European Child & Adolescent Psychiatry (2019) 28:625–633
https://doi.org/10.1007/s00787-018-1226-9

ORIGINAL CONTRIBUTION

# The predictive validity of the Strengths and Difficulties Questionnaire for child attention-deficit/hyperactivity disorder

**Kristin Romvig Overgaard[1] · Kathrine Bang Madsen[2] · Beate Oerbeck[1] · Svein Friis[1,3] · Carsten Obel[2]**

Received: 5 April 2018 / Accepted: 6 September 2018 / Published online: 15 September 2018


## Abstract

We need accurate screening measures for attention-deficit/hyperactivity disorder (ADHD) to ensure that children with the disorder are referred for assessment without raising concern for children with normal behaviour. The Strengths and Difficulties Questionnaire (SDQ) provides hyperactivity–inattention (HI), conduct, emotional and peer problem subscales and impact scores that may be used for screening. The aim of the study was to investigate the predictive validity of the Danish version of the parent SDQ HI subscale at the child age of 7 years for subsequent clinically diagnosed ADHD (age 8–15 years). Participants were part of the Danish National Birth Cohort ($N = 51{,}096$), and children with ADHD were identified through the Danish National Health registries ($n = 943$). Receiver operating characteristic analysis showed that the screening accuracy for the HI scores was good (area under the curve = .84). With Cox multivariate regression analysis, we found that SDQ HI subscale scores $\geq 7$ with impact gave a nearly 14-fold [hazard ratio (HR) = 13.59] increased risk for ADHD, while conduct and emotional problems indicated low risk (HRs of 1.62 and 1.67, respectively). For the HI subscale to be a sensitive measure for ADHD, a low cutoff (4) was needed, but gave many false screening positives (PPV = .02). Although the diagnostic accuracy of the parent version of the SDQ HI subscale for predicting ADHD was good, our results question the feasibility of screening the general child population for ADHD with only the parent SDQ HI subscale.

**Keywords** ADHD · SDQ · Cohort study · Patient registries

## Introduction

Attention-deficit/hyperactivity disorder (ADHD) is characterized by impairing symptoms of hyperactivity, impulsivity and attention deficit [1, 2]. The onset of the disorder occurs during childhood and is estimated to have a prevalence rate of about 5% [3, 4]. Co-occurring emotional and behavioural disorders are common [5–8]. Throughout life, ADHD is associated with an increased risk of other psychiatric disorders, educational and work failure, accidents, social problems, addictions and premature death, the latter especially in those diagnosed in adulthood [9–12]. If ADHD goes undetected in children, the outcome may worsen [13]. However, misclassifying children with age-appropriate activity, impulsivity and attention as screening positive for ADHD might cause unnecessary concern and use of health services, indicating that accurate screening is essential.

The Strengths and Difficulties Questionnaire (SDQ) has shown promise as a screening measure for detecting ADHD in both clinical and community samples [14–17]. The questionnaire is user-friendly, and consists of only 25 items with five subscales (each with five items) on hyperactivity–inattention (HI), conduct (CD), emotional (ED) and peer problems, and prosocial behaviour. The first four problem subscales may be summarized as a total difficulties score (http://www.sdqinfo.com) [18]. A cross-sectional community study ($N = 18{,}232$) that tested the accuracy of the SDQ in identifying ADHD (measured with a diagnostic interview) in 5–15-year-olds concluded that the HI subscale and total difficulties score performed equally well [both with areas under the curve (AUCs) > .90] in discriminating between children with ADHD and those without, and significantly

✉ Kristin Romvig Overgaard
kristin.romvig.overgaard@ous-hf.no

[1] Division of Mental Health and Addiction, Oslo University Hospital, Pb. 4959 Nydalen, 0424 Oslo, Norway

[2] Department of Public Health, Aarhus University, Aarhus, Denmark

[3] Institute of Clinical Medicine, University of Oslo, Oslo, Norway

better than the conduct problem subscale (AUCs > .80 for all age groups) [17].

Guidelines recommend the use of screeners that balance the ability to correctly classify the presence/absence of at least 70% of cases/non-cases [19]. One cross-sectional community study ($N = 6233$) suggested that a cutoff of four or five (out of ten) on the SDQ HI subscale gave the best balance between sensitivity and specificity for identifying ADHD (defined by another rating scale) [20], rather than the cutoff of ≥ 7 used in some epidemiological studies [21, 22]. Conversely, a clinical study found its highest diagnostic accuracy (positive likelihood ratio = 27.36) when using a cut-off ≥ 8 [15]. Neither of these two studies investigated whether the CD, ED and/or peer problem subscales added to diagnostic accuracy.

Cross-sectional studies have considerable limitations; a review from 2010 identified only three out of 48 studies of 4–12-year-old children ($N = 131{,}223$) that presented evidence for the predictive validity of the SDQ in a longitudinal design, and only one of parent HI ratings [16]. That study did not investigate the predictive validity of the ADHD diagnosis, but reported SDQ HI scores to be moderately stable over a 12-month period ($r = .77$) [23]. Responding to this lack of research, one longitudinal study found that the HI subscale scores at 3 years of age positively predicted ADHD 2 years later [24]. A recent cohort study of 5–7-year-old children from the County of Copenhagen, Denmark ($N = 2315$) found that combining high SDQ HI subscale scores (≥ 6) as well as impact from parents and teachers [according to the possible/probable algorithm for the prediction of a hyperactivity disorder (SDQHK), provided by the SDQ developers at http://www.sdqinfo.com] gave about a 20-fold increased risk of a subsequent ADHD diagnosis being registered in the National Health Registry at school age [25]. That same study found that the SDQ prediction algorithm for 'any probable diagnosis' gave a similar level of risk [hazard ratio (HR) = 16.34] for ADHD as the HI subscale (HR = 20.65), which was taken to indicate that psychopathology outside the core symptoms of ADHD increases the risk of later ADHD [25]. However, children given the SDQ prediction of 'any probable diagnosis' were less likely than children identified by the SDQHK possible/probable algorithm to later be diagnosed with ADHD (Positive predictive values (PPVs) of .10 and .33, respectively) [25].

