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

Paul Marshall, James Hoelzle & Molly Nikolas

Routledge
Taylor & Francis Group

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

Paul Marshall[a], James Hoelzle[b] and Molly Nikolas[c]

[a]Department of Psychiatry, Hennepin Healthcare Systems, Minneapolis, MN, USA; [b]Department of Psychology, Marquette University, Milwaukee, WI, USA; [c]Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, USA

ABSTRACT

Objective: Identify assessment measures that augment the clinical interview and improve the diagnostic accuracy of adult ADHD assessment.

Method: The sometimes limited research literatures concerning the diagnostic efficacies of the clinical interview, standard and novel ADHD behavior rating scales, performance and symptom validity testing, and cognitive tests are critically reviewed.

Results: Based on this qualitative review, both clinical interviews alone and ADHD behavior rating scales alone have adequate sensitivity but poor specificity in diagnosing ADHD. Response validity and symptom validity tests have reasonably good sensitivity and very good specificity in detecting invalid symptom presentation. Cognitive test batteries have inadequate sensitivity and specificity in identifying ADHD. Using cognitive tests in conjunction with behavior rating scales significantly improves the specificity of an assessment battery. Executive function behavior rating scales and functional impairment rating scales are unlikely to improve the diagnostic accuracy of ADHD assessment.

Conclusions: Based on this review, key clinical interview questions, behavior rating scales, symptom validity tests, and cognitive tests that have promise to enhance current assessment practices are recommended. These are the authors' personal opinions, not consensus standards, or guidelines promulgated by any organization. These measures are incorporated in a practical, somewhat abbreviated, battery that has the potential to improve clinicians' ability to diagnose adult ADHD.

Prevalence estimates indicate ADHD affects approximately 5% of adults, an estimated 11 million adults in the United States (Barkley, Murphy, & Fischer, 2008). There is

CONTACT Paul Marshall ✉ pmarsh7247@gmail.com 🖃 Department of Psychiatry, Hennepin Healthcare Systems, Minneapolis, MN, USA

ample reason to be concerned ADHD may be being over-diagnosed in at least a sub-set of the adult population (Hinshaw & Scheffler, 2014; Paris, Bhat, & Thombs, 2015) though some opine adult ADHD is still greatly under recognized and under treated (Adler & Alperin, 2015; Kooij, 2013). The number of postsecondary students and young adults seeking evaluation for ADHD has risen remarkably as knowledge of the nature of the disorder and the benefits of being diagnosed with ADHD has increased (Weyandt & DuPaul, 2013). In line with this, according to IMS Health, the number of prescriptions written for ADHD medications for patients ages 20–39 increased approximately 280% between 2007 and 2012, from 5.6 million to almost 16 million (Schwarz, 2013).

There are many reasons for the striking increase in adult ADHD diagnoses. ADHD stimulant medications are used to increase attention, improve academic performance, lessen psychological distress, and lose weight, as well as for recreational purposes (Tucha, Fuermaier, Koerts, Groen, & Thome, 2015). Students with ADHD are eligible to receive academic accommodations (e.g., extended test taking time, tutoring, and alternative courses) that can improve their grades. While having ADHD is stigmatizing for many, for some individuals having ADHD provides a more acceptable excuse for their difficulties (Suhr & Wei, 2013). Countless advertisements tout the ability of medications to improve academic performance, ameliorate strained personal relationships, alleviate depression, and contribute to professional success (Hinshaw & Scheffler, 2014).

In addition, many young adults are likely to be seen for an ADHD evaluation by health professionals who do not have particular expertise in this diagnostic process. For example, by their own admission, most primary care physicians (PCPs) feel they have inadequate knowledge and training to diagnose ADHD. In fact, only 34% of 400 PCPs surveyed felt they were "very or extremely knowledgeable" about adult ADHD, and only 13% felt they had received "very or extremely thorough" clinical training in making this diagnosis. In addition, 44% thought the diagnostic criteria were not clear, 72% indicated it was easier to diagnose ADHD in children than adults, and 75% rated the quality and accuracy of current ADHD diagnostic measures as either "poor" or "fair." Furthermore, 85% reported they would take a more active role in making this diagnosis if they had an easy to use and quickly administered screening tool that was appropriately developed and validated (Adler, Shaw, Stitt, Maya, & Morrill, 2009). Overall, there appears to be a critical need to examine and refine the current practices used in the assessment and diagnosis of ADHD in adulthood.

The following judicious literature review will systematically consider components of a multi-modal ADHD assessment. Relevant research pertaining to the diagnostic issues and accuracies of clinical interviews, self-report measures, and neuropsychological tests will be critically examined. This review will incorporate recommendations that might improve each component of an adult ADHD diagnostic assessment.

Before presenting this review, the primary statistics used to elucidate the clinical utility of cognitive tests and other assessment measures will be briefly described. Sensitivity is the percentage of people who have a condition (e.g., ADHD) that are predicted by the test/measure to have it or, to put it another way, the probability that the test/measure correctly identifies the presence of the condition. Specificity is the percentage of people that do not have the condition who are predicted by the

Table 1. Qualitative descriptors of sensitivity and specificity.

Sensitivity/Specificity 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

test to not have it; or, the probability that the test/measure correctly identifies the absence of the condition. Sensitivity and specificity statistics are useful in quantifying and comparing the diagnostic accuracy of different tests/measures. However, positive predictive power (PPP) and negative predictive power (NPP) are much more useful statistics in clinical decision making where the research findings regarding the diagnostic accuracy of the test/measure are applied to an individual patient. PPP statistics address the question, if the individual patient is identified by the test/assessment measure as having the condition, what is the probability the patient has the condition. NPP statistics address the question, if the individual patient is identified as not having the condition, what is the probability the patient does not have the condition (Ivnik et al., 2001).

Unlike sensitivity and specificity, calculation of PPP and NPP requires knowledge of the base rate of the condition (i.e., ADHD) in the population of interest (e.g., patients presenting for ADHD assessment). The clinician is most interested in the potential usefulness of a test/measure in making a diagnosis in a particular clinical setting. However, the clinical utility of a test/measure identified in a specific study will not be the same as in their clinical setting if the base rate of the condition differs across settings. Lange and Lippa (2017) have reviewed the complexities of using test/measure diagnostic accuracy statistics in a clinical setting. They have argued persuasively that the sensitivity and specificity of a test/measure in a clinical setting should not be interpreted in isolation but rather in the context of other diagnostic accuracy statistics including PPP and NPP.

Unfortunately, however, while the research studies reviewed in this manuscript routinely report sensitivity and specificity statistics, the majority do not report PPP, NPP, and other diagnostic accuracy statistics. Further, they do not consistently report sufficient data and other variables that would be required to conduct a meta-analysis. Consequently, only sensitivity and specificity findings are reported to provide at least some means of comparing the diagnostic accuracy of different test/assessment measures. Finally, Lange and Lippa (2017) provide the following recommended qualitative descriptors of the clinical utility of a test/measure based on its sensitivity and specificity (see Table 1).

Method

A systematic literature search was executed using the Medline and PsychInfo databases from 1998 through June 2019. To identify potentially relevant literature in the electronic database, we used the following search terms: "ADHD or attention deficit
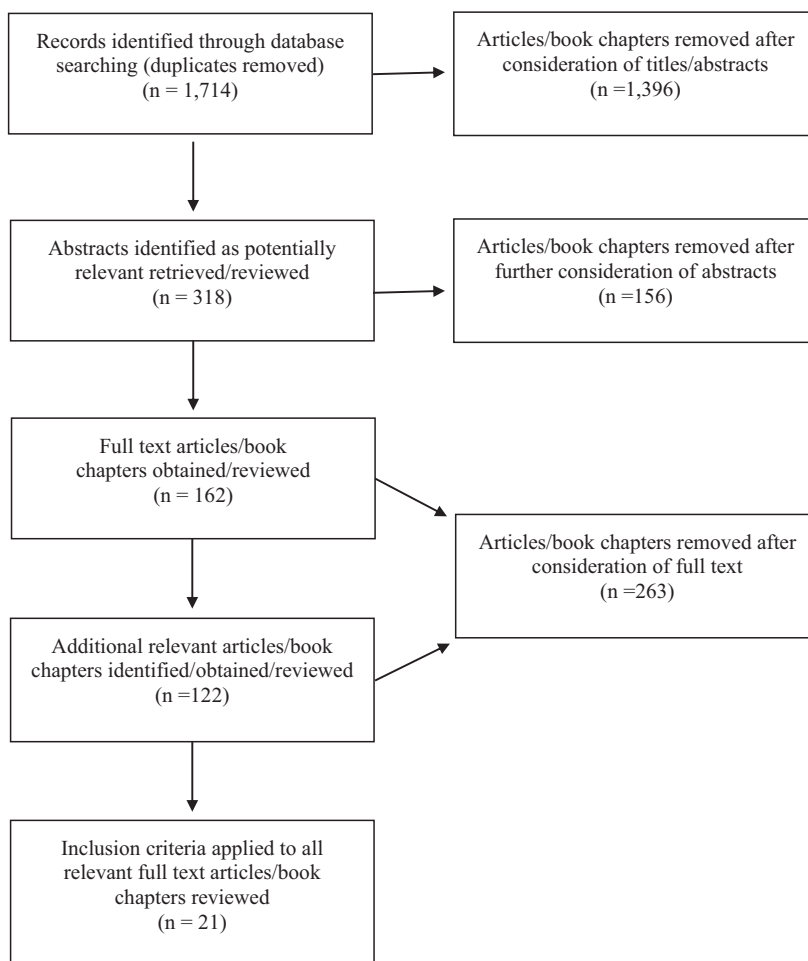
Figure 1. Flow diagram describing literature search and article selection.

hyperactivity disorder" AND "assessment or testing or evaluation" AND "adult" AND "diagnosis". Articles that were identified as electronic publications online were eligible for inclusion in this review.

The initial electronic database search identified 1,714 abstracts of journal articles and book chapter titles after duplicates were removed. These were all reviewed by the first author. The 318 abstracts that appeared potentially relevant to the assessment of *adult* ADHD were then retrieved and read by both the first and second authors. After this review was completed, the full text of 162 journal articles and book chapters whose abstracts suggested they were relevant –most of which had previously been obtained – were read. The bibliographies and citations of these journal articles and book chapters were also scrutinized for potentially relevant articles. As a result, the full text of an additional 122 articles were obtained and reviewed. The final phase of this literature search focused more narrowly on identifying articles that met the inclusion criteria (see Figure 1; Table 2). A summary of those 21 studies is presented in Table 3.

Table 2. Inclusion criteria.

-Published in English peer reviewed journals. -Participants are adults, age18 or older.
-ADHD is diagnosed using DSM criteria.
-ADHD diagnosed by widely used clinical or research semi-structured or structured interview.
-Group study investigating interviews, behavior rating scales, and/or neuropsychological tests for diagnosis,
   screening, or identification of ADHD.
-Comparison groups –adults diagnosed with ADHD versus control participants and/or participants diagnosed with
   psychiatric disorders.
-ADHD behavior rating scales and/or structured clinical interviews are independently validated.
-Neuropsychological tests are standardized and have normative data.
-Neuropsychological tests, behavior rating scales, and interviews are commercially available or in public domain.
-Results provide diagnostic accuracy statistics, at minimum sensitivity and specificity.

Articles not meeting inclusion criteria most commonly failed to report diagnostic classification statistics associated with tests and measures utilized during an adult ADHD assessment.

Grades of Recommendation, Assessment, Development, and Evaluation Working Group (GRADE; Ryan & Hill, 2016) guidelines were considered when determining the quality of studies included (see Table 3). Initial study ratings are based on study design (i.e., randomized clinical trials are assumed to be of higher quality than case studies) and are subsequently adjusted due to potential risk of bias, imprecision, inconsistency, indirectness, and publication bias. Each study reviewed was a cross-sectional design, and the most salient factor to consider when determining the quality of a specific study was whether the study included a clinical control group. Most commonly, the quality of studies that utilized a non-clinical control group were downgraded because the diagnostic utility of measures included in these studies have the potential to be somewhat inflated (i.e., it is easier to differentiate between clinical and control groups relative to differentiating between two clinical groups).

Clinical interview

Diagnostic issues

American Psychiatric Association guidelines stipulate the diagnosis of ADHD is to be made by conducting a thorough clinical interview and administering ADHD behavior rating scales (Hauk, 2013). There is no clearly defined "gold standard interview" for diagnosing adult ADHD (Haavik, Halmoy, Lundervold, & Fasmer, 2010). Nevertheless, this assessment process typically begins with a clinical interview which seeks to determine the presence of the core symptoms of adult ADHD and how these symptoms impact the patient's daily life. This usually involves asking the patient to provide examples of how these core symptoms have affected their social relationships as well as other daily activities across multiple settings (e.g., school, home, and the workplace).

The clinical interview also includes a review of the patient's family, developmental, medical, and psychiatric history. Symptoms of other disorders that might account for ADHD symptoms need to be ruled out, particularly depression and anxiety disorders, even though these conditions are often comorbid. Notably adults meeting diagnostic criteria for ADHD combined or inattentive types have higher rates of comorbidity relative to adults experiencing only hyperactive symptoms (Friedrichs, Igl, Larsson, & Larsson, 2012). Additionally, individuals with ADHD have increased risk of depression

Table 3. Articles providing diagnostic accuracy statistics of assessment instruments in diagnosing adult ADHD.

