Validation of the ADHD Rating Scale as a clirlician administered and scored instrument

Douglas E. Faries, Ilker Yalcin, Donald Harder and John H. Heiligenstein

J Atten Disord 2001; 5; 107

DOI: 10.1177/108705470100500204

Validation of the ADHD Rating Scale as a clinician administered and scored instrument

D. E. Faries, I. Yalcin, D. Harder, and J. H. Heiligenstein

There are several parent- and teacher-scored rating scales that are validated and available for assessing the severity of ADHD symptoms in pediatric populations. However, we are not aware of any symptom severity rating scales with published validation results when administered and scored by clinicians following a semi-structured interview with the parent or guardian.

In a recent clinical trial, in order to have consistency in assessing entry severity criteria and changes in symptom severity over time across patients, we used the ADHD Rating Scale, administered and scored by trained clinicians. Using data from this trial, we assessed the validity and reliability of the ADHD Rating Scale when completed by trained clinicians based on interviews with parents.

Results indicate that this version of the scale has acceptable levels of inter-rater reliability, test-retest reliability, internal consistency, convergent validity, discriminant validity, and responsiveness. The psychometric properties were comparable to other validated scales for assessing ADHD symptom severity. Overall, this study supports the use of the ADHD Rating Scale as a clinician administered and scored tool for assessing the severity of ADHD symptoms in pediatric patients.

Various parent and teacher rating scales for assessing Attention-Deficit Hyperactivity Disorder (ADHD) symptoms have been in use for a number of years, with many supported by research on their psychometric properties (Achenbach & Edlebrook, 1986; Barkley, 1990; Cohen, Becker, & Campbell, 1990; Biederman et al., 1993; Miller et al. 1995; Gadow & Sprafkin, 1996; Conners, 1997; Ullman, Sleator, & Sprague, 1997; DuPaul, Power, Anastopoulos, & Reid, 1998; March, Sullivan, & Parker, 1999). Conners (1999) provides an excellent overview of the use of rating scales for assessing ADHD symptoms. The Conners' Parent and Conners' Teacher rating scales (Conners, 1997) are among the most commonly used and thoroughly studied scales. The established psychometric properties include: internal and test-retest reliability; factorial, convergent, divergent, and discriminant validity; and, normative data.

Recently, DuPaul, Power, Anastopoulos, and Reid (1998) updated an earlier version of the ADHD Rating Scale (DuPaul, 1991). The new versions (both Parent and Teacher versions) have 18 items that correspond to the 18 ADHD symptoms contained in the DSM-IV diagnosis for ADHD. The total score from the scale is the sum of the scores for each of the 18 items. Their validation work includes: demonstration of internal consistency; convergent, discriminant, predictive, and factorial validity; and, test-retest reliability. In addition, based on a sample of 2000 US children, norms were developed for various age and gender groups, allowing for computation of *t*-scores (normalized scores) for each patient.

The above scales were validated based on parent or teacher scores. The importance of obtaining information on symptom severity from parents and teachers has long been established, however, there are also advantages for using a clinician-scored rating scale. For instance, in a clinical trial setting, the consistency provided by trained clinician raters can result in reduced variability, by allowing for a more homogenous patient population through consistent applications of entry severity criteria. In addition, using trained clinician raters can reduce variability by having them apply consistent judgements on severity ratings across patients. Also, clinicians can incorporate information from multiple sources and settings into a single score in a more consistent manner based on set clinical scoring criteria; this helps reduce the problem of statistical multiplicity that exists when separate measures are obtained. Clinicians are also trained to assess the impact of interventions already implemented to cope with ongoing

behavior problems (e.g. special schoolroom methods). Use of a clinician-rated scale also avoids the problems of needing to obtain ratings for children with multiple teachers, and the issue of school vacations. Clinician-rated instruments to assess change in illness severity are often required as primary efficacy outcome measures by regulatory authorities for most psychiatric disease states.

In a recent clinical trial, in addition to parent- and teacher-rated scales, we used the ADHD Rating Scale-IV-Parent as a clinician (investigator) administered and scored scale. For clarity, this scale will be referred to as the ADHDRS-PI. Clinicians completed the scale following an interview with the primary caregiver for the child (hereafter simply referred to as parent). Intuitively, we expected the clinician-scored version to perform fairly similarly to the validated parent version of the ADHD Rating Scale (ADHDRS-P). In this manuscript, we provide quantitative evidence for the validity and reliability of the clinician-scored version of the ADHD Rating Scale. This includes assessment of the internal consistency, test-retest reliability, inter-rater reliability, convergent validity, and responsiveness of the scale. These psychometric properties are assessed relative to other scales, including the Conners' Parent and Teacher scales. A head-to-head comparison of the parent- and clinician-scored versions of the ADHD Rating Scale is also included.

The format of the manuscript is as follows. First, a description of the clinical study design and rating scales used is provided. This is followed by a summary of the methods used for assessing the psychometric properties of the scale. The results are then provided along with a discussion of the findings.