Screening based on parent ratings is the easiest to implement in clinical and population settings and is thus considered more efficient and cost-effective than collecting information from teachers. However, there are few studies on the predictive validity of the parent SDQ for detecting ADHD in a population setting. In the present study, our overall aim was to investigate the predictive validity of the parent HI subscale (sum scores and a parent prediction category defined by high scores with impact) at the child age of 7 years for subsequent ADHD diagnosis (age 8–15 years). To achieve this, we investigated (1) the diagnostic accuracy of the HI subscale, other problem subscales, and total difficulties sum scores, (2) whether the SDQ HI, CD, ED and/or peer problem subscales (sum scores and parent prediction categories defined by high scores with impact) increased the risk for subsequent ADHD and (3) which SDQ HI subscale cutoff scores (alone and with impact) best balanced sensitivity and specificity for identifying subsequent ADHD.

## Method

### Participants and procedures

The present study was part of the Danish National Birth Cohort (DNBC) where general practitioners recruited women early in pregnancy between 1996 and 2003 [26]. When the child was 7 years old, a follow-up questionnaire about child health and development, including the parent version of the Strengths and Difficulties Questionnaire (SDQ) (http://www.sdq.com), was completed by the primary caregiver (99% mothers), either through the internet or on paper. A total of 57,273 participated in the 7-year follow-up. We excluded twins and siblings ($n = 5431$), and children with missing SDQ data ($n = 321$). Missing data were defined as present when at least one of the 25 items was missing. The proportion of children diagnosed with ADHD among the excluded due to missing SDQ data was not significantly different from that in the overall population (2.8% versus 2.7%, respectively, $p = .87$). The missing SDQ data were not gender specific (166 boys, 155 girls). The study population with complete SDQ data was 51,521. We then linked the DNBC with the Danish National Patient Register [27], the Danish Psychiatric Central Register [28] and the Register of Medicinal Product Statistics [29], and identified children diagnosed with ADHD (see below). We excluded a total of 425 children who received an ADHD diagnosis or redeemed ADHD medication before the age of 8 years, rendering a total of 51,096 children for analyses.

### ADHD diagnoses

ADHD diagnoses were identified through two different approaches. First, we used the Danish patient registries [27, 28], recording diagnoses given through outpatient or inpatient hospital contacts with clinicians who use the *International classification of diseases and related health problems*, 10th ed. (ICD-10) [30]. We defined ADHD to be present when ICD-10 Hyperkinetic Disorder diagnoses F90 or F98.8 were registered at ≥ 8 years of age. Hyperkinetic Disorder corresponds to the ADHD diagnosis in the *Diagnostic and statistical manual of mental disorders* (DSM-5)

[2]. The Danish National Patient Register was updated to October 2010 and the Danish Psychiatric Central Register to March 2013. Second, we used the Register of Medicinal Product Statistics with information about prescribed ADHD medication in Denmark. This was done to include ADHD patients diagnosed in private psychiatric practices in Denmark, which do not report diagnoses to the patient registries. Children were defined as being diagnosed with ADHD when they had two or more prescriptions of methylphenidate (N06BA04), or atomoxetine (N06BA09) in the register at ≥ 8 years of age. The Register of Medicinal Product Statistics included prescriptions redeemed until 31 December 2011. All children were followed from birth until a diagnosis of ADHD, medication prescription (defined above), death or emigration, whichever came first. The information on death and emigration was obtained from the Civil Registration System, which was updated on 18 October 2010. Of the 943 children with ADHD diagnoses, 79% were boys ($n = 746$) and 21% girls ($n = 197$) ($p < .001$), a ratio of 3.8:1. We made analyses to check for possible gender differences in diagnostic accuracy for all the problem subscale scores, but as the results were not significantly different for boys and girls, we chose to pool the data.

### Strengths and Difficulties Questionnaire (SDQ)

We used the Danish version of parent-reported SDQ [18]. The questionnaire consists of 25 questions rated on a three-point Likert scale (not true, somewhat true and certainly true; range 0–2). There are five SDQ subscales (each with five items) on hyperactivity–inattention (HI), conduct (CD), emotional (ED), peer problems and prosocial behaviour. The first four may be summarized to a total difficulties score. Additionally, an impact supplement of eight questions was provided to the parents. The first impact question asks whether the informant believes the child has a problem. The remaining questions assess chronicity, distress, social impairment and burden for others. The range of possible impact scores is 0–10. The questions on impact inquire about four domains: home life, friendships, classroom learning and leisure activities [31]. The psychometric properties of the SDQ have generally been found to be satisfactory [16, 18, 32, 33], with low correlations between subscales, indicating that they are relatively "uncontaminated" by one another [18].

In the present study, we used the HI subscale along with the other problem subscales: CD, ED and peer problems.

