

Attention deficit hyperactivity disorder: concordance of the adolescent version of the Composite International Diagnostic Interview Version 3.0 (CIDI) with the K-SADS in the US National Comorbidity Survey Replication Adolescent (NCS-A) supplement

JENNIFER GREIF GREEN,¹ SHELLI AVENEVOLI,² MATTHEW FINKELMAN,³
MICHAEL J. GRUBER,¹ RONALD C. KESSLER,¹ KATHLEEN R. MERIKANGAS,⁴
NANCY A. SAMPSON¹ & ALAN M. ZASLAVSKY¹

1 Department of Health Care Policy, Harvard Medical School, Boston, MA, USA

2 Division of Developmental Translational Research, National Institute of Mental Health, Bethesda, MD, USA

3 Tufts University School of Dental Medicine, Boston, MA, USA

4 Genetic Epidemiology Branch, Intramural Research Program, National Institute of Mental Health, Bethesda, MD, USA

Key words

attention deficit hyperactivity disorder, WHO Composite International Diagnostic Interview (CIDI), validity, National Comorbidity Survey Replication Adolescent Supplement (NCS-A)

Correspondence

Ronald C. Kessler, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115, USA.

Telephone (+1) 617-432-3587

Fax: (+1) 617-432-3588

Email: Kessler@hcp.med.harvard.edu

Received 17 September 2008;

revised 31 March 2009;

accepted 15 May 2009

Abstract

This paper evaluates the internal consistency reliability and concurrent validity of the assessment of Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) attention deficit hyperactivity disorder (ADHD) in the adolescent version of the World Health Organization (WHO) Composite International Diagnostic Interview Version 3.0 (CIDI). The CIDI is a lay-administered diagnostic interview that was carried out in conjunction with the US National Comorbidity Survey Adolescent Supplement, a US nationally representative survey of 10,148 adolescents and their parents. Internal consistency reliability was evaluated using factor and item response theory analyses. Concurrent validity was evaluated against diagnoses based on blinded clinician-administered interviews. Inattention and hyperactivity-impulsivity items loaded on separate but correlated factors, with hyperactivity and impulsivity items forming a single factor in parent reports but separate factors in youth reports. We were able to differentiate hyperactivity and impulsivity factors for parents as well by eliminating a subset who endorsed zero ADHD items from the factor analysis. Although concurrent validity was relatively weak, decomposition showed that this was due to low validity of adolescent reports. A modified CIDI diagnosis based exclusively on parent reports generated a diagnosis that had good concordance with clinical diagnoses [area under the curve (AUC) = 0.78]. Implications for assessing ADHD using the CIDI and the effect of different informants on measurement are discussed. Copyright © 2010 John Wiley & Sons, Ltd.

Introduction

The accuracy of structured assessments of attention deficit hyperactivity disorder (ADHD) has been a topic of considerable interest because of concerns about the potential for both over-diagnosis and misdiagnosis of the disorder (American Academy of Pediatrics, 2000; LeFever *et al.*, 1999; Sciotto and Eisenberg, 2007). In the Diagnostic and Statistic Manual of Mental Disorders, fourth edition (DSM-IV; American Psychiatric Association, 2000), ADHD is defined as a syndrome characterized by ongoing and atypical inattention or hyperactivity-impulsivity in at least two settings that begins prior to age seven. The DSM-IV presents three subtypes of ADHD: predominantly inattentive type (six or more symptoms of inattention), predominantly hyperactive-impulsive type (six or more symptoms of hyperactivity or impulsivity), and combined type (six or more symptoms of inattention and six or more symptoms of hyperactivity or impulsivity).

Recent debates about ADHD diagnosis have focused on the accuracy of current measurement tools (Furman, 2005) and the interpretation of discrepancies between multiple informants (Hartung *et al.*, 2005). Using a

nationally-representative dataset of adolescents and their parents, we contribute to these discussions by evaluating the internal consistency reliability and validity of a new multi-informant measure of ADHD, the ADHD module of the adolescent version of the World Health Organization (WHO) Composite International Diagnostic Interview Version 3.0 (CIDI 3.0; Merikangas *et al.*, 2009). Our analysis of the internal consistency of the CIDI symptom assessment addresses questions about the structure of ADHD that have been reflected in modifications of criteria with each revision of the DSM, where ADHD has been presented as consisting of between one and three dimensions (DuPaul *et al.*, 1998; Glutting *et al.*, 2005). Although the most recent revision of the DSM presents an empirically based two-factor model (Glutting *et al.*, 2005; Gomez *et al.*, 2003; Willcutt *et al.*, 1999), there remain questions about whether ADHD may be better conceptualized as a three-factor model that separates impulsivity from hyperactivity (Amador-Campos *et al.*, 2006).

We next turn to evaluating the concurrent validity of the CIDI ADHD module by comparing diagnoses based on the CIDI with independent diagnoses based on blinded clinical reappraisal interviews with the Schedule for Affective Disorders and Schizophrenia for School-Age

Children – Present and Lifetime Version (K-SADS-PL; Kaufman *et al.*, 1997). Several other clinical reappraisal studies of lay-administered ADHD interviews have found weak to moderate concordance of diagnoses with clinical re-interviews. A review of these studies found a wide range in the level of agreement between parent and/or child reports and clinician-assigned diagnoses ($\kappa = 0.09 - 0.60$) (Ezpeleta *et al.*, 1997). One reason for this wide variation is that some structured assessments are based only on reports by a single informant while others are based on reports by both youth and parents or teachers. Adolescent and informant reports of ADHD are often discrepant (Rubio-Stipec *et al.*, 1994). Methodological analysis generally finds that parent reports are more strongly associated with blinded clinical assessments of ADHD than are youth reports (Schwab-Stone *et al.*, 1996), a pattern that is often attributed to underreporting by youth (Rohde *et al.*, 1999). Based on this prior evidence, we consider validity separately for youth and parent reports.

Methods

Sample

The National Comorbidity Survey Replication Adolescent Supplement (NCS-A) is a US nationally-representative face-to-face survey of 10,148 adolescents ages 13–17 (Merikangas *et al.*, 2009). The NCS-A interviews were carried out by the professional interview staff of the Institute for Social Research at the University of Michigan between April 2001 and April 2003. A dual-frame sample was used that included a household sub-sample and a school sub-sample. The household sub-sample consisted of 904 adolescent residents of the households that participated in the National Comorbidity Survey Replication (NCS-R), a nationally representative household survey of adults (Kessler and Merikangas, 2004). The school sub-sample consisted of 9244 adolescents who were students in a probability sample of schools (public and private, day and boarding, selected with probabilities proportional to size) in the same counties as the NCS-R sample.

We began respondent recruitment by sending an informational letter and Study Fact Brochure to the parents of each target respondent. The letter contained an 800 phone number for parents to call if they had questions not covered in the Study Fact Brochure or if they wanted to opt out. A study interviewer then visited the household a few days later to talk with a parent and answer remaining questions before obtaining parental written informed consent. Only after obtaining this

consent did the interviewer approach the adolescent to obtain written informed assent. Each target respondent and parent was offered a \$50 incentive for participation. These recruitment and consent procedures were approved by the Human Subjects Committees of both the University of Michigan and Harvard Medical School.