Reference (GRADE)	Participants	Measure	Relevant Outcome
Barkley & Murphy, 2011 (III)	146 adults diagnosed with ADHD (mean age 32); 97 adults evaluated but not diagnosed with ADHD (mean age 38); 109 adult volunteers from community (mean age 36)	Structured Clinical Interview for ADHD; Adult ADHD Symptoms Scale self-report form; Deficits in Executive Functions Scale, Symptom Checklist 90-Revised	ADHD group had more severe ratings than the Clinical and Community control groups on all five executive function dimensions.
Dunlop et al., 2018 (III)	50 adult patients with primary psychiatric diagnosis of MDD (mean age 49), 55 healthy adult control subjects (mean age 44)	Structured Interview for DSM-IV, Mini-International Neuropsychiatric Interview Plus Adult ADHD Module, ASRS, Hamilton Depression Rating Scale	The ASRS had a positive predictive value of 21% and negative predictive value of 92%. suggesting a negative ASRS screen is effective in ruling ADHD out.
Dvorsky et al., 2016 (IV)	59 adults diagnosed with ADHD (mean age 20), 27 adults not diagnosed with ADHD (mean age 21)	CAADID, BASC-2, BAARS-IV self- and informant-report forms	Parental ratings of childhood ADHD symptoms of inattention were better than students' childhood and current ratings of ADHD symptoms in predicting their current ADHD diagnosis. Parental ratings of ADHD symptoms achieved acceptable thresholds for confirming and ruling out a diagnosis of ADHD.
Edebol et al., 2012 (III)	53 clinically diagnosed adults with ADHD (mean age 36), 45 clinically diagnosed adults with bipolar or borderline personality disorder (mean age 42), 29 adults clinically evaluated but not diagnosed with ADHD (mean age 35)	Qb Test- Plus, GAF, ASRS self-report form	The QbTest-Plus separated those diagnosed with ADHD, not diagnosed with ADHD, and diagnosed with bipolar or borderline personality disorder. Sensitivity of the QbTest-Plus was lower in the groups with bipolar disorder II, borderline personality disorder, and with disconfirmed diagnosis of ADHD.
Edebol et al., 2013 (IV)	55 adults clinically diagnosed with ADHD (mean age 33), 202 adults recruited from the general population screened for not having ADHD (mean age 31)	QbTest-Plus, GAF, ASRS self-report form	QbTest-Plus hyperactivity/motion tracking was the most specific symptom of ADHD. QbTest- Plus CPT inattention and impulsivity were the most and least sensitive symptoms of ADHD respectively.
Emser et al., 2018 (IV)	38 adults clinically diagnosed with ADHD (mean age 35) and 38 adult controls (mean age 32)	QbTest-Plus, CAARS self- and informant-report forms, TAP go/no go, divided attention, and sustained attention subtests	The QbTest-Plus and TAP subtests were less helpful in predicting an ADHD diagnosis than the CAARS forms. Omission errors on the TAP and the CAARS inattention symptom scale were the variables most predictive of diagnosing ADHD.
Harrison et al., 2019 (III)	249 adults clinically diagnosed with ADHD (mean age 21), 507 adults clinically assessed for ADHD or academic learning problems but not diagnosed with ADHD (mean age 22).	CAARS self-report form	The CAARS self-report form was neither sufficiently sensitive nor specific for diagnosing adult ADHD.

Table 3. Continued.

Reference (GRADE)	Participants	Measure	Relevant Outcome
Hirsch & Christiansen, 2017 (III)	206 adults clinically assessed and diagnosed with ADHD (mean age 32), 91 adults clinically assessed and not diagnosed with ADHD (mean age 32)	QbTest-Plus, CAARS self- and observer-report forms, TAP go/no go, divided attention, and sustained attention subtests, WURS, AKGT, ADHD-SB	The QbTest-Plus omission errors and reaction time variability measures were effective in discriminating between participants with and without ADHD.
Holst & Thorell, 2017 (III)	57 adults diagnosed with ADHD (mean age 26), 53 adults not diagnosed with ADHD (mean age 26)	DIVA, ASRS self-report form, WAIS Letter Number Sequencing and Digit Span subtests, CANTAB Finding the Phone Task subtests, DKEFS Color-Word Interference, Sorting, Tower, and Verbal Fluency subtests, Quick Delay Questionnaire	The ADHD group performed significantly more poorly in each six neuropsychological domains. Most group differences in testing were in the small range with exception of medium effect sizes for one measure in each of the verbal working memory, inhibition, fluency, and delay aversion domains.
Kamradt et al., 2014 (IV)	170 adults diagnosed with ADHD (mean age 24), 83 adults not diagnosed with ADHD (mean age 20)	A semi-structured diagnostic interview based on Adult ADHD Clinical Diagnostic Scale and work by Kessler et. al (2010), BAARS-IV self- and informant-reports, BFIS self- and informant-reports, BDEFS self- and informant-reports, WRAT, WAIS Digit Span subtest, DKEFS Color-Word Interference and Trail-Making subtests, Stop Task, Conners Continuous Performance Test (CPT)	Executive function related test variables predicted ADHD behavior rating scale symptom dimensions and related functional impairments. A combination of results from executive function tests and behavior rating scales have the greatest sensitivity in diagnosing ADHD.
Katz et al., 1998 (III)	58 adults diagnosed with only ADHD and/or diagnosed with ADHD and a co-morbid condition (mean age 29), 20 adults diagnosed with only major depression or dysthymia (mean age 35)	California Verbal Learning Test (CVLT), Paced Auditory Serial Addition Test (PASAT), Stroop Test, WAIS-Revised, Wechsler Memory Scale -Revised, and Conners CPT, Halstead- Reitan Speech Perception, Seashore Rhythm, Finger Tip Number Writing, Tactual Performance, and Trail-Making Test	A discriminate function equation including PASAT, CVLT, and Stroop test variables differentiated patients diagnosed only with ADHD or with a co-morbid disorder and patients diagnosed only with major depression or dysthymia.
Lovejoy et al., 1999 (IV)	26 adults clinically diagnosed with ADHD (median age 41) and 26 normal adult controls (median age 41)	Stroop Neuropsychological Screening Test, TMT, CVLT, Controlled Oral Word Association (COWA) test, WAIS-R Arithmetic and Digit Span subtests	Test measures considered separately had good specificity but inadequate sensitivity in diagnosing ADHD. A Summary Index score indicating performances on one or more tests were impaired had excellent sensitivity and reasonably good specificity.
Luty et al., 2009 (III)	107 adult patients undergoing substance abuse treatment, 42 subsequently diagnosed	WURS adult ADHD screening test, WHO Adult ADHD Self-report screener, CAARS	The CAARS was the most accurate of the three scales in diagnosing adult ADHD.

Table 3. Continued.

Reference (GRADE)	Participants	Measure	Relevant Outcome
	with ADHD and 65 not diagnosed with ADHD (group mean ages 38)		
Marshall et al., 2016 (III)	102 adults diagnosed with ADHD and 326 not diagnosed with ADHD (mean age both groups 26)	BAARS-IV self- and informant-reports, BFIS self- and informant-reports, BDEFS short form self- and informant-reports, the WAIS-IV Vocabulary, Block Design, Digit Span, Letter Number Sequencing (LNS), and Symbol Search subtests, Neuropsychological Assessment Battery Numbers and Letters Test, Word Memory Test, CVLT, PASAT, Salthouse Listening Span Test, Test of Variables of Attention (TOVA), Sentence Repetition Test, WMS-IV Spatial Addition subtest, Dot Counting Tests, b Test, and the DKEFS Color Word, Tower, Design Fluency, and Verbal Fluency subtests	A significant percentage of patients making a suspect effort will be diagnosed with ADHD using commonly used assessment methods: a clinical interview alone (71%), an interview and ADHD behavior rating scales combined (65%), and an interview, behavior rating scales, and most continuous performance tests combined (62%).
McCann et al., 2000 (III)	68 adults diagnosed with ADHD (mean age 34) and 73 not diagnosed with ADHD (mean age 38)	Structured clinical interview for DSM-IV, WURS	The WURS was sensitive in detecting ADHD but mis-classified about half of those who did not have ADHD.
Mostert et al., 2015 (IV)	133 adults clinically diagnosed with ADHD (mean age 35) and 132 normal adult controls (mean age 36)	DIVA, ADHD Rating Scale-IV self-report form, Digit Span, TMT, Verbal Fluency, Category Fluency, Delay Discounting, Time Estimation, Go/No go task, Sustained Attention Dots, Flanker task	The ADHD group exhibited impaired working memory and sustained attention, were more sensitive to delay aversion, and were variable in responding. However, the effect sizes of the individual tests in discriminating between the two groups was small to moderate (range: 0.05-0.70). A combination of the six most discriminating test variables still had inadequate sensitivity and specificity in differentiating the ADHD and control groups.
Nikolas et al., 2019 (III)	109 adults clinically diagnosed with ADHD (mean age 25), 52 adults diagnosed with depression (mean age 24), 85 healthy adult controls (mean age 23)	BAARS-IV self- and informant-reports, BFIS self- and informant-reports, BDEFS short form self- and informant-reports, WAIS-IV Vocabulary, Block Design, Digit Span, Letter Number Sequencing, and Symbol Search, Neuropsychological Assessment Battery Numbers and Letters	Measures of working memory, sustained attention, and response speed and variability best discriminated between the diagnosed with ADHD group and the depression and control groups combined. Single test measures performed poorly in identifying ADHD participants. However, a combined approach using self and informant ADHD symptoms ratings, family history of ADHD, and the TOVA reaction

Table 3. Continued.

Reference (GRADE)	Participants	Measure	Relevant Outcome
		Test, Word Memory Test, CVLT, PASAT, Salthouse Listening Span Test, Test of Variables of Attention, Sentence Repetition Test, WMS-IV Spatial Addition, Dot Counting Tests, b Test, DKEFS CWIT, Tower, Design Fluency, and Verbal Fluency	time variability measure correctly classified 87% of cases.
Pettersson et al., 2018 (IV)	60 adults clinically diagnosed with ADHD (mean age 28) and 48 adults clinically diagnosed without ADHD (mean age 33)	DIVA, ASRS self-report form, Beck Depression and Anxiety Inventories, Canadian Occupational Performance Measure, Rey Auditory Verbal Learning Test, PASAT, Conners CPT, the QbTest-Plus, the DKEFS Verbal Fluency and Trail Making Test, WAIS-IV Digit Span Backward and Coding	Each assessment measure exhibited poor discriminative ability except for the DIVA. A logistic regression analysis model including the DIVA and measures of inattention, impulsivity, and activity had similar sensitivity and better specificity than the DIVA alone.
Rapport, Van Voorhis, Tzelepis, & Friedman, 2001 (IV)	35 adults clinically diagnosed with ADHD and 32 adult controls (mean age of both groups 33)	Current Symptoms Scale self-report form, Gordon Diagnostic System CPT, Stroop test, Wisconsin Card Sorting Test, COWA, Design Fluency Test, TMT, Letter Number Sequencing, Rey-Osterreith Complex Figure, WAIS-R Vocabulary, Similarities, Picture Completion, and Picture Arrangement.	Patients with ADHD have specific deficits in response inhibition with intact abilities in other cognitive domains.
Soderstrom et al., 2014 (IV)	41 adults clinically diagnosed with ADHD (mean age 32) and 20 adults clinically diagnosed without ADHD (mean age 30)	ASRS self-report form, Current Symptoms Scale (CSS) self-report form, QbTest- Plus	The ASRS and CSS behavior rating scales had good sensitivity but poor specificity while variables from the QbTest-Plus exhibited poor sensitivity but good specificity. The discriminant validity of subjective ADHD self-rating scales alone and of an objective measure of sustained attention and motor activity alone is poor.
Solanto et al., 2004 (III)	70 adults clinically diagnosed with ADHD (mean age 35) and 33 adults clinically diagnosed with depressive or anxiety disorders (mean age 44)	CAARS self-report form, Brown Attention-Deficit Disorder Scale for Adults (BADDS) self- report form, Conners CPT	The BADDS and Conners CPT exhibited sufficient sensitivity and specificity in discriminating adults with ADHD from those with depressive or anxiety disorders.
Ustun et al., 2017 (IV)	193 adults undergoing ADHD evaluation (mean age 33) and 107 adult controls (mean age not specified)	ASRS self-report form screening scale, Adult ADHD Clinical Diagnostic Scale clinical interview	The ASRS screening scale identifies the vast majority of general population cases at a threshold that has high specificity and positive predictive validity.
van de Glind et al., 2013 (III)	136 adults seeking substance abuse treatment, 24 diagnosed	ASRS, CAAID	The ASRS is an effective screening tool missing few patients diagnosed with ADHD by clinical

(*continued*)

Table 3. Continued.