Study Design and Rating Scale Methods

This study of the validity and reliability of the ADHDRS-PI was conducted as part of a large clinical trial conducted at 24 sites in the United States investigating the efficacy and safety of atomoxetine, an experimental pharmacotherapy for ADHD. For more details regarding the use of atomoxetine for ADHD, see Spencer et al. (1998) and Spencer et al. (2001). The clinical trial design included a one-week, no-drug lead-in followed by 10 weeks of open label treatment with either atomoxetine or methylphenidate (Ritalin). The inclusion of the methylphenidate arm allowed for the validation of the rating scale using a compound known to be efficacious for ADHD (Garland, 1998; Goldman, Genal, Bezman, & Slanetz, 1998). Thus, this manuscript focuses on data from all patients during the lead-in period and the methylphenidate treated patients during the open label period.

Entry criteria required that patients meet DSM-IV criteria for a diagnosis of ADHD. Patients were required to meet criteria for ADHD following administration of the ADHD module from the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS). Determination of the patient's ADHD subtype was also based on the KSADS interview. The patients' ADHDRS-PI total scores must have been at least 1.5 SD above the ADHDRS-P norms provided by DuPaul et al. (1998). Patients were required to be at least 7, but less than 16 years of age, and have an IQ of at least 80. Patients were excluded if they had a history of psychosis or seizure disorder, had clinically significantly high blood pressure, were at risk for suicidality, or had been non-responders in the past to a trial of methylphenidate of sufficient dose and duration. Other than study medication, no psychotropic medications were allowed during the study.

Patients who were taking stimulant medication at Visit 1 were allowed to taper off during this period as long as they were off medications for at least 5 half-lives of the medication prior to Visit 2. At Visit 2, qualified patients were randomized to up to 10 weeks of open label treatment with either atomoxetine or methylphenidate. Patients were seen at approximately weekly intervals during this period.

Severity of ADHD was assessed using multiple measures. The primary efficacy measure, the ADHDRS-PI (scored by the clinician), was completed at each visit. As discussed in the Introduction, the ADHDRS-PI is an 18-item questionnaire with each item rated on a 0 to 3 scale (0 = never or rarely, 1 = sometimes, 2 = often, 3 = very often). The 18 items correspond to the 18 symptoms listed in the DSM-IV ADHD diagnosis of ADHD. The scales' link to the DSM-IV ADHD diagnostic criteria provides a measure of face validity for the scale as a measure focused on ADHD symptom severity. Specific comorbid symptoms of interest, such as oppositional behavior or depression, were captured using other rating scales. In this trial, the ADHDRS-PI was administered at each visit and it assessed symptoms over the past week. The clinician completed the scale following an interview with the parent. The clinicians were trained to select scores based on the frequency of the behavior (across multiple settings) and the degree of impairment, using a developmental comparison. The clinician scores were not to be a simple reading and recording of parent-chosen scores.

Parent-scored scales used in this study included the Conner's Parent Rating Scale (Conners, 1997) assessed at each visit, and the ADHDRS-P (DuPaul et al., 1998) completed at visits 1, 6, and 10. On visits when both the parent- and investigator-scored versions were to be

collected, a random alternating order of administration was used. Teacher-scored scales included the Conner's Teacher Rating Scale (CTRS; Conners, 1997). The CTRS was to be completed only for patients who had no more than two primary teachers and who would be in school for the duration of the treatment. The Clinical Global Impressions–Severity scale (Guy, 1976), a single item 7-point (reference; 1 = normal, not ill, 2 = minimally ill, 3 = mildly ill, 4 = moderately ill, 5 = markedly ill, 6 = severely ill, 7 = very severely ill) rating of the severity of symptoms in the opinion of the investigator based on the investigator's clinical experience with ADHD patients was also included at each visit.

Reliability and Validity Methods

The establishment of the reliability and validity of a rating scale is not a single test, rather it is a summary of its psychometric properties using multiple approaches. In this manuscript, the inter-rater reliability, internal consistency, test-retest reliability, convergent validity, discriminant validity, and responsiveness of the scale are investigated. The definitions and methods for each are described below. For general information on measures of reliability and validity, see Deyo, Diehr, and Patrick (1991), Guyatt, Patrick, and Feeny (1991), and Perrin et al. (1997).

Inter-Rater Reliability

A scale is said to have inter-rater reliability if ratings by different individuals at the same time on the same patient are similar. To assess the inter-rater reliability of the ADHDRS-PI, all clinical personnel who would be using the ADHDRS-PI were trained and assessed in a rater training session prior to the study. This involved discussion of the rating scale, specific guidelines for assigning scores for each item, and a review of at least one video-taped parent interview. Raters independently completed the ADHDRS-PI based on their assessments of the videotaped interview. After the interview the group results were shared and points of disagreement were discussed.

