

## Are all the 18 DSM-IV and DSM-5 criteria equally useful for diagnosing ADHD and predicting comorbid conduct problems?

Garcia Rosales, A., Vitoratou, S., Banaschewski, T., Asherson, P., Buitelaar, J. K., Oades, R. D., Rothenberger, A., Steinhausen, H-C., Faraone, S. V., & Chen, W.  
**2015 European Child and Adolescent Psychiatry, 24, 1325-1337**

This is the reformatted manuscript submitted prior to publication in its final form  
DOI 10.1007/s00787-015-0683-7

Correspondence: A.G.R. MRC Social Genetic Developmental and Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK rosales.garcia@kcl.ac.uk

### Abstract

In view of ICD-11 revision, we evaluate whether the 18 DSM-IV diagnostic items retained by DSM-5 could be further improved (i) in predicting ADHD 'case-ness' and impairment' and (ii) discriminating ADHD without CD (ADHD – CD) cases from ADHD with CD (ADHD + CD) cases.

In a multi-centre study sample consisting of 1497 ADHD probands and 291 unaffected subjects, 18 diagnostic items were examined for redundancy; then each item was evaluated for association with case-ness, impairment and CD status using Classical Test Theory, Item-Response Theory and logistic regression methods.

First, all 18 DSM-IV items contributed significantly and independently to the clinical diagnosis of ADHD. Second, not all the DSM-IV items carried equal weighting. "Often loses things", "forgetfulness" and "difficulty sustaining attention" mark severity for Inattentiveness (IA) items and "often unduly noisy", "exhibits a persistent pattern of restlessness", "leaves seat in class" and "often blurts out answers" for Hyperactivity/Impulsivity (HI) items. "Easily distracted", "inattentive to careless mistakes", "often interrupts" and "often fidgets" are associated with milder presentations. In the IA domain, "distracted" yields most information in the low-severity range of the latent trait, "careless" in the mid-severity range and "loses" in the high-severity range. In the HI domains, "interrupts" yields most information in the low-severity range and "motor" in the high-severity range. Third, all 18 items predicted impairment. Fourth, specific ADHD items are associated with ADHD + CD status.

The DSM-IV diagnostic items were valid and not redundant; however, some carried more weight than others. All items were associated with impairment.

### Keywords

Attention-deficit/hyperactivity disorder, Diagnostic criteria, DSM-IV, Conduct disorder Impairment

### Introduction

The 18 diagnostic items for attention deficit/hyperactivity disorder (AD/HD) in the fourth revision of the Diagnostic and Statistical Manual of the American Psychiatric Association (DSM-IV) have been retained by DSM-5 in their recent revision (2013). DSM-5, however, has elaborated some items with examples and

incorporated a revised age of onset criterion before the age of 12 years old, so as to improve the diagnosis of older adolescents and adults. The items were initially selected on the results of published literature and the field trials conducted specifically for the DSM-IV revision process (APA 1994). Considerable clinical experience and research evidence now support their use: they have yielded

reliable and consistent prevalence estimates across different national and cultural settings and have shown evidence of robust association with neurobiological measures [1]. There have, however, been suggestions from clinicians that the criteria might be improved for routine clinical use [2].

A number of issues with these items remain under-investigated, and the answers to these may potentially improve diagnostic efficiency and are also relevant in view of the impending ICD-11 revision.

First, the number of items is rather large for routine use by busy clinicians. Each of the 18 items requires separate evaluation and the time involved may make it less feasible to assess other aspects of a case. Indeed, clinicians could save time by a shorter list if there were inter-item redundancy. Item redundancy could arise from the wording of items, so that one item might depict the same underlying phenomenon as another. For example: *“often has difficulty sustaining attention in tasks or play activities”* might describe the same behaviour as *“often has difficulty organizing tasks and activities”*. If some items do indeed describe the same behaviour, then that single latent behavioural component might become over-represented and so bias the diagnosis. If so, a shortened list of criteria could be developed with correspondingly greater efficiency in the diagnostic process.

Second, impairment of function across different settings (e.g. home, school, leisure activities) is a necessary requirement for a DSM-5 diagnosis.

