

## Influence of assessment instrument on ADHD diagnosis

Maj-Britt Posserud · Anne Karin Ullebø ·  
Kerstin Jessica Plessen · Kjell Morten Stormark ·  
Christopher Gillberg · Astri Johansen Lundervold

Received: 21 February 2013 / Accepted: 11 June 2013 / Published online: 4 July 2013  
© Springer-Verlag Berlin Heidelberg 2013

**Abstract** We compared four instruments commonly used to screen for and diagnose Attention-Deficit/Hyperactivity Disorder (ADHD) in children. The Bergen Child Study included a DSM-IV ADHD symptom list and the Strengths and Difficulties Questionnaire (SDQ) as screen in Phase one. Phase two included the parent Development and Well-Being Assessment (DAWBA), whereas Phase three comprised in-depth clinical assessment, including the Schedule for Affective Disorders and Schizophrenia for School Aged Children (K-SADS). We compared ADHD as diagnosed by the four instruments in the children with normal intellectual functioning participating in all three phases ( $N = 234$ ). The DSM-IV ADHD symptom list showed moderate agreement with all other instruments ( $\kappa = 0.53$ – $0.57$ ), whereas there was fair agreement between the K-SADS-DAWBA ( $\kappa = 0.31$ ) and between SDQ-DAWBA ( $\kappa = 0.33$ ). The DAWBA diagnosed fewer children with ADHD than did the other instruments. Implications for use of the instruments are discussed.

**Keywords** ADHD · Assessment · Epidemiology · DAWBA · K-SADS · Agreement · SDQ

### Introduction

Medical diagnoses are often an approximation of the underlying problem, based on a functional description of the symptoms observed and described by the patient or a collateral informant. This is especially true for psychiatric diagnoses. Even with the recent advances made in the last decades in neuroscience, both with regard to imaging techniques and genetics, the precise underlying pathologic processes for each individual patient will probably remain obscure. This challenge has been met by operationalizing diagnostic criteria and developing standardised screening tools and diagnostic interviews [1]. However, agreement between different informants, assessment instruments and settings is usually only fair to moderate [2–4].

---

M.-B. Posserud (✉)  
Department of Child and Adolescent Psychiatry,  
Haukeland University Hospital, 5021 Bergen, Norway  
e-mail: maj-britt.posserud@uni.no

M.-B. Posserud · A. K. Ullebø · K. M. Stormark ·  
A. J. Lundervold  
Regional Centre for Child and Youth Mental Health and Child  
Well-Fare, Uni Health, Uni Research, Bergen, Norway

M.-B. Posserud · A. J. Lundervold  
Department of Biological and Medical Psychology, University  
of Bergen, Bergen, Norway

M.-B. Posserud · A. J. Lundervold  
K. G. Jebsen Center for Research on Neuropsychiatric Disorders,  
Bergen, Norway

K. J. Plessen  
Center for Child and Adolescent Mental Health Capital Region,  
Institute for Clinical Medicine, University of Copenhagen,  
Copenhagen, Denmark

K. J. Plessen  
Faculty of Medicine and Dentistry, Institute of Clinical  
Medicine, University of Bergen, Bergen, Norway

K. M. Stormark  
Department of Clinical Psychology, University of Bergen,  
Bergen, Norway

C. Gillberg  
Gillberg Neuropsychiatry Centre, Department of Neuroscience  
and Physiology, University of Gothenburg, Gothenburg, Sweden

Furthermore, most diagnostic instruments have advantages and drawbacks. If an instrument is thorough, it is usually quite lengthy. Although its content validity may be good, a large part of the patient population will not be able to complete the questionnaire or interview. If brief, it is easier for the patient to stay on task through the entire assessment, but content validity may be poorer. This is especially true for instruments assessing Attention-Deficit/Hyperactivity Disorder (ADHD), where difficulties in keeping focus over any length of time are a key problem.

In child and adolescent psychiatry, the challenge is even greater than in adult psychiatry, given that referral and diagnostic assessment rely on information provided by parents, teachers or other observers [5]. Thus, children depend on evaluations by other people, and it is only if others perceive a mental health problem that the child will be evaluated by a professional at all [4].

ADHD is one of the most common child psychiatric disorders. Prevalence rates vary considerably across nationalities and studies [6], but in a recent American study more than 5 % of the childhood population was currently *medicated* for ADHD [7]. ADHD exerts a vast toll on health services, especially child and adolescent mental health services (CAMH), and managing ADHD diagnostic work and follow-up is a major challenge for the limited resources in child and adolescent psychiatry worldwide. There is a large array of diagnostic tools available to aid the clinician in the assessment and follow-up of ADHD, and considering the previous discussion it is important to understand the bias the use of various instruments may impose on the ADHD diagnostic procedure.