Two statistical measures were used to assess the inter-rater reliability: percentage of agreement and average squared deviation from the mode. The percentage of agreement was computed simply as the proportion of ratings that were the same out of all possible pairs of raters on all possible items. The average squared deviation from the mode was computed for each rater by averaging (over the 18 items) the squared difference (Channon & Butler, 1998) between their ratings and the mode rating from the entire group. As opposed to percent agreement, the average squared deviation statistic more severely penalizes ratings that are more than 1 point from the mode. There are no published guidelines on what constitutes an acceptable level of percent agreement or average squared deviation. However, as a guideline, the percent agreement from four similar training tapes for a depression study involving the Hamilton Depression Rating Scale (Hamilton, 1960), the gold standard measure of depression severity in clinical trials, ranged from .55 to .70.

Internal Consistency

Measures of internal consistency assess the degree to which each item of a rating scale measures the same construct. Cronbach's alpha is a commonly used statistic to assess internal consistency (Cronbach, 1951). It is based on all possible correlations between two sets of items within a questionnaire. It ranges between 0 and 1, higher scores suggesting good internal consistency. If alpha is very low, the questionnaire is either too short or the items have very little in common. On the other hand, extremely high scores can indicate some redundancy in the rating scale. The commonly accepted minimal standard for group comparisons is 0.7 (Perrin et al., 1997). Cronbach's alpha was computed using Visit 1 scores for all patients.

Test-Retest Reliability

Test-retest reliability indicates the stability of ratings on the same individual at two different times while in the same condition. Changes in scores on the ADHDRS-PI from Visit 1 to Visit 2 (no-drug lead-in period) were used to assess test-retest reliability. The suggested time between Visit 1 and Visit 2 was 7 days, which is thought to be a reasonable timeframe for assessing test-retest reliability (Deyo et al., 1991). Patients who were still taking stimulant medication for ADHD when they arrived at Visit 1 were excluded from the test-retest calculations since the possibility of changes in symptoms from Visit 1 to Visit 2 due to discontinuation of drug could not be discounted.

Both Pearson's correlation coefficient and the intraclass correlation coefficient (ICC) were computed to quantify the correlation between Visit 1 and Visit 2 scores. Although Pearson's correlation coefficient is the most commonly used measure for assessing correlation and strength of the relationship between the scores, the ICC is the recommended measure in this situation (Deyo et al., 1991). The ICC is the ratio of the variability due to patients divided by the total variability (which contains the variability due to time). This statistic detects not only lack of correlation between scores at the different time points, as does Pearson's correlation coefficient, but also detects systematic changes from one time period to the next. For instance, if the score increased by 5 points from Visit 1 to Visit 2 for every single patient, the correlation coefficient would be 1.0, yet the ICC would be reduced. Thus, the ICC

is a measure that detects a lack of association and systematic differences between the time points. For further discussion about assessing test-retest reliability, see Deyo et al. (1991). Landis and Koch (1977) suggest that ICCs above 0.60 suggest satisfactory stability and ICCs greater than 0.80 are excellent. To further quantify shifts over time, *t*-test was used to test whether the mean scores obtained at Visits 1 and 2 were similar.

Convergent Validity

Convergent validity indicates a relationship between the scale under review and other scales thought to measure the same construct, in this case ADHD symptom severity. To assess convergent validity, the correlations (Pearson's) between the ADHDRS-PI total score, ADHDRS-P, CGI-severity, CPRS ADHD Index, and CTRS ADHD Index were computed. As noted previously, the CPRS and CTRS are validated parent and teacher instruments for assessing ADHD symptom severity, and the CGI is a commonly used investigator measure. Correlations were computed between baseline scores and between changes from baseline to endpoint scores. To control for any learning effects or biases in scoring, the order of administration of the ADHDRS-PI and ADHDRS-P was randomized.

Discriminant Validity

Discriminant validity indicates the ability of a scale to distinguish between different groups of subjects. For instance, a scale with discriminant validity will distinguish patients with and without a diagnosis of ADHD, or patients with and without significant hyperactive symptoms. In this study patients without any ADHD symptoms were not recruited, thus, a comparison between non-ADHD and ADHD patients was not possible. However, three other approaches to assessing the discriminant validity were used. First, a comparison between patients with a diagnosis of ADHD Inattentive subtype and a diagnosis of ADHD Combined (Inattentive plus Hyperactive/Impulsive symptoms) was conducted. Subtype was assessed as part of the KSADS diagnostic interview at the initial visit. Patients with an Inattentive subtype did not have sufficient hyperactive/impulsive symptomatology to meet the full Combined DSM-IV ADHD subtype. This provides an opportunity to assess whether the Hyperactive/Impulsive subscale of the ADHDRS-PI was able to distinguish between the ADHD subtypes.

