



Parent–clinician agreement in rating the presence and severity of attention-deficit/hyperactivity disorder symptoms

Ellen Nobel¹ · J. Agnes Brunnekreef¹ · Russell J. Schachar² · Barbara J. van den Hoofdakker^{1,3} · Pieter J. Hoekstra¹

Received: 31 May 2018 / Accepted: 16 August 2018
© Springer-Verlag GmbH Austria, part of Springer Nature 2018

Abstract

We determined the validity of a parent-report questionnaire as a research tool for rating attention-deficit/hyperactivity disorder (ADHD) symptoms in children. Using Cohen's kappa and Pearson correlation, we examined the agreement between parent reports of ADHD symptoms (using the Swanson, Nolan and Pelham Questionnaire-IV; SNAP-IV) and clinical judgment (using a semi-structured parent interview). Also, we explored factors that may be associated with the level of agreement, using regression analyses. We found moderate levels of agreement for severity of overall ADHD ($r=0.43$) and for hyperactive-impulsive symptoms ($r=0.54$), but no significant agreement for inattentive symptoms. On individual symptom level (range kappa = -0.05 – 0.22) and for the presence/absence of ADHD (kappa = 0.14), agreement was poor. Therefore, we conclude that parent-report questionnaires may be acceptable to rate the overall severity of ADHD symptoms in treatment effect studies, but not to detect the presence of ADHD in epidemiological studies.

Keywords ADHD · Agreement · SNAP-IV · PICS-IV

Introduction

There is clear consensus on how to diagnose attention-deficit/hyperactivity disorder (ADHD) in school-aged children; this should be done on the basis of interviews with multiple informants such as parents and teachers who provide information on the child's behavior in different settings, like at home, school, or with peers [e.g., The National Institute for Health and Care Excellence (NICE)] (National Collaborating Centre for Mental Health & Royal College of Psychiatrists' Research and Training Unit 2009). Moreover, the impairment of the symptoms is considered as crucial in establishing the presence and the severity of ADHD (American Psychiatric Association 2013). However, such

an extensive assessment is not always feasible in epidemiological or treatment studies, e.g., due to time and financial constraints. Parent-rated ADHD questionnaires are then frequently used. An important and unresolved question is how valid the use of these questionnaires is for determining the presence (as in epidemiological studies) and severity (as in treatment effect studies) of ADHD symptoms, in comparison with clinician ratings based on a semi-structured interview.

A well-known parent-based questionnaire is the Swanson, Nolan and Pelham-IV (SNAP-IV; Swanson 1992), which is frequently used as primary outcome measure in treatment studies (e.g., McCann et al. 2014; The MTA Cooperative Group 1991; Yang et al. 2015; Wagner et al. 2014) and for assessing ADHD in epidemiological studies (e.g., Tsai et al. 2017; Yamada et al. 2013; Ullebø et al. 2012). The SNAP-IV determines the severity of ADHD symptoms by asking parents to rate the frequency of symptoms albeit without taking impairment into account. The validity of the SNAP-IV as a measure to determine the severity of ADHD and especially the presence of ADHD symptoms is uncertain. Some studies (Alda and Serrano-Troncoso 2013; Bussing et al. 2008) concluded that the SNAP-IV is a reliable and valid *screening* instrument to identify children with ADHD concerns, but no studies have examined the psychometric properties of the SNAP-IV as a *diagnostic* tool. While another well-known

✉ Ellen Nobel
e.nobel@me.com

¹ Department of Child and Adolescent Psychiatry, University of Groningen, University Medical Center Groningen, Hanzeplein 1 XA10, 9713 GZ Groningen, The Netherlands

² Department of Psychiatry, University of Toronto, Toronto, ON, Canada

³ Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Groningen, The Netherlands

rating scale, the Conners' parent and teacher scales (Conners 2001) had good sensitivity (83.5%), but a disappointing level of specificity (35.7%) as a diagnostic tool for ADHD, clinical ratings with semi-structured interviews with parents and teachers scored higher on both criteria (sensitivity: 91.8%; specificity: 70.7%) (Conners 2001).

One possible way to examine the value of a parent ADHD questionnaire for clinical trials and epidemiological studies is to examine the agreement between the parent-rated ADHD symptoms and those rated by clinicians with a semi-structured clinical interview. However, besides studies investigating agreement in children with externalizing disorders in general (Bird et al. 1992; De Los Reyes and Kazdin 2005; Jensen and Weisz 2002; Klein et al. 2010), no studies have focused on a specific diagnosis such as ADHD, nor did they compare a parent questionnaire and a clinician judgment based on a semi-structured interview.