ADHD rating scales are easily accessible, quick to complete and score, and interviewer independent. Questionnaires are therefore often used in large-scale epidemiologic studies, to gather standardised information on symptoms and to monitor treatment in clinical work. In a review of evidence-based assessment for ADHD, the authors concluded that “diagnosing ADHD is most efficiently accomplished with parent and teacher rating scales” but that their use should be supplied by information regarding impairment, time of onset etc. [8]. The hyperactivity-inattention subscale of the Strengths and Difficulties Questionnaire (SDQ) has been validated against the Development and Well-Being Assessment (DAWBA) [9, 10], compared with the Child Behaviour Checklist (CBCL) and validated as screen against clinical diagnosis/semi-structured interviews in several studies [11–13]. It has also been compared with the DSM-IV ADHD symptom list [14]. The DSM-IV ADHD items from the Swanson, Nolan and Pelham Questionnaire-IV (SNAP-IV) have high face validity for the symptom constellation of the DSM-IV defined diagnoses of ADHD as it contains the same observable behavioural symptoms of inattention, impulsivity and

hyperactivity [15]. The main drawback of ADHD rating scales is the lack of clinical ascertainment of the diagnosis, and thereby insecurity as to whether respondents have appropriately understood the questions asked.

The DAWBA [16, 17] is a structured interview with open questions designed for interview or online self-completion (<http://www.dawba.com>). It has been translated into 28 languages, including Chinese, Arabic and many European languages, and been used in many large-scale epidemiologic studies. Similar to rating scales, the DAWBA does not include any direct assessment of the child. To withhold clinical validity, questions have been formulated to improve the informants’ understanding of what kind of behaviour is targeted, and follow-up open questions on the nature and the impact of the symptoms have been included. Goodman et al. [17] compared the agreement between the DAWBA and regular, non-standardised, case notes of 39 patients from a psychiatric clinic. The study used a three level agreement scale (“poor”, “partial” and “substantial”) and found a moderate to good agreement measured with Kendall’s tau  $\beta$  for ADHD. Further validation of the DAWBA regarding ADHD has to our knowledge not been undertaken, but there are recent studies examining the effect on clinical decision making and treatment of using the DAWBA interview in addition to clinical examination, finding that the DAWBA helped clinicians to some extent in diagnosing comorbid emotional problems, but did not change course of treatment or diagnostic processes regarding externalising disorders [18, 19]. Another study comparing the DAWBA with the Diagnostic Interview Schedule for Children (DISC-IV), and the Child and Adolescent Psychiatric Assessment (CAPA) showed that the DAWBA diagnosed less children in total, and reported lower ADHD prevalences than the DISC (9.2 vs. 14.0 %,  $p = 0.005$ ), but similar rates as the CAPA (9.2 vs. 10.6 %,  $p = 0.49$ ) [20]. Overall agreement was, however, relatively good ( $\kappa = 0.61$  for any disorder), and DAWBA was recommended whenever rather stringent diagnostic criteria is warranted [20].

The Schedule for Affective Disorders and Schizophrenia for School Aged Children—Present and Lifetime Version (K-SADS) [21] is a semi-structured interview to be carried out by mental health clinicians, as it involves clinical decisions based on the responses from parents and children, but also on the clinician’s observation and collection of other relevant information. Studies have shown that even experienced clinicians tend to omit vital information, e.g. information necessary to diagnose co-morbid disorders, when not using structured instruments, such as the K-SADS [22]. An important drawback of the K-SADS is that it is time- and resource consuming, limiting the number of subjects that can be included in research studies, and thus reducing the statistical power of the study. Still, it is a widely used instrument in research and clinical work

where it is often used as the golden standard due to its excellent content validity. However, the initial validity study did not validate against clinical diagnoses [21]. To make a clinical diagnosis of ADHD, the assessment should include formal cognitive testing, teacher/school reports, school observation, psychomotor assessment as well as medical examination, including testing of hearing and vision, but these are not an inherent part of the K-SADS but are provided in the anamnestic part of the interview with the parents. To our knowledge, only the Korean translation of the K-SADS interview has been validated against a clinical diagnosis of ADHD [23]. This study of 91 clinical cases did find good agreement of 0.70 with a clinical diagnosis of ADHD, with sensitivity of 0.77 and specificity of 0.95 for ADHD.

#### Aim of the study

We compared different instruments assessing ADHD in children participating in the Bergen Child Study (BCS). The SDQ, the DSM-IV ADHD symptom list from the SNAP-IV, the DAWBA and the K-SADS were all included to assess for ADHD in a subsample of the BCS. Being the instrument most similar to a clinical assessment, the K-SADS diagnosis was used as the reference value, and we investigated agreement across the four instruments.

#### Methods

The BCS is a population-based, longitudinal study of child mental health and development. One of the main aims of the study was to assess the overall prevalence of mental health problems in Norwegian school children, but also to provide national validations of commonly used instruments. A three-phase procedure applying different instruments was therefore used, providing the present study with four instruments targeting ADHD. Phase 1 included a parent and teacher questionnaire to all children in 2–4th grade (7–9 years old) ( $N = 9,340$ ) in 2002 (Fig. 1) [24, 25]. The questionnaire included the SDQ [9, 10], the DSM-IV ADHD items from the SNAP-IV [15] and other measures. DSM-IV items were adapted in accordance with SDQ and other instruments in the overall questionnaire to a three-point Likert scale (not true-somewhat true-certainly true, scored as 0-1-2). The questionnaire served as a general screen, and parents of all children, who were screen positive<sup>1</sup> were invited to the parent DAWBA (Phase 2)

<sup>1</sup> “Screen positive” was defined as scoring above the 90th percentile on either parent or teacher SDQ, or above the 98th percentile on other measures in the questionnaire, or moderate impact by any informant on the SDQ.

along with a large control group of screen negative children. Children diagnosed according to the DAWBA and control children were then invited to a more extensive clinical assessment, including the K-SADS and Wechsler Intelligence Scale for Children—third edition (WISC-III) (Phase 3). Control children were either screen negative in the questionnaire phase with no DAWBA diagnosis (DAWBA true negative) or they were screen positive in Phase 1, but received no diagnosis according to the DAWBA (DAWBA false positive). The second phase consisted of 1,009 DAWBA interviews (645 screen positive children, 44.6 % of invited families and 364 screen negative children, 49.7 % of invited families). Of the 234 children included in the present study, 68 had one or more DAWBA diagnoses (5 screen negative children, 63 screen positive children), 81 children were DAWBA true negative, whereas 85 children were DAWBA false positive.