Reference (GRADE)	Participants	Measure	Relevant Outcome
	with ADHD and 112 diagnosed by interview as well as ADHD behavior rating scale (age range for all patients 18-65)		interview. However, it over diagnoses ADHD.
Walker et al., 2000 (III)	30 adults diagnosed with ADHD (mean age 26), 30 adults diagnosed with mood and/or anxiety disorders (mean age 35), 30 normal adult controls (mean age 26)	WURS, Beck Depression Inventory, State-Trait Anxiety Inventory, National Adult Reading test, Conners CPT, TMT, COWA, Animal Fluency, Stroop Test, WAIS-R Digit Span, Arithmetic, and Digit Symbol	The ADHD group performed below controls on most neuropsychological tests. The predictive power of the tests was poor in discriminating patients with ADHD from patients with mood and anxiety disorders.

Note. All study designs were cross sectional. GRADE = Grades of Recommendation, Assessment, Development, and Evaluation Working Group. ADHD-SB- ADHD Self Rating Scale, AKGT-Amsterdam Short-Term Memory Test, ASRS -Adult Self Report Scale for Adult ADHD, BASC-2 - Behavior Assessment System for Children, College Version, BAARS-IV - Barkley Adult ADHD Rating Scale-IV, BDEFS - Barkley Deficits in Executive Functioning Scale, BFIS - Barkley Functional Impairment Scale, CAAID - Conners Adult ADHD Diagnostic Interview for DSM-IV, CAARS -Conners ADHD Rating Scale, long version, CANTAB- Cambridge, Neuropsychological Test, DKEFS-Delis Kaplan Executive Function System, CWIT- Color-Word Interference Test, DIVA-Diagnostic Interview for ADHD in Adults, CWIT GAF-Global Functioning Scale, QbTest-Plus - Quantified Behavior Test Plus, TAP - Test of Attentional Performance, WAIS - Wechsler Adult Intelligence Scale, WMS- Wechsler Memory Scale, WRAT- Wide Range Achievement Test, WURS-Wenders Utah Rating Scale.

(odds ratio: 2.7, 38% prevalence), anxiety (odds ratio: 3.2, 47% prevalence), and substance abuse (odds ratio: 3.0, 15% prevalence) (Kessler et al., 2006). Making a differential diagnosis between adult ADHD and comorbid psychiatric disorders is one of the most perplexing issues a clinician can encounter. This requires taking an often-lengthy longitudinal psychiatric history in which the onset, course, and persistence of key symptoms and related impairments are clarified (Adler & Alperin, 2015).

While a clinical interview is still an essential part of any ADHD evaluation, it is very important and helpful to have an informant (i.e., parent, sibling, or significant other) present at an evaluation to corroborate the patient's history and complete an ADHD behavior rating scale. However, such input is often not sought (Pazol & Griggins, 2012) and including informants is often not realistic in conventional clinical practice due to patient's privacy concerns and clinician's time and budgeting limitations (Gorlin, Dalrymple, Chelminski, & Zimmerman, 2016). Ideally a clinician will also gather additional archival records that can document potentially ADHD-related symptoms (Ramsey, 2015). Unfortunately, report cards, teacher evaluations, and past psychological test results are frequently unavailable in adult ADHD assessments (Roy-Byrne et al., 1997).

There are several problems with the validity of a clinical interview. First, the validity of the interview depends upon the patient providing a reasonably accurate and insightful self-report of potentially ADHD related symptoms not only for their adulthood but also for their childhood retrospectively. The accuracy of many adults' report of their possible ADHD related childhood difficulties is compromised by their poor recall of their childhood experiences (Wender, 1997) as well as the lack of objective means of determining whether their childhood behaviors were consistent with ADHD and extreme or impaired compared to other children (Murphy, Gordon, & Barkley,

2000). Some adults exhibit a positive illusory bias and are simply unaware of having their symptoms and impairments (Prevatt et al., 2012). Other adults lack insight into the causes of their behavioral problems, do not understand the way in which ADHD symptoms appear in adulthood, and may attribute them to personality or character traits as well as depression and anxiety (Barkley & Brown, 2008). Many adolescents and young adults with a history of childhood ADHD, as well as their parents, are not accurate when reporting their earlier ADHD symptoms during a semi-structured clinical interview (Miller, Newcorn, & Halperin, 2010). In general, it is significantly more difficult to diagnose ADHD in adults than children. This is because adults are more likely to have comorbid psychiatric and medical conditions as well as to have experienced stressful or traumatic events that can cause symptoms that mimic ADHD. Furthermore, it is more difficult to determine significant impairment in adults in the workplace relative to children in school (Murphy & Gordon, 2006).

A second threat to the validity of the clinical interview is the non-specificity of the behavioral symptoms of ADHD. Problems with concentration, attention, and over-activity can be due to multiple etiologies. Additionally, ADHD-related symptoms are commonly reported by college students (Lewandowski, Lovett, Codding, & Gordon, 2008). ADHD symptoms are non-specific, often multi-determined, and can be due to other psychological factors. As a result, for example, Barkley and Murphy (2011) found 45% of patients referred for ADHD assessment reported having sufficient symptoms to meet DSM-IV criteria although they were subsequently not diagnosed with ADHD for various reasons. Based on these findings, it is clearly problematic to simply consider the number of ADHD symptoms reported when formulating a diagnosis.

There is also controversy regarding what specific symptoms are most appropriate for diagnosing adult ADHD. Some research has found the 18 DSM ADHD criteria symptoms are not the most effective in differentiating adults with and without ADHD. Barkley et al. (2008) as well Fedele, Hartung, Canu, and Wilkowski (2010) have identified ADHD related symptoms that better differentiate control groups and adults diagnosed with ADHD than the 18 DSM criteria symptoms.

DSM-5 stipulates the 18 ADHD diagnostic symptoms must occur "often" and "interfere with, or reduce the quality of social, academic, or occupational functioning". In addition, clinicians are to specify if symptoms result in "impairment" in social, academic, or occupational functioning (American Psychiatric Association, 2013). Unfortunately, as is the case with many psychiatric disorders, what constitutes experiencing symptoms "often" and their causing significant "impairment" is unclear. Making the diagnosis via a clinical interview assumes that, if a patient's ADHD symptoms are sufficiently severe, they should be able to describe their ADHD symptoms, how long they have had them, and how they have impacted their life. Yet, many adults have difficulty doing so due, in part, to lack of insight into their behavior (Barkley & Brown, 2008). Unfortunately, studies of structured clinical interviews have also found that interviewer and patient characteristics have a significant effect on diagnosis and the test-retest reliability of interviews is low to moderate (Vaughn & Hoza, 2013). For these as well as other reasons, the results of the clinical interview are often unclear in establishing the diagnosis of ADHD in young adults.

Completing the commonly recommended, open ended, comprehensive clinical interview usually requires a minimum of 1–2 hours (Murphy & Gordon, 2006). Even structured interviews such as the Conners Adult ADHD Diagnostic Interview for the DSM-IV (CAADID; Epstein, Johnson, & Conners, 2000) and the DIVA (the Diagnostic Interview for ADHD in Adults, Kooj and Francken, 2007) take approximately 180 and 90 minutes, respectively (Gorlin et al., 2016). Regrettably, however, the reality is that many patients are being diagnosed on the basis of "extremely cursory" evaluations (Hinshaw & Scheffler, 2014). For example, a survey found only 20% of 1,216 PCPs and 35% of 708 psychiatrists completed an extended clinical interview during their adult ADHD assessment process (Goodman, Surman, Scherer, Salinas, & Brown, 2012).

Diagnostic accuracy

While there is ample research on the diagnostic accuracy of behavior rating scales and cognitive tests, there is a dearth of research on the accuracy of the clinical interview. This is because the results of a clinical interview itself are the primary – if not the sole basis – for the "gold standard" diagnosis of the ADHD criterion group in most research. Pettersson, Soderstrom, and Nilsson (2018) found the aforementioned DIVA had a sensitivity of 90% and specificity of 73%, in a group of adult outpatients presenting for ADHD assessment. Marshall, Hoelzle, Heyerdahl, and Nelson (2016) found that, of 102 patients later diagnosed with ADHD based not only on the interview but additional assessment, 39% had an interview consistent with ADHD, 45% had an indeterminate interview, and 16% had an interview inconsistent with their having this disorder. Those patients with an inconsistent interview but still diagnosed with ADHD had results on multiple behavior rating scales and cognitive tests that provided compelling, substantive evidence of their having ADHD.

Recommendations

There are numerous fine discussions of how to conduct a "gold standard" comprehensive ADHD clinical interview (e.g., Murphy & Gordon, 2006; Ramsey, 2015). We recommend considering using some additional means of potentially improving the clinical interview process. Zimmerman and colleagues (Gorlin et al., 2016) have developed a semi-structured 18 DSM IV symptom based diagnostic clinical interview for ADHD. Gorlin and Zimmerman (personal communication) have found that it takes only 20–25 minutes to complete a reportedly effective diagnostic interview. Thus, it could be used when time constraints do not allow for conducting a "gold standard" interview. The interview was validated in a sample of 1,194 consecutive patients evaluated in an outpatient psychiatric clinic. This is an appropriate sample for as many as 80% of adult patients diagnosed with ADHD meet diagnostic criteria for at minimum at least one other psychiatric disorder (Barkley et al., 2008).

The clinical interview might also be improved by being particularly thorough in clarifying the patient's difficulties with specific ADHD symptoms that research has suggested are the most discriminating in diagnosing adult ADHD. In a cross-validation study of the aforementioned clinical interview, Zimmerman, Gorlin, Dalrymple, and

Chelminiski (2017) reported that the answers regarding two of the 18 DSM ADHD symptoms in their clinical interview should be given relatively more weight in diagnosing ADHD i.e., endorsing either "difficulty sustaining attention" or "fidgets and squirms". The patient's answers to the combination of the two symptoms had a sensitivity of 90.7% and negative predictive value of 97.4%. Given the fact that problems with sustaining attention are very commonly reported, the most useful finding is that a patient *not* endorsing having significant problems with sustaining attention *or* fidgeting and squirming effectively rules out their having ADHD.

Ustun et al. (2017) found that one question pertaining to the DSM-5 symptom inattention (i.e., does not listen when spoken to directly), three questions pertaining to DSM-5 symptoms of hyperactivity and impulsivity (i.e., leaves seat inappropriately, has difficulty playing quietly/leisure time, and blurts out answers), and two questions pertaining to non-DSM executive dysfunction symptoms (i.e., puts things off to the last minute, depends on others to keep their life in order) were the most discriminating in a large clinical sample. They also created a rating scale based on these six symptoms (total score range 0-24). They found that a cut off score ≥ 14 was most appropriate in using the scale as a screening instrument as it had a sensitivity of 91.9% and specificity of 74%. On the other hand, a cutoff score ≥ 17 had a sensitivity of 76.3% and specificity of 92.9% making it more useful in minimizing false positive diagnoses.

Finally, it is very important to explore the family history of ADHD. A review of pertinent studies by Frazier and Youngstrom (2006) found there is an approximately 4-5-fold increase in the likelihood a patient has ADHD when they have a first degree relative with this disorder. Similarly, Nikolas, Marshall, and Hoelzle (2019) found a patient having a first-degree family member with a history of ADHD had an odds ratio of 3.5 of having ADHD. Furthermore, they noted such a history significantly increased the classification accuracy of a regression equation in discriminating between young adults with and without ADHD.

ADHD behavior rating scales

Diagnostic issues

The DSM-5 diagnostic criteria require that several ADHD-related symptoms must occur "often" and "interfere with, or reduce the quality of, social, academic, or occupational functioning". Although still relying on judgement, behavioral rating scales are more precise in quantifying symptom experiences (Barkley, 2011a) and are therefore potentially more helpful than a clinical interview in clarifying whether the patient experiences ADHD symptoms that meet these two specific criteria. For example, in the Barkley Adult ADHD Rating Scale IV (BAARS-IV, Barkley, 2011a), Barkley has operationalized the construct "often" by means of the following stipulation. That is, for a patient to be considered to have ADHD, their frequency of endorsement of the 18 DSM IVADHD symptoms must exceed that of all but 5% of the population. However, the patient's responses on such scales only reveal how typical or atypical their self-ratings of their behaviors are vis-a-vis the normal population but not individuals with psychiatric disorders. Furthermore, adult ADHD

behavior rating scales have the same weaknesses as most all behavior rating scales (Barkley, 2011b) and reflect a subjective impression of behavior rather than providing an objective measure of behavior.

Further complicating the diagnosis of ADHD in postsecondary students is the fact that most standardized ADHD behavior rating scales have adequate and representative norms for only the general adult population. This is unfortunate because such students are generally more intelligent and higher functioning in many respects than the general population. Consequently, students with ADHD may have scores in the average range on ADHD related measures while their scores would fall in the impaired range relative to the postsecondary student population (Weyandt & DuPaul, 2013).