The conditional (on adult participation in the NCS-R) response rate of adolescent respondents completing a face-to-face NCS-A interview in the household sub-sample was 86.8%. The corresponding response rate of adolescent respondents in the school sub-sample was 82.6%. The overall weighted (by sample size) NCS-A adolescent response rate across the two sub-samples was 82.9%. In addition to the adolescent face-to-face interviews, one parent of each adolescent was asked to complete an informant self-administered questionnaire (SAQ). An effort was made to have the parent complete the SAQ while the interviewer was in the household interviewing the adolescent. If this was not possible, the interviewer left the SAQ and a self-addressed pre-stamped envelope with the parent for later self-administration and mail return. Postcard and telephone reminders were used when the parent did not return the SAQ. An attempt was made to administer a truncated version of the SAQ, which included the ADHD section, to parents who never returned the paper-and-pencil version. Parents of 8470 NCS-A adolescent respondents completed the SAQ either in paper-and-pencil self-administration ($n = 6483$) or in telephone administration ($n = 1987$) by the end of the study. Extensive efforts were made to obtain as much parent report data as possible on ADHD symptoms in adolescents. The data were weighted for within-household probability of selection (only in the household sub-sample) and for residual discrepancies on the basis of socio-demographic and geographic variables between the samples and the population distributions of US residents in the 13–17 age range from the 2000 Census. More details on NCS-A weighting are reported elsewhere (Kessler *et al.*, 2009a; Kessler *et al.*, 2009b).

The NCS-A clinical reappraisal study was carried out in a quota sample of 321 adolescent–parent pairs (described in Kessler *et al.*, 2009c). The sample was confined to adolescents residing in households with telephones because the K-SADS clinical reappraisal interviews were administered by telephone. Telephone administration is now widely accepted in clinical reappraisal studies based on evidence of comparable validity to in-person administration (Kendler *et al.*, 1992; Rohde *et al.*, 1997; Sobin *et al.*, 1993). A great advantage of telephone administration is that a centralized and closely supervised clinical interview

staff can carry out the interviews throughout the entire sample area without the geographic restriction that is typically required for face-to-face clinical assessment. A disadvantage is that the small part of the population without telephones cannot be included in clinical calibration studies when interviews are done by telephone.

Respondents who met DSM-IV/CIDI criteria for one or more relatively uncommon disorder (e.g. agoraphobia, bipolar disorder, panic disorder, substance dependence with abuse) were sampled at a higher rate than respondents in a second sampling stratum who met criteria only for more common disorders. The lowest sampling fraction was for a third stratum made up of respondents who did not meet criteria for any lifetime DSM-IV/CIDI disorder. Each respondent in the clinical reappraisal study was given a \$50 incentive for participation (over and above the \$50 incentive for participation in the main survey).

Measures

The CIDI is a fully-structured diagnostic interview administered by trained lay interviewers to assess a wide range of DSM-IV and International Classification of Diseases (ICD-10) disorders. The adult version of the CIDI was developed for use in community epidemiological surveys (Kessler and Ustun, 2004). The adolescent version introduced some modifications in wording aimed at increasing the relevance of questions to adolescents. After a series of warm-up questions, the CIDI administers screening questions made up of diagnostic stem questions for a wide range of disorders. Positive responses are then probed in subsequent CIDI sections. The ADHD screening questions ask about a history of concentration problems prior to the age of seven that lasted a minimum of six months and that, in retrospect, seemed excessive compared to peers. A second screening question assesses a history of hyperactivity-impulsivity present before the age of seven that lasted a minimum of six months.

Adolescents who endorse the first screening question are subsequently entered into the ADHD section of the CIDI and asked retrospective questions about inattentiveness in childhood that correspond to each of the nine DSM-IV TR Criterion A symptoms of the inattentive type of ADHD. A similar set of nine questions that correspond to the nine Criterion A symptoms of the hyperactive-impulsive type of ADHD are administered to respondents who endorse the CIDI screening question about childhood hyperactivity-impulsivity. Respondents are skipped out of each question sequence when they either endorse

six questions or would not have a total of six even if they endorsed all remaining questions. Adolescents who endorse four or more of the nine questions in a given section are asked follow-up questions about role impairment associated with these symptoms. Subsequent questions in each series ask about age at onset and persistence of symptoms. Parents are administered an informant version of the same nine CIDI questions about Criterion A symptoms of inattentiveness and nine questions about hyperactivity-impulsivity. Parents who endorse any of these questions are asked about role impairment.

In the NCS-A, a dichotomous (yes-no) diagnostic classification of DSM-IV ADHD was generated from the CIDI adolescent and parent reports using an 'or' rule to count each symptom as present if it was endorsed by either the adolescent or the parent. In addition, independent diagnoses were generated separately based only on adolescent and parent reports. A series of exploratory analyses was also carried out to investigate the extent to which concordance of diagnoses based on CIDI reports with diagnoses based on blinded clinical reappraisal interviews could be improved by modifying the CIDI diagnostic classification rules in various ways described later in the paper.

The K-SADS-PL is a semi-structured clinician-administered research diagnostic interview designed to assess a range of child and adolescent DSM-IV disorders. The K-SADS interviewers in the NCS-A were experienced clinical interviewers who received 40 hours of training in the K-SADS from one of the developers of the K-SADS and were required to pass a certification test before beginning production interviewing. Completed clinical interviews were audio-taped and closely monitored for quality control. In addition, clinical interviewers attended bi-weekly quality control monitoring meetings to prevent interviewer drift. The K-SADS interviews were administered an average of two and a half months after the CIDI interviews (with the majority between one and a half to four months after CIDI interviews). This relatively long lag time between CIDI and K-SADS interviews was scheduled purposefully to be longer than the roughly two weeks between interviews that is more typical in clinical reappraisal studies in order to reduce the negative response set that often occurs in re-interviews due to respondent burden. In the case of ADHD, where age of onset is required to be in childhood to meet diagnostic criteria, this long lag time is unlikely to have had a meaningful effect on true lifetime prevalence.

A single clinician interviewed both the adolescent and the parent, in that order. A diagnosis based only on the information provided by the single respondent was

generated after each interview. A final diagnosis was then made, taking into consideration the information provided by both respondents. For the purposes of this study, the final clinician-rated diagnoses that included information obtained from both respondents are considered the gold standard. Ninety of the original 321 K-SADS cases were interviewed a second time with the K-SADS following a review of the quality of overall clinical interviews. A small number of clinical diagnoses were changed based on this process. In the very few cases where interviewer quality concerns existed but it was not possible to carry out a second K-SADS interview, diagnoses were changed by random imputation at rates estimated from the 90 re-interviewed cases (Rubin, 1981).