This study aimed to provide a first exploratory step in examining the suitability of the SNAP-IV for two types of studies: epidemiological studies and treatment effect studies, by determining the agreement with regard to the presence and severity of ADHD symptoms between the SNAP-IV and clinicians' ratings based on a semi-structured parent interview. For this, we chose the Parent Interview for Child Symptoms (PICS-IV; Schachar et al. 1995), as it closely follows the ADHD criteria of the diagnostic and statistical manual of mental disorders, fourth edition (DSM-IV; American Psychiatric Association 1994) of which it provides concrete operationalizations. With the PICS-IV, it is the clinician who judges the presence and severity of each symptom based on parent reports of the frequency and the associated impairment of the behavior. While the PICS-IV is also dependent on parental input, the involvement of clinician's weighing is a major difference with a parent questionnaire in which ratings are being done directly by the parents. It may be questioned whether parents are able to assess the presence of DSM-IV-based ADHD criteria, as the textual descriptions of the DSM are meant to be employed by individuals with appropriate clinical training and experience in diagnosis, which parents typically lack.

First, we examined agreement on two different levels, i.e., (1) meeting/not meeting overall DSM-IV symptom criteria for the presence of ADHD and (2) total severity of ADHD symptoms. The former is mostly used in epidemiological studies, and the latter in treatment effect studies. Additionally, to explore potential differences between agreement of parents and clinicians on specific symptoms, we also examined agreement on (3) the presence/absence of each of the 18 ADHD symptoms and (4) severity ratings of each of the 18 ADHD symptoms. We expected to find low levels of agreement on symptom level and higher agreement for meeting overall DSM-IV symptom criteria (American Psychiatric Association 1994) and total severity of ADHD symptoms

as used in epidemiological and treatment effect studies, respectively. In addition, we explored whether certain child variables (i.e., age, sex, number of ADHD symptoms, severity of oppositional defiant behaviors, and medication status) and parental factors (i.e., education level, parenting stress, depressive symptoms, and ADHD symptoms of the primary caretaker) would be associated with the level of agreement. In children with externalizing disorders, better agreement between clinicians and parents has been demonstrated for males, younger children, more impaired patients, and in those with greater symptom severity (Bird et al. 1992; Jensen and Weisz 2002; Klein et al. 2010). Some findings have indicated that higher parental stress, more parental anxiety, and depression symptoms are associated with less agreement between parents' ratings and those of other informants (De Los Reyes and Kazdin 2005).

Method

Participants

Patients were recruited from three different mental health care centers in the North of the Netherlands. Data were obtained from a baseline assessment of a sample of 70 children who participated in a randomized controlled trial in which the effect of a home-based behavioral parent training for school-aged children with ADHD and behavior problems was being evaluated. All patients had previously been offered routine treatments (i.e., pharmacological treatment and/or parent training) without sufficient response, often due to non-completion of the treatment.

Children with ADHD, aged 6–13 years, were eligible to participate in the study. At the time of referral to the study, all children met diagnostic DSM-IV-TR criteria for ADHD as confirmed by clinical interviews with parents and teacher. Comorbidity was not an exclusion criterion. The child had to have a full scale, verbal and performance IQ above 70. Table 1 contains the main sample characteristics.

The Medical Ethical Committee of the University Medical Center in Groningen has assigned ethical approval for the study (METC nr 2010.289). The trial is registered at <http://www.trialregister.nl>, home-based behavioral treatment for ADHD; NTR3021. All persons gave their informed consent prior to their inclusion in the study.

Measures

The ADHD section of the SNAP-IV is an 18-item ADHD symptom scale (nine inattentive and nine hyperactive/impulsive DSM-IV symptoms) that is rated on a four-point scale from not at all (0) to very much (3) with regard to the previous week (Swanson 1992), as rated by the primary

Table 1 Sample characteristics ($N=70$)

Number of boys (%)	50 (71.4%)
Mean age of child in years (SD; range)	8.9 (1.5; 6.0–12.1)
Mean intelligence of child ^a (SD; range)	94.0 (11.3; 71–119)
Number using ADHD medication (%)	61 (87.1%)
Number of highest finished educational level (%)	
Primary school	4 (5.7%)
High school	23 (32.9%)
Undergraduates programs	35 (50%)
College	8 (8%)
Mean parenting stress ^b (SD; range)	96.9 (17.2; 56–130)
Mean parental depression ^c (SD; range)	11.2 (8.5; 0–14)
Mean parental ADHD ^d (SD; range)	14.1 (9.6; 0–40)
Mean child ODD ^e (SD; range)	10.8 (4.3; 0–20)

SD standard deviation, ADHD attention-deficit/hyperactivity disorder, ODD oppositional defiant disorder

^aFull scale IQ. ^bTotal score on the Parenting Stress Index-Short Form (PSI-SF). ^cTotal score on the Beck Depression Inventory-II. ^dTotal score on the Adult ADHD rating scale. ^eSeverity of ODD symptoms as rated by the Parent Interview of Child Symptoms

caretakers of the children (i.e., mothers in 97% of all subjects). Reliability of this scale has been found to be high (coefficient alpha of 0.90 for inattentive symptoms and of 0.79 for hyperactive-impulsive symptoms; Bussing et al. 2008).