There is little doubt that ADHD symptoms can be impairing when present above a certain severity threshold. For example, in a study [3], Bussing et al. measured both symptoms and impairment and found that parents and patients diagnosed with childhood ADHD reported more functional impairment and lower quality of life than their healthy or

subthreshold ADHD counterparts. Similarly, Ramtekka et al. [4] found that patients who met the diagnostic threshold were substantially more impaired by their symptoms over their lifetime. It is not, however, clear whether all symptoms are related to impairment or whether any have particular importance for overall adjustment.

Third, the DSM-IV field trials provided symptom utility estimates for selection of diagnostic criteria of ADHD [5]. The selected items differed in positive and negative predictive powers, despite showing acceptable moderate to high values. The “often does not seem to listen when spoken directly” (“listen”) item, being an exception, has the lowest positive predictive power. Willcutt et al. [6] identified further evidence for “listen” and “is often easily distracted” (“distracted”) being weak items. Wolraich et al. [7], however, found “distracted” showing high item–total correlation, instead of “often blurts out an answer before a question has been completed” (“blurts”) and “often talks excessively” (“talks”) showing particularly low item–total correlations. Symptom utility of ADHD diagnostic criteria warrants further evaluation in the light of the DSM-5 publication.

Furthermore, both DSM-IV and DSM-5 assume equal weighting of 18 items with a diagnostic threshold set at the additive sum of items present. Alternatively, some items may be more discriminating or indexing greater severity in latent ADHD traits. We therefore evaluate for differential discriminating properties of the 18 items and whether specific candidates for a reduced list could be identified. The relationship between impairment and differential weighting of symptom severity as estimated by Item Response Theory (IRT) and impairment remains unexamined in the literature.

Finally, ADHD is frequently comorbid

with oppositional and conduct problems [8]. Furthermore, it may be helpful to the clinician to be aware of which ADHD symptoms are least confounded with CD when assessing ADHD + CD cases. One possible reason for the co-existence of these conditions is that the wording of criteria for each fails to differentiate their core expression properly and therefore creates overlap, arising from linguistic blurring artefact [9]. In the case of ADHD, particular attention needs to be paid to any item that potentially shares features with CD and signifies the presence of contamination (if the item concerned cross-loads onto both ADHD and the comorbid disorders). However, in a study by Newcorn et al. [10], patients with ADHD and comorbid oppositional defiant disorder/conduct disorder scored higher on impulsivity than their pure ADHD counterparts.

This was consistent with previous studies [11–13]. Nevertheless, there is evidence to indicate that the assumption of ADHD as the precursor of CD is largely an artefact, as Lahey et al. [14] showed that as once the genetic and environmental influences on conduct problems in early childhood were controlled, the predictive power of early ADHD symptoms on later conduct problems was significantly attenuated.

Alternatively, certain ADHD items may index the severity or specificity of ADHD (as the number of items increases the likelihood of the comorbid condition) or implicate a distinct familial subtype, as supported by previous findings by Christiansen et al. [15], or capture the presence of pathogenic substrates indexing aggression and anti-social tendency, reflecting a shared aetiology between conditions than other symptoms. Identifying these items can stimulate further research in elucidating the complex relationships of ADHD and CD association.

The objective of this study was to

evaluate the 18 items/symptoms in relation to their (a) *properties* and (b) *predictability* with respect to ADHD, impairment (secondary to ADHD defined by the perception of decrease in function as determined by parents/carers) and comorbid CD diagnosis. The item properties refer to (i) reliability, item redundancy and information provided (either by each item or in total), (ii) ability to discriminate between individuals with low and high ADHD levels and (iii) the assessment of each symptom's severity. With the term *predictability*, we refer here to the extent to which each item/symptom (i) contributes to a positive diagnosis of ADHD, (ii) increases the odds of impairment and/or (iii) associates with a comorbid CD diagnosis. Key questions for considering the need for taxonomic revisions are therefore:

(1) Are there any redundant or non-informative items for the diagnosis of ADHD? Are there DSM-5 items that are more discriminatory than others?

(2) Are there any items that are not associated with ADHD case-ness? Which items contribute the most?

(3) Which items contribute towards the parents' perception of impairment?

(4) Are there any items that are blurring the distinction between 'pure ADHD' and 'ADHD with comorbid CD'? That is, are there any distinctive diagnostic items between 'ADHD-only' and 'ADHD + CD'?

