Wender Utah Rating Scale: representation of past or present functioning?

Master Thesis
Department of Psychology en Health
Clinical Health Psychology
University of Tilburg

Author: Marianne Hartong
ANR: 116479
First Instructor: Prof. Dr. A.J.J.M. Vingerhoets
Second Instructor: Drs M.H.Y. van der Zee
June 2010
Abstract

Objective: To examine whether (a) the Wender Utah Rating Scale (WURS) is able to distinguish ADHD patients from normal controls based on their reported childhood experiences; (b) the WURS differentiates previously diagnosed ADHD patients based on its cut-off score of 46; and (c) correlation exists between the WURS scores and results on executive function measures.

Method: 23 adults previously diagnosed with ADHD and 50 normal controls completed a number of questionnaires and were administered neuropsychological tests of executive functioning.

Results: WURS differentiated significantly between ADHD patients and normal controls, but the cut-off score of 46 correctly discriminated only 52% of the ADHD patients. While results on neuropsychological tests distinguished between the two groups, there was no significant correlation with the WURS measure.

Conclusion: Results confirm earlier evidence that clinical diagnosis of ADHD should not be made dependent of self reported retrospective childhood behavior as measured by the WURS. Neuropsychological test results endorse the notion that ADHD patients suffer from general cognitive slowing.

Keywords: Wender Utah Rating Scale, ADHD, neuropsychological tests, executive functioning

1. Introduction

Traditionally Attention Deficit Hyperactivity Disorder (ADHD) is considered a childhood disorder with maladaptive functioning in the domains of attention, hyperactivity and impulsivity. With the publication of the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM-IV; American Psychiatric Association, 1994), the domains of hyperactivity and impulsiveness have been combined, and categorisation has resulted in three ADHD subtypes: (1) the predominantly hyperactive-impulsive type, (2) the predominantly inattentive type and (3) a combined type. An essential requisite for the diagnosis of ADHD is that the childhood onset of symptoms leading to impairment must have developed before seven years of age, applicable for both child and adult diagnostics. As acknowledged by Barkley, Murphy & Fischer (2008), first signs of the recognition of the persistence of ADHD into adulthood probably date back approximately 100 years to the perception by George Still (1902, in Barkley et al., 2008) that the (ADHD) symptoms he
identified in children, were of a chronic nature. However, real popularity of adult ADHD as research field only developed towards the end of the last century (Barkley et al., 2008). Reports on the prevalence of ADHD in adults and the extent to which ADHD persists from childhood into adulthood is largely dependent on research criteria, definitions employed and sources of information used (Barkley, Fischer, Smallish, & Fletcher, 2002; Barkley et al., 2008). Prevalence figures range from as low as 3-5% persistence into adulthood to 50% and more on the high end (Barkley et al., 2002; Rasmussen & Gillberg, 2001; Wenwei, 1996). Recent studies of the general population in the United States estimate that 4.1% (Kessler, Chiu, Demler, & Walters, 2005) to nearly 5% (Barkley et al., 2008) of the adult population suffer from ADHD, making ADHD a relatively common mental disorder within the adult population.

Considering the diagnostic criteria for (adult) ADHD, as formulated in the *DSM-IV*, there is considerable controversy concerning the requisite of disorder onset before the age of seven (Mackin & Horner, 2005; McCann, Scheele, Ward, & Roy-Byrne, 2000). In (clinical) practice it is not so much the fact that the disorder should be present from the childhood years that causes debate, but rather whether the age of seven is an adequate demarcation point and how this fact can be reliably established within the adult patient group. How do we know for certain that the patient had any signs of impairment in inattentiveness and/or hyperactivity-impulsivity before his or her seventh birthday? Especially adults who were never diagnosed having ADHD as a child are required to recall their inattentive and/or hyperactive-impulsive behavior as a child, from before they were seven years of age. This is an undertaking not without biases since adults show poor recall of their childhood ADHD symptoms (Wender, Reimherr, & Wood, 1981) and symptom manifestation may differ markedly between childhood and adulthood (Wender, 1997).