The PICS-IV is a semi-structured clinical interview with the parents that covers the full range of DSM-IV symptoms of ADHD and symptoms of oppositional defiant disorder (ODD), based on information restricted to direct observations in naturalistic contexts (daily situations) made by the parents. The clinician conducting the interview decides whether each symptom is present and rates its severity based on the frequency, the impairment that is associated with a particular symptom, and on the appropriateness of the behavior in relation to the child's age. Each symptom is rated by the clinician on a four-point scale over the previous half year, ranging from symptom is absent (0), to marked abnormality (3). The clinician has to score blank when he/she does not get enough detailed information to score the symptom (De Los Reyes and Kazdin 2005). The inter-rater reliability of the PICS-IV for an ADHD diagnosis has been shown to be good ($\kappa=0.73$). Kappas for individual ADHD symptoms have been shown to range from 0.50 to 0.96. Intra-class correlation coefficients for overall symptom scores were also excellent (inattentive symptoms: 0.93; hyperactivity/impulsivity symptoms: 0.97; Ickowics et al. 2006). Clinicians in the study were psychologists or nearly graduated psychologist students (seven in total) who had been intensively trained in the PICS-IV and supervised by an experienced clinician.

Education level was defined as the highest level of completed education of the primary caretaker. Parental stress by the primary caretaker was assessed with the Parenting Stress Index-Short Form (PSI-SF). This is a 36-item self-report measure where parents use a five-point scale to indicate the degree to which they agree with each statement (Abidin 1995). Depressive and ADHD symptoms of the primary caretaker were rated with, respectively, the Beck Depression Inventory-II (BDI-II; Beck et al. 1996; a 21-item self-report form with a four-point scale) and the Adult ADHD Rating Scale (AARS; Barkley and Murphy 1998; an 18-item self-report form with a four-point scale). Table 1 provides means, standard deviations, and ranges of all measurements.

The primary caretaker filled in all questionnaires at the same time; the PICS-IV was always done after completion of the SNAP-IV, thus avoiding potential confounding of the parent ratings through knowledge of the clinician's questions. The mean time between completion of the questionnaires and the PICS-IV was 0.74 weeks ($SD=1.45$; range 0–8 weeks); for the majority of the participants, all data were obtained within two weeks ($n=63$; 90%). With both the SNAP-IV and PICS-IV, the behavior of the child was evaluated as how the child is currently seen by the parents in the home setting, regardless of the possible use of ADHD medication when prescribed.

Data analyses

Agreement: The agreement between parents and clinicians in severity of inattention, hyperactivity-impulsivity, and overall ADHD symptoms was determined with the Pearson product-moment correlation. The severity of ADHD symptoms according to the SNAP-IV and the PICS-IV was the sum of all individual symptom ratings.

Cohen's kappas (Cohen 1960) were determined for parent–clinician agreement with respect to the ordinal severity ratings on each of the 18 ADHD symptoms and for parent–clinician agreement on the presence/absence of each ADHD symptom (ratings 0 or 1 versus ratings 2 or 3 on the SNAP-IV and PICS-IV). Furthermore, the agreement between parents and clinicians with respect to criterion A of the DSM-IV description of ADHD (i.e., the child has six or more inattentive symptoms and/or hyperactive-impulsive symptoms) was determined. Based on the number of symptoms on the SNAP-IV (six or more symptoms per domain, following the DSM-IV), the children were divided into four categories (ADHD combined type, ADHD predominantly inattentive type, ADHD predominantly hyperactive-impulsive type, and not fulfilling full ADHD criteria). The same was done regarding the number of symptoms on the PICS-IV. Agreement in ADHD types was again calculated with Cohen's kappas. Kappa scores were interpreted using Altman's (1994)

conventions: < 0.20 as poor, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as good, and 0.81–1.00 as very good level of agreement.

Participants who had more than 10% missing values per measurement were excluded from the analyses (loss of 3 participants). In case of less missing values (SNAP $n = 4$; PICS $n = 12$), these were replaced by the average rating-per-item subscale score.

Factors influencing agreement: first, we explored which child characteristics (i.e., number of ADHD symptoms and severity of oppositional defiant behavior) and parental variables (i.e., education level, parenting stress, depressive symptoms, and ADHD symptoms of the primary caretaker) might have been related to the level of agreement by determining Pearson product-moment correlations of all continuous variables with the difference scores between the ADHD total, ADHD inattentive, and ADHD hyperactive-impulsive scores on the SNAP-IV and PICS-IV. We first used t-tests to assess difference scores between the SNAP-IV and the PICS-IV between categories (i.e., sex of the child and the presence/absence of current ADHD medication of the child). Factors with a p value below 0.10 were subsequently analyzed in a multiple regression analysis. When running these analyses, we checked for multicollinearity using variance inflation factors.

Parenting stress, depressive symptoms, and ADHD symptoms of the primary caretaker were determined by calculating the total scores on the PSI-SF, BDI-II, and AARS, respectively. Missing values (less than 10% per questionnaire) were replaced by the average rating-per-item subscale score (PSI-SF $n = 8$; BDI $n = 5$; AARS $n = 5$). The presence of ADHD symptoms of the child was based on ratings of 2 or 3 ratings on PICS-IV ADHD items. The level of the child's ODD was based on the sum of all PICS-IV ODD symptoms each rated on a 0–3 scale.