## Materials and methods

### Sample

The *International Multicentre AD/HD Genetics* (IMAGE) project is an international collaborative study funded by National Institute of Mental Health (NIMH) to identify candidate genes, genetic markers and quantitative trait loci associated with AD/HD. The sample consisted of European Caucasian subjects recruited from 12 specialist host centres in eight countries:

Belgium, Germany, Holland, Ireland, Israel, Spain, Switzerland and the UK. Ethical approval for the study was obtained from ethical review boards within each country and informed consent obtained from children and families. All children (probands and siblings) were aged 5–17 years, have  $IQ \geq 70$ , of European Caucasian descent and with access to at least one biological parent for DNA collection. Exclusion criteria applying to both probands and siblings included autism, epilepsy,  $IQ < 70$ , brain disorders and any genetic or medical disorder associated with externalising behaviours that might mimic ADHD. A detailed description of the study design and diagnostic approach has been published elsewhere [16–18]. For this analysis, we included individuals with complete data from the Parental Account of Childhood Symptoms (PACS) interview and the Conners Parent and Teacher Rating Scales. ADHD case-ness was classified according to the IMAGE study algorithm (available upon request) and the Hypescheme (available upon request) [19]. These tools have allowed us to translate the information collected by PACS into individual DSM-IV items. These tools were developed, like other detailed interview schedules, to embody the DSM criteria in operational and behavioural terms, rather than using specifically the language of the DSM-IV. The validity of the PACS items is described by Chen and Taylor [20]. Hypescheme and the IMAGE algorithms were used to yield each DSM-IV criterion. Each DSM criterion was coded as a dichotomous variable (0 for absent, 1 for present) and a categorical diagnosis of ADHD was also recorded [19]. The Hypescheme is an operational criteria checklist and minimum dataset for the research diagnoses of ADHD. It consists of a computerised data system requiring completion of 111 items and an algorithm designed to generate all possible diag-

noses of attention-deficit hyperactivity disorders under the DSM-IV classification system.

## **Instruments**

### *PACS*

PACS [20] is a standardised investigator-based research diagnostic interview, designed to capture accurately and systematically the clinical phenotypes relating to ADHD (and hyperkinetic disorder) and other childhood psychiatric disorders. PACS is a semi-structured interview, which elicits specific descriptions of behaviours by probing specific social and task-related situations, rather than ascertaining global impressions that yield overall ratings. The PACS items do not have the same wording as the DSM, and PACS is an attempt to operationalise the DSM criteria. The interviewers elicit examples of problem behaviours, but make their own judgements and ratings of severity and frequency. For each item, a four-point rating of problem severity is generated to give a metric score range from 0 to 3. The frequency of problems is coded separately, but again on the same metric scale with defined cut-off thresholds. The structure and administration techniques of PACS have been designed to minimise opportunities for introducing responders' and raters' bias. PACS inter-rater reliability [20] has been reported as adequate in several samples and reliability checks have been maintained during the project. For the IMAGE project, PACS was translated into German, Dutch, Hebrew and Spanish. 'Impairment' was defined using a variable in PACS ascertaining parental global report of functional impairments attributable to the symptom domain of inattention and the symptom domain of hyperactivity/impulsivity, distinct from the symptom severity ascertained by PACS. We summed the two impairment items to yield a

summary variable. A child with a score of 2 or above in this summary variable (range 0–4) is designated as ‘impaired’.

In the IMAGE project, all probands were assessed by PACS, and only a proportion of siblings were assessed by PACS in accordance with surplus resource available to local teams. In this study, only individuals assessed by PACS (including both probands and siblings) were selected for analysis; there was a preponderance of probands in this subsample.

The diagnosis of CD was derived from a PACS algorithm (available upon request). The CD variable includes children and young people only diagnosed with CD (i.e. not with oppositional defiant disorder only).

#### *Item abbreviations*

The abbreviations of DSM items used in this report are adapted from those used in the DSM-IV field trial (Frick et al. [5]) and listed in *italic* as follows:

#### *Inattention items*

- (1) “Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or with other activities” (*Careless*).
- (2) “Often has trouble holding attention on tasks or play activities” (*Attention*).
- (3) “Often does not seem to listen when spoken to directly” (*Listen*).
- (4) “Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g. loses focus, side-tracked)” (*Instructions*).
- (5) “Often has trouble organizing tasks and activities” (*Disorganised*).
- (6) “Often avoids, dislikes, or is reluctant to do tasks that require mental effort over a long period of time (such as schoolwork or homework)” (*Unmotivated*).
- (7) “Often loses things necessary for tasks and activities (e.g. school materials, pencils, books, tools, wallets, keys, paper-

work, eyeglasses, and mobile telephones)” (*Loses*).