As there are no Danish norms available at present, the HI, CD and ED subscale scores, along with the impact scores, were dichotomized based on part of the prediction algorithm of the British norms (available at http://www.sdqinfo.com) used in a previous Danish study [25]. There is no available algorithm for the peer problem subscale; we were, therefore, not able to do a priori dichotomization in a sound way. Ideally, the SDQ prediction algorithm for hyperactivity disorders (SDQHK) should include information from both parent and teacher. With only the parent SDQ available, and to reduce false positives in the Cox analyses, we computed our parent (p) SDQ prediction categories on the strictest prediction algorithms as follows: hyperactivity disorders (p-SDQHK): HI score ≥ 7 and impact ≥ 2 or HI score ≥ 9 and impact ≥ 1, conduct disorders (p-SDQCD): CD score ≥ 5 and impact ≥ 2, and emotional disorders (p-SDQED): ED score ≥ 5 and impact ≥ 1.

### Demographics

Socio-occupational status was based on the current or most recent job within the last 6 months or on the type of education if the parent was attending school. Three categories were defined: (1) 'High' included parents in management or in jobs requiring higher education. (2) 'Middle' consisted of office workers, service workers, skilled manual workers and parents in the military and (3) 'Low' included unskilled workers and the unemployed. Women who could not be classified in this way (4.1%) were categorized according to their husband's socio-occupational status, defined by the same categories [34]. The frequencies of low and medium socio-occupational status were higher among children with ADHD (8% and 38%, respectively) than the cohort (3% and 28%; $\chi^2 = 164.8$, $p < .001$).

### Ethical approval

The parents provided written informed consent. The Danish Data Protection Agency and the DNBC Steering Committee approved the study.

### Statistical analyses

We used Stata/IC 11.2 for Windows (Stata Corp, College Station, TX) for all analyses. Differences between categorical variables were analysed with Chi square statistics, and differences between means of continuous variables were measured by *t* tests. Intercorrelations between the parent-reported SDQ subscales were calculated using Spearman coefficients because of the skewed distribution. Receiver operating characteristic (ROC) analyses were used to estimate areas under the curve (AUCs) to quantify the diagnostic accuracy. The ROC curve graphically represents the probability of true positive results of ADHD as a function of the probability of false-positive results. To evaluate the AUC values, we used the following guideline: < .70 = poor, .70–.79 = fair; .80–.89 = good; and .90–1.00 = excellent [35].

After checking presumption plots with proportional hazard rates for all comparison groups, we used the Cox

**Table 1** SDQ comparisons of mean scores for children (≥ 8 years of age) with and without ADHD diagnoses and Areas under the curve (AUC) for the SDQ subscales

| | Children with ADHD $n=943$ | | Children without ADHD $n=50,153$ | | $T$ | $p$ | AUC (99% CI) |
|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | | |
| Hyperactivity-inattention | 5.63 | 2.68 | 2.27 | 2.04 | −49.75 | <.001 | .83 (.82–.85) |
| Conduct problems | 2.61 | 1.77 | 1.15 | 1.21 | −36.27 | <.001 | .75 (.73–77) |
| Emotional problems | 2.52 | 2.21 | 1.58 | 1.72 | −16.53 | <.001 | .63 (.61–.65) |
| Peer problems | 2.04 | 2.08 | .68 | 1.19 | −34.32 | <.001 | .71 (.69–.74) |
| Impact scores | 1.39 | 1.96 | .15 | .68 | −51.49 | <.001 | .71 (.69–.74) |

*SDQ* Strengths and Difficulties Questionnaire, *ADHD* attention-deficit/hyperactivity disorder, *SD* standard deviation, *CI* confidence interval

proportional hazard models to analyse the risk for ADHD over time. Each predictor in the Cox regression analysis gives a measure of risk, a hazard ratio according to follow-up time. For a binary variable, a hazard ratio of 1 means that the risk of disorder is the same whether the participant has the characteristic or not. A hazard ratio greater than one indicates increased risk in the observation period for those with the characteristic; while a ratio less than 1, a decreased risk. We conducted the Cox regression analyses with the continuous SDQ subscale scores, as well as categorically (0–1) defined by the parent SDQ prediction categories. In the process of fitting the multivariate models, the predictors were first entered separately in a univariate Cox regression hazard model before entering them in a multivariate model. Adjustment for socio-occupational status was made in our final models. Regression analyses test whether combinations of predictors add significantly to the validity, and thereby complement ROC analyses. The overall risk time was 630,072 person-years. Because of the substantial sample size, the level of significance was set to .01, and the confidence intervals to 99%.

To optimize cutoff scores for clinical decisions, we estimated sensitivity (the probability of a measure to correctly classify a case as positive), specificity (the probability of a measure to correctly identify non-cases as negative), the positive predictive value (PPV, the probability of a true case given a positive test), and the positive likelihood ratio (+LR) (the probability of a child who has the disorder testing positive divided by the probability of a child who does not have the disorder testing positive) for each step of the HI subscale, as well as for an SDQ HI category with lower cutoff scores (HI score ≥ 4 and impact ≥ 1). LRs between 0 and 1 argue against the presence of the disorder; the closer they are to 0, the less likely the disorder. LRs greater than 1 argue for the presence of the disorder, whereas LRs that are equal to 1 lack diagnostic value [36]. LRs are derived from sensitivity and specificity and are independent of the proportion of the disorder in the sample [37]. The negative predictive values and negative LRs were not reported, as the proportion of ADHD in our sample was 1.8% ($n=943/51,096$), meaning that the probability of a true non-case given a negative test would be large at any cutoff in our sample.

## Results

### Background

There were significantly higher mean scores on the HI, CD, ED, peer problems and impact subscales for children with ADHD compared with those without ADHD (Table 1).

Intercorrelation coefficients between the parent-reported SDQ subscale scores were all low to moderate, both for children with and without ADHD (Table 2).