We also performed a number of sensitivity analyses where missing values were analyzed differently. First, when establishing the level of agreement between the severity and the presence of single symptoms and meeting criteria A of the DSM-IV, blank scores on the PICS-IV were treated as zero scores (instead of as missing values). Second, when establishing total scores on the different questionnaires, we replaced missing values in a sensitivity analysis with a zero score (instead of the average rating-per-item).

We considered p values < 0.05 as indicating nominal statistical significance. However, as we performed 49 independent tests, we controlled for multiple-comparison using the Benjamini–Hochberg procedure (false discovery rate). This procedure ranks the p values in ascending order (the smallest has a rank of 1), computes a critical value based on the rank and the number of tests, and considers results as significant of which the p value is smaller than the critical value. This procedure suggested considering a test as

significant if $p < 0.009$. Significance in tables and text has been determined with respect to this value.

Results

Parent clinician agreement

With regard to overall symptom severity (the sum of all individual symptom ratings; $n = 67$), the SNAP-IV severity of hyperactivity/impulsivity ($M = 15.9$; $SD = 5.6$; range 3–27) and total ADHD symptoms ($M = 32.1$; $SD = 9.6$; range 5–51) were significantly related to the PICS-IV severity of hyperactivity/impulsivity ($M = 14.2$; $SD = 5.0$; range 4.5–27) and total ADHD symptoms ($M = 25.8$; $SD = 6.9$; range 9–46). For the severity of the hyperactivity/impulsivity symptoms ($r = 0.54$, $p < 0.009$, 95% CI 0.35–0.69) and for the total ADHD symptoms ($r = 0.42$, $p < 0.009$, 95% CI 0.20–0.60), moderate effects were found. No statistically significant effect for the severity of inattention ($r = 0.33$, $p = 0.01$, 95% CI 0.01–0.53) was found. Also, the SNAP-IV severity of inattention ($M = 16.2$; $SD = 5.2$; range 2–27) was not significantly related to the PICS-IV severity of inattention ($M = 11.5$; $SD = 3.9$; range 3–22). A sensitivity analysis (where missing values were replaced with a zero score instead of average rating-per-item score) yielded similar results.

Furthermore, we determined Cohen's kappas at the level of meeting criteria A of the DSM-IV classification. We found no significant kappas with regard to agreement on types of ADHD ($\kappa = 0.14$; $p = 0.067$; $n = 52$, 95% CI -0.05 – 0.34) and on the presence or absence of any ADHD type ($\kappa = 0.19$; $p = 0.139$; $n = 52$, 95% CI -0.15 – 0.53). Table 2 presents Cohen's kappas for agreement between the SNAP-IV and PICS-IV regarding severity and the presence/absence of single ADHD symptoms. In contrast to the moderate agreement between parents and clinicians on the overall severity of ADHD symptoms, the levels of agreement between parents and clinicians on the severity of single ADHD symptoms and on the presence/absence of single symptoms were mostly poor. For one of the 18 ADHD symptoms ("talks excessively"), significant agreement with regard to severity of ADHD symptoms was found, and for five symptoms ("loses things"; "fidgeting/squirming"; "runs about/climbs excessively"; "talks excessively"; and "blurts out answers before questions were completed") concerning the presence/absence of symptoms. For all other ADHD symptoms, significant agreement was absent. A sensitivity analysis (where a blank score on the PICS-IV was transformed into a zero score instead of a missing value) resulted in similar outcomes.

Table 2 Level of agreement between the SNAP-IV and PICS-IV in the severity and presence of single ADHD symptoms

ADHD symptom	<i>n</i>	Severity	<i>p</i>	95% CI	Presence	<i>p</i>	95% CI
Fails to pay attention to details	68	− 0.004	0.958	− 0.14–0.13	0.29	0.014	0.06–0.52
Cannot sustain attention	70	− 0.05	0.339	− 0.15–0.15	− 0.07	0.409	− 0.35–0.11
Does not seem to listen when spoken to	69	0.15	0.038	0.00–0.30	0.17	0.147	− 0.05–0.39
Does not follow through on instructions	67	0.04	0.586	− 0.11–0.19	0.11	0.341	− 0.13–0.35
Disorganizing tasks and activities	66	0.08	0.180	− 0.05–0.21	0.09	0.410	− 0.13–0.31
Avoids tasks that require sustained mental effort	66	− 0.01	0.926	− 0.14–0.12	0.04	0.744	− 0.22–0.30
Loses things	70	0.08	0.235	− 0.06–0.22	0.46	< 0.001*	0.25–0.67
Distractibility	70	0.06	0.303	− 0.05–0.17	0.06	0.357	− 0.05–0.17
Forgetful	69	0.04	0.567	− 0.11–0.19	0.14	0.214	− 0.07–0.35
Fidgeting/squirming	70	0.06	0.428	− 0.10–0.22	0.34	0.005*	0.10–0.58
Difficulty staying seated	65	− 0.003	0.964	− 0.14–0.14	0.22	0.066	− 0.01–0.43
Runs about/climb excessively	70	0.10	0.150	− 0.04–0.24	0.38	0.001*	0.17–0.59
Not playing quietly	70	− 0.003	0.968	− 0.14–0.14	0.16	0.170	− 0.06–0.38
On the go/driven by a motor	70	0.15	0.031	0.00–0.30	0.26	0.027	0.02–0.50
Talks excessively	70	0.22	0.002*	0.07–0.37	0.42	< 0.001*	0.20–0.64
Blurts out answers before questions were completed	70	0.17	0.017	0.01–0.34	0.30	0.008*	0.09–0.51
Does not wait for his/her turn	65	0.11	0.078	− 0.02–0.24	0.17	0.084	− 0.02–0.36
Interrupts/intrudes on others	67	0.09	0.107	− 0.04–0.21	0.16	0.052	− 0.01–0.30