A diagnostic tool commonly used to obtain insight into childhood symptoms and their degree of severity, is the Wender Utah Rating Scale (WURS: Wender, 1985). From the original 61 items Ward, Wender, & Reimherr (1993) identified 25 items that retrospectively differentiated most significantly between adult ADHD patients and normal adult controls. Using a cut-off score of 46 or higher their research results found that the WURS correctly identified 86% of the ADHD patients and 99% of the normal controls. Since its introduction however, the WURS has met with a great deal criticism concerning its validity, sensitivity and specificity. A number of studies have found inconsistencies in the sensitivity and specificity of the cut-off score of 46. Findings suggest that it is not sensitive to ADHD and especially
problematic is the fact that it tends to (over-) misclassify individuals from non-patient groups by as much as 50% (McCann et al., 2000; Suhr, Zimak, Buelow, & Fox, 2009)).

Could it be that the WURS does not reflect childhood symptoms from as far back as before one’s seventh birthday, but rather that it more closely represents present symptoms or level of functioning?

As research on the issue of (adult) ADHD developed, aetiology has become defined along various lines of thinking and an important milestone was the redefinition of the frontal lobe hypothesis as a disorder of executive functioning (Sergeant, Geurts, & Oosterlaan, 2002). Barkley (1997) presumed that the core deficit in ADHD pertained to response inhibition and that this is linked to executive neuropsychological functions, such as working memory, self regulation and motor control. Subsequent research has yielded support for this model, whereby findings consistently show impairment of executive functions (Sergeant et al., 2002; Woods, Lovejoy, & Ball, 2002). Specifically for adults, Woods et al. (2002) found that tests differentiating adults with ADHD most distinctly from normal, healthy controls were the Stroop tasks, verbal letter fluency, auditory-verbal list learning and continuous performance tests. At the same time it is necessary to consider that, although both Stroop and CPT have repeatedly differentiated between ADHD patients and normal controls, not all studies have found these group differences (Epstein, Erkanli, Conners, Klaric, Costello, & Angold, 2003; Sergeant et al., 2002). Adult ADHD research has mainly focussed on the subject of inattention in the context of executive functioning, while Boonstra, Oosterlaan, Sergeant, & Buitelaar (2005) in their meta-analysis conclude that by and large, results on adult ADHD neuropsychological difficulties seem to point more in the direction of a general slowing on various cognitive responses. Patients show deficits in various areas of cognitive functioning, not only the executive functions.

As Mackin and Horner (2005) suggested, considering the dilemmas involved in obtaining an accurate picture surrounding childhood symptoms and the specificity of the WURS, it could be that there is a relationship between retrospective symptom report and present executive functioning. Question is, do individuals with current attention deficits report more childhood symptoms?

This then is the question central in the present study; whether the WURS is able to detect a difference in response of childhood experience between ADHD patients and normal controls and whether that difference is reflected by their scores on a number of executive function measures. Considering the results of prior studies (Suhr et al., 2009), this study will subsequently examine whether the standard WURS cut-off score of 46 will be sufficiently
sensitive to identify the ADHD patient group. Furthermore, it is expected that group
differences will show in the results of the various executive function measures. Finally, high
WURS scores are expected to go together with high scores on the executive function
measures and an earlier self reported age of onset.

2. Method

2.1. Participants

As part of a larger study in search of symptom validity tests in neuropsychological
research in ADHD patients, seventy-three respondents were recruited. All participants
responded to a (written) request to participate in the research program, either through their
mental health practitioner or directly by one of the test administrators. All participants were in
the age group 18 to 50 years. One third ($n = 23$) were patients previously clinically diagnosed
as having ADHD based on the DSM-IV criteria. These ADHD patients were recruited through
their treating psychologist or psychiatrist, who informed them of the study details. In the past,
patients had all been clinically diagnosed based on (hetero)anamnestic assessment of
symptoms in accordance with DSM-IV requirements. Use was made of a Dutch self-report
questionnaire (Kooij & Buitelaar, 1997) and a semi structured psychiatric interview.
Following consent from both psychologist and psychiatrist involved, diagnosis was set. The
other two thirds ($n = 50$) of the participants were normal controls, specifically not having
ADHD. The ADHD patients were recruited from within a mental health care institution, while
the normal controls consisted predominantly of psychology freshmen students. Respondents
were specifically requested to give their best performance on an ADHD retrospective
screening device and three executive function tests, all part of the larger test battery. Controls
and patients were screened for co-morbid symptoms of depression through use of the BDI
(Beck Depression Inventory, Beck et al., 1996), with exclusion following a score of 29 or
higher (severe depression). Both controls and ADHD patients were administered the SCL-90
(Symptom Checklist-90; Derogatis, 1994), but only controls were excluded if two or more
sub-scales fell in the high range. Screening of the ADHD patients with the SCL-90 was not
done as co-morbid psychiatric symptoms are expected within the ADHD patient group. The
ADHD patients were requested, with consent from their treating psychiatrist or physician, to
skip their medication on the day of testing. In addition, ADHD patients were asked at what
age they know they first showed ADHD related symptoms from which they experienced
hindrance as related to the impairment criteria of the DSM-IV.
ADHD patients received a gift voucher for their effort and were refunded their travel expenses, while university students were rewarded three grade points for their participation.