Second, although patients without symptoms of ADHD were not recruited, at endpoint a wide range of symptom severity levels were present as indicated by the CGI-severity scores. Thus, an analysis of variance was used to compare mean ADHDRS-PI total and *t*-scores by CGI-severity score. Comparison to the CGI-Severity scores also provides additional information regarding the clinical significance of specific ADHDRS-PI scores. Lastly, demonstrating lower correlation between measures that should not be logically related, as compared to measures which should be correlated, provides additional evidence of discriminant validity. Pearson's correlation coefficients between the ADHDRS-PI subscales and various other measures were computed.

Responsiveness

The responsiveness of a rating scale measures the scales ability to detect small but clinically significant changes in the patients symptom severity when a change has occurred. The Standardized Response Mean (SRM) is used to evaluate responsiveness (Stratford, Binkley, & Riddle, 1996). The SRM is defined as the mean change from baseline score divided by the standard deviation of the change scores. Higher values for the SRM indicate greater sensitivity to change. Data from the methylphenidate was used to compute the SRM since this is a proven efficacious compound for ADHD symptoms. Because this was an open label data without a placebo control, we could not assess responsiveness using effect sizes based on treatment differences from placebo.

Results

Inter-Rater Reliability

A rater training session was held prior to the start of the trial. Fifty-seven clinicians from 24 investigational sites were trained: 66% were MDs; 23% were PhDs; the rest had other clinical training and experience. Training included a presentation of the rating scale and a discussion of scoring including clinical anchor points for each item score. A videotaped interview with a parent of a child with ADHD was then shown and raters independently completed the ADHDRS-PI. Discussion of the ratings followed. From the scoring of the taped interview, there was a 73% percent agreement on item scores with an average squared deviation of 0.28. Thus, approximately 3/4ths of the time, different pairs of raters gave the same exact rating for an individual item. Agreement was similar for inattention and hyperactive/impulsive items. Over 80% of the clinicians independently arrived at a total score from 42 to 47. Although a training session of this type does not fully assess all aspects of rating (for instance the variability in interviewing skills), these results suggest the inter-rater reliability of the scale is reasonable for a multi-site clinical trial.

Subjects

Two-hundred and eighty-three patients entered into this clinical trial had at least one measurement from the ADHDRS-PI. Two-hundred twenty-eight patients met criteria for a DSM-IV diagnosis of ADHD and were randomized to a treatment group in this clinical trial. The mean age of the randomized patients was 10.4 years (range 7.0 to 15.8). Other patient characteristics for the group of randomized patients are summarized in Table 1.

Table 1. Summary of Patient Characteristics (*N* = 228)

Variable	*n*	%
Gender		
Male	211	92.5
Origin		
AF	26	11.4
CA	175	76.8
HP	17	7.5
Other	10	4.4
ADHD Subtype		
Hyper/Impulsive	3	1.6
Inattentive	52	22.9
Combined	172	75.8
Previous Stimulant Use	125	54.8
School Grade		
1–2	45	19.7
3–4	72	31.6
5–6	66	28.9
7–8	33	14.5
9–10	12	5.3

Abbreviations: AF = African-American; CA = Caucasian; HP = Hispanic.

The diagnosis of ADHD was assessed using the KSADS-E ADHD module, and screening for comorbid diagnoses was performed using the computerized Diagnostic Interview for Children and Adolescents – Revised (DICA-R; Reich, Welner, & Herjanic, 1990). All positive diagnoses obtained from the DICA-R were reviewed and either confirmed or overridden based on clinician judgement. Of all the patients, 58.9% had at least one comorbid diagnosis, which included: 42.6% with a diagnosis of oppositional defiant disorder; 17.6% with a psychosocial stressor; 10.1% with an elimination disorder; and, 6.2% with dysthymia.

A summary of the baseline ADHDRS-PI total and subscale scores by ADHD subtype is contained in Figure 1. There were only three hyperactive/impulsive ADHD subtype patients, thus, no summary statistics are provided for this group. On a *t*-score basis, the overall mean total score was 77.7. This indicates the mean baseline ADHD symptom severity was over 2.7 standard deviations above the norm for their age and gender.

Internal Consistency

Internal consistency was measured using Cronbach's alpha based on the 283 patients' data at Visit 1. Cronbach's alpha for the ADHDRS-PI was 0.86. For comparison, Cronbach's alpha for the CPRS ADHD Index and ADHDRS-P were .86 and .89, respectively. Thus, the average inter-item correlation was similar to that of validated parent and teacher scales. In addition, Cronbach's alpha was not sufficiently high to suggest redundancy in items.

Test-Retest Reliability

Test-retest reliability data was obtained from 147 randomized patients who had Visit 1 and Visit 2 measurements and who were not taking stimulant medication upon entry to the study. As these patients were not on medication, the expectation was for little change in group scores over the 1 week lead-in. The observed intra-class correlation coefficients and change scores for the rating scales are reported in Table 2. Results demonstrate the test-retest reliability of the ADHDRS-PI was both acceptable and similar to the other rating scales. The mean change in ADHDRS-PI total score from Visit 1 to Visit 2 was not statistically different from zero.