SNAP-IV = Swanson, Nolan and Pelham Questionnaire –IV, *PICS-IV* = Parent Interview of Child Symptoms, *ADHD* = attention-deficit/hyperactivity disorder, *n* = number of participants, Severity = kappa of agreement on symptom severity, Presence = kappa of agreement on presence/absence of symptom, 95% CI = 95% confidence interval

* $p < 0.009$ (according to Benjamini–Hochberg correction for multiple testing)

Factors influencing agreement

In explorative analyses (Pearson product-moment correlations for continuous variables and *t*-tests for categorical variables), the child's current medication use ($t = 1.81$; $p = 0.076$), parenting stress (PSI-SF; $r = 0.21$, $p = 0.090$), and the total number of ADHD symptoms according to the PICS-IV ($r = -0.24$; $p = 0.053$) were identified as candidate factors for agreement on total ADHD severity. For agreement on ADHD inattentive severity, candidate factors were parenting stress (PSI-SF; $r = 0.20$; $p = 0.098$) and number of inattentive ADHD symptoms according to the PICS-IV ($r = -0.31$; $p = 0.012$). Current medication use of the child ($t = 1.774$; $p = 0.081$), number of hyperactive-impulsive symptoms based on the PICS-IV ($r = -0.32$; $p = 0.010$), and total number of ADHD symptoms according to the PICS-IV ($r = -0.24$; $p = 0.051$) were the factors for agreement on ADHD hyperactive-impulsive score. Severity of comorbid oppositional defiant behavior, parental educational level, and parental psychopathology (depression and ADHD symptoms) were not related to agreement on total, hyperactive, or inattentive ADHD severity (p values > 0.10).

Parents rated ADHD severity higher than did clinicians, with the difference between the SNAP-IV-based total ADHD severity and the PICS-IV-based total ADHD severity being on average 6.3 points (SD = 9.1; range − 11.4–27.0);

only 20.9% of the primary caretakers scored lower on the SNAP-IV than did the clinician on the PICS-IV. The difference between the severity score on the inattentive ADHD symptoms according to the SNAP-IV- and the PICS-IV-based severity was on average 4.7 points (SD = 5.4; range − 7.4–16.0); only 16.4% of the parent scores were lower than the clinician scores. For the total severity on hyperactive-impulsive symptoms, the average difference between the two measurements was 1.7 points (SD = 5.1; range − 9.0–13.0). Here, 38.8% of the parents scored lower than the clinicians.

Subsequent regression analysis established that the more parenting stress the parents experienced (PSI-SF; $\beta = 0.13$, $p = 0.038$, 95% CI 0.01–0.26) and the less ADHD symptoms were scored by the clinician (PICS-IV; $\beta = -0.80$, $p = 0.025$, 95% CI − 1.50 to − 0.104), the larger was the difference between the parents and the clinician, with the primary caretakers scoring the ADHD severity higher than the clinicians. The child's medication status ($\beta = -4.57$, $p = 0.145$, 95% CI − 10.77–1.62) did not significantly influence parent–clinician differences. The regression analysis for the difference between SNAP-IV and PICS-IV on inattentive severity revealed that the lower the number of ADHD inattentive symptoms was scored by the clinician (PICS-IV; $\beta = -0.99$, $p = 0.008$, 95% CI − 1.72 to − 0.26), the bigger was the difference between the parents and the clinicians in severity, with the parents scoring higher than the clinicians.

No statistically significant results were found for the factor parenting stress ($\beta = -0.07$, $p = 0.063$, 95% CI -0.01 – 0.14). Furthermore, the factors child's current medication use of the child ($\beta = -2.64$, $p = 0.140$, 95% CI -6.17 – 0.89), number of hyperactive-impulsive symptoms based on the PICS-IV ($\beta = -0.84$, $p = 0.105$, 95% CI -1.86 – 0.18), and the total number of ADHD symptoms according to the PICS-IV ($\beta = 0.14$, $p = 0.700$, 95% CI -0.59 – 0.87) were not significantly associated with the difference score between the SNAP-IV and the PICS-IV on the hyperactive-impulsive severity. Multicollinearity was found not to be an issue, as the variance inflation factors remained well below the accepted boundary of 10 (Hair et al. 2010), with all variance inflation factors values being below 4.