2.2. Measures

2.2.1. Wender Utah Rating Scale (WURS)

The WURS (Wender, 1985) is a self-report measure, often used as a screening device for ADHD in adults, which retrospectively assesses ADHD relevant childhood behaviors and symptoms. The original version consists of 61 items. This study uses a short version which was developed when analysis showed 25 items to explain the greatest mean difference between ADHD patients and normal controls (Ward et al., 1993). Participants were instructed to rate each item depicting a behavior on a 5-point scale ranging from “not at all or very slightly” to “very much” as a measure of their childhood symptoms. The range of possible scores is from 0 to 100, with higher scores indicating more serious ADHD childhood symptoms.

2.2.2. Conners’ Continuous Performance Test (CPT)

The Conners’ CPT test (Conners, 1995) of vigilance is a computerized task of sustained attention, concentration and inhibition. It is an often applied measure to identify ADHD (Advokat, Martino, Hill, & Gouvier, 2007). The test requires the participant to indicate, through pressing of the computer space-bar, every time a letter appears on the computer screen and to inhibit this action when the letter X shows. Taking a total of 14 minutes, the monotonous task measures both commission and omission errors, but amongst others also inhibition and response time variability, two variables shown to discriminate ADHD adults from normal controls (Barkley, 1997; Walker, Shores, Trollor, Lee, & Sachdev, 2000).

2.2.3. Stroop Colour-Word Test (Stroop)

The central idea in Stroop tests (Stroop, 1935) is the interference of colour and words; it takes longer to call out the colour names of colour patches than it takes to read colour words. Hence it is even more difficult to read out colour names printed in a (different) coloured ink. In the test participants are instructed to read aloud three cards: from the first they must read the colour names (printed in black), from the second they must name the colour patches and from the third card they must name the colour ink the colour words are written in, although ink and word do not necessarily correspond. This last card is the so called
interference card as for instance the naming of the green ink will be interfered by the word RED that is printed.

2.2.4. Word Memory Test (WMT)

The Word Memory Test (WMT: Green, 2003) uses a list of 20 word pairs (e.g., rat-tail) to measure verbal learning and self-contained effort. Although symptom validity (effort) is the specific research strength of this test and the reason it was included in the test battery, the present study used only the measures of recognition and paired associates and (delayed) free recall tasks. The test was administered using the computerized version, whereby the experimenter concludes the test by filling out the paired associates and free recall tasks.

3. Results

3.1. Statistical analysis

First group differences are evaluated with regards to their descriptive characteristics. Subsequently the WURS scores are assessed in relation to the standard cut-off score of 46 and for possible group differences. Finally, the performance on the neuropsychological tests is evaluated and examined for correlation with the WURS.