Table 2. Test-Retest Reliability

Variable	*N*	ICC	Pearson's Correlation Coefficient	Mean Difference
ADHDRS-PI Total Score	147	.76*	.82*	-0.38
CPRS ADHD Index	148	.61*	.71*	-0.89*
CGI-Severity	150	.65*	.73*	0.01

*p-value < .05

Convergent Validity

To assess convergent validity, the correlations between the ADHDRS-PI (total score and subscale scores) and the ADHDRS-P, CPRS ADHD Index, CTRS ADHD Index, and the CGI-Severity score are reported in Figure 2. In addition to computing correlations on baseline scores, Figure 2 includes correlations between change scores, since the change score is the most commonly used measure of improvement. Each measure (ADHDRS-PI total score, inattentive subscale score, and hyperactive/impulsive subscale) is compared with the corresponding subscale with which it is expected to correlate (e.g. ADHDRS-PI hyperactive/impulsive subscale with the CPRS hyperactive subscale).

In general, correlations between the ADHDRS-PI and other measures of ADHD symptom severity were high, and all were statistically different from 0. The lowest correlation

Figure 1. ADHDRS-PI Baseline Total and Subscale Scores by Subtype

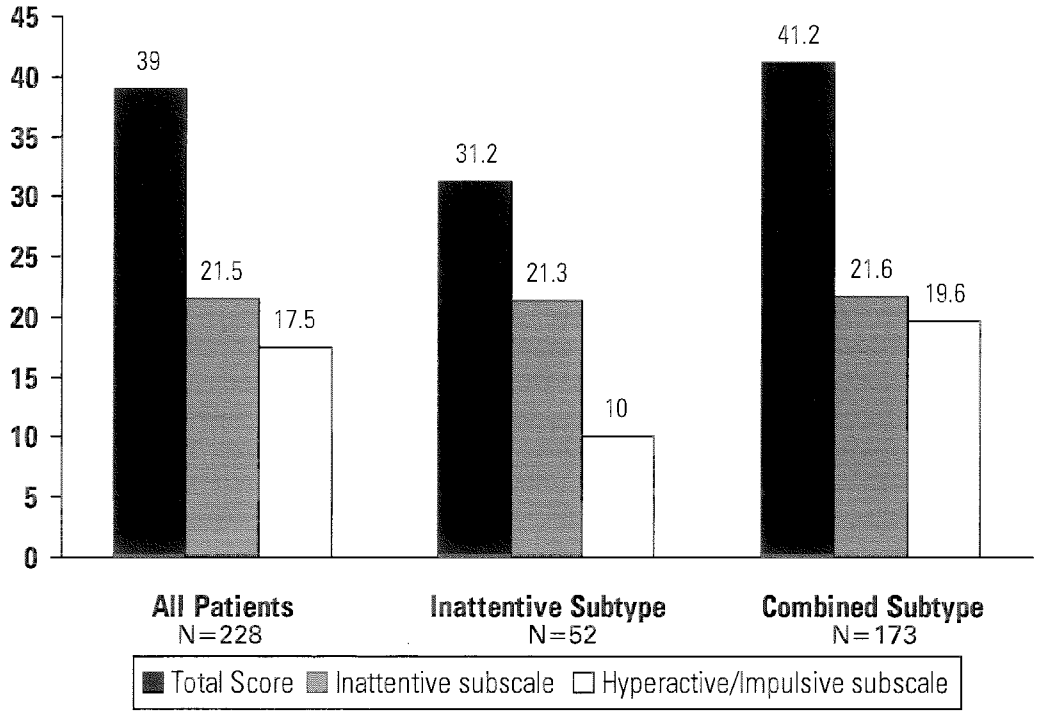
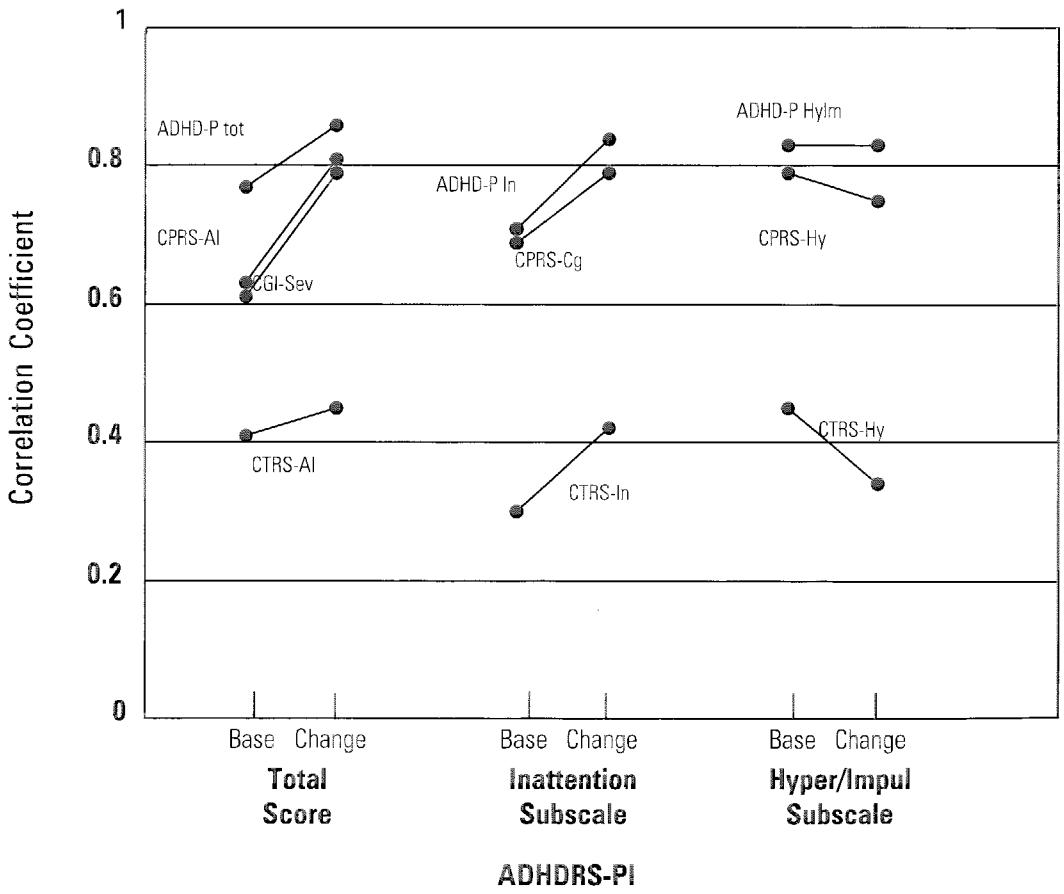


