. M. (2021). Non-credible symptom report in the clinical evaluation of adult ADHD : Development and initial validation of a new validity index embedded in the Conners' adult ADHD rating scales.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Journal of Neural Transmission (Vienna, Austria: 1996), *128*(7), 1045-1063. https://doi.org/10.1007/s00702-021-02318-y

Becke, M., Tucha, L., Weisbrod, M., Aschenbrenner, S., Tucha, O., & Fuermaier, A. B. M. (2022). Joint Consideration of Validity Indicators Embedded in Conners' Adult ADHD Rating Scales (CAARS). *Psychological Injury and Law*, *15*(2), 172-188. https://doi.org/10.1007/s12207-022-09445-1

Berger, I., Dakwar-Kawar, O., Grossman, E., Nahum, M., & Cohen Kadosh, R. (2021). Scaffolding the Attention-Deficit/Hyperactivity Disorder Brain Using Transcranial Direct Current and Random Noise Stimulation : A Randomized Controlled Trial. *Clinical Neurophysiology*. https://www.neuroscience.ox.ac.uk/publications/1152606

Brevik, E. J., Lundervold, A. J., Haavik, J., & Posserud, M.-B. (2020). Validity and accuracy of the Adult Attention-Deficit/Hyperactivity Disorder (ADHD) Self-Report Scale (ASRS) and the Wender Utah Rating Scale (WURS) symptom checklists in discriminating between adults with and without ADHD. *Brain and Behavior*, *10*(6), e01605. https://doi.org/10.1002/brb3.1605

Brown, T. E. (1996). Brown attention-deficit disorder scales. *TX: Psychological Corporation.*

Conners, C. K., Erhardt, D., & Sparrow, M. A. (1999). *Conner's Adult ADHD Rating Scales (CAARS).* New York : Multihealth Systems, Inc.

Courrégé, S. C., Skeel, R. L., Feder, A. H., & Boress, K. S. (2019). The ADHD Symptom Infrequency Scale (ASIS) : A novel measure designed to detect adult ADHD simulators. *Psychological Assessment*, *31*, 851-860. https://doi.org/10.1037/pas0000706

Dunlop, B. W., Wu, R., & Helms, K. (2018). Performance of the Adult ADHD Self-Report Scale-v1.1 in Adults with Major Depressive Disorder. *Behavioral Sciences (Basel, Switzerland)*, *8*(4), 37. https://doi.org/10.3390/bs8040037

Epstein, J. N., Erkanli, A., Conners, C. K., Klaric, J., Costello, J. E., & Angold, A. (2003). Relations Between Continuous Performance Test Performance Measures and ADHD Behaviors. *Journal of Abnormal Child Psychology*, *31*(5), 543-554. https://doi.org/10.1023/A:1025405216339

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Gift, T. E., Reimherr, M. L., Marchant, B. K., Steans, T. A., & Reimherr, F. W. (2021). Wender Utah Rating Scale : Psychometrics, clinical utility and implications regarding the elements of ADHD. *Journal of Psychiatric Research*, *135*, 181-188. https://doi.org/10.1016/j.jpsychires.2021.01.013

Harrison, A. G., & Armstrong, I. T. (2016). Development of a symptom validity index to assist in identifying ADHD symptom exaggeration or feigning. *The Clinical Neuropsychologist*, *30*(2), 265-283. https://doi.org/10.1080/13854046.2016.1154188

Harrison, A. G., & Edwards, M. J. (2023). The Ability of Self-Report Methods to Accurately Diagnose Attention Deficit Hyperactivity Disorder : A Systematic Review. *Journal of Attention Disorders*, 10870547231177470. https://doi.org/10.1177/10870547231177470

Harrison, A. G., Harrison, K. A., & Armstrong, I. T. (2022). Discriminating malingered attention Deficit Hyperactivity Disorder from genuine symptom reporting using novel Personality Assessment Inventory validity measures. *Applied Neuropsychology: Adult*, *29*(1), 10-22. https://doi.org/10.1080/23279095.2019.1702043

Harrison, A. G., Nay, S., & Armstrong, I. T. (2019). Diagnostic Accuracy of the Conners' Adult ADHD Rating Scale in a Postsecondary Population. *Journal of Attention Disorders*, *23*(14), 1829-1837. https://doi.org/10.1177/1087054715625299

Hinshaw, S. P., & Scheffler, R. M. (2014). *The ADHD explosion : Myths, medication, money, and today's push for performance* (p. xxxii, 254). Oxford University Press.

