Reliability and Validity of Self- and Other-Ratings of Symptoms of ADHD in Adults

Elizabeth E. Van Voorhees[1], **Kristina K. Hardy**[1], and **Scott H. Kollins**[1]
[1]Duke University Medical Center

Abstract

Objective—Few studies have examined concordance between raters of ADHD symptoms in adults; there is less information on how well rating scales function in distinguishing adult ADHD from other disorders. This study examined these variables using the Conners Adult ADHD Rating Scales (CAARS).

Method—The sample included 349 adults evaluated for attention problems. Correlations and kappa values were calculated using self- and observer-ratings of item-level symptoms; sensitivity, specificity, and discriminant validity of cluster scores in predicting clinician diagnoses were computed for 269 participants.

Results—Item-level concordance rates ranged from slight to fair. Cluster scores demonstrated a poor balance of sensitivity and specificity in predicting ADHD diagnosis; a high percentage of participants with internalizing disorders had scores in the clinical range.

Conclusion—Self- and observer- ratings on the CAARS provide clinically relevant data about attention problems in adults, but the instrument does not effectively distinguish between ADHD and other adult psychiatric disorders.

Keywords

Adult ADHD; reliability; rating scales; sensitivity; specificity

---

ADHD was initially conceptualized as a childhood disorder, a growing body of evidence suggests that the symptoms and impairment persist into adulthood (Barkley, Fischer, Smallish, & Fletcher, 2002; Barkley, Murphy, & Fischer, 2008; Davidson, 2008; Spencer, Biederman, Wilens, & Faraone, 1998). Diagnosis of the disorder in adults may be complicated: it may be difficult to establish the presence of symptoms in childhood through retrospective reports (March, Wells, & Conners, 1995; Shaffer, 1994); and the potential exists that symptoms of inattention or hyperactivity may be more accurately explained by any number of disorders (i.e. mood, anxiety, substance use, or sleep) that commonly present in adulthood. Though current clinical practice guidelines suggest that diagnostic evaluation should include a comprehensive interview and self- and observer- rating scales, limited information is available on how best to integrate discrepant data (AACAP, 1997).

Corresponding Author: Elizabeth E. Van Voorhees, 718 Rutherford St., Durham, NC 27710, e.vanv@duke.edu.

Specifically, few studies have examined concordance between different raters of ADHD symptoms in adults or the degree to which information provided by each rater contributes to differential diagnosis; there is even less information as to how well rating scales function in distinguishing adult ADHD from other commonly diagnosed adult disorders.

With respect to rater concordance, correlations between self- and observer-ratings of current adult ADHD symptoms have ranged from $r = 0.64$ to $r = 0.75$ (Barkley et al., 2008; Magnusson et al., 2006; Murphy & Barkley, 1996; Murphy & Schachar, 2000; Zucker, Morris, Ingram, Morris, & Bakeman, 2002). The most comprehensive study to date examined the reliability and validity of three adult ADHD rating scales in a sample of 120 adult outpatients with ADHD recruited from a psychomedical center in the Netherlands (Kooij et al., 2008). Patients and partners completed the ADHD Rating Scale (ADHD-RS), the Brown Attention-Deficit Disorder Scale (BADDS), and the Conners' Adult ADHD Rating Scale, long version (CAARS-LV); investigators also completed the ADHD-RS. Agreement between patient and partners across the various cluster scores of the three questionnaires ranged from $r = .386$ to $r = .637$. Patient-investigator ratings on the ADHD-RS were higher than partner-investigator ratings on both Inattentive ($r = .348$ vs. $r = .238$) and Hyperactive-Impulsive ($r = 0.440$ vs. $r = .242$) symptoms. Kappa values indexing concordance between patients and observers at the item level have ranged from moderate to poor (Downey, Stelson, Pomerleau, & Giordani, 1997; Zucker et al., 2002). These finding have led researchers to conclude that different scales may be appropriate to different assessment situations, as each appeared to contribute unique information, but that the relationship between rating scale, informant, and symptom presentation in adult ADHD has yet to be disentangled (Kooij et al., 2008).

The degree to which raters agree on symptoms is important in understanding the presence of symptoms and impairment, but patient/observer agreement may be less pertinent to diagnostic relevance. One study examining functional impairment associated with adult ADHD found that both "talks excessively" and "loses things necessary for tasks or activities" were endorsed with sufficient frequency in non-ADHD controls to preclude these items from being particularly informative in discriminating between the groups (Murphy & Barkley, 1996). Rating scales, in particular, may be limited in their ability to discriminate ADHD from other adult disorders. In one study of 82 adults presenting to a university-affiliated specialty clinic for evaluation for ADHD, 38 patients were ultimately diagnosed with ADHD and 44 received non-ADHD primary diagnoses, including Major Depressive Disorder, Bipolar Disorder, anxiety disorders, and substance abuse/dependence disorders. Clinicians diagnosed participants in this study based on a synthesis of information drawn from self-report, corroborating documents, and interviews of family members, and these diagnoses were compared to the results of three self-rating scales: the Adult Rating Scale (ARS), the Attention-Deficit Scale for Adults (ADSA), and the Symptom Inventory for ADHD (SI-ADHD). Depending upon criterion cutoff scores, false positives on the questionnaires were 36.4% on the ADSA, 46.5% on the SI-ADHD, and 67.4% on the ARS. Furthermore, between 56.5% and 73.9% of individuals who were diagnosed with Major Depressive Disorder or Dysthymia but not with ADHD produced false positives on these inventories (McCann & Roy-Byrne, 2004). A second study examining the sensitivity and specificity of the Brown ADD Scale and the CPT concluded that there was little evidence to suggest that either measure could contribute meaningfully to a differential diagnosis between ADHD and internalizing disorders (Solanto, Etefia, & Marks, 2004). Finally, Barkley and colleagues (2008) examined self-report of symptoms on the ADHD-RS in 146 participants referred to an adult ADHD specialty clinic and ultimately diagnosed with ADHD; 97 "clinical controls" referred to the same clinic but not diagnosed with ADHD; and 109 nonreferred "community controls." Self-rated symptoms reliably differentiated the community control group from both the clinical control and the ADHD groups, but did not

distinguish between the ADHD group and the clinical control group. This lack of discriminative validity between the two clinical groups was particularly pronounced in women (Barkley et al., 2008).