Discrepancies between self and informant reports on ADHD behavior rating scales are common and variable in their direction and raise the question regarding which should be given more weight in making the ADHD diagnosis. There are only moderate correlations between self and informant reports on ADHD behavior rating scales (Barkley, Knouse, & Murphy, 2011; Van Voorhees, Hardy, Kollins, 2011, Zucker, Morris, Ingram, Morris, & Bakeman, 2002). Dvorsky, Langberg, Molitor, and Bourchtein (2016) found parent ratings were superior to those of college students in predicting the latter's ADHD diagnosis. On the other hand, based on their substantial clinical experience, Murphy and Gordon (2006) have opined that the patient is a more reliable reporter than an informant. In general, it remains unclear which of these sources (the patient, parent, or significant other) is more valid in identifying the relative severity of ADHD symptoms and related functional impairment (Barkley et al., 2011).

Another diagnostic issue is how to best integrate information gleaned from patient and informant behavioral measures (e.g., ADHD behavior rating scales) as well as from the clinical interview. This issue has not been addressed in adults. However, a study of children diagnosed with ADHD reported as many as 50% of the patients were reclassified from one ADHD subtype to another when various sources of information were considered. The diagnosis also was affected by the specific algorithm used to combine the informant reports (Valo & Tannock, 2010).

The two traditional algorithms for combining behavior rating scale data from patients and informants typically employed by clinicians (consciously or not) are OR rules and AND rules. OR rules are the most lenient in diagnosing ADHD because they require a sufficient total number of ADHD symptoms be endorsed either by the patient OR their informant. The AND rules are more restrictive because they require both the patient AND the informant to endorse the patient having a sufficient number of inattentive and/or hyperactivity/impulsivity symptoms for the patient to be diagnosed with ADHD.

There are significant weaknesses associated with both the OR and AND rules for integrating information that lead to the OR rule and the AND rule likely resulting in the over-diagnosis and under-diagnosis of ADHD respectively (Martel, Nikolas, Schimmack, & Nigg, 2015). These researchers have proposed an alternative approach to integrating the results of behavioral rating scales completed by the patient and their informants. They recommend that ADHD symptoms be averaged (or summed which is equivalent) at the symptom domain (i.e., inattention, hyperactivity-impulsivity) and/or at the overall diagnostic category (ADHD) level for each rater. Then an average

across raters should be calculated and used to determine the symptom counts and diagnostic status. In their study of 725 children, these authors found that, while both the averaging approach and OR rule had good specificity (i.e., 91%), the averaging approach had better sensitivity than the OR rule (i.e., 83% versus 68%) in predicting ADHD diagnosis.

Diagnostic accuracy

Six studies have evaluated the effectiveness of ADHD behavior rating scales in differentiating between adults referred for assessment who were and were not diagnosed with this disorder. This is the most relevant comparison because clinicians are asked to diagnose ADHD in patients presenting for ADHD and psychiatric assessment, not the general population as represented by a normal control group. Taylor, Deb, and Unwin (2011) reviewed the psychometric properties of commonly employed adult self-report ADHD behavior rating scales. McCann and colleagues found the Wender Utah Rating Scale (WURS) had a sensitivity of 72% and specificity of 57% (McCann, Scheele, Ward, & Roy-Byrne, 2000).

Soderstrom, Pettersson, and Nilsson (2014) found the identical and immediate predecessor of the BAARS-IV (i.e., the Barkley Current Symptoms Scale -Self Report Form or BCSS, Barkley & Murphy, 2006) had a sensitivity of 85% and specificity of only 40%. They also found the Adult ADHD Self-Report Scale (ASRS, Kessler, Adler, Ames, Demler, et al., 2005) had a sensitivity of 90% but a specificity of only 35% in differentiating these two groups. Furthermore, Pettersson et al. (2018) reported the ASRS had a sensitivity of 92% and specificity of 27% in a subsequent study.

In college populations, Dvorsky et al. (2016) reported the sensitivity and specificity of the BAARS-IV self-report current inattention symptoms ratings were 89% and 30% respectively while those for the self-report childhood inattention symptom ratings were 65% and 40% respectively. On the other hand, Harrison, Nay, and Armstrong (2019) found the current CAARS ADHD Index score (Conners Adult ADHD Rating Scales, Conners, Erhardt, & Sparrow, 1999) (t score = 65) had a sensitivity of 64% and specificity of 86% in a postsecondary population.

There are seven studies reporting the diagnostic accuracy of ADHD scales in differentiating adults with ADHD from adults with psychiatric disorders. The Brown Attention-Deficit Disorder Scales (Brown, 1996) was found to have a sensitivity of 92% but specificity of only 33% in differentiating adults with ADHD (and some comorbid disorders) from adults with anxiety and depression disorders (Solanto, Etefia, & Marks, 2004). Nikolas and colleagues (2019, unpublished data) reported the BAARS-IV self-report current inattention summary score had a sensitivity of 76% and specificity of 71% in differentiating those with ADHD and depression. Similarly, Dunlop, Wu, and Helms (2018) noted the ASRS had a sensitivity of 60% and specificity of 69% in distinguishing patients diagnosed with major depression and ADHD versus patients with only a major depression diagnosis.

Three studies have examined the diagnostic accuracy of ADHD rating scales in discriminating between those diagnosed with ADHD and not diagnosed with ADHD in patients seeking treatment for substance use disorder. The ASRS had a sensitivity of

84% and specificity of 66% in distinguishing between these two groups when the ADHD diagnosis was determined via the CAADID clinical interview (van de Glind et al., 2013). Luty et al. (2009) found the CAARS had a sensitivity of 97% and specificity of 83%, the WHO Adult ADHD Self-Report Screener had a sensitivity of 89% and specificity of 83%, and the WURS – C had a sensitivity of 88% and specificity of 70% in discriminating between the two groups when the ADHD diagnosis was made based on information obtained during patient and informant clinical interviews. Lastly, Chiasson et al. (2012) reported that the ASRS had sensitivity of 100% and specificity of 81% when the diagnosis of ADHD was based on patient and informant clinical interviews.

Thus, with the possible exception of the CAARS, self-report ADHD behavior rating scales alone do not have good diagnostic accuracy as they result in far too many patients undergoing assessment incorrectly being diagnosed as having ADHD.

Recommendations

In sum, clinicians need to obtain ADHD behavior rating scales completed by the patient and a knowledgeable informant (Ramsey, 2015). Research to date suggests an averaging approach is the best means of integrating the ADHD behavior rating scales completed by the patient and informants. Research also suggests that, for several reasons, the most useful ADHD behavior rating scale to use in conjunction with the clinical interview is the CAARS. First, the CAARS – unlike the BAARS or ASRS – is composed of more than just the 18 DSM symptoms that would already have been evaluated in a gold standard clinical interview. Second, the CAARS appears to be the only rating scale that has adequate specificity in a young adult population (Harrison, Nay, & Armstrong, 2016). Third, the CAARS is the only rating scale that has validity scales to identify invalid symptom presentation (further described below).

Symptom and performance validity testing

Diagnostic issues

Several studies have documented many adults presenting for ADHD assessment clearly exaggerate or feign cognitive deficits and ADHD symptomatology. Marshall et al. (Marshall et al., 2016) found that 27%, Suhr et al. (Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008) found that 31%, and Nelson and Lovett (2019) found that 53% of young adults undergoing comprehensive ADHD assessment made an invalid symptom presentation. Similarly, in a significantly older adult population, 32% made such an invalid presentation (Hirsch & Christiansen, 2018).

Regrettably, it is well established that it is quite easy for an adult seeking an ADHD diagnosis to exaggerate or completely feign ADHD symptoms during a clinic interview and when completing the most commonly used behavior rating scales (Musso & Gouvier, 2014; Tucha et al., 2015). Marshall and colleagues (2016) also found that, of the 27% of their patients making an invalid symptom presentation, 71% would be diagnosed with ADHD based on a clinical interview alone, 65% based on the interview and ADHD behavior rating scales combined, and 62% based on the interview, behavior rating scales, and a continuous performance test combined.

Clinicians very likely have considerable difficulty detecting patients faking ADHD if measures to identify an invalid presentation in completion of behavior rating scales as well as in cognitive testing are not employed (Tucha et al., 2015). This is illustrated in a study by Booksh and colleagues (Booksh, Pella, Singh, & Gouvier, 2010) in which college students were asked to simulate ADHD symptoms during an assessment consisting of a structured clinical interview, behavioral rating scales, and cognitive testing. An independent psychologist was then asked to judge whether a student was a simulator, a normal control subject, or a patient previously diagnosed with ADHD. The psychologist misclassified 44% of the student simulators as having ADHD and 11% of them as being normal. Furthermore, in general, studies have indicated psychologists and psychiatrists are over confident in their ability to identify invalid cognitive and behavioral symptom presentations in their review of patient's clinical histories and cognitive testing (Dandachi-Fitzgerald, Merckelbach, & Ponds, 2017; Faust, Hart, Guilmette, & Arkes, 1988, Heaton, Smith, Lehman, & Vogt, 1978).

Research indicates that symptom validity tests (SVTs) and performance validity tests (PVTs) are the best available means for detecting invalid symptom presentations on behavior rating scales and inadequate effort in cognitive testing respectively in adult ADHD assessment (Sagar, Miller, & Erdodi, 2017; Tucha et al., 2015; Wallace et al., 2019). Numerous expert clinicians on adult ADHD assessment have recommended the use of SVTs and PVTs as part of a comprehensive adult ADHD assessment (e.g., Bordoff, 2017; Ramsey, 2015; Weyandt & DuPaul, 2013).

Diagnostic accuracy

While there are numerous PVTs that can be used in ADHD assessment, regrettably there is only one ADHD behavior rating scale that has an SVT. Two SVTs have been developed to identify invalid symptom reporting on the widely used adult ADHD behavior rating scale, the CAARS. Harrison and Armstrong (2016) derived an Exaggeration Index from CAARS symptoms and a few additional non-ADHD related symptoms. An Exaggeration Index cutoff score > 1 had a sensitivity of 51% and specificity of 88% while a score > 2 had a sensitivity of 34% and specificity of 94% in identifying an invalid symptom presentation. Using a different methodology, Suhr, Buelow, and Riddle (2011) created a CAARS Infrequency Index for which an Index score ≥ 21 had a sensitivity of 52% and specificity of 97%. Their scale was subsequently validated in a second sample (Cook, Bolinger, & Suhr, 2016). Barkley (2011b) has acknowledged the BAARS-IV is vulnerable to malingering but, unfortunately, it currently has no embedded SVT.

Several studies involving students asked to simulate ADHD have examined the efficacy of specific PVTs that are free standing or embedded in standard neuropsychological tests (Jasinski et al., 2011; Sollman, Ranseen, & Berry, 2010). However, only two studies have examined this issue in young adults presenting for ADHD evaluation. Suhr et al. (2008) found that, though having very high specificity, four commonly used embedded PVTs had very poor sensitivities (ranging from 19% to 4%) in detecting suspect effort in testing. In contrast, Marshall and colleagues found several measures with reasonably good sensitivity in their evaluation of the diagnostic accuracy of three

stand alone and six embedded PVTs in a much larger study of young adults undergoing ADHD evaluations (Marshall et al., 2010). The most effective PVTs were the Word Memory Test (Green, 2003) consistency score (sensitivity 64%, specificity 95%), the Test of Variables of Attention (Greenberg, Kindschi, Dupuy, & Corman, 2011) omission errors (sensitivity 63%, specificity 92%), the Conners Continuous Performance Test (Conners, 2008) omission errors (56% sensitivity, 87% specificity), and the b test (Boone, Lu, & Herzberg, 2002) E score (sensitivity 47%, specificity 93%).

It is important to note using failure on just one PVT to identify invalid symptom presentation results in a very large and unacceptable number of false positives (Marshall et al., 2010; Victor, Boone, Serpa, Buehler, & Ziegler, 2009). In contrast, failure of two or more PVTs and SVTs has been found to have an overall sensitivity of 50% and specificity of 93% (Sollman et al., 2010). Similarly, failure on two or more PVTs had a sensitivity of 48% and specificity of 100% (Jasinski et al., 2011) in simulation studies.

Recommendations

It is important to include at least four PVT and SVT measures because an individual's effort during testing can fluctuate significantly over the course of an assessment, and individuals differ in what cognitive abilities they choose to exaggerate or feign deficits (Boone, 2009; Marshall et al., 2010). Furthermore, as just noted, failure of two or more PVT and SVT measures has much greater diagnostic accuracy in identifying insufficient effort. Specifically, the use of at least one stand-alone PVT, the b Test, as well as a SVT embedded in the CAARS and PVTs embedded in the TOVA and other cognitive tests is recommended (see below).

In conclusion, there are multiple issues that clearly make the accurate diagnosis of adult ADHD based on clinical interviews, ADHD symptom related behavior rating scales, and review of relevant archival records a very difficult and demanding task. To reiterate, these complexities include the non-specificity of adult ADHD symptoms, the identification of the symptoms that are most appropriate for adult ADHD (and best discriminate between those with and without this disorder), the reliability and accuracy of patients' and informants' reports of ADHD symptoms, the identification of appropriate symptom thresholds for frequency and severity of ADHD symptoms, the determination of functional impairment, discrepancies between patient and informant reports, integration of multiple sources of assessment information, and patient misrepresentation of ADHD symptoms.

Neuropsychological testing

Diagnostic issues

Given that different types of symptoms are considered (e.g., subjective report of inattention problems versus performance on sustained attention tests), it is entirely possible the addition of neuropsychological tests to the clinical interview and behavior rating scales might improve the diagnostic accuracy of an adult ADHD assessment battery.