Analysis methods

Analysis of internal consistency of the ADHD symptom reports was carried out in the sample of 8470 adolescent-parent dyads that completed the CIDI. Tetrachoric factor analysis with Promax rotation was used to establish the dimensionality of the symptom reports. Item response theory (IRT) models (Hambleton *et al.*, 1991) were then used to evaluate the implicit assumption in the DSM-IV that each Criterion A symptom of ADHD has the same association as the others with true ADHD.

As there were a high percentage of respondents who did not endorse any ADHD items, in addition to the conventional two-parameter IRT model, we estimated a mixture model that is specifically designed for this situation and as a result tends to yield better fit (Finkelman *et al.*, submitted for publication). The mixture model conceptualized respondents as falling into two mutually exclusive categories, the first consisting of respondents conceptualized as not being in the ADHD spectrum and the second consisting of respondents in the ADHD spectrum whose responses (some of which, like those of respondents not in the ADHD spectrum, consisted entirely of negative responses) could be described adequately by the normally distributed latent liability assumed in the IRT model. The theoretical distinction between respondents not in the ADHD spectrum and respondents in the spectrum who endorsed no items was that the latter were assumed to be people who would have displayed at least some evidence of sub-threshold ADHD symptoms in a more in-depth question series. Although we could not tell which specific respondents were not in the ADHD spectrum, we were able to estimate the percentage of people in that first category derived under the mixture model using the EM algorithm (Dempster *et al.*, 1977), while new IRT slopes and thresholds were calcu-

lated in the second class. A second factor analysis was also carried out in the sub-sample of respondents in the second mixture model category.

A comparison of lifetime DSM-IV diagnoses of ADHD based on the CIDI and on the K-SADS was made in the full clinical reappraisal sample after weighting this sample to adjust for the over-sampling of CIDI cases and post-stratifying for small residual discrepancies between the weighted clinical reappraisal sample and the full weighted NCS-A sample on a wide range of matching variables. Diagnoses of DSM-IV ADHD based on the adolescents, parents, and combined were compared with diagnoses based on the K-SADS at both the aggregate and individual levels. At the aggregate level, we investigated whether prevalence estimates based on the CIDI are similar to those based on the K-SADS using McNemar χ^2 tests that take into account unequal sampling weights. At the individual level, we estimated CIDI sensitivity (SN; the percent of K-SADS positives detected by the CIDI), specificity (SP; the percent of K-SADS negatives classified as negative by the CIDI), positive predictive (PPV; the percent of CIDI positives classified as positive by the K-SADS), and negative predictive value (NPV; the percent of CIDI negatives classified as negative by the K-SADS). We also calculated two measures of overall concordance between CIDI and K-SADS diagnoses: Cohen's κ (Cohen, 1960) and the area under the receiver operating characteristic curve (AUC; Hanley and McNeil, 1982). Cohen's κ is reported because it is the most widely used measure of diagnostic concordance, but κ has the undesirable characteristic of being influenced by marginal distributions. AUC is also reported because it is an alternative measure of diagnostic concordance that is not influenced by marginal distributions (Kraemer *et al.*, 2003). All analyses of concordance were conducted using SAS 9.0 and SUDAAN 9.0.1 software programs (SAS Institute, 2002; Research Triangle Institute, 2005).

Results

Internal consistency reliability

Tetrachoric factor analyses were calculated separately for inattention (AD) and hyperactivity-impulsivity (HD) criteria using the symptom reports of all adolescents and parents who screened into the AD and HD modules. For AD, results indicated a three-factor solution (unrotated eigenvalues: 8.01, 4.18, 1.23, 0.86), with the three rotated (promax) factors corresponding to parent reports of AD and two factors for youth reports (Table 1). The existence of two adolescent factors rather than one might indicate that distractibility is somewhat distinct from executive

Table 1 Rotated (promax)¹ tetrachoric factor analysis (standardized regression coefficients) of parent and youth symptom reports separately for symptoms of inattention and hyperactivity-impulsivity ($n = 8470$)

Criteria	Factor 1	Factor 2	Factor 3
<i>Inattention</i>			
I. Youth reports			
Poor attention to detail	0.06	0.70	0.10
Difficulty sustaining attention	-0.13	0.75	0.02
Seem not to listen	0.05	0.74	0.11
Not follow through on instructions	0.17	0.06	0.78
Difficulty organizing tasks	-0.08	-0.01	0.91
Avoids tasks with sustained effort	-0.03	0.22	0.71
Loses things necessary for tasks	0.12	0.86	-0.19
Easily distracted	-0.03	0.75	0.09
Forgetful in daily activities	-0.07	0.76	0.10
II. Parent reports			
Poor attention to detail	0.93	0.08	-0.06
Difficulty sustaining attention	0.90	-0.06	0.07
Seem not to listen	0.91	-0.04	0.00
Not follow through on instructions	0.92	0.04	0.04
Difficulty organizing tasks	0.86	-0.08	0.09
Avoids tasks with sustained effort	0.93	-0.13	0.10
Loses things necessary for tasks	0.83	0.11	0.17
Easily distracted	0.91	0.00	0.04
Forgetful in daily activities	0.95	0.07	-0.07
<i>Hyperactivity-impulsivity</i>			
I. Youth reports			
Fidgets or squirms	0.08	0.70	0.14
Leaves seat	0.01	0.69	0.21
Runs about or climbs	0.09	0.77	0.02
Difficulty playing quietly	-0.12	0.85	-0.04
On the go or driven by a motor	0.02	0.74	-0.06
Talks excessively	0.00	0.17	0.73
Blurts out answers	0.04	-0.08	0.91
Difficulty waiting turn	-0.04	0.25	0.64
Interrupts or intrudes on others	-0.01	-0.05	0.89
II. Parent reports			
Fidgets or squirms	0.90	0.04	-0.03
Leaves seat	0.90	0.14	-0.13
Runs about or climbs	0.95	0.05	-0.06
Difficulty playing quietly	0.90	0.06	-0.14
On the go or driven by a motor	0.87	0.11	-0.06
Talks excessively	0.83	-0.11	0.14
Blurts out answers	0.88	-0.16	0.19
Difficulty waiting turn	0.89	0.04	0.05
Interrupts or intrudes on others	0.92	-0.11	0.07

¹Correlations among factors for inattention: F1–F2: 0.19; F1–F3: 0.19; F2–F3: 0.47. Correlations among factors for hyperactivity-impulsivity: F1–F2: 0.37; F1–F3: 0.20; F2–F3: 0.38.

function problems. However, it is difficult to interpret the two factors in any clear conceptual way. As a result, we collapsed all items from these two factors into a single youth factor for the subsequent IRT analysis. The factor analysis of HD symptoms also indicated a three-factor solution (unrotated eigenvalues: 8.44, 3.33, 1.55, 0.72). The rotated (promax) factor solution corresponded to a parent HD factor and two factors for adolescents that separated hyperactivity from impulsivity, with the exception of one DSM hyperactivity item (often talks excessively) that loaded on the impulsivity factor.