In sum, clinical analysis and synthesis of the range of interview, self-report, and collateral data is critical for accurate differential diagnosis of ADHD in adults. Research to date suggests that self- and observer-rating scales may each contribute uniquely to the determination of diagnosis and impairment, yet limited data regarding the reliability and discriminative validity of these scales makes is difficult for clinicians to determine their most appropriate use in the diagnostic decision-making process. As such, more data are needed not only on the association between self- and other-ratings of ADHD symptoms in adults, but also on the degree to which these symptoms are associated with clinician-determined symptoms and overall clinical diagnosis.

To begin to address this gap in the literature, this investigation was designed to examine the following: (a) the concordance of *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) (*DSM-IV*) criteria-specific ADHD symptoms between self- and observer-reports on the CAARS-LV; (b) the concordance of ADHD symptoms rated by self and by observer on the CAARS-LV with clinician-rated symptoms using Conners' Adult ADHD Diagnostic Interview (CAADID); (c) the association of self and observer ratings on the CAARS with ADHD diagnosis; and (d) the association of self and observer ratings on the CAARS with diagnoses other than ADHD.

Method

Participants

This sample included 349 adults ages 18–70 years referred to a medical center-affiliated ADHD clinic for evaluation for attentional difficulties. College-age participants were over-represented in the sample: while the mean age was 32 and the median age was 28, the modal age was 20. Participants rated their own behaviors on the Conners' Adult ADHD Rating Scale-Self: Long Version (CAARS-S). The Conners' Adult ADHD Rating Scale–Observer: Long Version (CAARS-O) was completed for all participants by friends ($n = 111$), parents ($n = 49$), spouses ($n = 115$), or others ($n = 74$). Gender data was not available for 8.8% of the sample; of those with available data, 38.5% were women. Race/ethnicity data were not available for 22.2% of the sample; of those with available data, 86.4% were Caucasian, 5.1% were African American, 1.8 % were Hispanic, 2.9 % were Asian, and 3.7% were biracial or "other."

A subset of 269 adults underwent a thorough ADHD assessment by a doctoral-level clinician. Diagnoses were determined through a synthesis of the following data: CAARS-S and CAARS-O; computerized Structured Clinical Interview for the *DSM-IV* (SCID); the Conners' Adult ADHD Interview for the *DSM-IV* (CAADID), Parts I and II; semi-structured clinical interview; and when available, psycho- educational test results, medical records, and school records. The primary diagnoses for these 269 participants were as follows: 26.3% ADHD, Combined Type; 33.1% ADHD, Predominately Inattentive Type; 8.9% ADHD, Not Otherwise Specified (NOS); 7.1% mood disorder (depression, dysthymia, bipolar disorders); 7.5% anxiety disorders (Generalized Anxiety Disorder [GAD], Posttraumatic Stress Disorder [PTSD], Anxiety NOS); 5.0% adjustment disorders; 1.4% substance use disorders; 6.1% Other Axis I diagnosis, and 4.6% no diagnosis. Demographic data of this subsample were similar to the larger sample.

Measures

Conners' Adult ADHD Rating Scales, Long Version (CAARS) (Conners, Erhardt, & Sparrow, 1999)—Both the self- and observer-rating forms of the CAARS were used in this investigation; the two versions are identical except that they are normed separately. The long version of the CAARS consists of 66 statements rated on a 0 to 3 scale with 0 = *not at all, never* to 3 = *very much, very frequently*. Eight cluster scores are derived from these items. Three of these scales correspond exactly to *DSM-IV* diagnostic criteria for ADHD and its subtypes: *DSM-IV* Inattentive; *DSM-IV* Hyperactive Impulsive; and *DSM-IV* Total. The other five scales index behaviors that have been found to be associated with ADHD but are not specifically defined in the *DSM-IV* criteria: Inattentive; Hyperactive; Impulsive; Problems with Self-Concept; and Conners' ADHD Index. *T*-scores are calculated for each scale based on age and gender. Test–retest reliability has been found to be acceptable, and the measure has been shown to be valid in distinguishing individuals with ADHD from healthy controls (Erhardt, Epstein, Conners, Parker, & Sitarenios, 1999).

Conners' Adult ADHD Diagnostic Interview (CAADID; Epstein, Johnson, & Conners, 2001)—This measure takes the form of a semi-structured interview that methodically and thoroughly records the age of onset, presence, persistence, and severity of each of the 18 potential ADHD symptoms. Prior to the interview patients complete CAADID Part I, a questionnaire that collects developmental information; information about academic, family, occupational, and personal functioning; and psychiatric history. Part II is the interview portion, which is administered by a clinician; this section assesses each symptom of ADHD in adulthood and in childhood, asking about specific examples of symptom manifestation at the different developmental levels. Kappa values for overall diagnosis using the CAADID have been found to be in the fair to good range for both adult and childhood symptoms, and concurrent validity has been reported for adult hyperactive-impulsive and for childhood inattentive symptoms (Epstein & Kollins, 2006).

Data Analyses

Missing Data Analyses for the CAARS Scales

SPSS 15.0 was used for all analyses. Less than 5% of the response data were missing for 17/18 of the *DSM-IV* symptom items for both self- and observer-report; the one exception to this was observer report of "leaves seat when not supposed to," for which 6.3% of the responses were missing.

Descriptive Analyses

Item-level frequency analyses were conducted using the entire sample (*n* = 349) for CAARS items that are directly analogous to *DSM-IV* ADHD symptoms; for these analyses, a symptom was considered to be absent if it was rated 0 or 1, and present if it was rated 2 or 3. Descriptive statistics including frequency of clinician-endorsed symptoms on the CAADID, and means of each of the CAARS cluster scores for self- and observer-report were computed using the portion of the sample (*n* = 269) upon which full diagnostic evaluations were conducted in our clinic.