Figure 2. Convergent Validity – Correlations of Baseline and Change Scores with the ADHDRS-PI



Abbreviations: tot = total score; Sev = Severity; In = Inattentive subscale; AI = ADHD Index; Hy = Hyperactive subscale; Hyim = Hyperactive/Impulsive subscale; Cg = Cognitive subscale.

occurred with the CTRS ADHD Index where only moderate correlations were observed. However, for comparison, the correlation between the CPRS ADHD Index and the CTRS ADHD Index was .27 at baseline and .29 at endpoint. In general, only moderate correlations between teacher and parent scales have been observed in previous studies (Conners, 1999; DuPaul et al., 1998).

The direct comparison between the ADHDRS-P and ADHDRS-PI revealed mean differences of less than 1 point at each of the 3 visits where both were obtained. At Visit 1, mean (standard deviation) for the clinician and parent scored versions were 38.1 (9.9) and 37.3 (10.6). As one might expect with the ongoing discussions between parents and investigators during the trial, differences in scores decreased over time. A difference of 1 point was not considered a clinically significant difference, as effective treatment in this study resulted in a mean decrease of over 17 points. There was no indication of an order effect regarding the administration of the ADHDRS-PI and the ADHDRS-P.

Discriminant Validity

Figure 1 provides a summary of the ADHDRS-PI subscale (inattentive and hyperactive/impulsive) scores for patients with each ADHD subtype. A statistically significant difference ($p < .001$) was noted between Inattentive subtype and Combined subtype ADHD patients on the hyperactive/impulsive subscale and total score. Differences between the subtypes on the inattentive subscale were small and not statistically significant. Thus, the scale did discriminate between patients whose diagnosis suggested the presence of clinically significant hyperactive symptoms (Combined subtype) and those whose diagnose did not (Inattentive subtype).

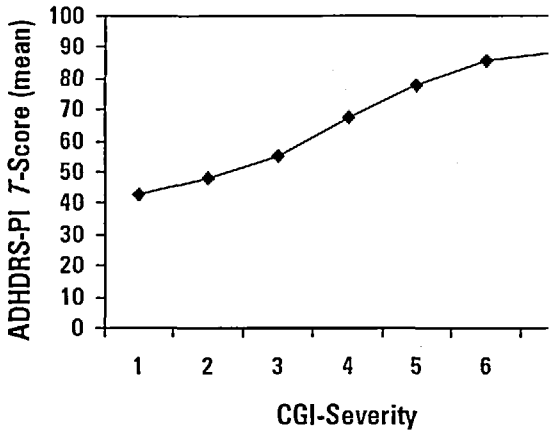
In addition, the correlations between scales thought to measure the same set of symptoms were provided previously in Figure 2. Table 3 provides a subset of this data, along with correlations between subscales expected to measure different symptom groups. As expected, correlations with the ADHDRS-PI inattentive subscale score were much higher with other cognitive/inattention subscales than with assessments of hyperactivity/impulsivity. In the same fashion, the ADHDRS hyperactive/Impulsive subscale performed as expecte

Table 3. Correlations between the ADHDRS-PI Subscales and Corresponding CPRS and ADHDRS-P Subscales