One-parameter (1PL) and two-parameter (2PL) IRT models were estimated for each of the two informants (adolescent and parent) on each of the two dimensions (AD and HD) (Table 2). Pearson chi-square statistics were calculated for the 1PL and 2PL models, comparing expected and observed outcomes. For both informants on the AD criteria and parents on the HD criteria, the 2PL model was a significantly better fit than the 1PL model. For adolescents on the HD criteria, the 1PL model was a significantly better fit than the 2PL model. Focusing first on the adolescent data, slopes for both the AD and HD

Table 2 IRT model item parameters for adolescent and parent CIDI inattention and hyperactivity-impulsivity items¹

	Youth reports		Parent reports	
	Slope threshold		Slope threshold	
I. Inattention				
Poor attention to detail	1.02	-0.13	3.33	0.88
Difficulty sustaining attention	0.80	0.33	2.37	1.14
Seem not to listen	1.10	-0.36	2.46	0.98
Not follow through on instructions	0.88	0.45	2.45	0.81
Difficulty organizing tasks	0.85	0.84	2.18	1.24
Avoids tasks with sustained effort	0.98	0.21	2.64	0.96
Loses things necessary for tasks	0.88	0.25	1.83	1.19
Easily distracted	1.08	-0.09	2.55	0.81
Forgetful in daily activities	1.14	0.06	3.11	1.03
(<i>n</i>)	(1136) ²		(8290) ²	
II. Hyperactivity-impulsivity				
Fidgets or squirms	0.91	-0.40	2.82	1.14
Leaves seat	0.91	0.33	2.83	1.22
Runs about or climbs	0.91	-0.22	3.39	0.98
Difficulty playing quietly	0.91	-0.10	2.53	1.34
On the go or driven by a motor	0.91	-0.10	2.13	1.12
Talks excessively	0.91	-0.05	1.34	1.16
Blurts out answers	0.91	0.36	1.73	1.41
Difficulty waiting turn	0.91	0.68	2.43	1.32
Interrupts or intrudes on others	0.91	0.24	1.72	1.10
(<i>n</i>)	(1512) ²		(8290) ²	

¹The two-parameter model fits better than the one-parameter model (which assumes constant slopes across items) for all scales other than the youth-report hyperactivity-impulsivity scale. Results are consequently reported for the one-parameter model for the latter scale and for the two-parameter model for the other scales.

²Sample sizes differ because the symptom questions were administered to all parents who knew the child before the age of seven ($n = 8290$) but not to all adolescents. Adolescents were administered an initial screening question separately for AD and HD that determined whether or not they would subsequently be administered the full set of symptom question. A total of 1136 adolescents endorsed the AD screening question, while 1512 endorsed the HD screening question.

factors are moderate (0.80–1.14 for AD and 0.91 for HD), indicating that none of the items is a strong indicator of the underlying dimension. (A slope of at least 1.0 is usually defined as the lower bound for an item that has good precision at its threshold on the underlying scale.) Thresholds were for the most part within one-third (\pm) of a standard deviation of the mean, indicating that most of the information in the scales is in a part of the severity distribution that is well below the clinical threshold. The conjunction of low slopes and sub-clinical thresholds indicates that the scale is not highly sensitive or specific in discriminating clinical cases from non-cases.

Slopes were considerably higher in the parent data for both AD and HD factors (1.83–3.33 for AD and 1.34–3.39 for HD), indicating that the items have excellent precision at their thresholds. It is noteworthy that the existence of significant slope differences across items for both AD and HD means that optimal scaling would weight items differentially to arrive at an estimate of underlying scale scores. This is different from the stipulation in the DSM that each Criterion A symptom of AD and of HD contributes equally to a diagnosis. Like the slopes, the thresholds of the parent items were a good deal higher than in the youth data (0.81–1.24 for AD and 0.98–1.41 for HD), indicating that the parent scales have much better precision than the youth scales.

The fact that a high proportion of respondents endorsed none of the ADHD symptom questions raises the possibility that the IRT assumption of a normally distributed latent liability might be violated. Based on this concern, we fitted separate two-class IRT mixture models for the adolescent and parent HD and AD data, where one class was stipulated to consist exclusively of respondents outside of the AD or HD spectrum; that is, to have no risk of reporting any ADHD symptoms. The other class was assumed to consist of respondents in the ADHD spectrum. Respondents in the latter class were assumed to have a normally distributed latent liability (including some proportion that would be expected to endorse none of the CIDI symptom questions). Relatively small proportions of adolescent respondents who completed the symptom questions were estimated in this model to be outside the spectrum for AD (8.3%) or HD (3.7%), while much larger proportions were estimated to be outside the spectrum for the parent AD (50.4%) and HD (54.5%) scales. This substantial difference between adolescents and parents is presumably due to the fact, noted earlier, that screening questions were used in the assessment of adolescents but not parents.

The two-class mixture model was a better fit than the standard IRT model for both adolescent and parent AD

and HD dimensions (based on Pearson χ^2 tests comparing expected and observed values for 2PL and 2PL-mixture models). Eliminating respondents not in the spectrum from the database, we replicated the factor analyses for AD and HD and again identified a three-factor solution for AD (unrotated eigenvalues: 8.44, 3.33, 1.55, 0.72) and a four-factor solution for HD (unrotated eigenvalues 6.67, 3.44, 1.78, 1.16, 0.77). The rotated (promax) factor solution for AD had a factor pattern very similar to the one in the original factor analysis; one factor included all parent reports and two factors included adolescent reports, where we could find no meaningful interpretation of the distinction between items in the two adolescent factors. (Detailed results are not reported, but are available on request.) The rotated (promax) solution for HD, in comparison, was different from the original solution in that it differentiated symptoms of hyperactivity from symptoms of impulsivity both in parent reports and in adolescent reports (Table 3). Adolescent and parent primary factor loadings were very similar, with the exception of the impulsivity item 'difficulty waiting turn,' which loaded on the impulsivity factor in the adolescent data, but the hyperactivity factor in the parent data.

Concordance of CIDI symptom reports with clinician ratings

Adolescent responses to the CIDI questions about Criterion A symptoms of AD generally underestimated K-SADS prevalence (Table 4). This was due to sensitivity being uniformly low (16.4–35.1%; i.e. a low proportion of adolescents classified by the K-SADS as having a history of the symptom reporting the symptom in the CIDI). Specificity, in comparison, was generally quite good (94.5–97.9%; i.e. a very high proportion of adolescents classified by the K-SADS as not having a history of the symptom denying the symptom in the CIDI). Concordance of adolescent symptom reports with K-SADS estimates was higher, in comparison, for Criterion A symptoms of HD, but this was as much because specificity decreased (89.3–96.5%) as because sensitivity increased (21.4–42.9%) (Table 4). Parent reports generally overestimated the prevalence of K-SADS Criterion A symptoms of both AD and HD (Table 4). This occurred because of both higher sensitivity and lower specificity than in adolescent reports. Similar patterns were found when we generated criterion-level estimates from the symptom reports, including Criterion A1 (six or more symptoms of inattention), Criterion A2 (six or more symptoms of hyperactivity-impulsivity), Criterion B (some symptoms present before the age of seven), and Criterion C (clinically