Analyses of Rater Concordance

Rater concordance for the CAARS was examined at both the item and cluster levels. Pearson correlations and kappa values for self- and observer-ratings were calculated for items corresponding to *DSM-IV* symptoms. Separate Pearson correlations were calculated between self-ratings and those of each of the most frequently represented observers: spouses (*n* = 15), friends (*n* = 111), and parents (*n* = 49). *Z*-scores were calculated and compared to

determine if significant differences existed in concordance rates between self-ratings and those of the three different groups of observers on each of the symptom-specific items.

Convergent and Discriminant Validity

Sensitivity reflects the proportion of cases in which the presence of the disorder is correctly identified; an index with a high sensitivity may be understood as having a low Type II error rate in detecting the disorder. Specificity, on the other hand, reflects the proportion of cases in which the absence of the disorder is correctly identified; an index with a high specificity may be seen as having a low Type I error rate.

Convergent and discriminant validity for the presence or absence of ADHD— Sensitivity and specificity were calculated for CAARS cluster scores. For these analyses, the criterion variable was the clinician's diagnosis based on all available information, and the predictor variable reflected the presence (1) or absence (0) of clinically relevant scale elevations (T 65) based on self- or observer-ratings. Crosstabs analyses were used to identify the number of cases for which the cluster score was in the clinical range and for which the clinician ultimately diagnosed the patient with ADHD (*true positives*); and to identify the number of cases for which the cluster score was *not* in the clinical range and the clinician did *not* ultimately diagnoses the patient with ADHD (*true negatives*). Percentages were then calculated by dividing number of true positives (*sensitivity*) or the number of true negatives (*specificity*) by the total sample.

As a second measure of convergent and discriminant validity, sensitivity, and specificity were also calculated using a cut-off score based on the *DSM-IV* symptom-specific items. In these analyses, clinician diagnosis remained the criterion variable, and the predictor variable was the presence (1) or absence (0) of 6 symptoms of Inattention, 6 symptoms of Hyperactivity/Impulsivity, or 6 symptoms of both Inattention and Hyperactivity/ Impulsivity.

Divergent validity for ADHD versus other adult psychopathology— The ability of the cluster scores to discriminate between ADHD and other adult psychopathology was examined. Crosstabs analyses were used to identify the number of cases for which the CAARS cluster scores was in the clinical range (T 65) based on self- and observer-ratings, but the clinician diagnosed the patient with a primary mood disorder, a primary anxiety disorder, or another diagnosis on Axis I; percentages were calculated based on these values.

Finally, mean scores within each of the three diagnosis groups (ADHD, mood disorders, and anxiety disorders) were calculated for each cluster for self- and observer-ratings. ANOVA was used to compare among the three diagnostic groups on each cluster score for self- and observer-ratings, and post hoc Bonferroni tests were used for pair wise comparisons.

Results

Descriptive Analyses

Symptom ratings across reporters were high in this clinical sample. Table 1 presents the frequency and frequency ranking of each symptom as endorsed by self and by observer on the CAARS, and as endorsed by clinician on the CAADID. Symptoms were more frequently rated as present by patients than by observers; clinician ratings were variable, and did not appear to be more consistent with either self or observer reports across items. Overall, inattentive symptoms were more commonly reported than hyperactive/impulsive symptoms. Frequency rankings were similar across patients and observers, but clinician rankings differed somewhat from both groups.

The three most commonly reported symptoms by both patients and observers were "has trouble keeping attention focused when working or at leisure" (self = 85.9%, other = 72.1%); "appears distracted when things are going on around him/her" (self = 87.3%, other = 65.8%); and "has trouble finishing job tasks or schoolwork" (self= 80.1%, other = 71.8%). The two least commonly reported were "leaves seat when not supposed to" (self = 26.0%, other = 19.5%) and "trouble doing leisure activities quietly" (self = 36.5%, other = 23.7%).

Consistent with self and observer ratings, "difficulty sustaining attention" (81.9%) and "distractible" (87.8%) were among the symptoms most commonly endorsed by clinicians. Most inconsistent with self- and observer-ratings, however, were clinicians' more frequent endorsement of "avoids tasks requiring mental effort" (clinician = 79.3%, self = 51.2%, other = 46.1%) and of "difficulty remaining seated" (clinician = 47.8%, self = 26.0%, other =19.5%).

Cluster scores for the CAARS were available for the participants who undertook the comprehensive diagnostic evaluation in our clinic ($n = 269$). The score distributions for all of the self- and observer-rated symptoms were roughly normally distributed, with the exception of observer ratings of Problems with Self Concept: these scores produced three distinct peaks at about $T = 50$ (average), $T = 65$ (above average), and $T = 75$ (very much above average). Consistent with the item-level frequency reporting, cluster scores based on self-ratings were generally higher than those based on observer-ratings; this was especially evident for the *DSM-IV* Inattentive Symptoms cluster and for the *DSM-IV* Index cluster.

Analyses of Rater Concordance

Pearson correlations and kappa values were calculated for self and observer ratings for each item corresponding to *DSM-IV* symptoms ($n = 349$). Pearson correlations were also calculated for the three most frequently represented observers: spouses ($n = 115$); friends ($n = 111$), and parents ($n = 49$). Results are presented in Table 2. Self-observer correlations ranged from $r = .24$ ("distractible") through $r = .46$ ("on the go/driven by a motor"). The average correlation for inattentive symptoms was $r = .33$; for hyperactive/impulsive symptoms was $r = .39$; and overall was $r = .36$. Correlations between self and the various raters were similar; the only item for which there was a significant difference among raters was "difficulty awaiting turn" for self-parent ($r = .61$) versus self-friend ($r = .34$) ($z = 2.0$, $p < .05$). Kappa values on individual items ranged from $\kappa = .11$ ("distractible") to $\kappa = .37$ ("loses things").

Concordance was higher at the level of symptom clusters. Correlation between self- and observer-ratings was slightly higher for the *DSM-IV* Hyperactive/Impulsive symptom cluster index ($r = .48$) than for the *DSM-IV* Inattentive ($r = .39$) and *DSM-IV* Total clusters ($r = .34$). The Conners' factor cluster scores demonstrated greater concordance among self- and observer-ratings, with correlations falling between $r = .51$ and $r = .59$.