A sensitivity analysis (where missing values were replaced with a zero score instead of average rating-per-item score) for the total ADHD severity and the hyperactive-impulsive severity yielded similar results. However, in the sensitivity analysis the inattentive severity of the number of inattentive symptoms according to the PICS-IV ($\beta = -0.99$, $p = 0.009$) fell just short of significance in association with the level of agreement with regard to inattentive severity.

Discussion

We investigated the agreement between parent reports of ADHD symptoms as measured by a parent questionnaire (the SNAP-IV) and clinicians' ratings of ADHD symptoms based on a semi-structured parent interview (the PICS-IV), to evaluate the suitability of parent questionnaires for two types of studies: epidemiological studies and treatment effect studies. First, we found no significant agreement between the SNAP-IV and the PICS-IV on ADHD type and on the presence or absence of ADHD in school-aged children. Based on these results, it is not recommended to use parent questionnaires to determine the presence of ADHD, as is commonly done in epidemiological studies (e.g., Tsai et al. 2017; Yamada et al. 2013; Ullebø et al. 2012). While there are a few changes in ADHD diagnostic criteria between the DSM-IV-TR and the current (fifth) edition of the DSM (DSM 5, American Psychiatric Association 2013), these are relatively minor and there is no reason why DSM 5-based parent questionnaires would compare differently to clinicians' interviews than those based on the DSM-IV-TR.

Second, in treatment effect studies, parent questionnaires such as the SNAP-IV are frequently used to determine the change over time in severity of ADHD symptoms (e.g., McCann et al. 2014; The MTA Cooperative Group 1991; Yang et al. 2015; Wagner et al. 2014). We found moderate levels of agreement with respect to total ADHD symptom severity, and therefore, these questionnaires seem to be suitable as severity rating in treatment effect studies. Most

parents (79.1%) rated ADHD severity of their child higher compared with the judgment of clinicians.

Parents appeared to be in line with clinicians in evaluating the presence and severity of the total severity of hyperactive and impulsive symptoms but not so in assessing the total severity of inattentive symptoms of their child. The same pattern was found for the agreement between parents and clinicians on single ADHD symptoms; the symptom with a significant level of agreement in severity was a hyperactive symptom, and also, four of the five symptoms with a significant level of agreement in the presence/absence were hyperactive-impulsive symptoms. Hyperactive-impulsive symptoms are generally more readily externally observable behaviors, while inattentive symptoms are less recognizable; this may explain that the agreement between parents and clinicians was better for hyperactive-impulsive symptoms than for inattentive symptoms. This assumption is confirmed by the significant agreement level on the inattentive symptom that is relatively concrete and observable: "often loses things." This symptom is not only more concrete, but it also shows less overlap with other inattentive symptoms, in contrast to the overlap between "cannot sustain attention" and "distractibility" and the overlap between "does not follow through on instruction," "forgetful," and "disorganizing tasks and activities."

These findings may implicate that in order to make parent questionnaires suitable for treatment effect studies, especially the inattentive symptoms of ADHD should be described in a more concrete, observable way instead of following exact DSM descriptions, which have never been designed to be rated by parents. In addition, not only a frequency rating of the behavior but also an impairment rating should preferably be given by the parent. Future studies might investigate whether such adjustments in a parent questionnaire would lead to better levels of agreement between a parent rating and a clinician's judgment. Some parent questionnaires do already contain some global form of impairment rating, e.g., the National Institute for Children's Health Quality ADHD rating scale (NICHQ Vanderbilt Assessment Scales 2002) asks parents to rate their child's performance in a variety of domains, yet do not break-down whether the presence of each ADHD symptom criterion is justified by it being associated with impairment. However, an additional advantage of a semi-structured interview over parent-report questionnaires in treatment effect studies is the possibility to establish blind measurements (Daley et al. 2014; Sonuga-Barke et al. 2013). It would also be worthwhile to investigate whether teachers show better agreement with clinicians with regard to inattentive symptoms, as these may perhaps be easier observed in a classroom setting, where sustained attention and engagement with tasks are critically important.

The number of child ADHD symptoms appeared to influence the level of agreement on the severity of inattentive

symptoms of ADHD between parents and clinicians. Fewer child ADHD symptoms were associated with more disagreement between parents and clinicians, whereas parents tend to estimate the severity of the inattentive symptoms higher than the clinician. Notably, we did not find factors that were related to the agreement on the severity of the more concrete, observable hyperactive-impulsive symptoms. Apparently, child factors have no significant influence on agreement when ADHD symptoms are described in concrete, objective, and observable terms such as most hyperactive and impulsive symptoms, in contrast to the less distinctive and observable inattentive symptoms. This may point to the necessity of clinical expertise to assess such symptoms, and again therefore, semi-structured interviews should be considered in treatment effect studies. Furthermore, we found no relation between level of agreement and parental educational level, parental psychopathology (depression and ADHD symptoms), and severity of comorbid oppositional defiant behavior.