Table 3 Rotated (promax)¹ tetrachoric factor analysis (standardized regression coefficients) for hyperactivity-impulsivity symptoms based on CIDI symptom reports in adolescent–parent pairs where both respondents were classified as being in the HD spectrum ($n = 1512$)

	Factor 1	Factor 2	Factor 3	Factor 4
I. Youth reports				
Fidgets or squirms	0.11	0.74	0.11	−0.05
Leaves seat	0.05	0.65	0.27	−0.05
Runs about or climbs	0.16	0.73	0.05	−0.01
Difficulty playing quietly	−0.21	0.87	−0.09	0.09
On the go or driven by a motor	0.04	0.77	−0.09	−0.00
Talks excessively	−0.09	0.18	0.72	0.07
Blurts out answers	−0.10	−0.02	0.84	0.16
Difficulty waiting turn	0.08	0.15	0.74	−0.16
Interrupts or intrudes on others	0.03	−0.15	0.94	−0.01
II. Parent reports				
Fidgets or squirms	0.95	−0.09	0.09	−0.13
Leaves seat	0.81	0.13	−0.13	0.11
Runs about or climbs	0.95	−0.02	0.00	0.01
Difficulty playing quietly	0.86	−0.00	−0.11	0.04
On the go or driven by a motor	0.82	0.06	0.00	−0.03
Talks excessively	−0.10	0.08	−0.03	0.93
Blurts out answers	0.16	−0.05	0.08	0.83
Difficulty waiting turn	0.67	0.03	0.08	0.29
Interrupts or intrudes on others	0.34	0.07	−0.01	0.63

¹Correlations among factors: F1–F2: 0.31; F1–F3: 0.12; F1–F4: 0.46, F2–F3: 0.39, F2–F4: 0.16, F3–F4: 0.19.

significant impairment) (Table 5). CIDI ratings based on adolescent reports overestimated the prevalence of Criterion A2, but underestimated the other criteria. CIDI ratings based on parent reports, in comparison, overestimated the clinical K-SADS diagnosis for all of these criteria except impairment.

The CIDI diagnostic algorithms

Consistent with the results reported in the last section, diagnoses based on adolescent CIDI reports substantially underestimated the prevalence of DSM-IV/K-SADS ADHD (3.0% versus 7.8%, $\chi^2 = 9.4$, $p < 0.05$). Individual-level concordance was also poor ($\kappa = 0.19$, $AUC = 0.57$) (Table 6). Diagnoses based on parent reports, in comparison, were much more consistent with DSM-IV/K-SADS prevalence (8.0% versus 7.8%, $\chi^2 = 0.2$, $p = 0.68$), and individual-level concordance was much better than for adolescent reports ($\kappa = 0.41$, $AUC = 0.71$). Diagnostic estimates based on composite (i.e. adolescent–parent) CIDI reports were inflated (11.9% versus 7.8%, $\chi^2 = 5.9$, $p < 0.05$), but had better concordance with K-SADS diagnoses

than those based on parent report alone ($\kappa = 0.43$, $AUC = 0.77$).

We explored several options for modifying CIDI ratings to improve concordance between diagnoses based on the CIDI and the K-SADS. First, we considered the possibility of modifying the symptom thresholds in the CIDI Criterion A symptom reports, but this did not improve concordance. Second, we considered the possibility of using predictive logistic regression analysis to improve concordance of CIDI symptom reports with K-SADS diagnoses. These analyses showed clearly that CIDI parent reports were significant predictors of K-SADS diagnoses while CIDI adolescent reports were not, after controlling for parent reports. Third, based on this result, we focused subsequent analyses on parent reports and considered ways in which we might improve concordance with K-SADS diagnoses by cross-classifying Criterion A1 and A2 CIDI reports and selecting a higher diagnostic threshold than the DSM-IV six of nine A1 or A2 symptoms to address the fact that CIDI parent reports overdiagnose ADHD. In addition, we explored the effects of eliminating Criterion B from the diagnostic algorithm, as

Table 4 Concordance (sensitivity and specificity) of CID/DSM-IV ADHD Criterion A inattention and hyperactivity-impulsivity symptoms with blinded K-SADS ratings in the NCS-A clinical reappraisal sample ($n = 321$)

	Prevalence													
	K-SADS				CID I				SN and SP					
	Composite		Youth		Parent		Composite		Youth		Parent		Composite	
	Percentage	(SE)	Percentage	(SE)	Percentage	(SE)	Percentage	(SE)	Percentage	(SE)	Percentage	(SE)	SN	SP
I. Inattention														
Poor attention to detail	14.9	(1.8)	5.2*	(1.3)	17.7	(1.7)	19.9*	(1.6)	18.9	97.2	40.3	86.2	48.6	85.1
Difficulty sustaining attention	18.7	(2.3)	5.6*	(1.4)	12.8*	(2.4)	17.0	(2.5)	20.7	97.9	28.2	90.7	44.4	89.3
Seem not to listen	10.8	(1.9)	8.7	(1.7)	15.9*	(1.9)	20.9*	(2.0)	35.1	94.5	45.2	87.6	67.6	84.7
Not follow through on instructions	12.5	(2.0)	6.6*	(1.4)	23.6*	(2.3)	27.0*	(2.3)	16.4	94.8	43.0	79.2	54.3	76.9
Difficulty organizing tasks	9.0	(1.5)	5.8	(1.2)	12.2	(2.1)	16.9*	(2.1)	20.9	95.7	35.9	90.2	48.6	86.2
Avoids tasks with sustained effort	11.3	(1.4)	6.8*	(1.5)	17.5*	(2.0)	21.3*	(2.1)	24.3	95.4	53.7	87.1	64.7	84.2
Loses things necessary for tasks	10.1	(1.5)	3.9*	(0.9)	14.7	(2.2)	16.4*	(2.3)	17.5	97.7	42.7	88.5	46.6	87.0
Easily distracted	17.8	(2.0)	7.6*	(1.6)	21.3	(2.7)	25.1*	(2.9)	23.4	95.8	51.7	85.2	61.6	82.9
Forgetful in daily activities	9.2	(1.6)	6.0	(1.4)	14.8*	(2.0)	18.2*	(2.2)	19.7	95.4	41.7	87.9	50.6	85.1
II. Hyperactivity-impulsivity														
Fidgets or squirms	11.6	(2.0)	9.2	(1.7)	10.7	(2.3)	16.6*	(2.7)	28.4	93.3	32.1	92.1	44.4	87.1
Leaves seat	11.9	(2.1)	6.3*	(1.4)	11.4	(2.0)	15.3	(2.3)	26.8	96.5	43.3	92.9	53.3	89.8
Runs about or climbs	6.5	(1.4)	11.3*	(1.8)	15.1*	(2.7)	22.2*	(3.0)	23.9	89.6	42.3	86.8	48.5	79.6
Difficulty playing quietly	4.8	(1.3)	6.8	(1.6)	10.6*	(2.0)	15.3*	(2.5)	31.9	94.5	63.6	92.1	72.9	87.6
On the go or driven by a motor	6.3	(1.6)	11.9*	(1.9)	13.3*	(2.7)	22.5*	(3.4)	28.9	89.3	46.0	88.9	64.2	80.3
Talks excessively	10.2	(1.8)	9.9	(1.8)	13.4	(1.9)	20.1*	(2.6)	37.9	93.3	60.6	91.9	77.8	86.5
Blurts out answers	6.5	(1.4)	8.8	(1.4)	10.3*	(1.7)	16.8*	(2.5)	42.9	93.6	43.7	92.0	72.8	87.1
Difficulty waiting turn	5.9	(1.5)	8.4	(1.8)	8.8	(1.7)	15.3*	(2.0)	30.5	93.0	46.8	93.6	61.5	87.5
Interrupts or intrudes on others	6.9	(1.5)	9.4	(1.7)	14.0*	(2.1)	19.4*	(2.6)	21.4	91.5	63.8	89.7	73.2	84.6