Convergent and Discriminant Validity

Convergent and discriminant validity for the presence or absence of ADHD—
Sensitivity of cluster scores based on self-ratings varied widely, with the *DSM-IV* Inattentive Symptoms Index providing the greatest sensitivity (0.95) and the Impulsivity/ Emotional Lability Index providing the least (0.39); see Table 3. However, these sensitivities were offset by the corresponding specificities: specificity of the *DSM-IV* Inattentive Symptoms Index was the lowest among the clusters (0.13), and the specificity of the Impulsivity/Emotional Lability Index was among the highest (0.75). The Conners' ADHD Index was the most effective in detecting both the true presence and true absence of the disorder, with a sensitivity of 0.65 and a specificity of 0.61.

To examine the iterative value of adding observer-ratings to self-ratings in predicting clinical diagnosis, sensitivity and specificity were recalculated based on scale elevations above $T = 65$ on both self and observer reports. The combination of self- and observer-ratings reduced the sensitivity of the scales to between 0.21(Impulsivity/Emotional Lability) and 0.63 (*DSM-IV* Inattentive Symptoms), but increased the specificity to above 0.70 for all but one of the indices. The *DSM-IV* Inattentive Symptoms Index continued to demonstrate a poor level of specificity (0.54) even when both reporters endorsed symptoms in the clinical range. Similarly, the two indices that maintained the highest specificity when both self and observer reports were in the clinical range (Hyperactivity/Restlessness Index [0.92] and Impulsivity/Emotional Lability Index [0.90]), also had the lowest sensitivity (0.26 and 0.21, respectively).

As a second measure of convergent and discriminant validity, the number of symptoms rated as present on the CAARS was compared with diagnosis as determined by the clinician. For these analyses ADHD was considered to be present even if it was not listed as the primary diagnosis; a mood disorder, anxiety disorder, or other disorder was identified as the primary diagnosis only in the absence of ADHD. Results indicated that of 210 individuals identified as having ADHD based on self-ratings of symptoms, 75.2% actually had the disorder, 6.7% had a primary mood disorder, 6.2% had a primary anxiety disorder, and 11.9% had another disorder or no disorder on Axis I. When both self- and observer-endorsement of symptoms was required, the number of individuals positive for ADHD based on CAARS ratings dropped to 131; of these, 81.7% actually had ADHD; 3.8% had a primary mood disorder; 6.1% had a primary anxiety disorder; and 8.4% had another diagnosis or no diagnosis on Axis I. In sum, adding observer data to self-report was most useful in diminishing the chance of incorrectly diagnosing a depressed individual with ADHD.

Divergent validity for ADHD versus other adult psychopathology—To examine the discriminant validity of the cluster ratings in identifying problems specific to ADHD, we calculated the percentage of individuals with mood disorders, anxiety disorders, and other/ no disorders that produced clinically-elevated cluster scores on the CAARS self-rating scales. We then recalculated these percentages for those with both self and observer cluster scores in the clinical range to examine the degree to which including collateral rating-scale data helped to specify the presence of ADHD. Findings are presented in Table 4.

Cluster scores based on self-ratings varied widely in their degree of specificity to ADHD diagnosis. Elevated scores on the Hyperactivity/Restlessness Index were least likely to be associated with disorders other than ADHD: clinical elevations on this scale were found in 24% of those with mood disorders, 13% of those with anxiety disorders, and 17% of those with other disorders or with no disorder. On the other hand, *DSM-IV* Inattentive Symptoms Index were least specific to ADHD: clinical elevations on this scale were observed in 94% of those with mood disorders, 87% of those with anxiety disorders, and 86% of those with other disorders or with no disorder. When observer-ratings in the clinical range were included, the pattern of association was similar but the percentages were lower. That is, clinical elevations on both self- and observer-ratings of the Hyperactivity/Restlessness Index remained lowest and were found in only 6% of those with mood disorders, 7% of those with anxiety disorders, and 9% of those with other/no disorder; clinical elevations in both raters on the *DSM-IV* Inattentive Symptoms Index remained highest and were found in 41% of those with mood disorders, 67% of those with anxiety disorders, and 39% of those with other/no disorder.

Finally, examination of the mean cluster scores of individuals with ADHD, mood disorders, and anxiety disorders confirmed that these scales were not effective at differentiating between ADHD and mood disorders. Whereas mean scores on self-ratings were

significantly different between ADHD and primary anxiety disorder on four of the eight scales, there were no significant differences between ADHD and primary mood disorders on any of the self-rated scales. With respect to observer ratings, the only significant difference observed was between ADHD and primary mood disorders on the Hyperactivity/ Restlessness Index; see Table 5.

Discussion

This investigation examined the reliability and construct validity of self- and observer-ratings on the CAARS using a large sample of adults referred to a university-affiliated ADHD clinic for assessment of attention problems. Our goal was to provide information that would help clinicians integrate data from multiple informants in the assessment of adult ADHD. Measures of concordance between self- and observer-ratings on the CAARS were examined at the item-level and at the level of cluster scores. Measures of concordance were also calculated between self- and observer-ratings on the CAARS, and clinician-identified symptoms on the CAADID. Using the diagnosis determined by the clinician through a comprehensive evaluation as the "gold standard," the sensitivity and specificity of CAARS self- and observer-rated cluster T-scores in detecting ADHD was evaluated. Finally, divergent validity was assessed by examining the rate at which item-level symptom counts and cluster T-scores in the clinical range for ADHD were also observed in other psychological disorders.

The majority of the participants in this sample (68.3%) were diagnosed with some form of ADHD; the large number of individuals with ADHD in our sample allowed us to examine concordance among raters to see how our findings compare with those previously reported. In addition, 27.1% of participants in our sample were found to have primary diagnoses other than ADHD on Axis I, and 4.6% had no Axis I diagnosis. Having a substantial percentage of our sample who presented for assessment of attention problems, but who were ultimately determined to have disorders other than ADHD, also allowed us to explore the discriminant validity of the CAARS within a sample of high-risk adults in an outpatient setting.