3.2. Descriptive analysis

Groups differed significantly on all descriptive statistics (Table 1). The control group was considerably younger and less educated than the ADHD group and consisted for only 16% of males, while this was 47% in the ADHD group. As expected, the ADHD group scored significantly higher on the BDI and SCL-90 screening devices. The self reported age-of-onset displayed a mean of $M=19.43$ ($SD=15.85$; range 2-47 years) well above the required criteria of before seven years of age. Only 32% of the ADHD patients reported an age-of-onset of seven years old or younger. None of those with an age-of-onset below 7 years were of the predominantly inattentive (ADD) subtype.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>AD(H)D (n = 23)</th>
<th>Control (n = 50)</th>
<th>Significance (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years ($M (SD)$)</td>
<td>31.91 (10.32)</td>
<td>21.78 (6.94)</td>
<td>&lt;.000</td>
</tr>
<tr>
<td>Gender (% male)</td>
<td>47.8</td>
<td>16.0</td>
<td>.011</td>
</tr>
<tr>
<td>Education, years ($M (SD)$)</td>
<td>15.22 (1.57)</td>
<td>14.54 (1.21)</td>
<td>.048</td>
</tr>
<tr>
<td>BDI</td>
<td>9.45 (5.90)</td>
<td>4.18 (3.95)</td>
<td>.001</td>
</tr>
<tr>
<td>SCL90</td>
<td>150.44 (41.46)</td>
<td>114.49 (25.89)</td>
<td>.002</td>
</tr>
</tbody>
</table>
3.3. Standard WURS cut-off analysis

Results show that the WURS differentiated significantly between groups. However, even the mean score of the ADHD group \((M = 45.87; SD = 19.97)\) did not meet the cut-off score of 46. As such, the cut-off score was only able to identify 52% of the ADHD patients, therefore not identifying 48% of the previously clinically diagnosed ADHD patients and consequently not displaying sufficient sensitivity. The cut-off resulted in a false positive rate of 6% in the control group.

3.4. Analysis of neuropsychological test performance

The ADHD patient group and the control group differed significantly on all measures of the Word Memory Test except the first measure of immediate recognition, on all measures of the Stroop Test, but on only two measures (commissions and variability) of the Conners’ Continuous Performance Test (Table 2). Because visual examination of the CPT data suggests that there is also a difference between groups on other measures and because skewness and kurtosis values indicated an asymmetrical, peaked distribution, the data were reanalyzed using the non-parametric Mann-Whitney U-test. Results confirm the visual inspection, whereby significant group differences are further found for the measures of omissions, hit reaction time standard error and detectability. These are all measures related to inattentiveness, whereby the ADHD patient group show more omission errors, more commission errors, a greater inconsistency in their response speed, more variability in response and they show less ability to distinguish and detect X and non-X stimuli correctly.

<table>
<thead>
<tr>
<th>Test Variable</th>
<th>ADHD (n = 23)</th>
<th>Control (n = 50)</th>
<th>Significance (p)</th>
<th>Significance M-W test (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMT Immediate Recognition</td>
<td>95.67 (7.17)</td>
<td>97.05 (12.84)</td>
<td>.410</td>
<td></td>
</tr>
<tr>
<td>WMT Delayed Recognition</td>
<td>95.80 (6.25)</td>
<td>99.00 (1.97)</td>
<td>.025*</td>
<td></td>
</tr>
<tr>
<td>WMT Paired Associates</td>
<td>86.74 (13.79)</td>
<td>96.40 (4.85)</td>
<td>.003*</td>
<td></td>
</tr>
<tr>
<td>WMT Free Recall</td>
<td>56.76 (18.30)</td>
<td>67.43 (13.59)</td>
<td>.007*</td>
<td></td>
</tr>
<tr>
<td>WMT Delayed Free Recall</td>
<td>61.76 (14.67)</td>
<td>70.22 (15.03)</td>
<td>.025*</td>
<td></td>
</tr>
<tr>
<td>CPT Omissions</td>
<td>3.09 (3.42)</td>
<td>2.18 (6.27)</td>
<td>.519</td>
<td>.021*</td>
</tr>
<tr>
<td>CPT Commission</td>
<td>18.96 (8.87)</td>
<td>14.40 (6.14)</td>
<td>.013*</td>
<td>.006*</td>
</tr>
<tr>
<td>CPT Hit RT</td>
<td>347.24 (61.91)</td>
<td>339.96 (39.22)</td>
<td>.608</td>
<td>.877</td>
</tr>
<tr>
<td>CPT Hit RT std error</td>
<td>5.46 (1.76)</td>
<td>4.64 (1.81)</td>
<td>.074</td>
<td>.031*</td>
</tr>
</tbody>
</table>
CPT Variability    8.15 (5.09)  5.68 (3.29)    .041*    .009*
CPT Detectability  .47 (.54)     .64 (.40)   .125   .005*
CPT Response style .52 (.57)     .46 (.38)   .624   .887
CPT Perseverations 1.09 (2.37)  .58 (2.03)   .351   .144
CPT Hit RT Block change .007 (.025) .014 (.14) .822   .071
CPT Hit SE Block change .0235 (.084) -.1190 (.852) .428   .239
CPT Hit RT ISI Change .0417 (.036)  .0438 (.034) .815   .742
CPT Hit SE ISI Change .0143 (.108)  .0012 (.109) .573   .443
Stroop Word        45.35 (5.42)  39.58 (5.62) .000*    
Stroop Colour      55.19 (8.69)  50.10 (9.59) .034*    
Stroop Word/Colour 84.30 (15.17)  73.22 (15.97) .007*    
Stroop Ratio       .6621 (.0761)  .6900 (.0686) .128