*The prevalence estimate based on the CID I differs significantly from the estimate based on the K-SADS at 0.05 level, two-sided test. SE, standard error; SN, sensitivity; SP, specificity.

Table 5 Concordance (sensitivity and specificity) of criterion-level assessments of DSM-IV ADHD based on the CIDI with those based on blinded K-SADS interviews in the NCS-A clinical reappraisal sample (*n* = 321)

DSM-IV Criteria ¹	Prevalence																
	K-SADS			CIDI			SN and SP										
	Composite	Youth	Parent	Composite	Youth	Parent	Composite	Youth	Parent	Composite	Youth	Parent					
Percentage	(SE)	Percentage	(SE)	Percentage	(SE)	Percentage	(SE)	Percentage	(SE)	Percentage	(SE)	Percentage	(SE)	Percentage	(SE)	Percentage	(SE)
A1 Inattention	10.9	(1.7)	4.3*	(1.2)	11.8	(1.5)	15.5	(1.6)	10.2	96.4	36.3	91.2	47.0	88.3			
A2 Hyperactivity-impulsivity	4.6	(1.2)	6.8	(1.5)	8.5*	(1.9)	14.2*	(2.7)	35.3	94.6	61.7	94.0	76.3	88.8			
A1 or A2	12.4	(1.8)	9.0	(1.7)	16.2	(2.2)	22.5*	(2.7)	21.0	92.7	47.4	88.2	59.2	82.7			
B Early onset	18.7	(2.2)	9.3*	(1.5)	25.5*	(2.9)	31.3*	(3.1)	23.6	93.9	56.7	81.7	66.8	76.8			
C Impairment	22.5	(2.3)	5.2*	(1.2)	18.9	(2.7)	22.0	(2.6)	13.7	97.3	50.0	90.1	56.2	87.9			

¹ Criterion A1: six or more inattention symptoms; Criterion A2: six or more hyperactivity-impulsivity symptoms; Criteria A1 or A2: six or more A1 or A2 symptoms; Criterion B: some symptoms were present before age seven years; Criterion C: some impairment present in two or more settings.

*The prevalence estimate based on the CIDI differs significantly from the estimate based on the K-SADS at 0.05 level, two-sided test. SE, standard error; SN, sensitivity; SP, specificity.

Table 6 Diagnostic concordance of the original and modified DSM-IV ADHD diagnoses based on the CIDI with diagnoses based on the composite K-SADS in the NCS-A clinical reappraisal sample ($n = 321$)

	Prevalence				Concordance										
	K-SADS		CIDI		SN		SP		PPV		NPV		Kappa		AUC ¹
	Percentage	(SE)	Percentage	(SE)	Estimated	(SE)	Estimated	(SE)	Estimated	(SE)	Estimated	(SE)	Estimated	(SE)	Estimated
Adolescent CIDI versus clinician K-SADS ¹	7.8	(1.6)	3.0*	(0.9)	15.9	(6.3)	98.1	(0.8)	41.6	(14.6)	93.2	(1.5)	0.19		0.57
Parent CIDI versus clinician K-SADS ¹	7.8	(1.6)	8.0	(1.6)	46.4	(12.9)	95.2	(1.4)	45.3	(11.2)	95.4	(1.5)	0.41		0.71
Original Composite CIDI versus clinician K-SADS ¹	7.8	(1.6)	11.9*	(1.7)	61.2	(11.7)	92.3	(1.6)	40.4	(8.6)	96.5	(1.3)	0.43		0.77
Modified CIDI versus clinician K-SADS ²	7.8	(1.6)	7.9	(1.6)	58.5	(11.3)	96.5	(1.0)	58.5	(9.7)	96.5	(1.4)	0.55		0.78

SN, sensitivity; SP, specificity; PPV, positive predictive value; NPV, negative predictive value; Kappa, Cohen's kappa; AUC, area under the receiver operating characteristic curve; SE, standard error.

¹Original CIDI diagnoses are reported for adolescent-only data, parent-only data, and composite (adolescent and parent) diagnoses, using ADHD Criteria A, B, and C. Clinician K-SADS is a composite of parent and adolescent K-SADS interviews, using imputed data based on clinical re-interviews (described in text).

²The modified CIDI diagnosis uses parent data only and requires 10 or more A1 and A2 symptoms in addition to Criterion C requiring 'a lot' of impairment in at least one area of functioning.

*The prevalence estimate based on the CIDI differs significantly from the estimate based on the K-SADS at 0.05 level, two-sided test.

parents significantly overestimated this criterion, and the effects of modifying the Criterion C measure of impairment because of its low sensitivity.

The scoring rule that best predicted K-SADS diagnoses required 10 or more endorsed symptoms out of the 18 in the A1 and A2 series in conjunction with CIDI Criterion C modified to require 'a lot' of interference in at least one area of role functioning. We could not improve concordance by setting separate thresholds for CIDI A1 and A2 symptom counts, including Criterion B, or introducing additional information from the CIDI adolescent reports. The new algorithm yielded CIDI prevalence estimates of DSM-IV ADHD that did not differ significantly from diagnostic estimates based on the K-SADS ($\chi^2 < 0.01$, $p = 1.00$). This is a considerable improvement over the original CIDI diagnoses.

Discussion

As described elsewhere (Kessler *et al.*, 2009c) there were several limitations in the design of this clinical reappraisal study that may impact results. These limitations include the telephone administration of the K-SADS (in contrast to the face-to-face CIDI administration); however, as mentioned earlier, there is evidence that telephone interviews are a valid method for clinical assessment (Rohde *et al.*, 1997). Second, we did not have any validating information from other sources (e.g. teachers, school records, behavioral observations) against which to compare the youth and parent reports in our sample. Without external validating information, we cannot relate often conflicting reports from adolescents and parents to behaviors observed by others outside the parent-child dyad. In the case of ADHD, the absence of data from school personnel is a particularly noteworthy limitation, as information from teachers is commonly obtained in research diagnoses of ADHD as a way to assess the presence of symptoms in multiple contexts and to corroborate reports from parents and adolescents (Collett *et al.*, 2003; Pliszka, 2007).