Across reporters, among the most frequently endorsed inattentive symptoms in this sample were "distractible" "difficulty sustaining attention," and "doesn't follow instructions or follow through"; this is consistent with previous findings of self- and observer- reports of adults with ADHD, as well as with symptoms commonly reported in control samples (Downey et al., 1997; Murphy & Barkley, 1996).

Also consistent with previous findings, concordance between self- and observer-ratings of individual items ranged from slight to fair, with κ values ranging from 0.11 to 0.37 (Downey et al., 1997; Zucker et al., 2002). Interestingly, those items with the lowest κ values were also those that were most frequently reported: "distractible" $\kappa = .11$; "difficulty sustaining attention" $\kappa = .15$; and "doesn't follow instructions or follow through" $\kappa = .20$. This may suggest that these characteristics are sufficiently common as to be difficult for reporters to reliably assess across situations (Murphy & Barkley, 1996). On the cluster indices, however, correlations between self- and observer-ratings rose to between $r = .39$ and $r = .59$; these concordance rates were similar to those previously reported by Kooij and colleagues (2008).

Among the different observers (friends, spouses, and parents), there was only one significant difference in agreement with self-ratings on symptom-specific items; this suggests that at least with respect to the core symptoms, various observers are likely to provide equally relevant data. Though some researchers have suggested that supervisors or others in a position to evaluate individuals in a structured setting may provide unique data (Belendiuk, Clarke, Chronis, & Raggi, 2007), others have found that employers actually tended to report

lower levels of ADHD symptomatology than either participants or other observers (Barkley et al., 2008). Unfortunately there were too few supervisors or coaches represented in our sample of observers to address this issue.

Participants generally reported greater symptomatology than did observers; this was reflected in a consistently higher frequency of *DSM-IV* symptom endorsement at the item-level, as well as in higher mean *T*-scores on all CAARS clusters. On the other hand, the frequency ranking of symptoms was fairly consistent between self and observer. This consistency in frequency ranking was not preserved in clinician ratings of the *DSM-IV* items, however: clinicians were more likely to conclude that avoidance of tasks requiring sustained mental effort was present than were either participants or observers, and they were less likely to conclude that being "on the go" or acting as if "driven by a motor" was present. This suggests that participants and observers may interpret the behaviors described on the rating scales similarly, and may be relatively close in the degree to which they believe these behaviors are common in the general population. However, for at least some of the items that are central to the diagnosis of ADHD, participants and observers appear to differ from clinicians in their conceptualization of the behavior or in their estimation of how "normal" it is. Given that one of the difficulties in diagnosing ADHD lies in determining how extreme the behavior is in relation to developmental and cultural norms, these findings suggest that it may be critical for clinicians to elicit concrete examples of behaviors and clear ratings of their frequency (i.e. daily, several times a day) in determining the degree to which a putative symptom is actually maladaptive. This conclusion is similar to Murphy and Barkley's (1996) observation that several of the *Diagnostic and Statistical Manual of Mental Disorders–Revised* (4th ed.) (*DSM-III-R*) criteria for ADHD were endorsed too frequently on both self- and observer-ratings scales to provide an adequate basis for discriminating between those with and without the disorder.

Clinically-elevated cluster *T*-scores demonstrated a relatively poor balance of sensitivity and specificity in predicting ADHD diagnosis. While self-rated *DSM-IV* Inattentive Symptoms *T*-scores were sensitive in detecting the presence of ADHD, they were not specific. On this index, even when both self- and observer-ratings were in the clinical range the specificity only improved to 0.54, while sensitivity fell to 0.63. Only *T*-scores in the clinical range on Conners' Index allowed for sensitivity and specificity above 0.60. As such, clinically-elevated *T*-scores on the CAARS clusters are relatively limited in the information they contribute to differentiating ADHD from other psychiatric disorders that commonly manifest in adulthood.

This lack of sensitivity and specificity in detecting ADHD in individuals with attention problems was underscored by the high percentage of participants with mood or anxiety disorders who produced cluster *T*-scores in the clinical range. *DSM-IV* Inattentive Symptoms were particularly highly elevated across groups, with 94% of those with mood disorders and 87% of those with anxiety disorders scoring at or above $T = 65$. The Conners' Hyperactivity/Restlessness Index was least represented across groups: only 24% of those with mood disorders and 13% of those with anxiety disorders scored in the clinical range. While including information from observer-ratings reduced the frequency with which clinically-elevated cluster scores were represented in these groups, 4 of the 8 indices were found to be at or above $T = 65$ in more than 25% of those with mood or anxiety disorders who did not have ADHD. Finally, an examination of mean cluster scores in those with ADHD, primary mood disorders, and primary anxiety disorders demonstrated that individuals with mood disorders are especially likely to be indistinguishable from those with ADHD on the CAARS.

Taken together, these findings suggest that while the CAARS is appropriate for screening for the presence of attention problems to determine whether or not a more thorough evaluation is necessary (Conners et al., 1999), it may be misleading if given substantial weight in the determination of differential diagnosis. This conclusion is supported by studies in smaller samples of adults with depression, anxiety, or substance use disorders that have reported cluster scores comparable to those found in ADHD (Belendiuk et al., 2007; Cleland, Magura, Foote, Rosenblum, & Kosanke, 2006; Solanto et al., 2004). Diagnosing ADHD in adults may require different clinical skills than diagnosing the disorder in children, both because the symptoms may manifest differently, and because attention problems are common to many disorders that peak in adolescence and adulthood.

On the other hand, it is possible that there is a subset of the items on the CAARS that would more effectively distinguish between ADHD and other adult psychiatric disorders. For example, in their analysis of the *DSM-IV* ADHD criteria endorsed by adults in clinical interviews, Barkley and colleagues (2008) utilized logistic regression to identify 4 of the current 18 *DSM-IV* criteria that maximally discriminated participants with ADHD from both community and clinical controls; the optimized subset of symptoms correctly classified 86% of ADHD and 47% of clinical control cases (Barkley et al., 2008). Given that in our sample the Conners' ADHD Index was more effective at discriminating between ADHD and other Axis I disorders than were *DSM-IV* cluster scores, it may be possible to use regression in a similar manner to identify a subset of these items that would be maximally sensitive and specific in distinguishing ADHD from other forms of adult psychopathology that impair attention. However, because individuals who are diagnosed with ADHD in adulthood are frequently clinically complex and often present with multiple comorbidities (Barkley et al., 2008), the boundaries and factor structure of ADHD as it manifests in adulthood have yet to be defined. To begin to address this issue, we are currently in the process of conducting factor analyses using a portion of this data set to examine how ADHD in adulthood compares to the child form of the disorder.