Ivnik, R. J., Smith, G. E., Cerhan, J. H., Boeve, B. F., Tangalos, E. G., & Petersen, R. C. (2001). Understanding the diagnostic capabilities of cognitive tests. *The Clinical Neuropsychologist*, *15*(1), 114-124. https://doi.org/10.1076/clin.15.1.114.1904

Jachimowicz, G., & Geiselman, R. E. (2004). *Comparison of Ease of Falsification of Attention Deficit Hyperactivity Disorder Diagnosis Using Standard Behavioral Rating Scales*.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health Organization adult ADHD

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

self-report scale (ASRS) : A short screening scale for use in the general population. *Psychological Medicine*, *35*(2), 245-256. https://doi.org/10.1017/S0033291704002892

Kessler, R. C., Adler, L., Barkley, R., Biederman, J., Conners, C. K., Demler, O., Faraone, S. V., Greenhill, L. L., Howes, M. J., Secnik, K., Spencer, T., Ustun, T. B., Walters, E. E., & Zaslavsky, A. M. (2006). The prevalence and correlates of adult ADHD in the United States : Results from the National Comorbidity Survey Replication. *The American Journal of Psychiatry*, *163*(4), 716-723. https://doi.org/10.1176/ajp.2006.163.4.716

Lange, R. T., & Lippa, S. M. (2017). Sensitivity and specificity should never be interpreted in isolation without consideration of other clinical utility metrics. *The Clinical Neuropsychologist*, *31*(6-7), 1015-1028. https://doi.org/10.1080/13854046.2017.1335438

Lee Booksh, R., Pella, R. D., Singh, A. N., & Drew Gouvier, W. (2010). Ability of College Students to Simulate ADHD on Objective Measures of Attention. *Journal of Attention Disorders*, *13*(4), 325-338. https://doi.org/10.1177/1087054708329927

Luty, J., Rajagopal Arokiadass, S. M., Sarkhel, A., Easow, J. M., Desai, R., Moorti, O. P., & El Hindy, N. (2009). Validation of self-report instruments to assess attention deficit hyperactivity disorder symptoms in adults attending community drug and alcohol services. *Journal of Addiction Medicine*, *3*(3), 151-154. https://doi.org/10.1097/ADM.0b013e31819343d0

Marshall, P., Hoelzle, J., & Nikolas, M. (2021). Diagnosing Attention-Deficit/Hyperactivity Disorder (ADHD) in young adults : A qualitative review of the utility of assessment measures and recommendations for improving the diagnostic process. *The Clinical Neuropsychologist*, *35*(1), 165-198. https://doi.org/10.1080/13854046.2019.1696409

McCann, B. S., Scheele, L., Ward, N., & Roy-Byrne, P. (2000). Discriminant Validity of the Wender Utah Rating Scale for Attention-Deficit/Hyperactivity Disorder in Adults. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *12*(2), 240-245. https://doi.org/10.1176/jnp.12.2.240

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

McGough, J. J., Sturm, A., Cowen, J., Tung, K., Salgari, G. C., Leuchter, A. F., Cook, I. A., Sugar, C. A., & Loo, S. K. (2019). Double-Blind, Sham-Controlled, Pilot Study of Trigeminal Nerve Stimulation for Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(4), 403-411.e3. https://doi.org/10.1016/j.jaac.2018.11.013

Mowlem, F. D., Skirrow, C., Reid, P., Maltezos, S., Nijjar, S. K., Merwood, A., Barker, E., Cooper, R., Kuntsi, J., & Asherson, P. (2019). Validation of the Mind Excessively Wandering Scale and the Relationship of Mind Wandering to Impairment in Adult ADHD. *Journal of Attention Disorders*, *23*(6), 624-634. https://doi.org/10.1177/1087054716651927

Murphy, K. R., & Gordon, M. (1998). Assessment of adults with ADHD. In *Attention-deficit hyperactivity disorder : A handbook for diagnosis and treatment, 2nd ed* (p. 345-369). The Guilford Press.