	CPRS		ADHDRS-P	
ADHDRS-PI	Cognitive	Hyperactive	Inattentive	Hyp/Imp
Inattentive Subscale	0.69	0.26	0.71	0.26
Hyp/Imp Subscale	0.20	0.79	0.22	0.83

Figure 3 summarizes the mean ADHDRS-PI *t*-score CGI-Severity score. Analysis of variance on endpoint indicated statistically significant differences in mean sc between each CGI-Severity level with the exceptio between CGI-Severity scores of 1 and 2 and between sc of 6 and 7. Thus, the scale was able to discriminate betv groups of patients rated minimally and mildly ill, betv groups rated mildly and moderately ill, between gr rated moderately and markedly ill, and between gr labeled markedly and severely ill. In the total sampl patients, CGI-Severity scores of 'mildly ill,' 'moderately and 'markedly ill' corresponded to mean *t*-scores of 5 67.4, and 77.9, respectively.

Figure 3. Mean ADHDRS-PI *T*-Scores by CGI-Severity S



Responsiveness

Mean baseline, mean change from baseline to endp the standard deviation in change scores, and the SRM each scale for patients who received methylphenidate presented in Table 4. As methylphenidate is a kn effective treatment for symptoms of ADHD, one w expect each scale to show a clinically and statistic significant difference from baseline. The standard response mean (SRM) produced by the ADHDRS-PI sc was similar to the SRMs using parent and clini measures of ADHD. The observed SRM for the tea rated measure (CTRS ADHD Index) was numerically sm However, only 11 patients had both a baseline a postbaseline teacher measurement and the variabilit scores from teachers was greater than from the pa version. Seventy percent of the patients had endp ADHDRS-PI *t*-scores of 65 or less (with 58% below 60) thus met criteria for having a clinically significant resp to methylphenidate. This is similar to the reported effi of stimulants in previous studies (Swanson et al. 1993

ble 4. SRM Based on Change from Baseline to Endpoint in the Methylphenidate Treatment Group

riable	*n*	Baseline Mean	Baseline *SD*	Change Mean	Change *SD*	*p*-val[a]	SRM[b]
HDRS-PI total	40	37.6	9.7	-17.8	14.7	<.001	1.21
HDRS-P total	36	36.3	10.2	-18.8	13.4	<.001	1.40
3S ADHD Index	40	25.9	6.9	-12.0	10.6	<.001	1.13
3S ADHD Index	11	24.2	8.0	-6.0	14.1	0.007	0.43
l-Severity	40	4.7	0.88	-1.7	1.51	<.001	1.13

thin treatment group *p*-values are from Wilcoxon Signed Rank test on mean change from baseline to endpoint (last visit carried forward) scores.
iM (Standardized Response Mean) is computed as the mean change from baseline to endpoint (last observation carried forward) divided by the standard riation of change from baseline to endpoint scores.

iscussion

this manuscript we have presented the results of a study the psychometric properties of the ADHD Rating Scale, ministered and scored by trained clinicians. Previous ork has demonstrated the validity of the scale as ministered by the parent or teacher. Based on this study, e clinician version also has acceptable levels of inter- ter reliability, internal consistency, test-retest reliability, nvergent validity, discriminant validity, and sponsiveness. The scale's properties were comparable other validated parent and teacher measures of ADHD mptom severity. Differences between scores on the parent d clinician versions were small and not clinically gnificant. Thus, the large amount of research done on e parent version of the rating scale is supportive of the lidity and reliability of the clinician version as well.

veral limitations of this work should be noted. First, the ethod of assessing inter-rater reliability does not capture e variability due to different interviewing skills. However, dependent scoring of videotaped interviews has become e standard method for obtaining inter-rater reliability data e to its practicality. Secondly, only a limited number of tients had both baseline and postbaseline teacher data. his only allowed for definitive claims regarding the sponsiveness of the ADHDRS-PI compared to parent- or ner investigator-rated scales. In addition, this work was t designed to assess whether the ADHDRS-PI adequately fferentiates from the ADHDRS-P when parent rspectives really do not represent the true severity of mptoms, nor to assess which version is superior. As ted by a reviewer, if the clinician simply read the scale to e parent and recorded their answers, one would achieve rfect correlation between the ADHDRS-PI and DHDRS-P. Thus, one should expect at least small fferences between the two versions, however, it is not ear exactly how much difference should be expected. fferences in our sample were less than one point, with a ghtly smaller variability noted for ADHDRS-PI scores.

Overall, this study supports the use of the ADHD Rating Scale as a clinician administered and scored tool for assessing the severity of ADHD symptoms in pediatric patients. There is certainly value in obtaining information from multiple informants, however, this investigator-rated scale provides another option for researchers engaged in ADHD research and for office-based clinicians involved in the treatment of ADHD.

References

Achenback, T. M., & Edlebrock, C. (1986). *Manual for the Teacher Report Form and the Child Behavior Profile.* Burlingon: University of Vermont.