Overall these findings point to the need for careful examination of self-reported symptoms of adult ADHD, and particularly of inattentive symptoms, in determining their relevance to a diagnosis. This is especially the case if information about childhood symptoms is unreliable because of lack of access to appropriate reporters, or because of patient difficulty in remembering details about childhood behaviors. Individual symptoms may not be highly concordant across reporters, and the attributions made about inattentive or impulsive behaviors are likely to vary among observers as well as across time and situation. Further, identification of symptoms as "present" or "absent" must be considered carefully in the context of etiology and functional impairment. In sum, the CAARS is an invaluable tool for identifying clinically significant problems with attention, but should be followed by a thorough clinical evaluation to determine differential diagnoses in adults seeking evaluation for ADHD.

Limitations and Future Directions

The findings reported here must be considered in light of several limitations. First, the comprehensive evaluations in our clinic include the CAARS-S and CAARS-O, and thus clinicians' diagnoses were not determined completely independently of these measures. However, the CAARS manual indicates that this measure is appropriately used as a screening measure (Conners et al., 1999), and the data from the CAARS was considered in the context of a wealth of other clinical and collateral information in determining diagnoses. We therefore believe that our conclusions remain relevant to clinicians using this measure as a tool in clinical practice.

Second, participants were drawn exclusively from referrals to a specialty ADHD clinic, and thus they may have been more likely to identify attention problems among their primary complaints than individuals recruited from a more general outpatient psychology or psychiatry setting. As such, our findings with respect to discriminant validity may be limited. On the other hand, almost a third of our sample was *not* diagnosed with ADHD; among those who were, just under half met criteria for at least one comorbid Axis I psychiatric diagnosis other than nicotine dependence. Considering that mood and anxiety disorders are characterized by significant attention problems (Castaneda, Tuulio-Henriksson, Marttunen, Suvisaari, & Lonnqvist, 2008), and that among individuals with ADHD comorbid conditions have been reported to be as high as 80% (Barkley et al., 2008), our findings may maintain some relevance to a wider range of mental health settings.

Another issue with respect to the representativeness of our sample is related to our clinic billing practices: As with many specialty outpatient clinics, our services are necessarily limited to those with some resources (i.e. insurance), and it is unclear how our patients compare to those who are likely to seek evaluations in settings serving a broader population. Given that research has demonstrated that adults with ADHD often present with significant functional deficits that affect job stability and performance (Murphy & Barkley, 1996), our sample may be largely comprised of individuals who have external resources or who have developed compensatory strategies that mitigate the impact of ADHD on their lives. As such, it will be important for future research to examine the psychometric properties of self- and observer-reports for individuals in settings with a broader clinical base (i.e. public health clinics), and for individuals whose lives may have been particularly affected by their symptoms (i.e. unemployment offices, prisons). In fact, despite the evidence that ADHD and substance abuse disorders are often highly comorbid (Biederman, Wilens, Mick, Faraone, & Spencer, 1998; Cleland et al., 2006; Levin, Evans, & Kleber, 1998; Murphy & Barkley, 1996), a relatively small percentage of our patients had current comorbid substance abuse disorders; this, again, supports the possibility that our sample may be comprised largely of a relatively mild subgroup of the larger population of adults with ADHD.

In sum, this study addresses an important gap in the literature by using a large clinical sample to examine the reliability and validity of the CAARS as a measure of ADHD in adults, and it provides unique and invaluable information to clinicians as to how this tool can best be used to determine the presence and diagnostic relevance of attention problems. Our findings suggest that the CAARS is a useful instrument when used as one source of information in an evaluation for ADHD in adults. Self- and observer-reports contribute unique information, and considering these two sources of information together may be useful in developing hypotheses regarding differential diagnoses. However, our findings also underscore the importance of using the CAARS only as a screening tool or as part of a more comprehensive evaluation; the guidelines for diagnosing ADHD in adults remain ambiguous, and clinical judgment will necessarily continue to play a significant role in determining how best to interpret the presence of attention problems in adulthood. As such, both self- and observer- ratings on the CAARS can be helpful in identifying problems and in forming hypotheses as to possible underlying pathology in adults being evaluated for attention problems, but the instrument does not, by itself, distinguish between ADHD and other adult psychiatric disorders.

Acknowledgments

Biographies

Elizabeth E. Van Voorhees, PhD is a Clinical Associate with the Department of Psychiatry at the Duke University Medical Center, and a licensed clinical psychologist conducting treatment and research at the Duke ADHD Program. Her research interests include gender differences in addiction among individuals with ADHD, and the interaction of extreme stress and neurophysiology in the development of substance use disorders.

Kristina Hardy, PhD is a licensed clinical psychologist, Assistant Clinical Professor with the Department of Psychiatry, and the Director of Clinical Services & Training for the Duke ADHD Program. She is also a clinician and researcher with the Division of Pediatric Hematology/Oncology and the Tisch Brain Tumor Center at Duke. Her research interests include the neuropsychological and psychosocial impact of pediatric illness, particularly with regard to attention problems and deficits in social cognition.

Scott H. Kollins, PhD is an Associate Professor of Psychiatry at the Duke Medical Center and the Director of the ADHD Program. He is a licensed clinical psychologist whose research interests are in the areas of psychopharmacology and the intersection of ADHD and substance abuse; he has published more than 60 papers in these areas and currently receives funding from NIMH, NIDA, NINDS, as well as from several industry sources.