Consistent with the DSM-IV, our factor analyses found separate, but correlated, factors for AD and HD symptoms of ADHD, although a more refined analysis that excluded respondents classified as not being in the ADHD spectrum found distinct factors for hyperactivity and impulsivity. Concordance of CIDI reports with blinded K-SADS diagnoses showed that parents were much more accurate reporters than adolescents. This is consistent with previous studies of ADHD measurement validity (Pelham *et al.*, 2005). However, parents overestimated the vast majority of ADHD symptoms compared

to blind clinician ratings. Recalibration of CIDI parent reports corrected for high prevalence rates based on parent report alone. The modified CIDI diagnostic classification yielded prevalences that did not significantly differ from K-SADS classifications and had good individual-level concordance with K-SADS diagnoses.

These results underscore the complexity of ADHD diagnostic assignment and illustrate the critical nature of multi-informant assessments for this disorder (Hoff *et al.*, 2002). Future attempts to improve the CIDI should explore the possibility of increasing the severity of parent symptom questions to reduce the current excessive rate of symptom reports. A challenge in doing this will be that Criterion A symptom-level sensitivity was seldom greater than 50% in the NCS-A, which means that false positives might be increased beyond an acceptable level by increasing the severity of these questions. Clearly, though, the clinical interviewers elicited additional information from parents that allowed the other cases to be detected, so the goal of revising the CIDI parent assessment should be to develop questions that capture as much of this information as possible in a fully-structured format.

It is less clear that improvements can be made in adolescent symptom reports, as positive responses to the CIDI adolescent Criterion A symptom questions were not strongly related to K-SADS ratings. Indeed, the great majority of adolescents judged by clinical interviewers to have a history of a given Criterion A symptom failed to endorse that symptom in the CIDI. When combined with the fact that a small minority of respondents classified by the K-SADS as not having a history of a particular Criterion A symptom endorsed the symptom in the CIDI, we are left with very poor item-level concordance. Pelham *et al.* (2005) reviewed previous methodological studies that consistently documented a similar pattern of low concordance between youth self-reports and clinician assessments of ADHD due to under-reporting of symptoms and severity of impairments. This research showed that youth with ADHD often have limited self-awareness of their symptoms (Zucker *et al.*, 2002), which means that there are fewer opportunities for eliciting more complete data with in-depth probing for adolescent than parent reports. As a result, future improvements in CIDI adolescent assessment of ADHD might profit from using one of the brief computerized tests that has been found to be useful in assessing ADHD in neuropsychological settings and other epidemiological studies (Conners *et al.*, 2003; Epstein *et al.*, 2003), although this would be more useful in assessing active than remitted cases.

Acknowledgments

The National Comorbidity Survey Replication Adolescent Supplement (NCS-A) is supported by the National Institute of Mental Health (NIMH; U01-MH60220) with supplemental support from the National Institute on Drug Abuse (NIDA), the Substance Abuse and Mental Health Services Administration (SAMHSA), the Robert Wood Johnson Foundation (RWJF; Grant 044780), and the John W. Alden Trust. The views and opinions expressed in this report are those of the authors and should not be construed to represent the views of any of the sponsoring organizations, agencies, or US Government. A complete list of NCS-A publications can be found at <http://www.hcp.med.harvard.edu/ncs>. Send correspondence to ncs@hcp.med.harvard.edu. The NCS-A is carried out in conjunction with the World Health Organization World Mental Health (WMH) Survey Initiative. We thank the staff of the WMH Data Collection and Data Analysis Coordination Centres for assistance with instrumentation, fieldwork, and consultation on data analysis. The WMH Data Coordination Centres have received support from NIMH (R01-MH070884, R13-MH066849, R01-MH069864, R01-MH077883), NIDA (R01-DA016558), the Fogarty International Center of the National Institutes of Health (FIRCA R03-TW006481), the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, and the Pan American Health Organization. The WMH Data Coordination Centres have also received unrestricted educational grants from Astra Zeneca, BristolMyersSquibb, Eli Lilly and Company, GlaxoSmithKline, Ortho-McNeil, Pfizer, Sanofi-Aventis, and Wyeth. A complete list of WMH publications can be found at <http://www.hcp.med.harvard.edu/wmh/>.

Declaration of interest statement

Dr Kessler has been a consultant for GlaxoSmithKline Inc., Kaiser Permanente, Pfizer Inc., Sanofi-Aventis, Shire Pharmaceuticals, and Wyeth-Ayerst; has served on advisory boards for Eli Lilly & Company and Wyeth-Ayerst; and has had research support for his epidemiological studies from Bristol-Myers Squibb, Eli Lilly & Company, GlaxoSmithKline, Johnson & Johnson Pharmaceuticals, Ortho-McNeil Pharmaceuticals Inc., Pfizer Inc., and Sanofi-Aventis. The remaining authors have no competing interests.

References

- Amador-Campos J.A., Forns-Santacana M., Guàrdia-Olmos J., Peró-Cebollero M. (2006) DSM-IV Attention Deficit Hyperactivity Disorder symptoms: Agreement between informants in prevalence and factor structure at different ages. *Journal of Psychopathology and Behavior Assessment*, **28**, 23–32, DOI: 10.1007/s10862-006-4538-x
- American Academy of Pediatrics. (2000) Clinical practice guideline: Diagnosis and evaluation of the child with attention-deficit/hyperactivity disorder. *Pediatrics* **105**, 1158–1170.
- American Psychiatric Association. (2000) *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, Text Revision, American Psychiatric Association.
- Cohen J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.
- Collett B.R., Ohan J.L., Myers K.M. (2003) Ten-year review of rating scales. V: Scales assessing attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**, 1015–1037, DOI: 10.1097/01.chi.0000081821.25107.e7
- Conners C.K., Epstein J.N., Angold A., Klaric J. (2003) Continuous performance test performance in a normative epidemiological sample. *Journal of Abnormal Child Psychology*, **31**, 555–562, DOI: 10.1023/A:1025457300409
- Dempster A.P., Laird N.M., Rubin D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, **39**, 1–38.
- DuPaul G.J., Anastopoulos A.D., Power T.J., Reid R., Ikeda M.J., McGoey K.E. (1998) Parent ratings of attention-deficit/hyperactivity disorder symptoms: Factor structure and normative data. *Journal of Psychopathology and Behavior Assessment*, **20**, 83–102.
- Epstein J.N., Erkanli A., Conners C.K., Klaric J., Costello J.E., Angold A. (2003) Relations between Continuous Performance Test performance measures and ADHD behaviors. *Journal of Abnormal Child Psychology*, **31**, 543–554, DOI: 10.1023/A:1025405216339
- Ezpeleta L., de la Osa N., Domenech J.M., Navarro J.B., Losilla J.M. (1997) Diagnostic agreement between clinicians and the Diagnostic Interview for Children and Adolescents – DICA-R – in an outpatient sample. *Journal of Child Psychology and Psychiatry*, **38**, 431–440, DOI: 10.1111/j.1469-7610.1997.tb01528.x
- Finkelman M., Green J.G., Gruber M., Zaslavsky A.M. (Submitted for publication) A zero- and k-inflated mixture model for health questionnaire data.
- Furman L. (2005) What is attention-deficit hyperactivity disorder (ADHD)? *Journal of Child Neurology*, **20**, 994–1002.
- Glutting J.J., Youngstrom E.A., Watkins M.W. (2005) ADHD and college students: Exploratory and confirmatory factor structures with student and parent data. *Psychological Assessment*, **17**, 44–55, DOI: 10.1037/1040-3590.17.1.44
- Gomez R., Burns G.L., Walsh J.A., de Moura M.A. (2003) A multitrait-multisource confirmatory factor analytic approach to the construct validity of ADHD rating scales. *Psychological Assessment*, **15**, 3–16, DOI: 10.1037/1040-3590.15.1.3