While not the main focus of the study, our findings confirm existing recommendations to use semi-structured interviews to diagnose children suspected to have ADHD in clinical practice rather than to rely on parent-report questionnaires, preferably by collecting data from multiple settings. Indeed, the NICE-guideline states that a diagnosis of ADHD should not be made on the basis of rating scales alone (National Collaborating Centre for Mental Health & Royal College of Psychiatrists' Research and Training Unit 2009). This recommendation also aligns with the finding that questionnaires alone fail to systematically assess all relevant clinical features of a disorder (Parker and Corkum 2016; Raiker et al. 2017).

Limitations

Some limitations of this study need to be acknowledged. The most important limitation is the relatively modest sample size and the presence of missing data. Moreover, we possibly included a diagnostically more complex and mostly medicated group. This selection bias may have resulted in over-estimation of ADHD symptom severity by parents. Another limitation of the study is that the SNAP-IV and PICS-IV give different instructions in judging the severity of an ADHD symptom and with regard to setting. That is, the SNAP-IV asks parents to assess how frequent a specific symptom is present at home or at school in the previous week, whereas in the PICS-IV the clinician is asked to rate the severity of a specific symptom in specific situations at home in the previous half year, taking into account the frequency, impairment, and age appropriateness of the behavior. To improve the comparability of the scorings on a parent questionnaire and a semi-structured interview in future studies, it would be preferable to give parents and clinicians the

same timescale and context. This difference in the timescale can also influence the agreement with regard to the presence/absence of ADHD, and therefore, our conclusion that the SNAP-IV is less acceptable for epidemiological studies needs to be interpreted carefully.

Furthermore, it should be remembered that the reliability and validity estimates from the SNAP-IV and PICS-IV stem from English versions and may differ from the Dutch versions that we used. However, the PICS-IV is a semi-structured interview in which clinicians do not use literal wordings, whereas translation of the SNAP-IV was in line with the official DSM-IV-TR Dutch translation. Thus, we are confident that the influence of translation is negligible. Finally, when interpreting the level of agreement based on correlation coefficients, one should keep in mind that these are insensitive against linear differences in severity between both methods. Furthermore, while the PICS-IV has been previously shown to have good inter-rater reliability (Ickowicz et al. 2006), we did not assess inter-rater reliability for the PICS-IV assessors of the current study, as individual children were only rated by one clinician instead of multiple.

Conclusions and future directions

In conclusion, a DSM-based parent questionnaire appears to be less valid to establish whether an ADHD diagnosis is present or absent, as is commonly done in epidemiological studies. Therefore, the use of parent questionnaires for diagnostic purpose should be reconsidered, and we recommend the use of continuous ratings of ADHD severity rather than categorical measures of ADHD when parent ratings are being used in epidemiological samples. Moreover, parent questionnaires appear to be appropriate instruments to rate the overall severity of hyperactive-impulsive ADHD symptoms in children in treatment effect studies, but less suitable to rate the severity of individual symptoms. This is especially true for children with fewer ADHD symptoms. Future research should examine the test–retest reliability of parent questionnaires, as the consistency of measurement across time is a critical indicator of a measure's utility as a treatment outcome index. A semi-structured interview in which a clinician evaluates the presence and severity of the ADHD behavior may be a better alternative, because clinicians are better at rating the more internal inattention symptoms and, moreover, because these instruments can be used in a blinded way. Parent questionnaires can, however, be improved by describing the inattention symptoms in more concrete, observable behaviors of a child. Future research should also focus on the influence of the use of different clinicians in scoring ADHD symptoms. Unfortunately, this could not be done reliably in the current study, given the small sample size.

Acknowledgements The authors thank the participating families, clinicians, and the health care centers Accare, Jonx Lentis, and GGZ In de Bres.