References

AACAP. Practice parameters for the assessment and treatment of children, adolescents, and adults with attention-deficit hyperactivity disorder. Journal of the American Academy of Child and Adolescent Psychiatry. 1997; 36:85–121. [PubMed: 9000785]

Barkley RA, Fischer M, Smallish L, Fletcher K. The persistence of attention-deficit/hyperactivity disorder into young adulthood as a function of reporting source and definition of disorder. Journal of Abnormal Psychology. 2002; 111:279–289. [PubMed: 12003449]

Barkley, RA.; Murphy, KR.; Fischer, M. ADHD in adults: What the science says. New York: Guilford; 2008.

Belendiuk KA, Clarke TL, Chronis AM, Raggi VL. Assessing the concordance of measures used to diagnose adult ADHD. Journal of Attention Disorders. 2007; 10:276–287. [PubMed: 17242423]

Biederman J, Wilens TE, Mick E, Faraone SV, Spencer T. Does attention-deficit hyperactivity disorder impact the developmental course of drug and alcohol abuse and dependence? Biological Psychiatry. 1998; 44:269–273. [PubMed: 9715358]

Castaneda AE, Tuulio-Henriksson A, Marttunen M, Suvisaari J, Lonnqvist J. A review on cognitive impairments in depressive and anxiety disorders with a focus on young adults. Journal of Affective Disorders. 2008; 106:1–27. [PubMed: 17707915]

Cleland C, Magura S, Foote J, Rosenblum A, Kosanke N. Factor structure of the Conners Adult ADHD Rating Scale (CAARS) for substance users. Addictive Behaviors. 2006; 31:1277–1282. [PubMed: 16169157]

Conners, CK.; Erhardt, D.; Sparrow, E. Conners' Adult ADHD Rating Scales (CAARS) technical manual. North Tonawanda, NY: Multi-Health Systems; 1999.

Davidson MA. ADHD in adults: A review of the literature. Journal of Attention Disorders. 2008; 11:628–641. [PubMed: 18094324]

Downey KK, Stelson FW, Pomerleau OF, Giordani B. Adult attention deficit hyperactivity disorder: Psychological test profiles in a clinical population. Journal of Nervous and Mental Disease. 1997; 185:32–38. [PubMed: 9040531]

Epstein, JN.; Johnson, D.; Conners, CK. Conners' adult ADHD diagnostic interview for DSM-IV (CAADID) technical manual. North Tonawanda, NY: Multi-Health Systems; 2001.

Epstein JN, Kollins SH. Psychometric properties of an adult ADHD diagnostic interview. Journal of Attention Disorders. 2006; 9:504–514. [PubMed: 16481667]

Erhardt D, Epstein JN, Conners CK, Parker JDA, Sitarenios G. Self-ratings of ADHD symptoms in adults II: Reliability, validity, and diagnostic sensitivity. Journal of Attention Disorders. 1999; 3:153–158.

Kooij JJS, Boonstra AM, Swinkels SHN, Bekker EM, de Noord I, Buitelaar JK. Reliability, validity, and utility of instruments for self-report and information report concerning symptoms of ADHD in adult patients. Journal of Attention Disorders. 2008; 11:445–458. [PubMed: 18083961]

Levin FR, Evans SM, Kleber HD. Prevalence of adult attention-deficit hyperactivity disorder among cocaine abusers seeking treatment. Drug Alcohol Depend. 1998; 52:15–25. [PubMed: 9788002]

Magnusson P, Smari J, Sigurdardottir D, Baldursson G, Sigmundsson J, Kristjansson K, et al. Validity of self-report and informant rating scales of adult ADHD symptoms in comparison with a semistructured diagnostic interview. Journal of Attention Disorders. 2006; 9:494–503. [PubMed: 16481666]

March JS, Wells KW, Conners CK. Attention-deficit hyperactivity disorder, part I: Assessment and diagnosis. Journal of Practical Psychiatry and Behavioral Health. 1995; 1:219–228.

McCann BS, Roy-Byrne P. Screening and diagnostic utility of self-report attention deficit hyperactivity disorder scales in adults. Comprehensive Psychiatry. 2004; 45:175–183. [PubMed: 15124147]

Murphy K, Barkley RA. Attention deficit hyperactivity disorder adults: Comorbidities and adaptive impairments. Comprehensive Psychiatry. 1996; 37:393–401. [PubMed: 8932963]

Murphy P, Schachar R. Use of self-ratings in the assessment of symptoms of attention deficit hyperactivity disorder in adults. American Journal of Psychiatry. 2000; 157:1156–1159. [PubMed: 10873926]

Shaffer D. Attention deficity hyperactivity disorder in adults. American Journal of Psychiatry. 1994; 151:633–638. [PubMed: 7909410]

Solanto MV, Etefia K, Marks DJ. The utility of self-report measures and the continuous performance test in the diagnosis of ADHD in adults. CNS spectrums. 2004; 9:649–659. [PubMed: 15337862]

Spencer T, Biederman J, Wilens TE, Faraone SV. Adults with attention-deficit/hyperactivity disorder: A controversial diagnosis. Journal of Clinical Psychiatry. 1998; 59:59–68. [PubMed: 9680054]

Zucker M, Morris MK, Ingram SM, Morris RD, Bakeman R. Concordance of self- and informant ratings of adults' current and childhood attention-deficit/hyperactivity disorder symptoms. Psychological Assessment. 2002; 14:379–389. [PubMed: 12501563]

Table 1

Frequency and Frequency Rank of ADHD Symptoms as Reported by Self, Observer, and Clinician

Symptom	Self		Observer		Clinician	
	Frequency	Rank	Frequency	Rank	Frequency	Rank
Loses things	59.8	10	59.4	7	60.7	10
Difficulty sustaining attention	85.9	2	72.1	1	81.9	2
Forgetful	71.8	6	63.9	5	65.9	7
Difficulty listening	69.7	7	53.9	9	64.1	8
Makes careless mistakes	74.6	5	60.3	6	73.3	5
Avoids sustained effort tasks	51.2	13	46.1	11	79.3	3
Does not follow through	80.1	3	71.8	2	77.4	4
Easily distracted	87.3	1	65.8	3	87.8	1
Disorganized	76.0	4	65.8	3	71.1	6
Talks too much	48.1	14	35.5	14	30.0	17
Excessively loud	36.5	17	23.7	17	23.3	18
Leaves seat	26.0	18	19.5	18	47.8	13
Trouble waiting	40.3	15	30.1	16	50.4	12
"On the go"	62.4	9	56.6	8	45.2	14
Fidgets	64.6	8	48.0	10	62.2	9
Runs about	53.6	12	43.9	12	57.8	11
Blurts out	57.7	11	36.0	13	44.4	15
Interrupts	38.6	16	30.6	15	39.8	16