- Hambleton R.K., Swaminathan H., Rogers H.J. (1991) *Fundamentals of Item Response Theory*, Sage Publications.
- Hanley J.A., McNeil B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.
- Hartung C.M., McCarthy D.M., Milich R., Martin C.A. (2005) Parent–adolescent agreement on disruptive behavior symptoms: A multitrait-multimethod model. *Journal of Psychopathology and Behavior Assessment*, **27**, 159–168.
- Hoff K.E., Doepke K., Landau S. (2002) Best practices in the assessment of children with Attention Deficit/ Hyperactivity Disorder: Linking assessment to intervention. In: *Best Practices in School Psychology IV* (eds Thomas A., Grimes J.), Volume 2, pp. 1129–1150, National Association of School Psychologists.
- Kaufman J., Birmaher B., Brent D., Rao U., Flynn C., Moreci P., Williamson D., Ryan N. (1997) Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**, 980–988.
- Kendler K.S., Neale M.C., Kessler R.C., Heath A.C., Eaves L.J. (1992) A population-based twin study of major depression in women. The impact of varying definitions of illness. *Archives of General Psychiatry*, **49**, 257–266.
- Kessler R.C., Avenevoli S., Costello E.J., Green J.G., Gruber M.J., Heeringa S., Merikangas K.R., Pennell B.-E., Sampson N.A., Zaslavsky A.M. (2009a) National Comorbidity Survey Replication Adolescent Supplement: II. Overview and design. *Journal of the American Academy of Child and Adolescent Psychiatry*.
- Kessler R.C., Avenevoli S., Costello E.J., Green J.G., Gruber M.J., Heeringa S., Merikangas K.R., Pennell B.-E., Sampson N.A., Zaslavsky A.M. (2009b) Design and field procedures in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *International Journal of Methods in Psychiatric Research*, **18**, 69–83.
- Kessler R.C., Avenevoli S., Green J., Gruber M.J., Guyer M., He Y., Jin R., Kaufman J., Sampson N.A., Zaslavsky A.M., Merikangas K.R. (2009c) National Comorbidity Survey Replication Adolescent Supplement: III. Concordance of DSM-IV/CIDI Diagnoses with clinical reassessments. *Journal of the American Academy of Child and Adolescent Psychiatry*, **48**, 386–399.
- Kessler R.C., Merikangas K.R. (2004) The National Comorbidity Survey Replication (NCS-R): Background and aims. *International Journal of Methods in Psychiatric Research*, **13**, 60–68.
- Kessler R.C., Ustun T.B. (2004) The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, **13**, 93–121.
- Kraemer H.C., Morgan G.A., Leech N.L., Gliner J.A., Vaske J.J., Harmon R.J. (2003) Measures of clinical significance. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**, 1524–1529, DOI: 10.1097/01.chi.0000091507.46853.d1
- LeFever G.B., Dawson K.V., Morrow A.L. (1999) The extent of drug therapy for attention deficit-hyperactivity disorder among children in public schools. *American Journal of Public Health*, **89**, 1359–1364.
- Merikangas K.R., Avenevoli S., Costello E.J., Koretz D., Kessler R.C. (2009) National Comorbidity Survey Replication Adolescent Supplement: I. Background and measures. *Journal of the American Academy of Child and Adolescent Psychiatry*, **48**, 367–369.
- Pelham W.E. Jr, Fabiano G.A., Massetti G.M. (2005) Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, **34**, 449–476.
- Pliszka S. (2007) Practice parameter for the assessment and treatment of children and adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **46**, 894–921.
- Research Triangle Institute. (2005) *SUDAAN: Professional Software for Survey Data Analysis, version 9.0.1*, Research Triangle Institute.
- Rohde L.A., Biederman J., Knijnik M.P., Ketzner C., Chachamovich E., Viera G.M., Pinzon V. (1999) Exploring different information sources for DSM-IV ADHD diagnoses in Brazilian adolescents. *Journal of Attention Disorders*, **3**, 91–96, DOI: 10.1177/108705479900300203.
- Rohde P., Lewinsohn P.M., Seeley J.R. (1997) Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. *American Journal of Psychiatry*, **154**, 1593–1598.
- Rubin D.B. (1981) The Bayesian bootstrap. *The Annals of Statistics*, **9**, 13–134
- Rubio-Stipec M., Canino G.J., Shrout P., Dulcan M., Freeman D., Bravo M. (1994) Psychometric properties of parents and children as informants in child psychiatry epidemiology with the Spanish Diagnostic Interview Schedule for Children (DISC.2). *Journal of Abnormal Child Psychology*, **22**, 703–720.
- SAS Institute. (2002) *SAS 9.1.3 Help and Documentation*, SAS Publishing.
- Schwab-Stone M.E., Shaffer D., Dulcan M.K., Jensen P.S., Fisher P., Bird H.R., Goodman S.H., Lahey B.B., Lichtman J.H., Canino G., Rubio-Stipec M., Rae D.S. (1996) Criterion validity of the NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3). *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**, 878–888.
- Scitout M.J., Eisenberg M. (2007) Evaluating the evidence for and against the overdiagnosis of ADHD. *Journal of Attention Disorders*, **11**, 106–113.

- Sobin E., Weissman M.M., Goldstein R.B., Adams P. (1993) Diagnostic interviewing for family studies: Comparing telephone and face-to-face methods for the diagnosis of lifetime psychiatric disorders. *Psychiatric Genetics*, **3**, 227–233.
- Willcutt E.G., Pennington B.F., Chhabildas N.A., Friedman M.C., Alexander J. (1999) Psychiatric comorbidity associated with DSM-IV ADHD in a nonreferred sample of twins. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**, 1355–1362.
- Zucker M., Morris M.K., Ingram S.M., Morris R.D., Bakeman R. (2002) Concordance of self- and informant ratings of adults' current and childhood attention-deficit/hyperactivity disorder symptoms. *Psychological Assessment*, **14**, 379–389.