(8) “Is often easily distracted” (*Distracted*).

(9) “Is often forgetful in daily activities” (*Forgetful*).

#### *Hyperactivity items*

“Often fidgets with or taps hands or feet, or squirms in seat” (*Fidgets*). “Often leaves seat in situations when remaining seated is expected” (*Seat*). “Often runs about or climbs in situations where it is not appropriate (adolescents or adults may be limited to feeling restless)” (*Runs/climbs*). “Often unable to play or take part in leisure activities quietly” (*Quiet*). “Is often “on the go” acting as if “driven by a motor”” (*Motor*). “Often talks excessively” (*Talks*). “Often blurts out an answer before a question has been completed” (*Blurts*). “Often has trouble waiting his/her turn” (*Wait*). “Often interrupts or intrudes on others (e.g. butts into conversations or games)” (*Interrupts*).

#### **Statistical analyses**

##### *Item properties*

The reliability of the assessment was evaluated for the whole sample (including both probands and siblings) via the Kuder–Richardson (KR-20) [21] formula (*internal consistency* coefficient for binary items), according to the classical test theory (CTT).

Item response theory methods were then also applied to the whole sample (including both probands and siblings). IRT models were fitted for inattentiveness (IA) and hyperactivity–impulsiveness (HI). As opposed to the CTT, the IRT methodology allows for the reliability assessment at any desired point on the latent trait spectrum (local reliability), either for each item and/or for the total set of items (*item and test information*, respectively, according to the IRT terminology) instead of the CTT overall assessment in cases and controls. That is, in IRT the reliability is assessed as

a function of the trait and is typically depicted via the item information curves (IIC; for each item separately) and the total information curve (TIC; for the total set of items). The IRT model also provides information regarding the ability of each item to discriminate between persons with different levels of the trait as well as in terms of symptom severity. In particular, the two-parameter logistic IRT model (2PL) was employed here, according to which the probability of a symptom  $j$  to be present, for a person with the latent trait  $\theta$ , is given by

$$P_j(\theta) = \frac{\exp[\alpha_j(\theta - \beta_j)]}{1 + \exp[\alpha_j(\theta - \beta_j)]}$$

where  $\theta$  stands either for IA or HI,  $\alpha_j$  is the *discrimination* parameter and  $\beta_j$  is the *difficulty* parameter. This probability can be plotted over the origin of the trait  $\theta$  (where  $\theta$  is a latent variable, that is, not directly observable but measurable via the symptoms endorsement patterns of the studied sample) and is assumed a priori to follow a normal distribution (with zero mean and standard deviation equal to one). The resulting curve is referred to as the item characteristic curve (ICC). The discrimination parameter corresponds to the slope of ICC curve. A steep slope (large  $\alpha_j$ ) indicates that a small increment in the trait value corresponds to a large increment in the probability of endorsing a particular symptom. Therefore, the corresponding item can discriminate between subjects with different trait levels. Conversely, a small  $\alpha_j$  indicates that subjects with large differences in their trait levels have similar probabilities of endorsing the symptom  $j$ . The “difficulty” parameter, on the other hand, equals the trait levels of a person with 50 % chance to endorse a symptom. The higher the “difficulty” parameter, the higher are the IA (or HI) levels required for a 0.5 probability of a symptom presence. In other words, the symptom is more likely

to be present in subjects with high levels of the trait and therefore is more severe. The IRT model focuses on the importance of each individual symptom. IRT analyses model the “item difficulty” parameter and “item discrimination” parameter of each item in relation to the latent trait within the context of the rest of the items. The “difficulty” parameter (which is defined at the 50 % chance of endorsing a symptom) reflects the severity of a symptom, while the “discrimination” parameter refers to the ability of an item to discriminate between low and high scorers (with respect to the trait). The model parameters can inform us that the presence of a particular symptom is associated with high trait levels, therefore suggesting that this symptom is more likely to be present in cases rather than controls. However, the item parameters do not provide a direct binary outcome (case–control) towards case-ness. The MPLUS [22] software was used for the IRT analysis.