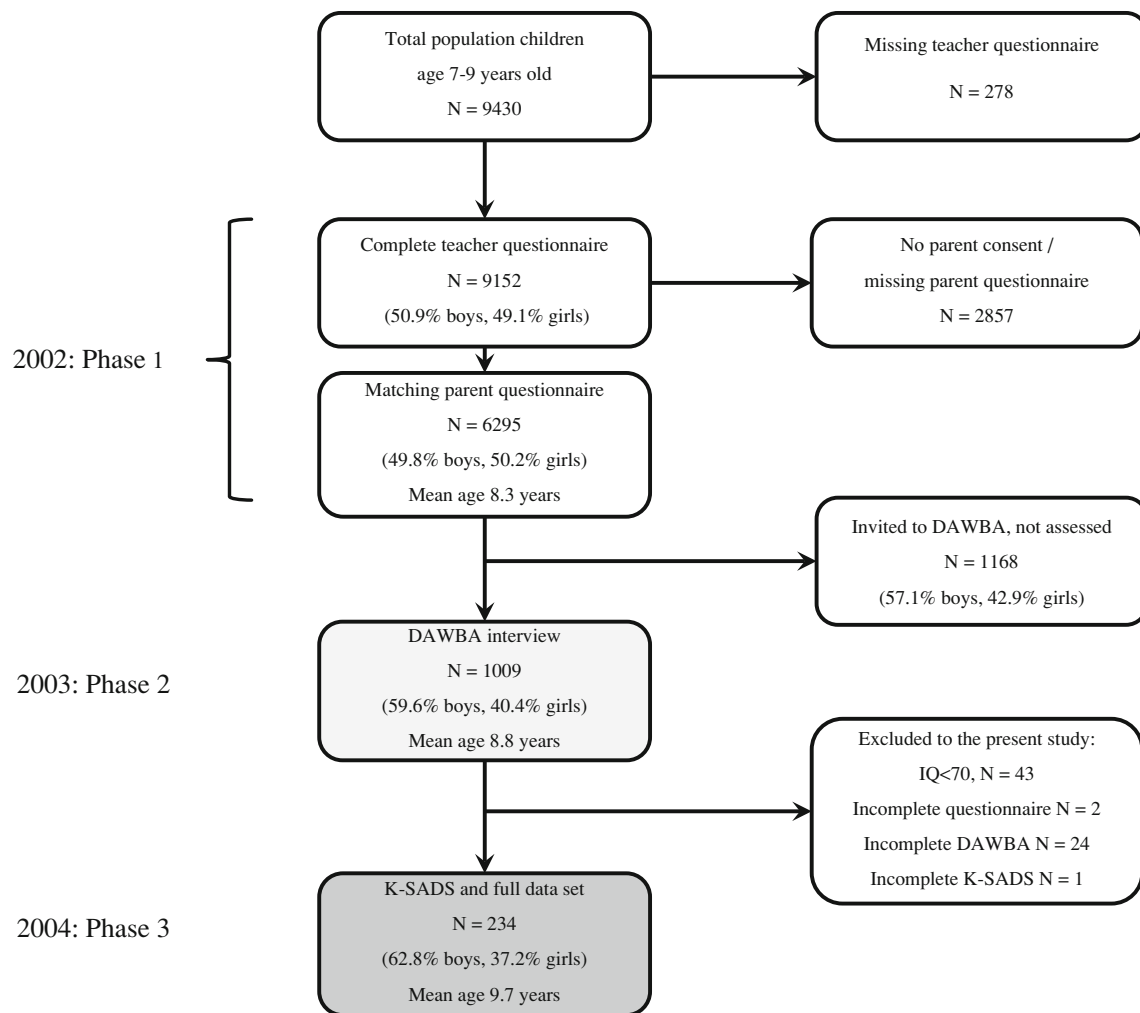
The BCS was approved by the Western Norway Regional Committee on Medical and Health Research Ethics and the Norwegian Data Inspectorate.

#### Participants

Children included in the present study participated in all three phases of the BCS and had no missing data on any of the key variables. Children with a full scale IQ < 70 were excluded ( $N = 43$ ) as ADHD diagnosis in children with intellectual disability is a complex clinical task and the psychometric properties of the instruments involved may differ in this group. One child did not complete the K-SADS, 24 children had incomplete/missing DAWBA, and two children had incomplete questionnaires, leaving 234 children with complete data on all key variables (Fig. 1).