**Table 2** Spearman correlation in the total population for the parent Strengths and Difficulties Questionnaire problem subscales and impact scores

| Variables | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Children with ADHD ($n=943$) | | | | | |
| 1. Hyperactivity–inattention | – | | | | |
| 2. Conduct problems | .43* | – | | | |
| 3. Emotional problems | .22* | .30* | – | | |
| 4. Peer problems | .33* | .31* | .35* | – | |
| 5. Impact scores | .53* | .40* | .44* | .50* | – |
| Children without ADHD ($n=50,153$) | | | | | |
| 1. Hyperactivity–inattention | – | | | | |
| 2. Conduct problems | .40* | – | | | |
| 3. Emotional problems | .17* | .19* | – | | |
| 4. Peer problems | .18* | .19* | .26* | – | |
| 5. Impact scores | .25* | .24* | .26* | .27* | – |

*$p<.05$

## Diagnostic accuracies

Parent HI scores discriminated well between true positive and negative ADHD cases {AUC = .84 [95% confidence interval (CI) = .82–.85]}, and similarly to that of the total difficulties scores (AUC = .84, 95% CI = .82–.85). Please see Fig. 1 for the HI ROC curve.

The AUC values for the other subscale problem and impact scores were all poor to fair (Table 1).

## Risk analyses

Higher HI, CD, ED and peer problem scores all increased the risk of subsequent ADHD diagnosis in the univariate Cox regression analyses. In the final multivariate analysis, for each step of the parent HI scores, the relative risk of subsequent ADHD increased by 1.49, and the CD and peer problem scores increased risk by 1.16 and 1.18, respectively, while the ED scores did not contribute significantly (Table 3).

Using the parent SDQ prediction categories in the regression models, children with high HI score with impact (p-SDQHK = 1) had about a 20-fold increased risk for ADHD (HR = 20.52, CI = 16.83–24.99) in the univariate analysis compared to children without ADHD. The p-SDQHK = 1 gave a nearly 14-fold increased risk of ADHD in the multivariate analysis when the p-SDQCD, p-SDQED and socio-occupational status were included in the model (HR = 13.59 (CI = 10.35–17.82)) (Table 3). The p-SDQCD and p-SDQED gave significant contributions to ADHD in the multivariate regression analysis but much smaller contributions than those of the p-SDQHK category (HRs of 1.62 and 1.67, respectively) (Table 3).

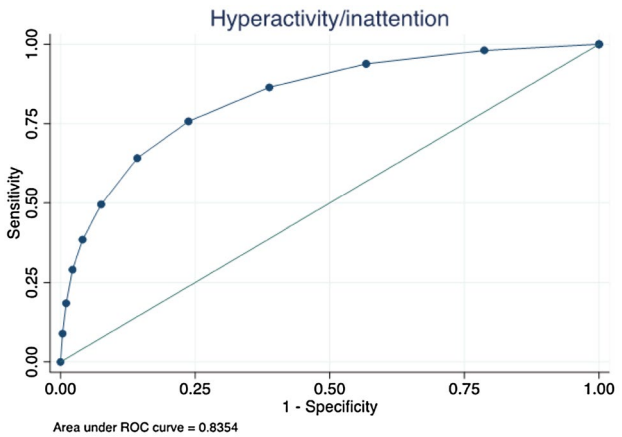


**Fig. 1** The receiver operating characteristic (ROC) curve for the parent SDQ hyperactivity–inattention subscale

## Prediction analyses

For the SDQ HI subscale scores, a cutoff of 7 gave a sensitivity of 38.5% and specificity of 95.9%, with a .09 probability of correctly identifying a child with ADHD given a score above the cut-off (PPV). The best balance between sensitivity and specificity was found at a cutoff score of 4 (sensitivity = 75.7%, specificity = 76.2%), with a PPV of .02 (Table 4).

For the p-SDQHK category, the sensitivity and specificity of identifying ADHD were 24.5% (CI = 21.0–28.3%) and 98.6% (CI = 98.5–98.8%), respectively, with a PPV of .25 (CI = .21–.29) and a + LR of 18.71 (CI = 15.75-22.31). Lowering the cutoff to HI score ≥ 4 and impact ≥ 1 gave sensitivity of 44.4% (CI = 40.8–49.3%), specificity of 95.3%

**Table 3** Univariate and multivariate Cox regression analyses of parent SDQ subscale scores and categories

| Predictors | Univariate | | | | Multivariate* | | | |
|---|---|---|---|---|---|---|---|---|
| | HR | SE | 99% CI | *p* | HR | SE | 99% CI | *p* |
| SDQ scores | | | | | | | | |
| Hyperactivity–inattention | 1.65 | .02 | 1.60– 1.70 | < .001 | 1.49 | .02 | 1.44–1.54 | < .001 |
| Conduct problems | 1.76 | .03 | 1.68–1.83 | < .001 | 1.16 | .02 | 1.09–1.22 | < .001 |
| Emotional problems | 1.28 | .02 | 1.23–1.32 | < .001 | | | | |
| Peer problems | 1.56 | .002 | 1.50–1.62 | < .001 | 1.18 | .02 | 1.13–1.23 | < .001 |
| SDQ categories | | | | | | | | |
| p-SDQHK | 20.52 | 1.55 | 16.83–24.99 | < .001 | 13.59 | 1.41 | 10.35–17.82 | < .001 |
| p-SDQCD | 17.27 | 1.91 | 12.98–22.96 | < .001 | 1.62 | .22 | 1.13–2.31 | < .001 |
| p-SDQED | 6.75 | .62 | 5.33–8.56 | < .001 | 1.67 | .19 | 1.25–2.24 | < .001 |