#### *Predictability*

Logistic regression (LR) was employed to evaluate the effect of the presence of a symptom on the odds of receiving an ADHD diagnosis (response variable). A separate model was fitted for each item, adjusted for gender and age. In order to include in the LR model the information provided by the rest of the items, the total IA and HI sum scores were computed after omitting the item of interest and were inserted in the model as covariates. Under this approach, the odds ratio (OR) reported for each item is conditional on the summarised information of the remaining items as well as on age and gender. The (between items) comparison of the ORs is essentially the comparison of the effect size of the symptoms on the ADHD diagnosis.

As opposed to the IRT methodology, where the complete response pattern of an individual is taken under consideration

(that is, the specific symptoms present in each case), the LR approach followed here isolates the effect of each symptom conditional on the latent trait, as measured by the remaining items. The same analysis was replicated for the parents' evaluation of impairment and for the CD diagnosis. Separate results for each gender are also provided for clarity. LR analyses were conducted in R [23].

## Results

### Sample characteristics

The recruited sample consisted of 1383 children initially referred with ADHD (as 'probands') and 405 siblings of the probands. Of the 1788 individuals, 1240 subjects (89.7 %) amongst the probands were classified as ADHD cases by research protocol and 257 (63.5 %) subjects among their siblings, therefore yielding 1497 (i.e. 1240 + 257) affected cases and 291 (i.e. 148 + 143) unaffected individuals. The mean age was 11 years (SD = 2.8) with no significant differences between the two groups ( $t$  test:  $t = 1.433$ ,  $df = 572.7$ ,  $p = 0.153$ , equal variances not assumed) or between genders ( $t$  test:  $t = 0.059$ ,  $df = 465.0$ ,  $p = 0.096$ , equal variances not assumed). With respect to gender, the percentage of males was significantly higher in the proband group (86.5 versus 62.5 %,  $\chi^2 = 117.5$ ,  $df = 1$ ,  $p < 0.001$ ) than in the siblings group (Table 1).

For ADHD + CD analysis, a slightly smaller sample was available for analysis due to missing data on CD status in some individuals. 338 of 1360 (24.9 %) ADHD cases and 46 of 307 (15 %) unaffected siblings had a CD diagnosis. Different subsets of the initial sample were used in each section, corresponding to each research question posed.