Several meta-analyses have reviewed the hundreds of studies examining the utility of individual neuropsychological tests in differentiating patients diagnosed with ADHD from control groups (Alderson, Kasper, Hudec, & Patros, 2013; Boonstra, Oosterlaan, Sergeant, & Buitelaar, 2005; Boonstra, Kooij, Oosterlaan, Sergeant, & Buitelaar, 2010; Frazier, Demaree, & Youngstrom, 2004; Hervey, Epstein, & Curry, 2004; Kofler et al., 2013; Skodzik, Holling, & Pedersen, 2013). These studies have included numerous tests of attention, response inhibition, executive functions, memory, working memory, cognitive processing speed, motor speed, and intelligence. The ability of a test to differentiate between groups is typically expressed in terms of an effect size (d'). By convention, effect sizes approaching .30 are considered small, between .40-.70 are considered medium, and .80 or greater are considered large (Cohen, 1988). The vast majority of these individual tests have small to medium effect sizes with most pooled effect sizes falling in the medium range (Pievsky & McGrath, 2018). In sum, the modest effect sizes of the vast majority of *individual* cognitive tests clearly indicate that many adults with ADHD perform in the normal range and only a minority of them will render an impaired performance on any specific test (Nigg, Willcutt, Doyle, & Sonuga-Barke, 2005). These findings, as well as fundamental conceptual concerns about the ability of cognitive tests to assess ADHD cognitive symptoms (Barkley, 2011b), have led many to recommend neuropsychological testing not be used in diagnosing adult ADHD (Barkley, 2019; Barkley et al., 2008; Solanto, 2015).

However, individuals with ADHD are *consistently inconsistent* in their performance on neuropsychological tests over time (Kofler et al., 2014) as they can often rally to focus their attention for brief periods of time on any one particular test measure (Leimkuhler, 1994). Furthermore, there is clearly pervasive and significant cognitive heterogeneity in patients diagnosed with ADHD as currently specified. ADHD symptoms are most likely caused by the additive and interactive combination of several cognitive deficits, none of which are necessary or sufficient to cause ADHD when they occur alone (Willcutt, 2015). Therefore, it is important to also consider the effectiveness of cognitive test batteries rather than individual tests in differentiating patients diagnosed with ADHD from control and clinical control groups.

Diagnostic accuracy

Some of the few studies of the utility of cognitive test batteries in diagnosing adult ADHD have had more promising results, particularly in differentiating patients diagnosed with ADHD from normal control group participants. Rapport, Van Voorhis, Tzelapis, and Friedman (2001) found a discriminant function analysis based on a battery of seven cognitive tests had a sensitivity of 58.8% and specificity of 81.3%. The Quantified Behavior Test Plus (QBTP+) is continuous performance test (CPT) with multiple measures of not only sustained attention but also hyperactivity (i.e., tracking of head movement during the test). The QBTP+ had a sensitivity of 87% and specificity of 85% (Edebol, Helldin, & Norlander, 2013). Mostert and colleagues (2015) also reported a regression model based on the results of their test battery had a sensitivity of 64.9% and specificity of 82.1%.

More promisingly, Lovejoy et al. (1999) found a clinically impaired performance on any one of six tests in a battery had a sensitivity of 96% and specificity of 85% while a clinically impaired performance on any two of the six tests had a sensitivity of 69% and specificity of 96%. Furthermore, a regression model based on a battery of seven cognitive tests had a sensitivity of 93% and specificity of 90% (Walker, Shores, Trollor, Lee, & Sachdev, 2000).

It is important to note, however, that the relatively greater diagnostic accuracy of the Lovejoy et al. (1999) and Edebol et al. (2013) test batteries may have been due to unique characteristics of their ADHD groups. The Lovejoy ADHD group included *only* patients who were currently taking ADHD stimulant medications and had reported that these medications were "very helpful" in managing ADHD symptomatology. Furthermore, 65% of the ADHD group had first degree relatives who were diagnosed with ADHD. In the Edebol et al. study (2013), 94% of the ADHD group had been diagnosed with ADHD combined type and only 4% with ADHD inattentive type. Thus, the sensitivity of their test battery in diagnosing patients with ADHD inattentive type is essentially unknown. This is very problematic since ADHD inattentive type is the most prevalent subtype, affecting 45% of the adult ADHD population (Woo & Rey, 2005).

Like virtually all individual tests, batteries of cognitive tests have not fared well in discriminating between patients with ADHD versus psychiatric patients. The aforementioned Walker et al. study (2000) regression model had poor specificity (80%) and Katz and colleagues' (Katz, Wood, Goldstein, Auchenbach, & Geckle, 1998) discriminant function analysis had even less specificity (40%) in distinguishing patients with ADHD from those with depression. The QBT + continuous performance test had a similarly unacceptable specificity (36%) in differentiating patients with ADHD versus bipolar disorder (Edebol, Helldin, & Norlander, 2012). In contrast, however, Holst and Thorell (2017) found an executive function test battery-based regression equation had a sensitivity of 66.7% and specificity of 81.4% in discriminating between patients with ADHD versus patients with psychiatric mood disorders.

Unfortunately, only two studies have examined the ability of a test battery to differentiate between adult patients diagnosed with ADHD versus those evaluated for but not diagnosed with ADHD. To reiterate, this is the most relevant comparison because clinicians are asked to diagnose ADHD in patients presenting for ADHD assessment, not the general population as represented by a normal control group. Edebol and colleagues (2012) found the QBT + test battery had a sensitivity or 59% and specificity of 41% while Hirsch and Christiansen (2017) reported it had sensitivity of 90% and specificity of 45% in differentiating between these two groups.

In summary, most individual cognitive tests have poor sensitivity though some have reasonable specificity in identifying those diagnosed with ADHD versus normal control participants. The notable exceptions are CPT tests that usually have good sensitivity but poor specificity (Riccio & Reynolds, 2006). Batteries of cognitive tests have greater and potentially more useful levels of sensitivity than individual tests. Lovejoy and colleagues' research in particular suggests that using the criteria of psychometrically defined clinical impairment based on a battery of tests rather than single test performances holds promise for significantly increasing the sensitivity of cognitive test measures to correctly

diagnose adult ADHD. However, research also suggests that cognitive tests have not only limited sensitivity but also inadequate specificity when trying to make a differential diagnosis between patients with ADHD and those with other psychiatric disorders. Thus, far too many individuals with depression and other psychiatric diagnoses will plausibly be diagnosed with ADHD using *only* the cognitive tests evaluated to date.

Just four studies to date have examined the effectiveness of using neuropsychological testing in addition to ADHD behavior rating scales in diagnosing adult ADHD. Soderstrom et al. (2014) reported that a discriminant function analysis based on the self-report ASRS and BCSS rating scales as well as the QBT + inattention and impulsivity measures had a sensitivity of 75% and a specificity of 62% in differentiating between patients diagnosed with adult ADHD and clinical control participants. Pettersson et al. (2018) found that a regression model based on the ASRS and a cognitive test battery as well as a clinical interview (the DIVA) had a sensitivity of 90% and specificity of 81% in a group of adult outpatients presenting for ADHD evaluations. It is important to note, however, that the results of these studies by Pettersson, Soderstrom, and Nilsson are confounded by the fact the QBT+, ADHD behavior rating scales, and interview were considered in making the original ADHD diagnosis in the ADHD criterion group.

Emser et al. (2018) utilized machine learning paradigms with objective data (e.g., scores on the QBT+) and symptom rating data (i.e., the CAARS) to determine measures that best predicted ADHD diagnoses among adults. They found that the results of a go/no go task, a divided attention task, and a sustained attention task as well as the QBT + taken together had a sensitivity of .82 and specificity of .76. However, the combined use of CAARS data and objective test data had a sensitivity of .90 and specificity of .90. When used together, the symptom rating data made stronger contributions to the prediction of ADHD diagnosis relative to test data. These findings are similar to Nikolas et al. (2019) who used logistic regression and also found optimal diagnostic utility with approaches combining symptom rating data and test data. Future work incorporating additional indices as well as other measurement methods and utilizing a variety of prediction analytics may be able to improve identification of ADHD as well as illuminate the nature of its heterogeneity.

The research to date clearly demonstrates the limited diagnostic utility of both individual tests and batteries of cognitive tests. However, this conclusion may be premature given three major study design limitations. First and foremost, experts acknowledge that the criteria for inclusion in ADHD criterion groups are "highly uncertain" (Gordon, Barkley, & Lovett, 2006; Murphy & Gordon, 2006). Thus, the limited diagnostic utility of tests and other measures may be a function of the inclusion of many individuals without ADHD in the criterion group (Suhr et al., 2008). Second, most all studies to date have *not* included measures of performance or symptom validity to detect inadequate effort in testing and/or an invalid presentation in completion of behavior rating scales. Third, reaching definitive conclusions regarding the diagnostic utility of cognitive testing has been also made difficult by the use of a variety of different tests across studies to measure the same cognitive constructs (e.g., sustained attention, executive functions, and working memory). Many of these tests are not sufficiently difficult and precise to be sensitive in identifying adult ADHD-

related cognitive deficits (e.g., Alderson et al., 2013). Certainly, the diagnostic utility of cognitive tests might be enhanced with further research.

In conclusion, the four studies concerning the diagnostic utility of employing both ADHD behavior rating scales and cognitive testing suggest that this approach is the most effective means of diagnosing adult ADHD. Specifically, the addition of cognitive testing to ADHD behavior rating scales very significantly increases the *specificity* of an assessment battery thereby significantly reducing the number of patients misdiagnosed as having ADHD. Therefore, the inclusion of some cognitive testing is recommended for any adult ADHD assessment. Finally, given the low sensitivity of cognitive tests, if the test results are not abnormal and are inconsistent with the results of the interview and behavior rating scales, it is the test results that should be disregarded (Mapou, 2019).

Recommendations

According to DSM-5 criteria, the core symptoms of ADHD are inattention, impulsivity (poor response inhibition), and hyperactivity. Family and twin studies have identified three cognitive phenotypes that reflect the familial-genetic risk in ADHD. They are slow and highly variable reaction times on tests of sustained attention, commission errors on go/no-go tasks (indicative of difficulties with response inhibition), and errors on working memory tests (Pinto, Asherson, Ilott, Cheung, & Kuntsi, 2016).

Continuous performance tests are considered a key component of any ADHD assessment because they assess attention, vigilance, processing speed, impulsivity, and response inhibition (Advokat, Martino, Hill, & Gouvier, 2007; Fuermaier, Fricke, de Vries, Tucha, & Tucha, 2019). Furthermore, greater variability in response time on CPTs is clearly related to attention lapses and distractibility (Adams, Roberts, Milich, & Fillmore, 2011). The TOVA and the Conners CPT (Conners, 2008) are the two CPTs widely used in ADHD assessment. The TOVA is recommended for several reasons. Unlike the Conners CPT, the TOVA 8.0 provides cut off scores for four embedded suspect effort indices (a.k.a. Performance Validity Index) based on normative data (Greenberg, 2011). Additionally, Marshall and colleagues (2010) have identified cutoff scores for three additional TOVA embedded indices based on young adults who clearly made an invalid symptom presentation while undergoing ADHD assessment. It is particularly important for a CPT test to have such a PVT because performance on this test can be impaired not only by intentional exaggeration or feigning of ADHD symptoms, but also by occasionally occurring non-volitional factors such as acute, unusual levels of fatigue (e.g., due to inadequate sleep and mild illness). Unfortunately, it is not clear whether the TOVA or Conners CPT has better diagnostic accuracy because there have been no studies directly comparing their diagnostic accuracy. Notably, Nikolas et al. (2019) did find TOVA reaction time variability had a clinically significant Odds Ratio (3:1) and improved the diagnostic accuracy of a regression equation based on a comprehensive ADHD assessment. They also found TOVA reaction time variability was the best predictor of central ADHD symptoms as measured by behavior rating scales. Specifically, it predicted inattention (i.e., the BAARS-IV

Inattention/Memory Scale summary score) and executive function deficits (i.e., the Barkley Deficits in Executive Functioning Scale percentile (BDEFS; Barkley, 2011b).

Working memory processes enable the temporary storage, maintenance, and manipulation of information that is necessary to guide behavior (Barkley, 2007). In studies involving both children and adults, Willcutt, Doyle, Nigg, Faraone, and Pennington (2005) found that verbal working memory tests had a weighted mean positive effect size of $d = .55$ in differentiating between those with and without ADHD. The Salthouse Listening Span Task (Salthouse, 1994) is recommended because it is much more difficult and, hence, more sensitive in detecting relatively mild deficits in verbal working memory compared to other commonly used verbal working memory measures (e.g., Letter Number Sequencing; Digit Span). The learning trials and short delay free recall portions of the California Verbal Learning Test-II (CVLT, Delis et al., 2000) are also recommended since they evaluate individual's verbal working memory and short- term focused attention. In their meta-analysis of adult ADHD studies, Hervey et al. (2004) found the CVLT learning trials 1-5 had a large positive effect size ($d = 0.91$).