Nahum, M., Dakwar-Kawar, O., Mairon, N., Hochman, S., Berger, I., & Cohen-Kadosh, R. (2023). *Transcranial Random Noise Stimulation combined with Cognitive Training for Treating ADHD : A Randomized, Sham-Controlled Clinical Trial*. https://doi.org/10.21203/rs.3.rs-2550466/v1

Ogundele, M. O., Ayyash, H. F., & Banerjee, S. (2011). Role of computerised continuous performance task tests in ADHD. *Progress in Neurology and Psychiatry*, *15*(3), 8-13. https://doi.org/10.1002/pnp.198

Paz, Y., Friedwald, K., Levkovitz, Y., Zangen, A., Alyagon, U., Nitzan, U., Segev, A., Maoz, H., Koubi, M., & Bloch, Y. (2018). Randomised sham-controlled study of high-frequency bilateral deep transcranial magnetic stimulation (dTMS) to treat adult attention hyperactive disorder (ADHD) : Negative results. *The World Journal of Biological Psychiatry*, *19*(7), 561-566. https://doi.org/10.1080/15622975.2017.1282170

Pettersson, R., Söderström, S., & Nilsson, K. W. (2018). Diagnosing ADHD in Adults : An Examination of the Discriminative Validity of Neuropsychological Tests and Diagnostic Assessment Instruments. *Journal of Attention Disorders*, *22*(11), 1019-1031. https://doi.org/10.1177/1087054715618788

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Potts, H. E., Lewandowski, L. J., & Lovett, B. J. (2021). The Multidimensional ADHD Rating Scale : A measure of symptoms, impairment, and symptom validity. *Journal of Clinical and Experimental Neuropsychology*, *43*(4), 426-436. https://doi.org/10.1080/13803395.2021.1942795

Potts, H. E., Lewandowski, L. J., & Lovett, B. J. (2022). Identifying Feigned ADHD in College Students : Comparing the Multidimensional ADHD Rating Scale to Established Validity Measures. *Journal of Attention Disorders*, *26*(12), 1622-1630. https://doi.org/10.1177/10870547221092095

Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology*, *18*(4), 379-395. https://doi.org/10.1093/arclin/18.4.379

Ramsay, J. R. (2015). Psychological assessment of adults with ADHD. In *Attention-deficit hyperactivity disorder : A handbook for diagnosis and treatment, 4th ed* (p. 475-500). The Guilford Press.

Riccio, C. A., & Reynolds, C. R. (2001). Continuous Performance Tests Are Sensitive to ADHD in Adults but Lack Specificity. *Annals of the New York Academy of Sciences*, *931*(1), 113-139. https://doi.org/10.1111/j.1749-6632.2001.tb05776.x

Sayal, K., Prasad, V., Daley, D., Ford, T., & Coghill, D. (2018). ADHD in children and young people : Prevalence, care pathways, and service provision. *The Lancet Psychiatry*, *5*(2), 175-186. https://doi.org/10.1016/S2215-0366(17)30167-0

Skeel, R. L., Lesica, S., Fust, B., Garnett, A., & Bolen, L. (2022). Validation of an adult ADHD measure of feigning in a sample including individuals with depression and anxiety symptoms. *Applied Neuropsychology: Adult*, *0*(0), 1-7. https://doi.org/10.1080/23279095.2022.2158335

Smallwood, J., & Schooler, J. W. (2015). The Science of Mind Wandering : Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology*, *66*(1), 487-518. https://doi.org/10.1146/annurev-psych-010814-015331