M-W-test = Mann-Whitney U-test

Group differences on the Stroop test apply for all its measures, but not on the calculated ratio score. The ratio score is a method of quantifying Stroop Interference as suggested and used by Lansbergen, Kenemans, & Van Engeland (2007) in their meta-analysis. The interference ratio $I_r = C/CW$, controls for colour naming (C) so that only the interference from the incongruent word (CW) is measured. Independent-samples t-test analyses yielded a not significant group difference of $p = .128$. Even when controlling for the word reading (W) through analysis of covariance (ANCOVA), results ($p = .104$) are not significant and only a small 3.7% of the variance of the interference is explained by group membership. Therefore ADHD patients do not seem to suffer from deficient interference control on the Stroop task.

Independent-samples t-tests revealed a significant difference between the ADHD patient group and the normal control group for all Word Memory Test measures, except the immediate recognition task. Healthy, not impaired controls are expected to have a mean score on the paired associates’ measure that is considerably closer to their means on the immediate and delayed recognition measures, in comparison to impaired groups (Green, 2003). This is the case in the present study when comparing the normal control group with the ADHD patient group. These results indicate that ADHD patients have more problems retaining information which may possibly be due to diminished semantic clustering (Roth, Wishart, Flashman, Riordan, Huey, & Saykin, 2004).

3.5. Analysis of WURS and neuropsychological test correlations

When not distinguishing between groups, Pearson’s Product-Moment Correlations between WURS and executive function tests yielded small to medium ($r = -.291$, $p < .05$,.
significant correlations on the Word Memory Test measures, except immediate recall, and the Stroop Word measure ($r = .267, p < .05$). However, correlations computed for ADHD patients and normal controls separately, found no significant correlation between the WURS and scores on neuropsychological tests for ADHD patients. For normal controls only a small significant correlation was found for the WMT delayed recall measure. Furthermore, no association was found between the WURS score and the reported age-of-onset.

4. Discussion

The aim of this research study was to determine whether the WURS is able to detect a difference in retrospective childhood experiences between adult ADHD patients and normal controls and whether that difference is reflected by their scores on neuropsychological test measures. The results indicate that although the WURS differentiates significantly between ADHD patients and normal controls, the cut-off score of 46 is insufficiently sensitive. Only 52% of the ADHD patient group was correctly identified. Although considerably less than the 72% McCann et al. (2000) found in their clinically referred sample, it is in line with a large university population where Suhr et al. (2009) found a WURS sensitivity of only 17% for the ADHD group. Although lowering the cut-off score seems a logical consequence, this would compromise the specificity of the WURS measure (Hill, Pella, Singh, Jones, & Gouvier, 2008; Suhr et al., 2009).