Finally, in addition to commission errors on a CPT test, response inhibition is assessed by the Stroop test. The Delis-Kaplan Executive Function System (DKEFS) Color Word Interference Test (CWIT) is recommended for it is, in essence, a better designed and normed variant of the Stroop test (Delis, Kaplan, & Kramer, 2001). In their meta-analysis of executive function tests in adult ADHD research, Boonstra et al. (2005) found that the Stroop test interference condition had a large positive effect size ($d = 0.89$) while the color naming and word naming trials had medium positive effect sizes with $d > $s of 0.60 and 0.62 respectively. More recently, Halleland, Haavik, and Lundervold (2012), Holst and Thorell (2017) and Nikolas et al. (2019) have reported the DKEFS CWIT inhibition/switching trial, a measure of set shifting, has demonstrated significant discriminative validity.

Additional assessment measures that might improve the adult ADHD assessment process

Logically, both executive function and functional impairment scales might improve the diagnostic accuracy of an assessment battery. It has become increasingly clear in the past decade that executive function (EF) behaviors are as central to ADHD as sustained attention with an even greater impact on functional impairment, particularly in adulthood. In fact, Barkley (2015) has cogently proposed that ADHD is primarily a disorder of executive function rather than attention deficits. Furthermore, Kessler et al. (2010) have noted that EF related behavioral problems (e.g., difficulties in organizing, planning ahead, prioritizing, completing tasks on time, and making mistakes) are the most specific and consistent predictors of DMS-IV based adult ADHD diagnoses.

It also makes sense to include executive function rating scales because they assess very critical EF behaviors not assessed by EF neuropsychological tests. As Barkley (2011b) has noted, EF neuropsychological tests assess the moment-to-moment, "instrumental" level of EF but are ineffective in assessing the "adaptive", "tactical", and "strategic" EF levels used in carrying out social, educational, vocational, and other activities of daily living over longer time frames. Toplak, West, and Stanovich (2013) have also observed and posited that EF neurocognitive tests and EF behavior rating

scales assess different constructs. EF neurocognitive tests provide important information about the immediate efficiency of information processing mechanisms in the brain (i.e., attention, working memory, long term memory), whereas behavior rating scales provide information about the longer-term effectiveness and success of EF related actions in the pursuit of rational goals.

Very few studies have examined the diagnostic accuracy of executive function behavior rating scales. Barkley and Murphy (2011) found patients with adult ADHD report having much more significant EF impairment than normal control groups, and to a lesser extent, clinical control groups on the immediate and virtually identical predecessor of the BDEFS. Barkley (2011b) also found ADHD-EF index score derived from this BDEFS predecessor was effective in discriminating between adults diagnosed with ADHD and a normative control group (positive predictive power 94%, negative predictive power 87%). However, 96% of the clinical control group (i.e., individuals presenting for ADHD evaluation but subsequently not diagnosed with ADHD) would also be diagnosed with ADHD based on the ADHD-EF index score.

Kamradt, Ullsperger, and Nikolas (2014) investigated the sensitivity and specificity of eight EF test measures and BDEFS subscales in discriminating between young adults diagnosed with ADHD and a community control group. The sensitivity of individual test measures ranged from 11-23% while their specificity ranged from 89 to 96%. Rather similarly, the sensitivity of the individual BDEFS subscales ranged from 22 to 23% while their specificity ranged from 87-98%. Finally, they examined the utility of combining both the EF test measures and the BDEFS rating scales in making the diagnosis of ADHD. This approach had an overall, reasonably good sensitivity of 84% but very low specificity of 44%. Thus, a battery combining EF test measures and behavior ratings was reasonably effective in ruling ADHD in but very ineffective in ruling ADHD out.

Finally, it is important to acknowledge an additional factor when considering the potential diagnostic utility of EF behavior rating scales in diagnosing adult ADHD. Marshall et al. (2016) found young adults exaggerating or faking ADHD symptoms were indistinguishable from those diagnosed with ADHD on the two BDEFS indices considered to be most important in making the diagnosis of ADHD: the summary score and total symptom count.

Until DSM-5, DSM ADHD criteria stipulated that ADHD symptoms needed to cause "clinically significant impairment" in social, academic, or occupational functioning. Under DSM-5, these criteria have been relaxed somewhat to "there is clear evidence that the symptoms interfere with, or reduce the quality of social, academic, or occupational functioning." Yet DSM-5 also continues to ask the clinician to delineate whether symptoms result in mild, moderate, or severe "impairment" in functioning. Though still relying on subjective judgement, behavioral rating scales are more precise in quantifying symptom experiences and therefore potentially more helpful than a clinical interview in clarifying the degree to which symptoms impair a patient's functioning in these domains.

Barkley (2011c) created the Barkley Functional Impairment Scale (BFIS) to assist in this task. A study using a prototype of the BFIS found adults diagnosed with ADHD had higher self-rated and informant rated total functional impairment scores than both a normal control group and a clinical control group with other psychiatric disorders (Barkley et al., 2008). Logistic regression analyses revealed current self-report rating scores on three BFIS life activity domains were most effective in differentiating

adults with ADHD from a normal control group. These life activity domains were occupational functioning, educational activities, and money management with medium and large odds ratios (OR) of 2.45, 6.39, and 3.95, respectively. On the other hand, the domains that best differentiated adults with ADHD and a clinical control group, educational activities and money management, had only small ORs of 1.90 and 1.50, respectively. These results suggest that the BFIS has limited discriminative validity in diagnosing ADHD in patients presenting for ADHD assessment. Nikolas et al. (2019, unpublished data) found that the BFIS mean impairment percentile had sensitivities and specificities of 19% and 32% respectively in differentiating between patients diagnosed with ADHD versus control participants and 81% and 22% between patients with ADHD versus individuals with depression.

Further limiting the diagnostic utility of the BFIS and other functional impairment scales is the fact they have no embedded SVTs and can be completed in an invalid manner without detection. Barkley (2011c) has warned that this could happen on the BFIS and Marshall et al. (2016) found that this was the case. Bryant et al. (2018) had similar findings with respect to the World Health Organization Disability Schedule (WHODAS, World Health Organization, 2012), another commonly used measure of functional impairment. Finally, individuals instructed to feign ADHD could not be differentiated from genuine patients diagnosed with ADHD in their reports on the Weiss Functional Impairment Rating Scale (Fuermaier et al., 2018).

Recommendations

EF behavior rating scales are highly correlated with behavior rating scales consisting of the 18 DSM-IV ADHD symptoms. In fact, the behavior rating scales of DSM ADHD symptoms are so highly correlated with EF behavior rating scales that they approach, if not meet, standards of collinearity (Barkley, 2011b). This has led Barkley (2012, 2015) to conclude EF behavior rating scales and ADHD behavior rating scales may well be identifying the same psychological construct. Thus, his conclusion as well as the aforementioned studies indicate the addition of an EF scale to an ADHD behavior rating scale is unlikely to improve the diagnostic accuracy of an assessment battery. Finally, findings regarding the BFIS in particular, as well as the WHODAS and WFIRS, suggest that adding a functional impairment scale to a battery will also not significantly improve diagnostic accuracy.

A proposed adult ADHD diagnostic battery

Numerous experts have proposed lines of research that should improve our ability to diagnose ADHD but will undoubtedly take several years to fully explore (e.g., Heidbreder, 2015; Koziol & Stevens, 2012; Weyandt & DuPaul, 2013; Willcutt, 2015). Many clinicians have expressed an immediate, pressing need for means to improve the adult ADHD assessment process. Therefore, it appears appropriate to recommend a relatively brief, easy to administer, and inexpensive diagnostic battery based on the research conducted to date. Based on the recommendations previously noted, the proposed battery would include (a) the semi-structured diagnostic interview module for the assessment of adult ADHD recommended by Gorlin et al. (2016), (b) the CAARS completed not only by the

patient but an informant who knows their current behavior very well (e.g., a partner) and, if possible, a parent, (c) the TOVA, (d) the Salthouse Listening Span Task, (e) the CVLT-II, (f) the DKEFS Color-Word Interference test, and (g) the b test.[1]

This entire assessment battery should take approximately two hours for – most importantly – the patient to complete. Administration and subsequent scoring of the various assessment measures done by a psychometrist or assistant should take no more than two hours. Finally, administration of the clinical interview as well as review and interpretation of the assessment results should take the clinician less than one hour and thirty minutes. Hence, the entire ADHD assessment battery should take approximately five hours. This is shorter than the six-eight hours required to do a comprehensive ADHD assessment consisting of a complete review of medical records, a thorough diagnostic interview, neuropsychological testing, and a patient feedback session (Pazol & Griggins, 2012). According to current pricing (October 2019), the initial cost of purchasing assessment manuals, tests, and scoring software would be $2,213. The subsequent cost of the assessment measures would be $44.30 per administration.

Conclusions

In summary, adults are increasingly referred for neuropsychological evaluation to determine the presence of ADHD. There are numerous challenges associated with this differential diagnosis, including but not limited to non-specific symptoms, difficulties associated with recalling childhood symptoms, and the ease with which ADHD symptoms are misrepresented. While numerous studies have been conducted to understand adult ADHD, this qualitative review highlights ways that this body of literature is limited. While aspects of the proposed battery have empirical support, nevertheless, it will be critically important to evaluate its utility in future research. At a minimum, it is essential that prospective research be conducted investigating whether utilizing the battery results in more accurate diagnoses than standard practice procedures (i.e., a clinical interview and completion of self-report measures). Additionally, efforts should be directed towards understanding whether the battery differentially predicts ADHD subtypes and how it might clarify the impact of comorbid psychological conditions on symptom reporting and neuropsychological performances. Addressing these important questions and more is likely to result in more accurate clinical decisions and ultimately improve patient outcomes. Finally, the proposed battery and other recommendations to improve the diagnostic process are the authors' personal opinions, not consensus standards, or guidelines promulgated by any organization.

Disclosure statement

No potential conflict of interest was reported by the authors.

Note

1. The semi-structured diagnostic interview module for the assessment of adult ADHD is included in the article by Gorlin et al., (2016). The CAARS, TOVA, CVLT-II, and DKEFS Color Word Interference, and b tests are all commercially available. The Salthouse Listening Span Task is in the public domain and can be obtained at no cost by contacting the first author at pmarsh7247@gmail.com.

References

Adams, Z., Roberts, W., Milich, R., & Fillmore, M. (2011). Does response variability predict distract-ibility among adults with attention-deficit/hyperactivity disorder? *Psychological Assessment*, *23*(2), 427–436.

Advokat, C., Martino, L., Hill, B., & Gouvier, W. (2007). Continuous performance test (CPT) of college students with ADHD, psychiatric disorders, cognitive deficits, or no diagnosis. *Journal of Attention Disorders*, *10*(3), 253–256. doi:10.1177/1087054706292106

Adler, L., & Alperin, S. (2015). Diagnosing ADHD in adults in the primary care setting. In R. Barkley, *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment* (4th ed., pp. 501–509). New York, New York: Guilford Press.

Adler, L., Shaw, D., Stitt, D., Maya, E., & Morrill, M. (2009). Issues in the diagnosis and treatment of adult ADHD by primary care physicians. *Primary Psychiatry*, *16*(5), 57–63.

Alderson, R., Kasper, L., Hudec, K., & Patros, C. (2013). Attention deficit/hyperactivity disorder (ADHD) and working memory in adults: A meta-analytic review. *Neuropsychology*, *27*(3), 287–302. doi:10.1037/a0032371

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Barkley, R. (2007). *Working memory, thought, and action*. New York, New York: Oxford University Press.

Barkley, R. (2019). Neuropsychological testing is not useful in the diagnosis of ADHD: Stop it (or prove it)!. *The ADHD Report*, *27*(2), 1–8. doi:10.1521/adhd.2019.27.2.1

Barkley, R. (2011a). *Barkley adult ADHD rating scale-IV (BAARS-IV)*. New York, New York: Guilford Press.

Barkley, R. (2011b). *Barkley deficits in executive functioning scale (BDEFS Scale)*. New York, New York: Guilford Press.

Barkley, R. (2011c). *Barkley functional impairment scale (BFIS Scale)*. New York, New York: Guilford Press.

Barkley, R. (2012). *Executive functions: What they are, how they work, and why they evolved*. New York, New York: Guilford Press.

Barkley, R. (2015). Executive functioning and self-regulation viewed as an extended phenotype: Implications of the theory of ADHD and its treatment. In R. Barkley, *Attention Deficit hyperactivity disorder: A handbook for diagnosis and treatment* (4th ed., pp. 405–434). New York, New York: Guilford Press.

Barkley, R., & Brown, T. (2008). Unrecognized attention-deficit/hyperactivity disorder in adults presenting with other psychiatric disorders. *CNS Spectrums*, *13*(11), 977–984. doi:10.1017/S1092852900014036

Barkley, R., Knouse, L., & Murphy, K. (2011). Correspondence and disparity in the self- and other ratings of current and childhood ADHD symptoms and impairments in adults with ADHD. *Psychological Assessment*, *23*(2), 437–446. doi:10.1037/a0024001

Barkley, R., & Murphy, K. (2011). The nature of executive function (EF) deficits in daily life activities in adults with ADHD and their relationship to performance on EF tests. *Journal of Psychopathology and Behavioral Assessment*, *33*(2), 137–158. doi:10.1007/s10862-011-9217-x

Barkley, R., & Murphy, K. (2006). *Attention-deficit hyperactivity disorder - A clinical workbook* (3rd ed.). New York, New York: Guilford Press.