#### Procedures

DSM-IV ADHD items were dichotomised and ADHD was deemed present based on the information from the questionnaire when six or more symptoms were present on the inattention and/or hyperactivity subscale from both parents and teachers. This definition generated prevalence rates that correspond well with international estimates of ADHD prevalence [26]. The SDQ algorithm for hyperkinetic disorders (ICD-10 hyperkinetic and DSM-IV ADHD) was applied according to the internet site, <http://www.sdqinfo.com>, for children below the age of 11 (no self-report). All DAWBA interviews were conducted by trained interviewers (medical doctors or medicine/psychology/social science students) and scored by two trained child psychiatrists. All diagnoses were finally reviewed and discussed with a third child psychiatrist, who was one of the main developers of the DAWBA. Inter-rater agreement measured by Kappa was 1.0 for ADHD [25]. The K-SADS



**Fig. 1** Flowchart of three phases of the first wave of the Bergen Child Study. *DAWBA* Development and Well-Being Assessment, *K-SADS* Schedule for Affective Disorders and Schizophrenia for School Aged Children

interviews were performed by a team of psychologists and one medical doctor (M.P.). The K-SADS diagnosis of ADHD was dichotomised so that “in remission” (one case) and “definite” disorders were counted as ADHD. “Possible” ADHD and “previous” ADHD were scored as no ADHD. As K-SADS was part of the assessment in Phase 3 and following the DAWBA, there was a time lag between the DAWBA interview (majority performed in 2003) and the K-SADS (majority performed in 2004) of minimum 6 months but less than 2 years (Fig. 1).

#### Statistical analyses

Analyses were conducted using PASW statistical software package, version 17.0. Agreement was measured using Cohen’s kappa ( $\kappa$ ) and Kendall’s tau  $\beta$  for comparison. Cohen’s kappa was evaluated as slight ( $\kappa = 0–0.20$ ), fair ( $\kappa = 0.21–0.40$ ), moderate ( $\kappa = 0.41–0.60$ ), good/substantial ( $\kappa = 0.61–0.80$ ) and perfect ( $\kappa = 0.81–1$ ) [27].

#### Results

##### ADHD diagnosis and agreement

The DAWBA interview identified fewer children in total with an ADHD diagnosis compared to the K-SADS and to the two questionnaires (Table 1). Few girls were identified with an ADHD diagnosis by any of the instruments, and the DAWBA identified no girls as having ADHD, leading to low or non-existing agreement for girls (Tables 1, 2). Agreement between the various instruments was fair to moderate ( $\kappa = 0.31–0.57$ ) for the total sample and for boys, with K-SADS–DAWBA agreement being the lowest ( $\kappa = 0.31$ ) (Table 2). The ADHD DSM-IV symptom list had moderate agreement with all other instruments, ( $\kappa = 0.53–0.57$ ), whereas the SDQ and the DAWBA had fair agreement ( $\kappa = 0.33$ ). The analyses of level of agreement were repeated for an ICD-10 diagnosis of hyperkinetic disorder generated from the DAWBA, with a

**Table 1** The number of children categorised as ADHD by the four instruments in the Bergen Child Study by gender (total  $N = 234$ , 147 boys and 87 girls)

Instrument	Boys		Girls	
	ADHD	Not ADHD	ADHD	Not ADHD
SDQ	44	103	5	82
DSM-IV	28	119	2	85
DAWBA	16 (14) <sup>a</sup>	131	0	87
K-SADS	31	116	6	81

ADHD Attention Deficit Hyperactivity Disorder, SDQ Strengths and Difficulties Questionnaire, DSM-IV Diagnostic Statistical Manual—IV diagnostic criteria for ADHD, DAWBA Development and Well-Being Assessment, K-SADS Schedule for Affective Disorders and Schizophrenia for School Aged Children

<sup>a</sup> DSM-IV DAWBA diagnosis with ICD-10 Hyperkinetic Disorder in parenthesis

**Table 2** Agreement on the ADHD diagnosis between the four instruments in the Bergen Child Study

Instruments	Cohen’s Kappa ( $\kappa$ )			Kendall’s tau- $\beta$ Total sample
	Total sample	Boys	Girls	
K-SADS–DAWBA	0.31	0.33	–	0.35
K-SADS–SDQ	0.49	0.45	0.52	0.50
K-SADS–DSM-IV	0.54	0.60	–0.04**	0.53
DAWBA–SDQ	0.33	0.33	–	0.40
DAWBA–DSM-IV	0.57	0.58	–	0.61
DSM-IV–SDQ	0.53	0.53	0.26*	0.56

Agreement between DAWBA and other instruments could not be calculated in girls due to no DAWBA cases

ADHD Attention Deficit Hyperactivity Disorder, SDQ Strengths and Difficulties Questionnaire, DSM-IV Diagnostic Statistical Manual—IV diagnostic criteria for ADHD, DAWBA Development and Well-Being Assessment, K-SADS Schedule for Affective Disorders and Schizophrenia for School Aged Children

\* Not significant

\*\*  $p < 0.01$ ,  $p < 0.001$  for all other values

Kappa agreement of  $\kappa = 0.28$  with SDQ,  $\kappa = 0.33$  with K-SADS (DSM-IV ADHD diagnosis) and  $\kappa = 0.50$  with the DSM-IV ADHD symptom list. The higher agreement with the DSM-IV symptom list was owing to none of the DAWBA cases being DSM-IV screen negative. However, there were many false screen positive cases. The agreement between DAWBA and K-SADS regarding ICD-10 hyperkinetic disorder was expected to be lower than for the DAWBA generated ADHD (DSM-IV) disorder as the K-SADS only reports ADHD defined by DSM-IV. The agreement between the SDQ prediction of hyperactivity-inattention and the DAWBA was fair, both for the ICD-10 and the DSM-IV criteria.