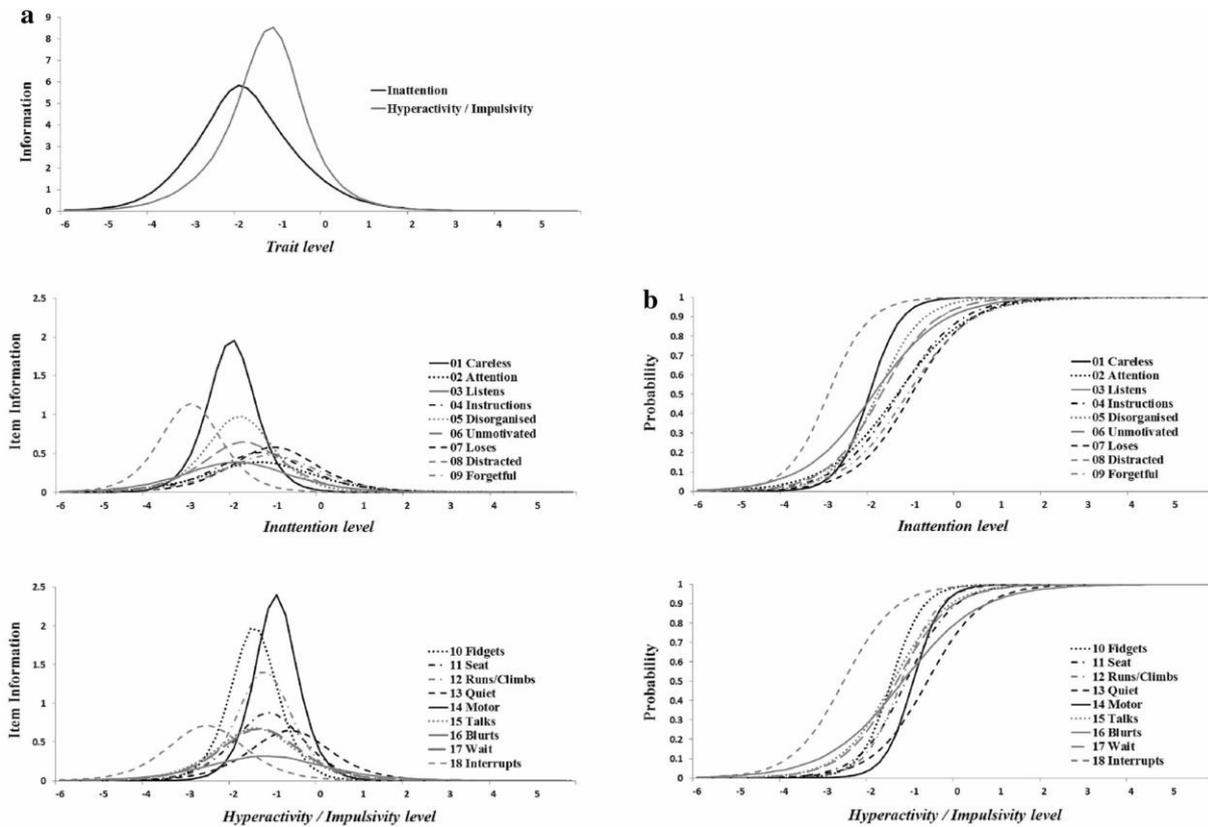
### Item properties

*Are there any redundant items for the diagnosis of ADHD?* Amongst the DSM-IV IA items, the item-total correlations (point biserial coefficients,  $r$ ) varied from 0.27 to 0.44, indicating that none of the items can be considered as inconsistent ( $r < 0.2$ ) or redundant ( $r > 0.7$ ) for the assessment of the trait. The KR-20 for IA items was 0.7. When any of the items were omitted, this coefficient decreased, further supporting non-redundancy. Similarly, amongst HI items, the item-total correlations varied from 0.32 to 0.62. The KR-20 for HI items was 0.8 again, indicating robust internal consistency. When any of the items were omitted, this coefficient decreased, indicating therefore non-redundancy.

The IRT model, in contrast, allows for the evaluation of the information provided by each item, at any level of the trait. Figure 1 presents the total information for selected trait levels (average trait levels, 0, as well as 1, 2, and 3 standard deviations (SDs) above and below average). The total

**Table 1** Number of IA and HI symptoms present in cases and controls

ADHD status		Control (N = 291)	Case (N = 1497)	Total sample
IA	Mean	6	8	8
	SD	2	1	2
	Median	6	8	8
	Range	0–9	0–9	0–9
HI	Mean	5	8	7
	SD	3	2	2
	Median	5	8	8
	Range	0–9	0–9	0–9
Total symptoms	Mean	11	16	15
	SD	4	2	3
	Median	12	17	16
	Range	0–18	6–18	0–18



**Fig. 1 a** Total and item information curves, depicting the amount of information provided at each point of the trait continuum, by the total set and by each item separately ( $n = 1788$ ). **b** Item characteristic curves per trait. Curves on the *right side* of the trait continuum indicate more severe symptoms and *steeper slopes* indicate more discriminative symptoms ( $n = 1788$ )

information curves represent the information across the entire trait continuum. The IA scale peaks at  $-2$  SDs and the HI one at  $-1$  SD.

For the IA symptoms, *Careless* appears to be the most reliable item (i.e. showing the highest peak) especially for subjects with trait values between  $-1$  and  $-3$  SDs (with a peak at  $-2$  SDs). The *Distracted* item, however, is the only reliable symptom at a level as low as  $-3.5$  SDs. The *Loses* item is associated with lower information values than the previous two items; however, it provides the highest information amongst all items for subjects with trait levels between average and  $-1$  SD. The symptoms *Listen* and *Attention* provided the lowest information values across the continuum.

For the HI symptoms, *Motor*, *Fidgets* and *Runs/climbs* were associated with

high information, at HI levels ranging from  $-2.5$  to  $-0.5$  SDs. *Interrupts* had lower values, constituting the only reliable item for subjects with HI levels close to  $-3$  SDs. *Wait* and *Talks* items have also low information. The least informative symptom was *Blurts*, which was associated with the lowest values across most of the spectrum. Item information curve, Fig. 1a, also showed that, in the IA domain, *Distracted* yields most information in the low-severity range of the latent trait, *Careless* in the mid-severity range and *Loses* in the high-severity range. In the HI domain, *Interrupts* yields most information in the low-severity range and *Motor* in the high-severity range.