Table 2

Concordance of ADHD Symptoms on the Self- and Observer-Reports on the CAARS

Symptom	Overall r (n = 349)	Friend r (n = 111)	Parent r (n = 49)	Spouse r (n = 115)	Overall Kappa
Loses things	0.43	0.37	0.54	0.52	0.37
Difficulty sustaining attention	0.25	0.30	0.26	0.29	0.15
Forgetful	0.34	0.33	0.14	0.45	0.21
Difficulty listening	0.33	0.14	0.24	0.49	0.28
Makes careless mistakes	0.34	0.40	0.26	0.25	0.27
Avoids task requiring sustained mental effort	0.37	0.33	0.45	0.40	0.31
Does not follow through	0.27	0.34	0.18	0.25	0.20
Easily distracted	0.24	0.25	0.31	0.26	0.11
Disorganized	0.37	0.41	0.32	0.43	0.27
Talks too much	0.43	0.31	0.39	0.52	0.33
Excessively loud	0.32	0.39	0.22	0.29	0.27
Leaves seat	0.36	0.46	0.47	0.43	0.29
Trouble waiting	0.43	0.34	0.61	0.40	0.30
"On the go"	0.46	0.47	0.57	0.47	0.34
Fidgets	0.43	0.41	0.50	0.42	0.36
Runs about	0.45	0.49	0.55	0.45	0.36
Blurts out	0.30	0.30	0.12	0.32	0.21
Interrupts	0.33	0.18	0.27	0.48	0.33

Note: CAARS = Conners Adult ADHD Rating Scales.

Table 3

Sensitivity and Specificity of Elevated CAARS Cluster *T*-Scores (*T* > 65) in Predicting of ADHD Diagnosis

Scale	Self		Self & Other	
	Sensitivity	Specificity	Sensitivity	Specificity
Inattention/memory problems	0.79	0.49	0.56	0.78
Hyperactivity/restlessness	0.47	0.83	0.26	0.93
Impulsivity/emotional lability	0.39	0.75	0.21	0.90
Problems with Self-concept	0.44	0.62	0.29	0.77
DSM-IV inattentive symptoms	0.95	0.13	0.63	0.54
DSM-IV hyper/imp symptoms	0.58	0.70	0.27	0.85
Conners' index	0.65	0.61	0.43	0.83
DSM-IV index	0.91	0.27	0.53	0.72

Note: CAARS = Conners Adult ADHD Rating Scales; *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.).

Table 4

Percentage of Individuals Diagnosed With Disorders Other Than ADHD With Clinically-Elevated ($T > 65$) CAARS-LV Cluster Scores

Scale	Self			Self & Other		
	Mood D/Os ($n = 17$)	Anxiety D/Os ($n = 15$)	Other[a] ($n = 36$)	Mood D/Os ($n = 17$)	Anxiety D/Os ($n = 15$)	Other[a] ($n = 33$)
Inattention/memory problems	59	60	42	29	33	12
Hyperactivity/restlessness	24	13	17	6	7	9
Impulsivity/emotional lability	35	33	19	6	13	12
Problems with self-concept	59	33	28	35	27	15
DSM-IV inattentive symptoms	94	87	86	41	67	39
DSM-IV hyper/imp symptoms	29	27	33	18	20	12
Conners' index	53	47	31	12	33	12
DSM-IV index	76	67	75	29	40	24

Note: CAARS-LV = Conners' Adult ADHD Rating Scale, long version; *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.).

[a] Includes the following disorders as primary diagnosis on Axis

I: adjustment; cognitive NOS; disruptive behavior NOS; eating; learning; PDD; substance use; V code; no diagnosis/diagnosis deferred.

Table 5

Mean CAARS Cluster Scale Scores for ADHD, Primary Mood Disorders, and Primary Anxiety Disorders

Scale	Self			Other		
	ADHD (n = 149)	Mood D/Os (n = 17)	Anxiety D/Os (n = 15)	ADHD (n = 144)	Mood D/Os (n = 17)	Anxiety D/Os (n = 15)
Inattention/memory problems	73.37 (11.20)	67.95 (9.25)	65.65 (11.10)	68.81 (11.50)	62.37 (10.89)	63.23 (10.94)
Hyperactivity/restlessness	63.47 (9.85)[*]	56.55 (11.01)	55.35 (10.93)[*]	61.34 (12.71)[*]	51.50 (10.26)[*]	57.60 (11.42)
Impulsivity/emotional lability	62.23 (11.48)	61.08 (14.05)	59.26 (14.18)	58.50 (11.62)	54.35 (11.75)	55.17 (9.35)
Problems with self-concept	62.13 (12.25)	66.74 (10.31)	62.42 (10.42)	60.65 (12.04)	65.88 (12.32)	61.34 (13.60)
DSM-IV inattentive symptoms	85.97 (11.33)[*]	83.34 (12.19)	75.37 (13.15)[*]	68.63 (11.06)	63.52 (11.02)	66.32 (11.41)
DSM-IV hyper/imp symptoms	68.33 (12.60)[*]	61.03 (13.57)	57.60 (13.57)[*]	61.04 (13.32)	55.14 (11.69)	56.00 (10.67)
Conners' index	69.40 (8.93)	67.31 (8.98)	64.34 (9.51)	66.34 (11.47)	61.19 (9.79)	64.86 (11.35)
DSM-IV index	81.04 (11.18)[*]	74.77 (13.17)	68.98 (12.92)[*]	66.81 (11.45)	59.87 (9.97)	62.86 (10.57)

[*] $p < .05$.