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Söderström, S., Pettersson, R., & Nilsson, K. W. (2014). Quantitative and subjective behavioural aspects in the assessment of attention-deficit hyperactivity disorder (ADHD) in adults. *Nordic Journal of Psychiatry*, *68*(1), 30-37. https://doi.org/10.3109/08039488.2012.762940

Song, P., Zha, M., Yang, Q., Zhang, Y., Li, X., & Rudan, I. (2021). The prevalence of adult attention-deficit hyperactivity disorder : A global systematic review and meta-analysis. *Journal of Global Health*, *11*, 04009. https://doi.org/10.7189/jogh.11.04009

Suhr, J. A., Buelow, M., & Riddle, T. (2011). Development of an infrequency index for the CAARS. *Journal of Psychoeducational Assessment*, *29*, 160-170. https://doi.org/10.1177/0734282910380190

Suhr, J., & Wei, C. (2013). Symptoms as an Excuse : Attention Deficit/Hyperactivity Disorder Symptom Reporting as an Excuse for Cognitive Test Performance in the Context of Evaluative Threat. *Journal of Social and Clinical Psychology*, *32*(7), 753-769. https://doi.org/10.1521/jscp.2013.32.7.753

Taylor, A., Deb, S., & Unwin, G. (2011). Scales for the identification of adults with attention deficit hyperactivity disorder (ADHD) : A systematic review. *Research in Developmental Disabilities*, *32*(3), 924-938. https://doi.org/10.1016/j.ridd.2010.12.036

van de Glind, G., van den Brink, W., Koeter, M. W. J., Carpentier, P.-J., van Emmerik-van Oortmerssen, K., Kaye, S., Skutle, A., Bu, E.-T. H., Franck, J., Konstenius, M., Moggi, F., Dom, G., Verspreet, S., Demetrovics, Z., Kapitány-Fövény, M., Fatséas, M., Auriacombe, M., Schillinger, A., Seitz, A., … Levin, F. R. (2013). Validity of the Adult ADHD Self-Report Scale (ASRS) as a screener for adult ADHD in treatment seeking substance use disorder patients. *Drug and Alcohol Dependence*, *132*(3), 587-596. https://doi.org/10.1016/j.drugalcdep.2013.04.010

Vaughn, A. J., Epstein, J. N., Rausch, J., Altaye, M., Langberg, J., Newcorn, J. H., Hinshaw, S. P., Hechtman, L., Arnold, L. E., Swanson, J. M., & Wigal, T. (2011). Relation Between Outcomes

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

on a Continuous Performance Test and ADHD Symptoms Over Time. *Journal of Abnormal Child Psychology*, *39*(6), 853-864. https://doi.org/10.1007/s10802-011-9501-y

Ward, M. F., Wender, P. H., & Reimherr, F. W. (1993). The Wender Utah Rating Scale : An aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *The American Journal of Psychiatry*, *150*(6), 885-890. https://doi.org/10.1176/ajp.150.6.885

Weaver, L., Rostain, A. L., Mace, W., Akhtar, U., Moss, E., & O'Reardon, J. P. (2012). Transcranial Magnetic Stimulation (TMS) in the Treatment of Attention-Deficit/Hyperactivity Disorder in Adolescents and Young Adults : A Pilot Study. *The Journal of ECT*, *28*(2), 98. https://doi.org/10.1097/YCT.0b013e31824532c8

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Note

The MARS and the ASIS are still underdevelopment therefore not yet publicly nor commercially available. The WURS and ASRS are publicly available, the MEWS is available without charge by contacting philip.asherson@kcl.ac.uk. The ASRS, the CAARS and the WURS are commercially available.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Appendices

Appendix 1. Articles providing diagnostic accuracy for behaviour rating scale used in the assessment of adult ADHD.