**Table 3** Instrument properties in detecting K-SADS defined ADHD

Instrument	Sensitivity	Specificity	PPV	NPV
SDQ	0.68	0.88	0.51	0.94
DSM-IV	0.54	0.95	0.67	0.92
DAWBA	0.27	0.97	0.63	0.88

ADHD Attention Deficit Hyperactivity Disorder, SDQ Strengths and Difficulties Questionnaire, DSM-IV Diagnostic Statistical Manual—IV diagnostic criteria for ADHD, DAWBA Development and Well-Being Assessment, K-SADS Schedule for Affective Disorders and Schizophrenia for School Aged Children, PPV Positive Predictive Value, NPV Negative Predictive Value

Sensitivity and specificity regarding K-SADS defined ADHD diagnosis

The two questionnaires (SDQ and DSM-IV symptom list) had the best screening properties for ADHD as defined by the K-SADS, with a sensitivity of 0.54 and a specificity of 0.95 for the ADHD DSM-IV symptom list, and sensitivity of 0.68 and specificity of 0.88 for the SDQ. The DAWBA had low sensitivity (0.27), but showed a high specificity (0.97) (Table 3).

Informant and ADHD diagnosis

We examined the relationship between how each informant rated the child on the DSM-IV ADHD symptom list and ADHD diagnoses according to the DAWBA and the K-SADS. We defined a child as screen positive for ADHD if the informant rated the child as having a minimum of six symptoms on at least one subscale. All children who received a diagnosis of ADHD by the DAWBA were scored screen positive by the parents, and only two children had been rated negative for ADHD by their teacher, whereas the K-SADS diagnosed ADHD in eight children with screen negative parental questionnaire and eleven children with teacher screen negative questionnaires (Table 4).

K-SADS–DAWBA case disagreement

K-SADS diagnosed 37/234 children with ADHD, whereas the DAWBA diagnosed 16 children with DSM-IV ADHD. The two instruments only fully agreed in 10 children, but 27/37 of the children with a K-SADS ADHD diagnosis were considered to be “cases” by the DAWBA (Table 5). The 27 children with a K-SADS but no DAWBA diagnosis of ADHD were diagnosed by the DAWBA as pervasive developmental disorder (PDD; 4 children), conduct disorder [CD; 10 children (3 girls)], emotional disorder [2 children (1 girl)], attachment disorder (1 boy), and 10 of the 27 children received no diagnosis in the DAWBA (2 girls).



**Table 4** Number of children with an ADHD diagnosis according to DAWBA and K-SADS versus informant rating on the DSM-IV checklist

Screen status on the DSM-IV checklist	DAWBA ADHD	K-SADS ADHD
Parent negative	0	8
Parent positive	16	29
Teacher negative	2	11
Teacher positive	14	26

*ADHD* Attention Deficit Hyperactivity Disorder, *DAWBA* Development and Well-Being Assessment, *K-SADS* Schedule for Affective Disorders and Schizophrenia for School Aged Children

**Table 5** Number of children diagnosed with ADHD or not, comparison of K-SADS versus the DAWBA

DAWBA	K-SADS	
	ADHD	Not ADHD
ADHD	10	6
Not ADHD	27 (6 girls)	191 (81 girls)

*ADHD* Attention Deficit Hyperactivity Disorder, *DAWBA* Development and Well-Being Assessment, *K-SADS* Schedule for Affective Disorders and Schizophrenia for School Aged Children

Six children were diagnosed with ADHD by the DAWBA, but not by the K-SADS. According to the WISC-III and the K-SADS, five of those children had borderline IQ ( $IQ \leq 80$ ) and/or learning disorders (one of whom was scored as “possible” ADHD by the K-SADS). The sixth child had motor problems and previous encopresis.

All children with a diagnosis of PDD by the DAWBA also received an independent PDD diagnosis in Phase 3 [28], but using the K-SADS the exclusion criteria for ADHD were not applied, meaning that they were diagnosed as having ADHD even though they also received a PDD diagnosis. These four children could possibly have received an ADHD diagnosis by the DAWBA as well, had the exclusion rules not been applied by the DAWBA.

Eight of the 37 children receiving a diagnosis of ADHD in the K-SADS were on methylphenidate or atomoxetine, versus none of the children without a K-SADS ADHD diagnosis. The DAWBA identified four of the eight children receiving methylphenidate or atomoxetine as having ADHD. One of the medicated children received a DAWBA diagnosis of PDD, whereas three children using methylphenidate received no diagnosis of either ADHD or PDD according to the DAWBA, two of whom received no DAWBA diagnosis at all, and one child the DAWBA diagnoses of posttraumatic stress disorder and specific phobia.

## Discussion

In the present study, we examined the agreement of several frequently used instruments to screen for and to diagnose ADHD in children. We defined the K-SADS interview as reference as its procedure was the most similar to clinical practice. The agreement was fair to moderate across all instruments. The DAWBA was the most restrictive diagnostic instrument, diagnosing fewer children as having ADHD than the other instruments. Both the DSM-IV ADHD symptom list, the SDQ and the K-SADS labelled twice or more as many children as having an ADHD.