Barkley, R. A. (1990). *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment.* New York: Guilford.

Biederman, J., Faraone, S. V., Doyle, A., Lehman, B. D., Kraus, I., Perrin, J., et al. (1993). Convergence of the Child Behavior Checklist with structured interview-based psychiatric diagnoses of ADHD children with and without comorbidity. *Journal of Child Psychology and Psychiatry, 34* (7), 1241–1251.

Channon, E., & Butler, A. (1998, November). *Comparing investigators' use of rating scales such as PANSS in multi-investigator studies of schizophrenia.* Poster presentation at the 11th European College of Neuropsychopharmacology (ECNP) Congress.

Cohen, M., Becker, M. G., & Campbell, R. (1990). Relationships among four methods of assessment of children with attention deficit-hyperactivity disorder. *Journal of School Psychology, 28,* 189–202.

Conners, C. K. (1997). *Conners' rating scales: Revised technical manual.* North Towanda, NY: Multi-Health Systems.

Conners, C. K. (1999). Clinical use of rating scales in diagnosis and treatment of attention-deficit/ hyperactivity disorder. *Pediatric Clinics of North America, 46* (5), 857–870.

Cronbach, L. F. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

Deyo, R. A., Diehr, P., & Patrick, D. L. (1991). Reproducibility and responsiveness of health status measures: Statistics and strategies for evaluation. *Controlled Clinical Trials, 12,* 142S–158S.

DuPaul, G. J. (1991). Parent and teacher ratings of ADHD symptoms: Psychometric properties in a community-based sample. *Special Issue: Child Psychopharmacology; Journal of Clinical Child Psychology, 20,* 245–253.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *ADHD Rating Scale II: Checklists, norms, and clinical interpretation.* New York: Guilford.

Gadow, K. D., & Sprafkin, J. (1996). *Child and adolescent symptom inventories.* Odessa, FL: Psychological Assessment Resources.

Garland, E. J. (1998). Pharmacotherapy of adolescent attention deficit hyperactivity disorder: Challenges, choices, and caveats. *J Psychopharmacology, 12,* 385–395.

Goldman L. S., Genel M., Bezman R. J., & Slanetz P. J. (1998). Diagnosis and treatment of attention-deficit/ hyperactivity disorder in children and adolescents. *JAMA, 279,* 1100–1107.

Guy, W. (1976). *ECDEU assessment manual for psychopharmacology, revised.* Bethesda, MD: US Department of Health, Education, and Welfare.

Guyatt, G., Patrick, D., & Feeny, D. (1991). Glossary. *Controlled Clinical Trials, 12,* 274S–280S.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23,* 56–62.

Jacobsen, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19.

Landis, J., & Koch, G. (1977). The measuremen observer agreement for categorical data. *Biomet 33,* 159–174.

March, J. S., Sullivan, K., & Parker, J. (1999). Test-r reliability of the Multidimensional Anxiety S for Children. *Journal of Anxiety Disorders, 13* 349–358.

Miller, L. S., Klein, R. G., Piacentini, J., Abikoff, H., Shal R., Samoilov, A., et al. (1995). The New York Tea Rating Scale for Disruptive and Antisocial Behavior. *Journal of the American Academy of C and Adolescent Psychiatry, 34* (3), 359–370.

Perrin, E. S., Aaronson, N. K., Alonso, J., Burnam, A., L K., Patrick, D. L., et al. (1997, March). Instrument re criteria. *Medical Outcomes Trust.*

Reich, W., Welner, Z., & Herjanic, B. (1990). *Diagnc Interview for Children and Adolescents–Revi.* North Tonawanda, NY: Multi-Health Systems.

Stratford, P. W., Binkley, J. M., & Riddle, D. L. (1996). He status measures: Strategies and analytic method assessing change scores. *Physical Therapy, 76* (109–123.

Spencer, T., Biederman, J., Wilens, T., Prince, J., Hatch, Jones, J., et al. (1998). Effectiveness and tolerab of tomoxetine in adults with attention def hyperactivity disorder. *American Journa Psychiatry, 155,* 693–695.

Spencer, T., Biederman, J., Heiligenstein, J., Wilens Faries, D., Prince, J., et al. (2001). An open-label, ranging study of tomoxetine in children with atten deficit hyperactivity disorder. *Journal of Child Adolescent Psychopharmacology, 11*(3), 251-26

Ullman, R., Sleator, E., & Sprague, R. (1997). *ADL Comprehensive Teachers Rating Scale, e* Champaign, IL: MetriTech.

Douglas E. Faries, Ilker Yalcin, Donald Harder, John H. Heiligenstein are with Eli Lilly and Company.

Correspondence concerning this article should addressed to Douglas Faries, Ph.D., Eli Lilly and Comp Lilly Corporate Center, Indianapolis, Indiana 46 Phone: (317) 276-9801, Fax: (317) 433-6590, Electronic n faries_douglas_e@lilly.com