*Are some DSM-5 criteria more discriminatory?*

Among IA items, the most discriminating symptom (i.e. the item characteristic curve with the steepest gradient) was “careless” and the least discriminating ones were *Listen* and *Attention*. In other words, an increment on the IA levels corresponds to a substantial increase in the probability of endorsing the *Careless* symptom, but this is not the case for the least discriminating items. Amongst the HI items, the most discriminating item was *Motor* and the least discriminating one was *Blurts*.

*Loses* was the most severe symptom amongst IA items, meaning that subjects had a 50 % chance to endorse this item with IA levels of  $-1$  SD, and *Distracted* was the least severe symptom. Not all IA items carry equal weight of severity. Items are listed according to decreasing severity as follows: *Loses* ( $-1.08$ ), *Forgetful* ( $-1.17$ ), *Instructions* ( $-1.38$ ), *Attention* ( $-1.41$ ), *Unmotivated* ( $-1.78$ ), *Disorganised* ( $-1.87$ ), *Listen* ( $-1.96$ ), *Careless* ( $-2.05$ ) and *Distracted* ( $-3.00$ ).

Among the HI symptoms, the most severe symptom was *Quiet* and the least severe was *Interrupts*. Items are listed according to decreasing severity as follows: *Quiet* ( $-0.69$ ), *Motor* ( $-1.02$ ), *Seat* ( $-1.81$ ), *Blurts* ( $-1.26$ ), *Runs/climbs* ( $-1.29$ ), *Wait* ( $1.40$ ), *Talks* ( $1.48$ ), *Fidgets* ( $1.52$ ) and *Interrupts* ( $-2.62$ ).

Item information curves showed that, in the IA domain, *Distracted* yields most information in the low-severity range of the latent trait, *Careless* in the mid-severity range and *Loses* in the high-severity range. In the HI domain, *Interrupts* yields most information in the low-severity range and *Motor* in the high-severity range.

In the current sample, subjects with average traits (0) had more than 80 % chance to endorse any of the symptoms. This reflects that the current sample is a clinical one, with high ADHD prevalence.

Predictability

*Are there any items that are not associated with ADHD caseness? Which items contribute the most to caseness?*

We estimated the ORs for caseness for each item. The comparison group, in this section, consisted of siblings who did not have an ADHD diagnosis (of any type). The cases consisted of ADHD probands of combined or IA or HI subtypes, depending on whether the symptom of interest was an IA or an HI symptom. Cases whose siblings were in the control group were excluded from this analysis. Therefore, none of the cases and/or controls was a relative of any other subject in the sample, to minimise potential bias (Table 2).

For the IA items, the symptoms of 1026 cases and 145 controls were analysed, adjusted for age and gender. Before adjusting for the total IA and HI scale scores, the ORs were highly significant for all nine items. After controlling also for the two scale scores, only five items remained significant, namely *Careless*, *Attention*, *Listen*, *Instructions*, and *Disorganised*. The largest OR corresponded to *Careless*, where the odds of being classified as a case were nine times higher for the subjects who endorse the symptom compared to the ones who do not, adjusted for age, gender and trait sum scores.

For the HI items (1055 cases and 145 controls), all ORs were highly significant prior to adjusting for the IA and HI scores. After controlling for the traits, only six items remained significant, namely *Interrupts*; *Fidgets*; *Runs/climbs*; *Motor*; *Talks* and *Wait*. Amongst these, the OR of the *Interrupts* item was the highest (13). *Which items contribute towards the parents’ perception of impairment?*

All symptoms significantly increased the odds of impairment by 2 to 4.6 times, before adjusting for the IA and HI total scores. After adjustment, only *Listen*, *Seat* and *Quiet* remained significantly associated with impairment. Interestingly, while the presence of the former two

symptoms increased the odds of impairment, the presence of the *Quiet* symptom reduced the odds in subjects with the same trait scores, gender, and age. The reduction was even higher in the female sample, but not significant for the males. Finally, when only the males were considered, a reduction in the scores was also present under the presence of the *Loses* symptom (Table 3).