Barkley, R., Murphy, K., & Fischer, M. (2008). *ADHD in adults: What the science says*. New York, New York: Guilford Press.

Bordoff, B. (2017). The challenges and limitations of diagnosing and pharmacologically treating ADHD in university students. *Psychological Injury and Law*, *10*(2), 114–120. doi:10.1007/s12207-017-9288-4

Booksh, R., Pella, R., Singh, A., & Gouvier, W. (2010). Ability of college students to simulate ADHD on objective measures of attention. *Journal of Attention Disorders*, *13*, 325–338. doi:10.1177/1087054708329927

Boone, K. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, 23(4), 729–741. doi: 10.1080/13854040802427803

Boone, K., Lu, P., & Herzberg, D. (2002). *The b test*. Los Angeles, CA: Western Psychological Services.

Boonstra, A., Oosterlaan, J., Sergeant, J., & Buitelaar, J. (2005). Executive functioning in adult ADHD: A meta-analytic review. *Psychological Medicine*, 35(8), 1097–1108. doi:10.1017/S003329170500499X

Boonstra, A. M., Kooij, J. J. S., Oosterlaan, J., Sergeant, J. A., & Buitelaar, J. K. (2010). To act or not to act, that's the problem: Primary inhibition difficulties in adult ADHD. *Neuropsychology*, 24(2), 209–221. doi:10.1037/a0017670

Brown, T. (1996). *Brown attention-deficit disorder scales*. San Antonio, Texas: The Psychological Corporation.

Bryant, A., Lee, E., Howell, A., Morgan, B., Cook, C., Patel, K., … Suhr, J. (2018). The vulnerability of self-reported disability measures to malingering: A simulated study. *The Clinical Neuropsychologist 32(1), 109–118*.doi:10.1080/13854046.2017.1346145

Chiasson, J., Stavro, K., Rizkallah, E., Lapierre, L., Dussault, M., Legault, L., & Potvin, S. (2012). Questioning the specificity of the ASRS-v1.1 to accurately detect ADHD in substance abusing populations. *Journal of Attention Disorders*, 16(8), 661–663. doi:10.1177/1087054711425768

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Conners, K. (2008). *Conners Continuous Performance Test III*. Tonawanda, NY: Multi-Health Systems, Inc.

Conners, K., Erhardt, D., & Sparrow, E. (1999). *Conners adult ADHD Rating Scales Technical Manual*. Tonawanda, NY: Multi-Health Systems, Inc.

Cook, C., Bolinger, E., & Suhr, J. (2016). Further validation of the Conner=s adult attention deficit/hyperactivity rating scale infrequency index (CII) for detection of non-credible report of attention deficit/hyperactivity disorder symptoms. *Archives of Clinical Neuropsychology*, 31(4), 358–364.

Dandachi-Fitzgerald, B., Merckelbach, H., & Ponds, R. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology*, 39(3), 257–264. doi:10.1080/13803395.2016.1223278

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system*. San Antonio, TX: The Psychological Corporation.

Dunlop, B., Wu, R., & Helms, K. (2018). Performance of the adult ADHD Self-Report Scale-v1.1 in adults with major depressive disorder. *Behavioral Sciences*, 8(4), 37. 8040037. doi:10.3390/bs

Dvorsky, M., Langberg, J., Molitor, S., & Bourchtein, E. (2016). Clinical utility and predictive validity of parent and college student symptom ratings in predicting an ADHD diagnosis. *Journal of Clinical Psychology*, 72(4), 401–418. doi:10.1002/jclp.22268

Edebol, H., Helldin, L., & Norlander, T. (2013). Measuring adult attention deficit hyperactivity disorder using the quantified behavior test plus. *Psych Journal*, 2(1), 48–62. doi:10.1002/pchj.17

Edebol, H., Helldin, L., & Norlander, T. (2012). Objective measures of behavior manifestations in adult ADHD and differentiation from participants with bipolar II disorder, borderline personality disorder, participants with disconfirmed ADHD as well as normative participants. *Clinical Practice & Epidemiology in Mental Health*, 8, 134–143. doi:10.2174/1745017901208010134

Emser, T. S., Johnston, B. A., Steele, J. D., Kooij, S., Thorell, L., & Christiansen, H. (2018). Assessing ADHD symptoms in children and adults: Evaluating the role of objective measures. *Behavioral and Brain Functions*, 14(1), 1–14. doi:10.1186/s12993-018-0143-x

Epstein, J., Johnson, D., & Conners, C. (2000). Conners adult ADHD diagnostic interview for DSM-IV. North Tonawanda, NY: Multi-Health Systems.

Faust, D., Hart, K., Guilmette, T., & Arkes, H. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice*, 19(5), 508–515. doi:10.1037/0735-7028.19.5.508

Fedele, D., Hartung, C., Canu, W., & Wilkowski, B. (2010). Potential symptoms of ADHD for emerging adults. *Journal of Psychopathology and Behavioral Assessment*, 32(3), 385–396. doi:10.1007/s10862-009-9173-x

Frazier, T., Demaree, H., & Youngstrom, E. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, 18(3), 543–555. doi:10.1037/0894-4105.18.3.543

Frazier, T., & Youngstrom, E. (2006). Evidence based assessment of attention deficit/hyperactivity disorder: Using multiple sources of information. *Journal of the American Academy of Child & Adolescent Psychiatry*, 45(5), 614–620. doi:10.1097/01.chi.0000196597.09103.25

Friedrichs, B., Igl, W., Larsson, H., & Larsson, J.-O. (2012). Coexisting psychiatric problems and stressful life events in adults with symptoms of ADHD – A large Swedish population-based study of twins. *Journal of Attention Disorders*, 16(1), 13–22. doi:10.1177/1087054710376909

Fuermaier, A. B. M., Fricke, J. A., de Vries, S. M., Tucha, L., & Tucha, O. (2019). Neuropsychological assessment of adults with ADHD: A Delphi consensus study. *Applied Neuropsychology: Adult*, 26(4), 340–354. doi:10.1080/23279095.2018.1429441

Fuermaier, A., Tucha, O., Koerts, J., Butzbach, M., Weisbrod, M., Aschenbrenner, S., & Tucha, L. (2018). Susceptibility of functional impairment scales to non-credible responses in the clinical evaluation of adult ADHD. *The Clinical Neuropsychologist*, 32(4), 671–680. doi:10.1080/13854046.2017.1406143

Goodman, D., Surman, C., Scherer, P., Salinas, G., & Brown, J. (2012). Assessment of physician practices in adult attention-deficit/hyperactivity disorder. *The Primary Care Companion for CNS Disorders*, 14(4), 1–18. doi:10.4088/PCC.11m01312

Gordon, M., Barkley, R., & Lovett, B. (2006). Tests and observational measures. In R. Barkley, ed. *Attention-deficit hyperactivity disorder – A handbook for diagnosis & treatment* (3rd ed., pp. 369–388). New York, New York: The Guilford Press.

Gorlin, E., Dalrymple, K., Chelminski, I., & Zimmerman, M. (2016). Reliability and validity of a semi-structured DSM-based diagnostic interview module for the assessment of Attention Deficit Hyperactivity Disorder in adult psychiatric outpatients. *Psychiatry Research*, 242, 46–53. doi:10.1016/j.psychres.2016.05.020

Green, P. (2003). *The word memory test*. Edmonton/Seattle: Green's Publishing Inc.

Greenberg, L. (2011). *The test of variables of attention*. (Version 2.0). (Computer software.) The TOVA Company: Los Angeles, California.

Greenberg, L., Kindschi, C., Dupuy, T., & Corman, C. (1996). *Test of variables of attention*. Los Alamitos, California: Universal Attention Disorders.

Haavik, J., Halmoy, A., Lundervold, A., & Fasmer, O. (2010). Clinical assessment and diagnosis of adults with attention-deficit/hyperactivity disorder. *Expert Review of Neurotherapeutics*, 10(10), 1569–1580. doi:10.1586/ern.10.149

Halleland, H., Haavik, J., & Lundervold, A. (2012). Set shifting in adults with ADHD. *Journal of the International Neuropsychological Society*, 18(4), 728–737. doi:10.1017/S1355617712000355

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

Harrison, A., Nay, S., & Armstrong, I. (2019). Diagnostic accuracy of the Conners Adult ADHD Rating Scale in a postsecondary population. *Journal of Attention Disorders*, 23(14), 1829–1837. doi:10.1177/1087054715625299

Hauk, L. (2013). AAP releases guidelines on diagnosis, evaluation, and treatment of ADHD. *American Family Physician*, 87(1), 61–62.

Heaton, R. K., Smith, H. H., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, 46(5), 892–900. doi:10.1037//0022-006X.46.5.892

Heidbreder, R. (2015). ADHD symptomatology is best conceptualized as a spectrum: A dimensional versus unitary approach to diagnosis. *ADHD Attention Deficit and Hyperactivity Disorders*, 7(4), 249–269. doi:10.1007/s12402-015-0171-4

Hervey, A., Epstein, J., & Curry, J. (2004). Neuropsychology of adults with attention-deficit/hyperactivity disorder: A meta-analytic review. *Neuropsychology*, *18*(3), 485–503. doi:10.1037/0894-4105.18.3.485

Hinshaw, S., & Scheffler, R. (2014). *The ADHD Explosion: Myths, Medication, Money, and Today's Push for Performance*. New York: Oxford University Press.

Hirsch, O., & Christiansen, H. (2018). Faking ADHD? Symptom validity testing and its relation to self-reported, observer-reported symptoms, and neuropsychological measures of attention in adults with ADHD. *Journal of Attention Disorders*, *22*(3), 269–280. doi:10.1177/1087054715596577

Hirsch, O., & Christiansen, H. (2017). Factorial structure and validity of the quantitative behavior test plus (Qbt+). *Assessment*, *24*(8), 1037–1049. doi:10.1177/1073191116638426

Holst, Y., & Thorell, L. (2017). Neuropsychological functioning in adults with ADHD and adults with other psychiatric disorders: The issue of specificity. *Journal of Attention Disorders*, *21*(2), 137–148. doi:10.1177/1087054713506264

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

Jasinski, L. J., Harp, J. P., Berry, D. T. R., Shandera-Ochsner, A. L., Mason, L. H., & Ranseen, J. D. (2011). Using symptom validity tests to detect malingered ADHD in college students. *The Clinical Neuropsychologist*, *25*(8), 1415–1428. doi:10.1080/13854046.2011.630024

Kamradt, J. M., Ullsperger, J. M., & Nikolas, M. A. (2014). Executive function assessment and adult attention-deficit/hyperactivity disorder: Tasks versus ratings on the Barkely Deficits in Executive Functioning Scale. *Psychological Assessment*, *26*(4), 1095–1105. doi:10.1037/pas0000006

Katz, L., Wood, D., Goldstein, G., Auchenbach, R., & Geckle, M. (1998). The utility of neuropsychological tests in evaluation of attention deficit/hyperactivity disorder (ADHD) versus depression in adults. *Assessment*, *5*(1), 45–51. doi:10.1177/107319119800500107

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E. V. A., … Walters, E. E. (2005). The World Health Organization adult ADHD self-report scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine*, *35*(2), 245–256. doi:10.1017/S0033291704002892

Kessler, R. C., Adler, L., Barkley, R., Biederman, J., Conners, C. K., Demler, O., … Zaslavsky, A. M. (2006). The prevalence and correlates of adult ADHD in the United States: Results from the National Comorbidity Survey Replication. *American Journal of Psychiatry*, *163*(4), 716–723. doi:10.1176/ajp.2006.163.4.716

Kessler, R. C., Green, J. G., Adler, L. A., Barkley, R. A., Chatterji, S., Faraone, S. V., … Van Brunt, D. L. (2010). Structure and diagnosis of adult attention deficit/hyperactivity disorder: Analysis of expanded symptom criteria from the adult ADHD clinical diagnostic scale. *Archives of General Psychiatry*, *67*(11), 1168–1178. doi:10.1001/archgenpsychiatry.2010.146

Kofler, M., Alderson, M., Raiker, J., Bolden, J., Sarver, D., & Rapport, M. (2014). Working memory and intraindividual variability as neurocognitive indicators in ADHD: Examining competing model predictions. *Neuropsychology*, *28*(3), 459–471. doi:10.1037/neu0000050

Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., & Kolomeyer, E. G. (2013). Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clinical Psychology Review*, *33*(6), 795–811. doi:10.1016/j.cpr.2013.06.001

Kooij, J. (2013). *Adult ADHD: Diagnostic assessment and treatment* (3rd ed.). New York, New York: Springer.