The parent prediction categories were: hyperactivity-inattention disorders (p-SDQHK): HI score ≥ 7 and impact ≥ 2 or HI score ≥ 9 and impact ≥ 1; parent conduct disorders (p-SDQCD): CD score ≥ 5 and impact ≥ 2, and parent emotional disorders (p-SDQED): ED score ≥ 5 and impact ≥ 1

*HR* Hazard ratio, *SE* standard error, *CI* confidence interval, *SOC status* socio-occupational status

*Adjusted for socio-occupational status

**Table 4** Sensitivity (Se), specificity (Sp), positive predictive values (PPVs) and positive likelihood ratios (+LR) for the Strengths and Difficulties Questionnaire hyperactivity–inattention subscale (SDQ HI) scores

| SDQ HI score | Se (%) | Sp (%) | PPV | +LR (99% CI) | *N* with ADHD |
|---|---|---|---|---|---|
| 1 | 98.1 | 21.3 | .004 | 1.25 (1.23–1.26) | 925 |
| 2 | 93.9 | 43.3 | .01 | 1.65 (1.61–1.69) | 885 |
| 3 | 86.4 | 61.2 | .02 | 2.23 (2.15–2.31) | 815 |
| 4 | 75.7 | 76.2 | .02 | 3.21 (3.05–3.38) | 714 |
| 5 | 64.2 | 85.8 | .04 | 4.54 (4.24–4.86) | 605 |
| 6 | 49.5 | 92.5 | .06 | 6.65 (6.06–7.29) | 467 |
| 7 | 38.5 | 95.9 | .09 | 9.47 (8.36–10.64) | 363 |
| 8 | 29.1 | 97.8 | .14 | 13.24 (11.37–15.38) | 274 |
| 9 | 18.5 | 99.0 | .21 | 17.56 (14.11–21.78) | 174 |
| 10 | 8.9 | 99.6 | .31 | 23.39 (16.45–32.73) | 84 |

The proportion of children with ADHD diagnosis (≥8 years of age) was 1.8%

*CI* Confidence interval

(CI = 95.2–95.6%), a PPV of .15 (CI = .14–.17) and +LR of 9.75 (CI = 8.80–10.89).

## Discussion

The main findings of our prospective cohort study support the parent SDQ hyperactivity–inattention (HI) subscale at the child age of 7 years as a valid measure for predicting clinically diagnosed ADHD. This is in line with earlier studies that have reported the SDQ to be a valid tool to discriminate psychopathology [16], including ADHD [17]. The accuracy level of the parent HI subscale in the present study was good, and consistent with earlier studies of schoolchildren that have reported good to excellent screening accuracy [15, 20]. It is noteworthy that the brief HI subscale and the total difficulties scores performed similarly well (both AUC values were .84), a finding consistent with Algorta et al.'s study [17].

The consistently high AUCs for the SDQ in cross-sectional community samples may in part be due to few children with ADHD being compared with many healthy children, a known problem discussed previously [17, 38]. That we did not find excellent accuracy in our study might be because of the longitudinal design ensured by excluding children diagnosed before the age of 8 years, as this will have weakened our screening accuracy compared with cross-sectional studies [17, 20]. This may also be because our registered ADHD diagnoses were made by clinicians, as the inter-rater agreement between clinical diagnoses and research diagnosis made by diagnostic interviews are known to be moderate (kappa = .49) [39]. By use of national patient registries to obtain ADHD diagnoses we ensured a large sample, but at the cost of lower diagnostic accuracy compared to studies that use parent diagnostic interviews. However, with clinical diagnoses from registries we probably reduced the problem of same-rater bias which possibly affect studies that compare parent information from a rating scale with a diagnostic interview.

For each step of the SDQ HI subscale, we found significantly increased risk for subsequent ADHD. When including high HI subscale scores and impact as defined by the parent SDQ prediction category for hyperactivity disorders (p-SDQHK), we found about a 20-fold increased risk of subsequent ADHD in univariate analysis. This was a similar result to a study where combined high parent and teacher SDQ HI subscale scores and impact (as defined by the SDQHK prediction algorithm) at the child age of 5–7 years gave an approximately 20-fold increased risk of subsequent ADHD diagnoses [25]. The same study found that the SDQ prediction algorithm for 'any probable diagnosis' (http://www.sdqinfo.com) showed screening properties for ADHD that were very similar to those of the HI subscale, which was taken to indicate that psychopathology outside the core symptoms of ADHD constitutes an increased risk of later ADHD [25]. Although we found that both the SDQ CD and ED prediction categories significantly predicted ADHD, the p-SDQHK was clearly the strongest predictor in the multivariate analyses. This was in accordance with a large population study that found that a clinical diagnosis of ADHD was most strongly predicted by ADHD symptoms assessed by screening scales at 3 and 5 years of age (odds ratios of 3.23 and 10.30, respectively) [40].

Predictive validity studies of the SDQ are scarce, but other community studies have reported similar findings to ours. One study followed SDQ HI scores over time and found moderate stability ($r = .77$) over a 12-month period [23]. Another study reported that the SDQ HI subscale scores at 3 years of age significantly predicted ADHD 2 years later [24]. In addition, a third longitudinal community study that did not investigate ADHD specifically, but considered 'any psychiatric disorder', from 4 to 6 years of age, found that the SDQ total scores discriminated well between any disorder and no disorder (AUC = .85) [41].