Reference	Participants	C-G	P/N	Measure	Relevant Outcome
Becke et al., 2021	1001 credible neurotypical adult controls (median age 49), 22 overreporting neurotypical adult controls (median age 32), 100 credible adults with ADHD (median age 34), 22 non-credible adults with ADHD (median age 31,5) 242 adults instructed to simulate (median age 20).	No	Yes	WURS-25, ASRS, CAARS, CII, ACI, TOMM, GET.	The CAARS-ACI index was found to be useful in distinguishing genuine patients with ADHD from simulators and symptom over report but did not detect non-credible adults diagnosed with ADHD with adequate accuracy.
Brevik et al., 2020	646 adults with ADHD (mean age 34), 908 adult controls (mean age 29)	No	No	WURS, ASRS.	Both the WURS and the ASRS exhibit high diagnostic accuracy, with the WURS showing superior discriminatory properties.
Courrégé et al., 2019	Study 1: 30 adult controls, 17 adults thinking they have ADHD, 28 adults diagnosed with ADHD, 31 adult simulators (group mean age 33) Study 2: 42 adult controls, 31 adults thinking they have ADHD, 29 adults diagnosed with ADHD, 31 adult simulators (group mean age 35). Study 3: 48 adult controls, 33 adults thinking they have ADHD, 26 adults diagnosed with ADHD, 38 adult simulators (group mean age 36).	No	Yes	ASIS, BAARS-IV	The ASIS indicated strong internal consistency and convergent validity with the BAARS-IV. Moreover, the ASIS was found to be highly sensitive and specific to malingering.
Gift et al., 2021	137 adults with ADHD (mean age 31), 230 adults with major depression or	Yes	Yes	WURS, WURS-25	The full WURS was more effective in distinguishing ADHD from MDD and

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Reference	Participants	C-G	P/N	Measure	Relevant Outcome
	generalized anxiety disorder (MDD / GAD) (mean age 38), 120 adult controls (mean age 34)				GAD, compared to the WURS-25. On the other hand, the WURS-25 showed good separation of ADHD subjects from normal controls. However, the WURS performed better in differentiating ADHD patients from psychiatric controls, with higher sensitivity and specificity
Harrison et al., 2022	111 adults with ADHD (mean age 22), 66 clinical controls (mean age 23.8), 36 definite malingerers (mean age 23), 117 adults without diagnosis (mean age 23)	No	Yes	PAI, CII, EI, WMT, MSVT, TOMM, VSVT, TOVA.	The PAI algorithms proposed to identify feigned ADHD had low positive predictive value, hence they appear inadequate as symptom validity measures.
Mowlem et al., 2019	Study 1: 41 adults with ADHD (mean age 28), 47 adult controls (mean age 29). Study 2: 81 adults with ADHD (mean age 33), 30 adult controls (29)	No	No	MEWS, BRS, CAARS, ALS-SF, WFIRS-S	The MEWS has been found to be a dependable and credible tool to measure excessive mind wandering (MW) in adults diagnosed with ADHD. The study also revealed that individuals with ADHD reported substantially higher levels of MW when compared to their non-ADHD counterparts.
Skeel et al., 2022	101 adult controls, 99 adults with elevated DASS, 104 adults with diagnosed ADHD, 75 adult simulators (group mean age 42)	Yes	Yes	ASIS, BAARS-IV, DASS	The infrequency scale of the ASIS was deemed a trustworthy and effective measure for detecting ADHD feigning, even when individuals with symptoms of depression and anxiety were included in the sample.
Becke et al., 2022	856 adult controls (median age 50), 72 adults with ADHD (median age 35), 135 adult simulators (median age 39).	No	Yes	WURS-25, ASRS, CAARS, CII, ACI, TOMM, GET.	The combination of specific items from the CAARS Infrequency Index (CII) and the ADHD Credibility Index (ACI) in the CII-ACI-Compound index achieves a balance between sensitivity and specificity, but it does not