Koziol, L., & Stevens, M. (2012). Neuropsychological assessment and the paradox of ADHD. *Applied Neuropsychology: Child*, *1*(2), 79–89. doi:10.1080/21622965.2012.694764

Lange, R., & Lippa, S. (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. doi:10.1080/13854046.2017.1335438

Lewandowski, L. J., Lovett, B. J., Codding, R. S., & Gordon, M. (2008). Symptoms of ADHD and academic concern in college students with and without ADHD diagnoses. *Journal of Attention Disorders*, *12*(2), 156–161. doi:10.1177/1087054707310882

Leimkuhler, M. (1994). Attention deficit disorder in adults and adolescents: Cognitive, behavioral, and personality styles. In J. Ellison, *Psychotherapists guide to neuropsychiatry* (pp. 175–216). New York: Plenum Press.

Lovejoy, D., Ball, J., Keats, M., Stutts, M., Spain, E., Janda, L., & Janusz, J. (1999). Neuropsychological performance of adults with attention deficit hyperactivity disorder (ADHD): Diagnostic classification estimates of measures of frontal lobe/executive functioning. *Journal of the International Neuropsychological Society*, *5*(3), 222–233. doi:10.1017/S1355617799533055

Luty, J., Arokiadass, S., Sarkhel, A., Easow, J., Desai, R., Moorti, O., & 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. doi:10.1097/ADM.0b013e31819343d0

Mapou, R. (2019). Neuropsychological testing is not useful in the diagnosis of ADHD, but…. *The ADHD Report*, *27*(2), 8–12. doi:10.1521/adhd.2019.27.2.8

Marshall, P., Hoelzle, J., Heyerdahl, D., & Nelson, N. (2016). The impact of failing to identify suspect effort in patients undergoing adult attention-deficit/hyperactivity disorder (ADHD) assessment. *Psychological Assessment*, *28*(10), 1290–1302. doi:10.1037/pas0000247

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., & Barker, J. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention-deficit/hyperactivity disorder. *The Clinical Neuropsychologist*, *24*(7), 1204–1237. doi:10.1080/13854046.2010.514290

Martel, M., Nikolas, M., Schimmack, U., & Nigg, J. (2015). Integration of ratings from multiple informants in ADHD diagnosis: A psychometric model with clinical utility. *Psychological Assessment*, *27*(3), 1060–1071. doi:10.1037/pas0000088

McCann, B., 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. doi:10.1176/jnp.12.2.240

Miller, C., Newcorn, J., & Halperin, J. (2010). Fading memories: Retrospective recall inaccuracies in ADHD. *Journal of Attention Disorders*, *14*(1), 7–14. doi:10.1177/1087054709347189

Mostert, J. C., Onnink, A. M. H., Klein, M., Dammers, J., Harneit, A., Schulten, T., … Hoogman, M. (2015). Cognitive heterogeneity in adult attention deficit/hyperactivity disorder: A systemic analysis of neuropsychological measurements. *European Neuropsychopharmacology*, *25*(11), 2062–2074. doi:10.1016/j.euroneuro.2015.08.010

Murphy, K., & Gordon, M. (2006). Assessment of adults with ADHD. In R. Barkley, *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment* (3rd ed., pp. 425–450). New York, New York: Guilford Press.

Murphy, K., Gordon, M., & Barkley, R. (2000). To what extent are ADHD symptoms common? A reanalysis of standardization data from DSM-IV checklist. *ADHD Report*, *8*, 1–5.

Musso, M., & Gouvier, W. (2014). "Why is this so hard?" A review of detection of malingered ADHD in college students? *Journal of Attention Disorders*, *18*(3), 186–201. doi:10.1177/1087054712441970

Nelson, J., & Lovett, B. (2019). Assessing ADHD in college students: Integrating multiple evidence sources with symptom and performance validity data. *Psychological Assessment*, *31*(6), 793–804. doi:10.1037/pas0000702

Nigg, J., Willcutt, E., Doyle, A., & Sonuga-Barke, E. (2005). Causal heterogeneity in attention deficit hyperactivity disorder: Do we need neuropsychologically impaired subtypes? *Biological Psychiatry*, *57*(11), 1224–1230. doi:10.1016/j.biopsych.2004.08.025

Nikolas, M., Marshall, P., & Hoelzle, J. (2019). The role of neuropsychological tests in the assessment of adult attention-deficit disorder. *Psychological Assessment*, *31*(5), 685–698. doi:10.1037/pas0000688

Paris, J., Bhat, V., & Thombs, B. (2015). Is adult attention-deficit hyperactivity disorder being over diagnosed? *The Canadian Journal of Psychiatry*, *60*(7), 324–328. doi:10.1177/070674371506000705

Pazol, R., & Griggins, C. (2012). Making the case for a comprehensive ADHD assessment model on a college campus. *Journal of College Student Psychotherapy*, *26*(1), 5–21. doi:10.1080/87568225.2011.605689

Pettersson, R., Soderstrom, S., & Nilsson, K. (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. doi:10.1177/1087054715618788

Pievsky, M., & McGrath, R. (2018). The neurocognitive profile of attention-deficit/hyperactivity disorder: A review of meta-analyses. *Archives of Clinical Neuropsychology*, *33*(2), 143–157. doi:10.1093/arclin/acx055

Pinto, R., Asherson, P., Ilott, N., Cheung, C., & Kuntsi, J. (2016). Testing for the mediating role of endophenotypes using molecular genetic data in a twin study of ADHD traits. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *171*(7)B, 982–992. doi:10.1002/ajmg.b.32463

Prevatt, F., Proctor, B., Best, L., Baker, L., Van Walker, J., & Wright-Taylor, N. (2012). The positive illusory bias. Does it explain self-evaluations in college students with ADHD? *Journal of Attention Disorders*, *16*(3), 235–243. doi:10.1177/1087054710392538

Ramsey, J. (2015). Psychological assessment of adults with ADHD. In R. Barkley, *Attention-deficit hyperactivity disorder- A handbook for diagnosis & treatment* (4th ed., pp. 475–500). New York, New York: Guilford Press.

Rapport, L. J., Van Voorhis, A., Tzelepis, A., & Friedman, S. R. (2001). Executive functioning in adult attention-deficit hyperactivity disorder. *The Clinical Neuropsychologist*, *15*(4), 479–491. doi:10.1076/clin.15.4.479.1878

Riccio, C., & Reynolds, C. (2006). Continuous performance tests are sensitive to ADHD in adults but lack specificity: A review and critique for differential diagnosis. *Annals of the New York Academy of Sciences*, *931*(1), 113–139. doi:10.1111/j.1749-6632.2001.tb05776.x

Roy-Byrne, P., Scheele, L., Brinkley, J., Ward, N., Wiatrak, C., Russo, J., … Varley, C. (1997). Adult attention-deficit hyperactivity disorder: Assessment guidelines based on clinical presentation to a specialty clinic. *Comprehensive Psychiatry*, *38*(3), 133–140. doi:10.1016/S0010-440X(97)90065-1

Ryan, R., & Hill, S. (2016). *How to GRADE the quality of evidence*. Retrieved from http://cccrg.cochrane.org/author-resources.

Sagar, S., Miller, C., & Erdodi, E. (2017). Detecting feigned attention-deficit hyperactivity disorder (ADHD): Current methods and future directions. *Psychological Injury and Law*, *10*(2), 105–113. doi:10.1007/s12207-017-9286-6

Salthouse, T. (1994). The aging of working memory. *Neuropsychology*, *8*(4), 535–543. doi:10.1037/0894-4105.8.4.535

Schwarz, A. (2013, December 14). The selling of attention deficit disorder. *The New York Times*.

Skodzik, T., Holling, H., & Pedersen, A. (2013). Long term memory performance in adult ADHD: A meta-analysis. *Journal of Attention Disorders*, *20*(10), 1–18. doi:10.1177/1087054713510561

Soderstrom, S., Pettersson, R., & Nilsson, K. (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. doi:10.3109/08039488.2012.762940

Solanto, M. (2015). Executive function deficits in adults with ADHD. In R. Barkley, ed. *Attention-deficit hyperactivity disorder-A handbook for diagnosis & treatment* (4th ed., pp. 256–266). New York, New York: Guilford Press.

Solanto, M., Etefia, K., & Marks, D. (2004). The utility of self-report measures and the continuous performance test in the diagnosis of ADHD in adults. *CNS Spectrums*, *9*(9), 649–659. doi:10.1017/s1092852900001929

Sollman, M., Ranseen, J., & Berry, D. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, *22*(2), 325–335. doi:10.1037/a0018857

Suhr, J., Buelow, M., & Riddle, T. (2011). Development of an infrequency index for the CAARS. *Journal of Psychoeducational Assessment*, *29*(2), 160–170. doi:10.1177/0734282910380190

Suhr, J., Hammers, D., Dobbins-Buckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. The *Archives of Clinical Neuropsychology*, *23*(5), 521–530. doi:10.1016/j.acn.2008.05.003

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. doi: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. doi:10.1016/j.ridd.2010.12.036

Toplak, M., West, R., & Stanovich, K. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? The *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143. doi:10.1111/jcpp.12001

Tucha, L., Fuermaier, A., Koerts, J., Groen, Y., & Thome, J. (2015). Detection of feigned attention deficit hyperactivity disorder. *Journal of Neural Transmission*, *122*(S1), 123–s134. doi:10.1007/s00702-014-1274-3

Ustun, B., Adler, L., Rudin, C., Faraone, S., Spencer, T., Berglund, P., … Kessler, R. (2017). The World Health Organization adult attention-deficit/hyperactivity disorder self-report screening scale for DSM-5. *Jama Psychiatry*, *74*(5), 520–526. doi:10.1001/jamapsychiatry.2017.0298

Valo, S., & Tannock, R. (2010). Diagnostic instability of DSM-IV ADHD subtypes: Effects of informant source, instrumentation, and methods for combining symptom reports. *Journal of Clinical Child & Adolescent Psychology*, *39*(6), 749–760. doi:10.1080/15374416.2010.517172

van de Glind, G., van den Brink, W., Koeter, M. W., Carpentier, P. J., van Emmerik-van Oosrtmerssen, K., & Kaye, S. (2013). Validity of the adult ADHD Self-Report Scale (ASRS) as a screener for adult ADHD in treatment seeking substance abuse disorder patients. *Drug and Alcohol Dependence*, *132*(3), 587–596. doi:10.1016/j.drugalcdep.2013.04.010

Van Voorhees, E. E., Hardy, K. K., & Kollins, S. H. (2011). Reliability and validity of self- and other-ratings of symptoms of ADHD in adults. *Journal of Attention Disorders*, *15*(3), 224–234. doi:10.1177/1087054709356163

Vaughn, A., & Hoza, B. (2013). The incremental utility of behavioral rating scales and a structured diagnostic interview in the assessment of attention-deficit/hyperactivity disorder. *Journal of Emotional and Behavioral Disorders*, *21*(4), 227–239. doi:10.1177/1063426611427910

Victor, T., Boone, K., Serpa, J., Buehler, M., & Ziegler, E. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, *23*(2), 297–313. doi:10.1080/13854040802232682

Walker, A., Shores, A., Trollor, J., Lee, T., & Sachdev, P. (2000). Neuropsychological functioning in adults with attention deficit hyperactivity disorder. *Journal of Clinical and Experimental Neuropsychology*, *22*(1), 115–124. doi:10.1076/1380-3395(200002)22:1;1-8;FT115

Wallace, E., Garcia-Willingham, N., Walls, B., Bosch, C., Balthrop, K., & Berry, D. (2019). A meta-analysis of malingering detection measures for attention-deficit/hyperactivity disorder. *Psychological Assessment*, *31*(2), 265–270. doi:10.1037/pas0000659

Wender, P. (1997). Attention deficit hyperactivity disorder in adults: A wide view of a widespread condition. *Psychiatric Annals*, *27*(8), 556–562. doi:10.3928/0048-5713-19970801-08

Weyandt, L., & DuPaul, G. (2013). *College Students with ADHD*. New York: New York: Springer.

Willcutt, E. (2015). Theories of ADHD. In R. Barkley, ed. *Attention-deficit hyperactivity disorder- A handbook for diagnosis & treatment* (4th ed., pp. 391–404). New York, New York: Guilford Press.

Willcutt, E., Doyle, A., Nigg, J., Faraone, S., & Pennington, B. (2005). Validity of the executive function theory of attention-deficit hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, *57*(11), 1336–1346. doi:10.1016/j.biopsych.2005.02.006

Woo, B., & Rey, J. (2005). The validity of the DSM-IV subtypes of attention-deficit/hyperactivity disorder. *Australian & New Zealand Journal of Psychiatry*, *39*, 344–353. doi:10.1080/j.1440-1614.2005.01580.x

World Health Organization. (2012). *Measuring health and disability: Manual for WHO Disability Assessment Schedule (WHODAS 2.0)*. Geneva: World Health Organization, 2010.

Zimmerman, M., Gorlin, E., Dalrymple, K., & Chelminiski, I. (2017). A clinically useful screen for attention-deficit/hyperactivity disorder in adult psychiatric outpatients. *Annals of Clinical Psychiatry*, *29*(3), 160–166.

Zucker, M., Morris, M., Ingram, S., Morris, R., & Bakeman, R. (2002). Concordance of self- and informant-ratings of adult's current and childhood attention-deficit/hyperactivity disorder symptoms. *Psychological Assessment*, *14*(4), 379–389. doi:10.1037//1040-3590.14.4.379