In the present study, we found no cutoff that gave satisfactory values for sensitivity, specificity and PPV. An SDQ HI cutoff score of 7 gave a sensitivity of 38%, specificity of 96% and PPV of .09, meaning that many children with ADHD would be missed. Lowering the HI cutoff level to 4 gave a good balance between sensitivity (76%) and specificity (76%), but increased the number of false positive diagnoses, resulting in a very low probability of correctly identifying a child with ADHD (PPV = .02). The low proportion of children with ADHD in our sample (1.8%) increased the risk of causing concern without reason when lowering the cutoff level. Knowing the proportion of the disorder in the population being screened is important as it particularly affects the positive and negative predictive values. In theory, however, sensitivity, specificity and likelihood ratios generalize across samples. A clinical study with a 54.1% proportion of children with ADHD reported a sensitivity level of 57% for detecting ADHD with an SDQ HI cutoff ≥ 7 [15], whereas a community study with a 5% proportion of ADHD (326 with ADHD out of 6233 children) reported sensitivity and specificity similar to our study, and somewhat higher PPVs (.17 at cut-off 4, and .44 at cut-off 7) [20]. Combining high HI scores and impact (p-SDQHK) improved the PPV (.25) in our study but gave low sensitivity. However, the PPV was only somewhat lower than in a study that combined high parent and teacher HI scores and impact (SDQHK algorithm) in a population with an ADHD proportion of 2.9% (PPV = .33) [25]. This raises the question of the feasibility of routine screening for ADHD with the SDQ HI subscale in the general population. Our results indicate that screening with the parent HI subscale should be limited to groups where concern has been raised by parents or teachers.

Strengths of the present study were the large cohort sample, the longitudinal design and the use of clinical data from Danish registers on diagnoses and prescriptions of central stimulants. With data from registers, only death and migration cause attrition. However, there were several limitations. One was attrition to the DNBC cohort, where a previous study found that the cohort is not representative in terms of socio-economic factors [42]. However, a study of the representativeness of childhood psychiatric diagnoses reported that children with a registered ADHD diagnosis were only modestly underrepresented in the cohort, and that children using ADHD medication were present in the DNBC to the same extent as in the general population [43]. Another limitation was the use of registered clinical ADHD diagnoses, as diagnosis and medication use depend on cultural differences in both the referral to psychiatric facilities and the diagnostic practices among child psychiatrists [44]. Although the Danish clinicians diagnose by the ICD-10, we do not have information about the methodology used to reach diagnostic conclusions. However, a validity study of hyperkinetic disorder (HD) in the Danish Psychiatric Central Research Registry (for patients aged 4–15) confirmed that HD was present in 86.8% of the patients by going through their patient records [45]. That study concluded that for most HD diagnoses, the assessment forming the basis for diagnosis reflected a thorough and high-quality psychiatric assessment based on a multi-informant approach. Unfortunately, we did not have information about other diagnoses than ADHD and could not investigate discriminant validity. This lack of diagnostic information meant that we could not consider comorbidity to ADHD, although one recent patient registry study found low proportions of registered comorbidity [46]. Nonetheless, unregistered comorbidity may have given rise to some false-positive cases in our study. We did not obtain ratings from teachers and could, therefore, not apply the SDQ prediction algorithms as intended by the developers. This may limit the predictive values. However, not all studies have found improved sensitivity and specificity by combining parent and teacher SDQ ratings [41, 47]. In addition, we do not have information about those children treated in the private practices that did not redeem medication. Hence, we may have missed some children in our analyses, which may have weakened our findings to some degree. Finally, we did not have available a screening measure with the full DSM ADHD symptom list. Therefore, we could not rule out that screening with an ADHD symptom list might have been more accurate than the short SDQ HI subscale. However, one study that defined ADHD by a diagnostic interview and compared the screening properties of the SDQ and an ADHD symptom list found that SDQ was a more sensitive measure than the ADHD symptom list (68% versus 54%), while the specificity was good to excellent for the two questionnaires (88% versus 95%) [48].

In sum, the parent SDQ HI subscale scores discriminated well between children with and without ADHD and gave a higher risk for subsequent ADHD than the SDQ conduct, emotional and peer problem scores. Although the sensitivity analyses were similar to earlier community studies with increased sensitivity at lower scores (≥ 4), they resulted in many false screening positive diagnoses. Including impact in addition to obtaining high HI scores reduced the number of false positive predictions, but this rate would still be a problem if the SDQ HI subscale was to be used in the general population with a low proportion of ADHD. Thus, our results raise the question of the feasibility of screening the general child population for ADHD with only the parent SDQ HI subscale. However, in a clinical setting, where concerned parents ask whether their child might have ADHD, the proportion of children with the disorder is expected to be higher than in the general community. In such a setting, the parent SDQ HI ratings may be more useful for correctly identifying children with ADHD, although a clinical assessment is always needed.

**Acknowledgements** The Danish National Research Foundation established the Danish Epidemiology Science Centre, which initiated and created the Danish National Birth Cohort. The cohort is also a result of a major grant from the Danish National Research Foundation. Additional support for the Danish National Birth Cohort was obtained from the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Augustinus Foundation, and the Health Foundation.

## Compliance with ethical standards

**Conflict of interest** The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. Professor emeritus Svein Friis has received an honorarium as a data consultant for RAND Corporation for a project sponsored by the Janssen-Cilag pharmaceutical company.