The rate of ADHD identified by the K-SADS may seem high (37 of 234), but note that the sample is a quasi-case-control sample, consisting of children who were in two-thirds screen positive on the BCS screening questionnaires for psychopathology, where almost half of the screen positive children also had a clinical DSM-IV DAWBA diagnosis (any kind).

All instruments identified few girls with a diagnosis of ADHD. The agreement values are, therefore, to be regarded as unstable, and for the DAWBA they could not be calculated, as no girls were diagnosed with ADHD by the DAWBA. This low number of girls captured by the BCS study suggests that the screening and diagnostic procedures in the study were inadequate to identify ADHD symptoms in girls. Using combined teacher and parent information as a requirement for the questionnaire ADHD diagnosis undoubtedly contributed to this regarding the SDQ and the DSM-IV checklist ADHD, as we have previously shown that teachers have a very high gender bias of 6.2:1 for the combined type ADHD as opposed to parental reports with 2.9:1 [26]. This does not, however, explain the low ascertainment in the DAWBA and the K-SADS. Previous studies have shown that special focus on girls' problems is needed to enable proper assessment for girls [29], and that very few adult women with ADHD are diagnosed in childhood [30]. Yet other studies have shown that even when ADHD symptoms and diagnosis are registered in a girl, she is more likely to receive treatment for emotional problems [31], suggesting that professionals also struggle to perceive ADHD symptomatology adequately in girls.

Our finding that the DAWBA was the most restrictive instrument fits well with several epidemiological studies using the DAWBA reporting ADHD prevalence rates in the lower range [16, 25]. Angold et al. [20] also found that the DAWBA diagnosed less children as having ADHD than the DISC and that the DAWBA was in general more restrictive than both the CAPA and the DISC. The DAWBA in the BCS estimated the prevalence of ADHD to 1.7 % [25]. It has been suggested that the low prevalence is due to the Nordic advantage of good social services and low income differences [32], but when estimating ADHD

prevalence based on the DSM-IV symptom list instead of on the DAWBA, we found a prevalence of 6.8 % in the same population, i.e. considerably higher [26]. Moreover, we diagnosed a higher number of cases with ADHD with the K-SADS than with the DAWBA. Although the prevalence of the ICD-10 defined hyperkinetic disorder (HKD) is much less prevalent than the DSM-IV defined ADHD [6], the DAWBA seems to identify HKD and DSM-IV ADHD at a similar low rate [25]. This demonstrates the caveat of concluding on etiologic reasons for differences between studies using different methodology.

Although often used as the golden standard in research, the K-SADS interview has received very little validation studies itself. However, in the only study validating the K-SADS against a clinical diagnosis, Kim et al. [23] showed good agreement of 0.70 between a clinical ADHD diagnosis and the K-SADS generated diagnosis for the Korean translation of K-SADS. The present study showed moderate agreement between the K-SADS and the questionnaires, but only fair agreement with the DAWBA, mainly due to the low number of children diagnosed with ADHD by the DAWBA. The overall magnitude of children diagnosed with ADHD was most similar between the K-SADS ( $N = 37$ ) and the DSM-IV symptom list ( $N = 30$ ), whereas SDQ labelled more children ( $N = 49$ ) as having ADHD and the DAWBA only 16 children as having ADHD.

The low rate of ADHD diagnosed by the DAWBA is probably related to the design of the DAWBA. It was designed not to overdiagnose child and adolescent mental health problems, but to provide the government in the UK with relevant parameters to plan their health services according to the *CAMH needs* of children in the UK society. However, the DAWBA estimated ADHD prevalence in Bergen was 1.7 % (in 2002–2004), whereas the prescription reports from 2010 showed that in the county of Hordaland (where Bergen city comprises 54 % of the population), 2.4 % of children and adolescents (age 10–14) received medication for ADHD (<http://www.reseptregisteret.no/Prevalens.aspx>). This could suggest that the DAWBA underestimates the rate of children with ADHD, and this is also supported by the identification of four children in the sample currently receiving ADHD medication, who were not diagnosed as ADHD by the DAWBA (one of these children also had ASD).

Screening questionnaires are on the other hand limited by the uncritical listing of symptoms without including measures of functional impairment, generating sometimes disparately high prevalence estimates. The 6.8 % ADHD prevalence generated from the DSM-IV checklist is well above the 2.4 % of children who were using ADHD medication in Hordaland in 2010 and is probably an overestimate of clinically significant ADHD (although a

recent report showed that 5.4 % of US children were currently on ADHD medication, and bearing in mind that not all children fulfilling ADHD criteria should be treated with medicine) [7]. Some children may have many ADHD symptoms, but lack clinical impairment, and some children may have ADHD symptoms due to other problems, e.g. emotional problems, psychosocial problems and learning difficulties. The questionnaires in this study and the K-SADS were, however, in the same range concerning the rate of children labelled as ADHD. As ADHD is often comorbid with emotional problems, psychosocial problems and learning difficulties it may be necessary to consider comorbid conditions rather than to think one diagnosis or the other [33].

None of the cases of ADHD diagnosed by the DAWBA had screen negative parent questionnaires. This supports the hypothesis that the quality of the parental report is crucial for the diagnostic precision of the DAWBA. If the parent is unable to give an informative account of their child's behaviour, the DAWBA will probably not be appropriate to diagnose a child with ADHD. The same pattern was found in a recent prevalence study of diagnoses within the autism spectrum (PDD diagnoses) [28]. This could also explain some of the differences between the results in a study by Foreman et al. [34], who examined the DAWBA in a sample of worried parents, and the results in the present study, where parents were invited as part of a population study and possibly had no previous concerns at all.