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Reference	Participants	C-G	P/N	Measure	Relevant Outcome
					show significantly improved results compared to the use of the ACI alone.
Potts et al., 2021	Study 1: 39 adults with ADHD, 56 malingering adults, 62 adult controls (group mean age 19) Study 2: 49 adults with ADHD, 68 malingering adults, 70 adult controls (group mean age 18)	No	Yes	MARS	The MARS indexes could detect individuals with ADHD at high sensitivity rates but had moderate specificity with regard to non-ADHD controls. The infrequency index was effective in identifying genuine ADHD from malingered ADHD.
Potts et al., 2022	34 adults with ADHD (mean age 18), 34 malingering adults (mean age 18)	No	Yes	MARS, CAT-A, WMT	The infrequency scale of the MARS had a higher sensitivity rate of with close to optimal specificity compared to other tests for detecting feigned ADHD.
Harrison et al., 2019	507 clinical controls (mean age 22) and 249 adults with ADHD (mean age 21)	Yes	Yes	CAARS	The CAARS is reliable for screening (specificity) but not diagnosing (sensitivity). It often misidentified individuals has having ADHD when they presented other psychological conditions.

Note. C-G = Clinical Group. "Yes" was attributed if there was the presence of a clinical group. P/N = predictive value. "Yes" was attributed if PPV & NVP were reported along with sensitivity and specificity. WURS = Wender Utah Rating Scale. WURS-25 = Wender Utah Rating Scale, Short-form. ASRS = ADHD self-report scale. CAARS = Conner's Adult ADHD Rating Scales. CII = CAARS Infrequency Index. ACI = ADHD Credibility Index. WMT = Word Memory Test. TOMM = Test of Memory Malingering, GET = Groningen Effort Test. ASIS = ADHD Symptom Infrequency Scale. BAARS-IV = Barkley Adult ADHD Rating Scale-Fourth Edition. PAI = Personality Assessment Inventory. EI = Exaggeration Index. MSVT = Medical Symptom Validity Test. VSVT = Victoria Symptom Validity Test. TOVA = Test of Variable of Attention. MEWS = Mind Excessively Wandering Scale. BRS = Barkley ADHD Rating Scale. ALF-SF = Affective Liability Scale-Short Form. WFIRS-S = Weiss Functional Impairment Rating Scale-Self Report. DASS = Depression Anxiety Stress Scales. MARS = Multidimensional ADHD Rating Scale. CAT-A = Clinical Assessment of Attention Deficit-Adult.

medRxiv preprint doi: https://doi.org/10.1101/2023.05.30.23290708; this version posted December 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Appendix 2. Psychometric statistics retrieved from the studies reviewed, arranged by measurement, and shorted by scale.

Measurement of ADHD						
Scale	**Sensitivity %**	**Specificity %**	**PPV**		**NPV**	
			Value %	**BR %**	**Value %**	**BR %**
WURS-61	84-95	93-94	88-94	.	84-91	.
WURS-25	62-90	85-94	73-88	.	79-91	.
ASRS	80	88	.	.	.	.
CAARS	64	86	51	.	71	.
MARS	86-92	58-67	60-65	38-40	86-92	38-40
MEWS	90	90	.	.	.	.
Measurement of Feigning						
Scale	**Sensitivity %**	**Specificity %**	**PPV**		**NPV**	
			Value %	**BR %**	**Value %**	**BR %**
ACI	30	98	69-95	10-50	58-92	10-50
CII	46	95	51-90	10-50	63-94	10-50
ACI-CII	41	87	15-87	5-50	60-67	5-50
ASIS	71-79	89-90	65-71	29	92-93	29
MARS	62-65	88-92	63-92	25-50	69-87	25-50
PAI Item-FAA	45	78	19	19	92	19
PAI Scale-FAA	36	79	17	19	91	19

Note. PPV = Positive predictive value. NPV = Negative predictive value. BR = base rate. WURS-61 = Wender Utah Rating Scale, Long form. WURS-25 = Wender Utah Rating Scale, Short-form. ASRS = ADHD self-report scale. CAARS = Conner's Adult ADHD Rating Scales. MARS = Multidimensional ADHD Rating Scale. MEWS = Mind Excessively Wandering Scale. ACI = ADHD Credibility Index. CII = CAARS Infrequency Index. ACI-CII = ACI-CII Compound index. ASIS = ADHD Symptom Infrequency Scale. PAI: Personality Assessment Inventory; Item-FAA: Feigned Adult ADHD index using PAI items; Scale-FAA: Feigned Adult ADHD index using PAI Scales.