## References

1. Thapar A, Cooper M (2016) Attention deficit hyperactivity disorder. Lancet 387(10024):1240–1250. https://doi.org/10.1016/S0140-6736(15)00238-X
2. American Psychiatric Association (APA) (2013) Diagnostic and statistical manual of mental disorders: DSM-5. APA, Washington, DC
3. Polanczyk G, Rohde LA (2007) Epidemiology of attention-deficit/hyperactivity disorder across the lifespan. Curr Opin Psychiatry 20(4):386–392. https://doi.org/10.1097/YCO.0b013e3281568d7a
4. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA (2015) Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry 56(3):345–365. https://doi.org/10.1111/jcpp.12381
5. Kraut AA, Langner I, Lindemann C et al (2013) Comorbidities in ADHD children treated with methylphenidate: a database study. BMC Psychiatry 13:11. https://doi.org/10.1186/1471-244X-13-11
6. Ghanizadeh A (2009) Psychiatric comorbidity differences in clinic-referred children and adolescents with ADHD according to the subtypes and gender. J Child Neurol 24(6):679–684. https://doi.org/10.1177/0883073808331086
7. Steinhausen HC, Novik TS, Baldursson G et al (2006) Co-existing psychiatric problems in ADHD in the ADORE cohort. Eur Child Adolesc Psychiatry 15(Suppl 1):I25–I29. https://doi.org/10.1007/s00787-006-1004-y
8. Pliszka SR (2003) Psychiatric comorbidities in children with attention deficit hyperactivity disorder: implications for management. Paediatr Drugs 5(11):741–750. https://doi.org/10.2165/00148581-200305110-00003
9. Chorozoglou M, Smith E, Koerting J, Thompson MJ, Sayal K, Sonuga-Barke EJ (2015) Preschool hyperactivity is associated with long-term economic burden: evidence from a longitudinal health economic analysis of costs incurred across childhood, adolescence and young adulthood. J Child Psychol Psychiatry 56(9):966–975. https://doi.org/10.1111/jcpp.12437
10. Dalsgaard S, Ostergaard SD, Leckman JF, Mortensen PB, Pedersen MG (2015) Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. Lancet 385(9983):2190–2196. https://doi.org/10.1016/S0140-6736(14)61684-6
11. Dalsgaard S, Leckman JF, Mortensen PB, Nielsen HS, Simonsen M (2015) Effect of drugs on the risk of injuries in children with attention deficit hyperactivity disorder: a prospective cohort study. Lancet Psychiatry 2(8):702–709. https://doi.org/10.1016/s2215-0366(15)00271-0
12. Erskine HE, Norman RE, Ferrari AJ et al (2016) Long-term outcomes of attention-deficit/hyperactivity disorder and conduct disorder: a systematic review and meta-analysis. J Am Acad Child Adolesc Psychiatry 55(10):841–850. https://doi.org/10.1016/j.jaac.2016.06.016
13. Lee HJ, Cho S, Shin MS (2008) Supporting diagnosis of attention-deficit hyperactive disorder with novelty detection. Artif Intell Med 42(3):199–212. https://doi.org/10.1016/j.artmed.2007.11.001
14. Becker A, Woerner W, Hasselhorn M, Banaschewski T, Rothenberger A (2004) Validation of the parent and teacher SDQ in a clinical sample. Eur Child Adolesc Psychiatry 13(Suppl 2):ii11–ii16. https://doi.org/10.1007/s00787-004-2003-5
15. Carballo JJ, Rodriguez-Blanco L, Garcia-Nieto R, Baca-Garcia E (2014) Screening for the ADHD phenotype using the Strengths and Difficulties Questionnaire in a clinical sample of newly referred children and adolescents. J Atten Disord. https://doi.org/10.1177/1087054714561858
16. Stone LL, Otten R, Engels RC, Vermulst AA, Janssens JM (2010) Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. Clin Child Fam Psychol Rev 13(3):254–274. https://doi.org/10.1007/s10567-010-0071-2
17. Algorta GP, Dodd AL, Stringaris A, Youngstrom EA (2016) Diagnostic efficiency of the SDQ for parents to identify ADHD in the UK: a ROC analysis. Eur Child Adolesc Psychiatry 25(9):949–957. https://doi.org/10.1007/s00787-015-0815-0
18. Goodman R (2001) Psychometric properties of the Strengths and Difficulties Questionnaire. J Am Acad Child Adolesc Psychiatry 40(11):1337–1345. https://doi.org/10.1097/00004583-200111000-00015
19. Sheldrick RC, Garfinkel D (2017) Is a positive developmental-behavioral screening score sufficient to justify referral? A review of evidence and theory. Acad Pediatr 17(5):464–470. https://doi.org/10.1016/j.acap.2017.01.016
20. Ullebo AK, Posserud MB, Heiervang E, Gillberg C, Obel C (2011) Screening for the attention deficit hyperactivity disorder phenotype using the Strength and Difficulties Questionnaire. Eur Child Adolesc Psychiatry 20(9):451–458. https://doi.org/10.1007/s00787-011-0198-9
21. Cuffe SP, Moore CG, McKeown R (2009) ADHD and health services utilization in the National Health Interview Survey. J Atten Disord 12(4):330–340. https://doi.org/10.1177/1087054708323248
22. Dopfner M, Breuer D, Wille N, Erhart M, Ravens-Sieberer U, BELLA Study Group (2008) How often do children meet ICD-10/DSM-IV criteria of attention deficit-/hyperactivity disorder and hyperkinetic disorder? Parent-based prevalence rates in a national sample—results of the BELLA study. Eur Child Adolesc Psychiatry 17(Suppl 1):59–70. https://doi.org/10.1007/s00787-008-1007-y
23. Hawes DJ, Dadds MR (2004) Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. Aust NZ J Psychiatry 38(8):644–651. https://doi.org/10.1080/j.1440-1614.2004.01427.x
24. Croft S, Stride C, Maughan B, Rowe R (2015) Validity of the Strengths and Difficulties Questionnaire in preschool-aged children. Pediatrics 135(5):e1210–e1219. https://doi.org/10.1542/peds.2014-2920
25. Rimvall MK, Elberling H, Rask CU, Helenius D, Skovgaard AM, Jeppesen P (2014) Predicting ADHD in school age when using the Strengths and Difficulties Questionnaire in preschool age: a longitudinal general population study, CCC2000. Eur Child Adolesc Psychiatry 23(11):1051–1060. https://doi.org/10.1007/s00787-014-0546-7