The obvious advantages of the DAWBA of being easy to administer and demanding little input from CAMHs and other services has resulted in the DAWBA becoming a widely used instrument for the assessment of child mental health problems, in clinics and in research projects. Foreman et al. [34] even suggested the DAWBA could substitute the clinical examination by a child psychiatrist regarding ADHD, based on their results reporting an agreement of  $\kappa = 0.62$  for DSM-IV ADHD, and a positive predictive value (PPV) of  $>0.8$  with the clinical diagnosis. Another two recent studies showed very little enhancement of the DAWBA to the clinical work, except in improved identification of internalising comorbid conditions [18, 19]. In our study, most children with a K-SADS diagnosis of ADHD but no DAWBA ADHD were labelled as having CD by the DAWBA, whereas the majority of children diagnosed with ADHD by the DAWBA, but not by the K-SADS were children with borderline intellectual functioning. Differentiating ODD, ADHD and borderline intellectual functioning is a challenging task, and no golden standard exists, therefore we fully agree with Angold et al. [20]; “The issue is not “which interview is right,” but “what are each interview's properties and which is most suitable for a particular application?” (p. 515). Our study

indicates that the DAWBA is generally more restrictive, more prone to label a child as CD whereas the K-SADS labels the same child as ADHD, and that some children regarded as only having borderline low intellectual functioning by the K-SADS are labelled as ADHD by the DAWBA. Our study shows that parental information is crucial for the DAWBA diagnosis, indicating that the DAWBA should be used and interpreted cautiously with less engaged or less able parents.

When it comes to screening for ADHD, our results suggest that in situations where it is important to have few false screen positive cases, the DAWBA is the best choice, but in situations, where high sensitivity is crucial, i.e. where it is important not to miss any cases, the SDQ seems more useful.

### Limitations

We used the K-SADS as the reference value in the present paper, as it resembles a clinical work-up most closely. The interviews were, however, performed by different clinicians, and the results are probably less reliable than the DAWBA results. Unfortunately, we do not have a kappa measure for the agreement between the K-SADS interviewers, but meetings for discussion of complex cases were arranged on a regular basis, and the final decisions were always given to the senior psychologist, who had the main responsibility for the K-SADS part during the assessment period. In addition, the time lag between the questionnaires, the DAWBA and the K-SADS was at least 6 months and varied, reducing the direct comparability of the instruments. ADHD is a neurodevelopmental disorder where the overall prevalence is reported to decrease with age. This should have resulted in fewer children being identified by the K-SADS when compared with the DAWBA. In K-SADS, the duration of the disorder is also scored, and none of the ADHD cases were scored as having lasted less than 3 years. Furthermore, we also had questionnaire data collected prior to the DAWBA interview, that show both higher agreement with the later interview (K-SADS) and higher prevalence than the earlier interview, the DAWBA. This indicates that performing the interviews more concurrently in time would, if anything, had led to an even lower agreement between the two interviews.

### Conclusion

In psychiatry research, low agreement between instruments and informants is common. The present study showed that this was true also for commonly used ADHD assessment tools. The agreement between the four instruments in the study was fair to moderate. Large variance in which

children are considered as having ADHD and in ADHD prevalence rates between studies using differing methodology should therefore be expected. The DAWBA was restrictive in diagnosing ADHD, with only 35 to 50 % of all cases diagnosed by the other instruments reaching diagnostic status on this instrument. Prevalence rates based on the DAWBA are likely to be correspondingly lower than prevalence rates from studies using the other instruments. The ADHD DSM-IV symptom checklist list had the best agreement with all other instruments ( $\kappa = 0.53\text{--}0.57$ ). The DAWBA is recommended for screening if high specificity is desired (0.97), whereas the SDQ had the highest sensitivity (0.68), and the DSM-IV checklist the best trade-off between sensitivity (0.54) and specificity (0.95) and the best positive predictive value (PPV = 0.67). Few girls were identified by all instruments; the DAWBA diagnosed none, indicating that major focus on girl symptomatology is warranted.

**Acknowledgments** We thank the Bergen Child Study research group that has worked together for many years, and whose joint effort has made the study possible. We are grateful to the support of the Regional Centre for Child and Youth Mental Health and Child Welfare, Uni Health, for hosting the study for all these years. Thanks to Anna Spyrou for proof-reading the final manuscript. We also thank the Bergen City Council for collaborating to the study and allowing us to perform the study through the schools. But above all, we thank all the teachers, parents and children in Bergen, for participating in study. The present study was funded by the Western Regional Health Authorities, and the Regional Centre for Child and Youth Mental Health and Child Welfare, Uni Health, Uni Research.