Part of the DSM-IV criteria is the age-of-onset of seven years or younger, which in this sample would correctly diagnose only 32% of the ADHD patient population. This age discrepancy raises questions concerning the accuracy of the original ADHD diagnosis. However, in the present study just one question was posed exclusively to the patient himself, while in the diagnostic process information is included from multiple sources, including close relatives and school results. Furthermore, in comparison to their parents and partners, adults with ADHD are inclined to underreport the severity of their symptoms (Kooij, Boonstra, Swinkels, Bekker, Noord, & Buitelaar, 2008). In addition Applegate, Lahey, Hart, Biederman, Hynd, Barkley, et al. (1997) found within their sample of youths with the predominantly inattentive type, a large portion who did not meet the age-of-onset criteria, while in this case information was gathered from their parents. As Barkley and Biederman (1997) conclude, there is no clinical reason or scientific justification to uphold the age-of-onset criteria. Especially with adolescents and adults it seems an unjustified restriction in the clinical
diagnosis of the AD(H)D subtypes, which appears to be a left over of the supposition that it concerns a disorder with distinct childhood onset.

Of the neuropsychological test, several measures found significant differences between the ADHD patient group and the normal controls. Group differences found in the CPT applied especially for the inattentiveness measures. This is in line with studies that show that hyperactivity and impulsiveness may decrease with age, whereas inattention seems more likely to persist (Advokat et al., 2007). It is then all the more strange that there is no correlation between the WURS and CPT, especially since McCann et al. (2000) found that the WURS inattentive symptoms best discriminate adult ADHD patients.

The Stroop was advocated by Woods et al. (2002) as being the most prominent and reliable measure that differentiates adult ADHD patient from normal controls. Results from the present study support that claim on all three Stroop measures, but not for the interference ratio. The meta-analysis of Boonstra and colleagues (2005) yielded the same results. They therefore stress the importance of using a correction method on the so called interference or Colour /Word card, like the interference ratio used by Lansbergen et al. (2007) as applied in the present study. As Lansbergen et al. observed, the present study too found that the ADHD group was systematically slower in base-word reading. They indicate that this underestimates the true interference level when no correction is made. However, despite the correction, the present study found no interference difference. This might be caused by the fact that our participant sample consisted of diagnosed ADHD patients, while much research makes use of samples of clinically referred patients, where tests are part of the diagnostic process. In light of the benefits patients seeking clinical evaluation may attribute to obtaining an ADHD diagnosis (Van Egmond, Kummeling, & Van Balkom, 2005), we suspect the malingering rate in these samples to be proportionally higher. The Stroop test can be easily faked or over exaggerated within acceptable limits, thus inflating test results. We found no association between the WURS score and Stroop test results. This further confirms our suspicion that apart from being an indication of general cognitive slowing, characteristic of ADHD patients (Boonstra et al., 2005), the Stroop test generally and the interference measure specifically, is not indicative of a distinct executive function deficit.

The question put forward at the beginning of this study was whether retrospective symptom report, as measured by the WURS, could be related to present executive functioning, as measured by neuropsychological tests. The current research findings do not support this claim. On the other hand, rather than disqualifying the neuropsychological tests, the impression emerges that the Wender Utah Rating Scale is not adequately sensitive nor
sufficiently specific to diagnose ADHD symptoms exclusively for ADHD patients. Additionally, being a disorder known for its high rate of co-morbidity, it seems almost impossible to determine if specific symptoms, or test results for that matter, are indicative of cognitive difficulties exclusively related to the ADHD or whether they are due to the existence of co-morbid disabilities (Boonstra et al., 2005).

The main conclusion that can be drawn considering the present results is that the WURS should not be used in clinical situations to determine the validity of an AD(H)D diagnosis in adults. Further research among ADHD adult patients is needed to confirm the presence and therefore the requirement of childhood of symptoms and impairment. The specific age of onset criterion is another aspect that jeopardizes legitimate adult ADHD diagnosis.

In view of the above implications, the limitations of the present study must be considered. Most of the shortcomings are related to the sample composition. The fact that the patient group was considerably smaller than the control group confines the power of the results. The groups were not matched, differing significantly on essential characteristics such as age and gender. Even though difference in years of education showed marginal significance, we should note that test results were not controlled for by participant IQ, while research has shown that cognitive functioning is significantly lower in ADHD patients (Frazier, Demaree, & Youngstrom, 2004). The homogeneity of the control group, consisting mainly of university freshmen, may therefore have inflated the results of their scores on the neuropsychological tests. Within the test battery no IQ measure was available and while the debate surrounding ADHD and IQ has not been resolved, future research would do well controlling for level of cognitive functioning.

5. Literature