26. Olsen J, Melbye M, Olsen SF et al (2001) The Danish National Birth Cohort—its background, structure and aim. Scand J Public Health 29(4):300–307. https://doi.org/10.1177/14034948010290040201
27. Andersen TF, Madsen M, Jorgensen J, Mellemkjoer L, Olsen JH (1999) The Danish National Hospital Register. A valuable source of data for modern health sciences. Dan Med Bull 46(3):263–268
28. Munk-Jorgensen P, Mortensen PB (1997) The Danish Psychiatric Central Register. Dan Med Bull 44(1):82–84
29. Kildemoes HW, Sorensen HT, Hallas J (2011) The Danish National Prescription Registry. Scand J Public Health 39(7 Suppl):38–41. https://doi.org/10.1177/1403494810394717
30. World Health Organization (WHO) (1990) International classification of diseases and related health problems (ICD-10), 10th edn. WHO, Geneva
31. Goodman R (1999) The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. J Child Psychol Psychiatry 40(5):791–799
32. Niclasen J, Teasdale TW, Andersen AM, Skovgaard AM, Elberling H, Obel C (2012) Psychometric properties of the Danish Strength and Difficulties Questionnaire: the SDQ assessed for more than 70,000 raters in four different cohorts. PLoS One 7(2):e32025. https://doi.org/10.1371/journal.pone.0032025
33. Niclasen J, Skovgaard AM, Andersen AM, Somhovd MJ, Obel C (2013) A confirmatory approach to examining the factor structure of the Strengths and Difficulties Questionnaire (SDQ): a large scale cohort study. J Abnorm Child Psychol 41(3):355–365. https://doi.org/10.1007/s10802-012-9683-y
34. Nohr EA, Bech BH, Vaeth M, Rasmussen KM, Henriksen TB, Olsen J (2007) Obesity, gestational weight gain and preterm birth: a study within the Danish National Birth Cohort. Paediatr Perinat Epidemiol 21(1):5–14. https://doi.org/10.1111/j.1365-3016.2007.00762.x
35. Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240(4857):1285–1293
36. McGee S (2002) Simplifying likelihood ratios. J Gen Intern Med 17(8):646–649
37. Youngstrom EA (2014) A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. J Pediatr Psychol 39(2):204–221. https://doi.org/10.1093/jpepsy/jst062
38. Youngstrom E, Meyers O, Youngstrom JK, Calabrese JR, Findling RL (2006) Comparing the effects of sampling designs on the diagnostic accuracy of eight promising screening algorithms for pediatric bipolar disorder. Biol Psychiatry 60(9):1013–1019. https://doi.org/10.1016/j.biopsych.2006.06.023
39. Rettew DC, Lynch AD, Achenbach TM, Dumenci L, Ivanova MY (2009) Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. Int J Methods Psychiatr Res 18(3):169–184. https://doi.org/10.1002/mpr.289
40. Oerbeck B, Overgaard KR, Pripp AH, Reichborn-Kjennerud T, Aase H, Zeiner P (2017) Early predictors of ADHD: evidence from a prospective birth cohort. J Atten Disord. https://doi.org/10.1177/1087054717696765
41. Hamre Sveen T, Berg-Nielsen TS, Lydersen S, Wichstrom L (2016) Screening for persistent psychopathology in 4-year-old children. Pediatrics 138(4):e20151648. https://doi.org/10.1542/peds.2015-1648
42. Jacobsen TN, Nohr EA, Frydenberg M (2010) Selection by socioeconomic factors into the Danish National Birth Cohort. Eur J Epidemiol 25(5):349–355. https://doi.org/10.1007/s10654-010-9448-2
43. Madsen KB, Hohwu L, Zhu JL, Olsen J, Obel C (2017) Social selection in cohort studies and later representation of childhood psychiatric diagnoses: the Danish National Birth Cohort. Scand J Public Health. https://doi.org/10.1177/1403494817726619
44. Madsen KB, Ersboll AK, Olsen J, Parner E, Obel C (2015) Geographic analysis of the variation in the incidence of ADHD in a country with free access to healthcare: a Danish cohort study. Int J Health Geogr 14:24. https://doi.org/10.1186/s12942-015-0018-4
45. Mohr-Jensen C, Koch SV, Lauritsen BM, Steinhausen H-C (2016) The validity and reliability of the diagnosis of hyperkinetic disorders in the Danish Psychiatric Central Research Registry. Eur Psychiatry 35:16–24. https://doi.org/10.1016/j.eurpsy.2016.01.2427
46. Oerbeck B et al (2017) ADHD, comorbid disorders and psychosocial functioning: how representative is a child cohort study? Findings from a national patient registry. BMC Psychiatry 17(1):23. https://doi.org/10.1186/s12888-017-1204-7
47. Overgaard KR, Oerbeck B, Friis S, Pripp AH, Biele G, Aase H, Zeiner P (2018) Attention-deficit/hyperactivity disorder in preschoolers—the accuracy of a short screener. J Am Acad Child Adolesc Psychiatry 57(6):428–435. https://doi.org/10.1016/j.jaac.2018.03.008
48. Posserud MB, Ullebo AK, Plessen KJ, Stormark KM, Gillberg C, Lundervold AJ (2014) Influence of assessment instrument on ADHD diagnosis. Eur Child Adolesc Psychiatry 23(4):197–205. https://doi.org/10.1007/s00787-013-0442-6