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

### References

1. Moller HJ (2005) Problems associated with the classification and diagnosis of psychiatric disorders. *World J Biol Psychiatry* 6(1):45–56
2. Achenbach TM, McConaughy SH, Howell CT (1987) Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychol Bull* 101(2):213
3. De Reyes Los 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(4):483–509
4. Yeh M, Weisz JR (2001) Why are we here at the clinic? Parent-child (dis)agreement on referral problems at outpatient treatment entry. *J Consult Clin Psychol* 69(6):1018–1025
5. Mash EJ, Hunsley J (2005) Evidence-based assessment of child and adolescent disorders: issues and challenges. *J Clin Child Adolesc Psychol* 34(3):362–379
6. Faraone SV et al (2003) The worldwide prevalence of ADHD: is it an American condition? *World Psychiatry* 2(2):104–113
7. Kim J et al (2011) Health behaviors and obesity among US children with attention deficit hyperactivity disorder by gender and medication use. *Prev Med* 52(3–4):218–222



8. Pelham WE Jr, Fabiano GA, Massetti GM (2005) Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *J Clin Child Adolesc Psychol* 34(3):449–476
9. Goodman R (1997) The strengths and difficulties questionnaire: a research note. *J Child Psychol Psychiatry* 38(5):581–586
10. Goodman R et al (2000) Using the strengths and difficulties questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry: J Mental Sci* 177:534–539
11. Becker A et al (2004) Validation of the parent and teacher SDQ in a clinical sample. *Eur Child Adolesc Psychiatry* 13(Suppl 2):III11–III16
12. Goodman R, Scott S (1999) Comparing the strengths and difficulties questionnaire and the child behavior checklist: is small beautiful? *J Abnorm Child Psychol* 27(1):17–24
13. Klasen H et al (2000) Comparing the German versions of the strengths and difficulties questionnaire (SDQ-Deu) and the child behavior checklist. *Eur Child Adolesc Psychiatry* 9(4):271–276
14. Ulllebo AK et al (2011) Screening for the attention deficit hyperactivity disorder phenotype using the strength and difficulties questionnaire. *Eur Child Adolesc Psychiatry* 20(9):451–458
15. Swanson JM et al (2001) Clinical relevance of the primary findings of the MTA: success rates based on severity of ADHD and ODD symptoms at the end of treatment. *J Am Acad Child Adolesc Psychiatry* 40(2):168–179
16. Ford T, Goodman R, Meltzer H (2003) The British child and adolescent mental health survey 1999: the prevalence of DSM-IV disorders. *J Am Acad Child Adolesc Psychiatry* 42(10):1203–1211
17. Goodman R et al (2000) The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 41(5):645–655
18. Aebi M et al. (2012) The use of the development and well-being assessment in clinical practice: a randomized trial. *Eur Child Adolesc Psychiatry* 21(10):559–567
19. Ford T et al (2013) Can standardized diagnostic assessment be a useful adjunct to clinical assessment in child mental health services? A randomized controlled trial of disclosure of the development and well-being assessment to practitioners. *Soc Psychiatry Psychiatr Epidemiol* 48(4):583–593
20. Angold A et al (2012) Psychiatric diagnostic interviews for children and adolescents: a comparative study. *J Am Acad Child Adolesc Psychiatry* 51(5):506–517
21. Kaufman J et al (1997) Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 36(7):980–988
22. Ambrosini PJ (2000) Historical development and present status of the schedule for affective disorders and schizophrenia for school-age children (K-SADS). *J Am Acad Child Adolesc Psychiatry* 39(1):49–58
23. Kim YS et al (2004) The reliability and validity of Kiddie-Schedule for affective disorders and schizophrenia-present and lifetime version-Korean version (K-SADS-PL-K). *Yonsei Med J* 45(1):81–89
24. Posserud MB, Lundervold AJ, Gillberg C (2006) Autistic features in a total population of 7–9-year-old children assessed by the ASSQ (Autism Spectrum Screening Questionnaire). *J Child Psychol Psychiatry* 47(2):167–175
25. Heiervang E et al (2007) Psychiatric disorders in Norwegian 8- to 10-year-olds: an epidemiological survey of prevalence, risk factors, and service use. *J Am Acad Child Adolesc Psychiatry* 46(4):438–447
26. Ulllebo AK et al (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
27. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174
28. Posserud M-B, Lundervold AJ, Gillberg C (2009) Validation of the autism spectrum screening questionnaire in a total population sample. *J Autism Dev Disord* 39(1):126–134
29. Kopp S, Kelly KB, Gillberg C (2010) Girls with social and/or attention deficits: a descriptive study of 100 clinic attenders. *J Atten Disord* 14(2):167–181
30. Halmoy A et al (2009) Occupational outcome in adult ADHD: impact of symptom profile, comorbid psychiatric problems, and treatment: a cross-sectional study of 414 clinically diagnosed adult ADHD patients. *J Atten Disord* 13(2):175–187
31. Graetz BW et al (2006) Gender comparisons of service use among youth with attention-deficit/hyperactivity disorder. *J Emot Behav Disord* 14(1):2–11
32. Heiervang E, Goodman A, Goodman R (2008) The Nordic advantage in child mental health: separating health differences from reporting style in a cross-cultural comparison of psychopathology. *J Child Psychol Psychiatry* 49(6):678–685
33. Gillberg C (2010) The ESSENCE in child psychiatry: early symptomatic syndromes eliciting neurodevelopmental clinical examinations. *Res Dev Disabil* 31(6):1543–1551
34. Foreman D, Morton S, Ford T (2009) Exploring the clinical utility of the development and well-being assessment (DAWBA) in the detection of hyperkinetic disorders and associated diagnoses in clinical practice. *J Child Psychol Psychiatry* 50(4):460–470