Funding This study has been financially supported by the Netherlands Organization for Health Research and Development (ZonMw, nr 15700.3010).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standard of the national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Abidin RR (1995) Parenting stress index (PSI) manual, 3rd edn. Pediatric Psychology Press, Charlottesville
- Alda JA, Serrano-Troncoso E (2013) Attention-deficit hyperactivity disorder: agreement between clinical impression and the SNAP-IV screening tool. *Actas Esp Psiquiatr* 41(2):76–83
- Altman DG (1994) Practical statistics for medical research. Chapman and Hall, London
- American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington (DC)
- American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Association, Washington, DC
- Barkley RA, Murphy KR (1998) Attention deficit hyperactivity disorder: a clinical workbook, 3rd edn. Guilford, New York
- Beck AT, Steer RA, Brown GK (1996) Beck depression inventory—second edition manual. The Psychological Corporation, San Antonio
- Bird HR, Gould MS, Staghezza B (1992) Aggregating data from multiple informants in child psychiatry and epidemiological research. *J Am Acad Child Adolesc Psychiatry* 31:78–85
- Bussing R, Fernandez M, Harwood M, Hou W, Wilson GC, Eyberg SM, Swanson JM (2008) Parent and teacher snap-iv ratings of attention deficit/hyperactivity disorder symptoms: psychometric properties and normative ratings from a school district sample. *Assessment* 5(3):317–328
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:121–153
- Conners CK (2001) Conners' rating scales: revised technical manual. Multi-Health Systems, Toronto
- Daley D, Van der Oord S, Ferrin M, Danckaerts M, Doepfner M, Cortese S, Sonuga-Barke EJ, (European ADHD Guidelines Group) (2014) Behavioral interventions in attention-deficit/hyperactivity disorder: a meta-analysis of randomized controlled trials across multiple outcome domains. *J Am Acad Child Adolesc Psychiatry* 53(8):835–847
- De Los Reyes A, Kazdin AE (2005) Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychol Bull* 131:483–509
- Hair JF, Black WC, Babin BJ, Anderson RE (2010) Multivariate data analysis, 7th edn. Pearson Prentice Hall, Upper Saddle River
- Ickowicz A, Schachar RJ, Sugerman R, Chen SX, Millette C, Cook L (2006) The parent interview for child symptoms: a situation-specific clinical research interview for attention-deficit hyperactivity and related disorders. *Can J Psychiatry* 51(5):325–328
- Jensen AL, Weisz JR (2002) Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnosis for clinic-referred children and adolescents. *J Consult Clin Psychol* 70:158–168
- Klein JB, Lavigne JV, Seshadri R (2010) Clinician-assigned and parent-report questionnaire-derived child psychiatric diagnosis: correlates and consequences of disagreement. *Am J Orthopsychiatry* 80(3):375–385
- McCann DC, Thompson M, Daley D, Barton J, Laver-Bradbury C, Hutchings J, Coghill D, Stanton L, Maishman T, Dixon L, Caddy J, Chorooglou M, Raftery J, Sonuga-Barke EJ (2014) Study protocol for a randomized controlled trial comparing the efficacy of a specialist and a generic parenting program for the treatment of preschool ADHD. *Trials* 15:142–162
- National Collaborating Centre for Mental Health & Royal College of Psychiatrists' Research and Training Unit (2009) Attention deficit hyperactivity disorder: The NICE guideline on diagnosis and management of ADHD in children, young people and adults. The British Psychological Society and The Royal College of Psychiatrists, Leicester/London
- NICHQ Vanderbilt Assessment Scales (2002) nichq.org/~media/files/resources/nichqvanderbiltassessment-full.ashx. Accessed 23 Aug 2018
- Parker A, Corkum P (2016) ADHD diagnosis: as simple as administering a questionnaire or a complex diagnostic process? *J Atten Disord* 20(6):478–486
- Raiker JS, Freeman AJ, Perez-Algorta G, Frazier TW, Findling RL, Youngstrom EA (2017) Accuracy of Achenbach scales in the screening of attention-deficit/hyperactivity disorder in a community mental health clinic. *J Am Acad Child Adolesc Psychiatry* 56(5):401–409
- Schachar RJ, Ickowicz A, Wachsmuth R (1995) Parent interview for child symptoms (P.I.C.S.-4). The Hospital for Sick Children, Toronto (**unpublished document**)
- Sonuga-Barke EJ, Brandeis D, Cortese S, Daley D, Ferrin M, Holtmann M, Stevenson J, Danckaerts M, Van der Oord S, Döpfner M, Dittmann RW, Simonoff E, Zuddas A, Banaschewski T, Buitelaar J, Coghill D, Hollis C, Konofal E, Lecendreux M, Wong ICK, Sergeant J, (European ADHD Guidelines Group) (2013) Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. *Am J Psychiatry* 170(3):275–289
- Swanson J (1992) School-based assessments and interventions for ADD students. K.C. Publications, Irvine
- The MTA Cooperative Group (1991) A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 56:1073–1086
- Tsai CJ, Chen YL, Lin HY, Gau SS (2017) One-year trajectory analysis for ADHD symptoms and its associated factors in community-based children and adolescents in Taiwan. *Child Adolesc Psychiatry Mental Health* 11:28. <https://doi.org/10.1186/s13034-017-0165-4>
- Ullebø AK, Posserud MB, Heiervang E, Obel C, Gillberg C (2012) Prevalence of the ADHD phenotype in 7- to 9-year-old children: effects of informant, gender and non-participation. *Soc Psychiatry Psychiatr Epidemiol* 47(5):763–769
- Wagner DJ, Vallerand IA, McLennan JD (2014) Treatment receipt and outcomes from a clinic employing the attention-deficit/

- hyperactivity disorder treatment guideline of the children's medication algorithm project. *J Child Adolesc Psychopharmacol* 24(9):472–480
- Yamada DC, Porter AA, Conway JL, LeBlanc JC, Shea SE, Hancock- Friesen CL, Warren AE (2013) Early repair of congenital heart disease associated with increased rate of attention deficit hyperactivity disorder symptoms. *Can J Cardiol* 29(12):1623–1628
- Yang R, Gao W, Li R, Zhao Z (2015) Effect of Atomoxetine on the cognitive functions in treatment of attention deficit hyperactivity disorder in children with congenital hypothyroidism: a pilot study. *Int J Neuropsychopharmacol* 18:1