*Are there any items that are blurring the distinction between ADHD without comorbid CD (ADHD – CD) and ADHD with comorbid CD (ADHD + CD)? That is, are there any distinctive diagnostic items between ADHD – CD and ADHD + CD?*

Among cases, 24.9 % were diagnosed

with comorbid CD. A series of LR models were carried out comparing ADHD + CD with ADHD – CD, first for each item adjusted only for age and gender and then by adjusting also for the IA and HI scores (computed after excluding the item of interest).

In the total sample, the items associated with increased elevated risk with CD are *Unmotivated*, *Loses*, *Seat*, *Runs/climbs*, *Quiet*, *Motor* and *Wait*. When adjusted for the IA and HI scores, the presence of the *Unmotivated* and *Loses* remained significant and indicated an increment in the odds of having comorbid CD. On the contrary, the presence of the symptoms *Listen* and *Talks* reduced the odds of comorbid CD (Table 4).

**Table 2**

Odds ratios (standard error) for ADHD “caseness” (case/control–response variable), in the presence of each item/symptom

- Cases: 1120 (males: 975) / controls: 145 (males: 73)
- Nagelkerke’s  $R^2$  was 0.7 for the total sample models, 0.6 for males and 0.7 for females
- Significant ORs are denoted in bold
- The largest values are denoted in bold italics

\* Not enough cases to estimate the OR

**a** All models are adjusted for age and gender **b** All models are adjusted for age, IA and HI scores

	Total sample <sup>a</sup>				Stratified by gender <sup>b</sup>			
	Unadjusted for IA and HI scores		Adjusted for IA and HI scores		Males		Females	
	OR (SE)	p value	OR (SE)	p value	OR (SE)	p value	OR (SE)	p value
01 Careless	<b>82.7 (1.6)</b>	<0.001	<b>9.0 (2.2)</b>	<b>0.006</b>	<b>16.0 (2.6)</b>	<b>0.004</b>	<b>2.5 (3.7)</b>	<b>0.477</b>
02 Attention	<b>10.8 (1.2)</b>	<0.001	<b>3.7 (1.4)</b>	<0.001	<b>5.8 (1.4)</b>	<0.001	<b>1.9 (1.7)</b>	<b>0.224</b>
03 Listen	<b>9.1 (1.3)</b>	<0.001	<b>3.6 (1.4)</b>	<0.001	<b>4.6 (1.5)</b>	<0.001	<b>2.9 (1.8)</b>	<b>0.066</b>
04 Instructions	<b>8.0 (1.2)</b>	<0.001	<b>3.4 (1.4)</b>	<0.001	<b>3.2 (1.5)</b>	<b>0.003</b>	<b>4.1 (1.7)</b>	<b>0.010</b>
05 Disorganised	<b>17.5 (1.3)</b>	<0.001	<b>3.5 (1.5)</b>	<b>0.004</b>	<b>2.8 (1.7)</b>	<b>0.058</b>	<b>4.9 (2.2)</b>	<b>0.046</b>
06 Unmotivated	11.5 (1.3)	<0.001	2.1 (1.5)	0.069	1.9 (1.7)	0.223	2.2 (1.9)	0.197
07 Loses	6.7 (1.2)	<0.001	1.4 (1.4)	0.279	1.1 (1.5)	0.888	2.5 (1.8)	0.105
08 Distracted	32.6 (2.0)	<0.001	7.7 (5.0)	0.204	7.3 (5.0)	0.217	*	
09 Forgetful	4.7 (1.2)	<0.001	1.4 (1.4)	0.287	1.6 (1.5)	0.222	1.1 (1.8)	0.876
10 Fidgets	<b>35.7 (1.3)</b>	<0.001	<b>3.2 (1.6)</b>	<b>0.015</b>	<b>3.4 (1.6)</b>	<b>0.008</b>	<b>3.4 (1.6)</b>	<b>0.008</b>
11 Seat	9.7 (1.2)	<0.001	1.6 (1.4)	0.223	1.5 (1.4)	0.224	1.5 (1.4)	0.224
12 Runs/climbs	<b>22.7 (1.3)</b>	<0.001	<b>3.1 (1.5)</b>	<b>0.006</b>	<b>3.9 (1.5)</b>	<b>0.001</b>	<b>3.9 (1.5)</b>	<b>0.001</b>
13 Quiet	8.8 (1.2)	<0.001	1.5 (1.4)	0.220	1.6 (1.4)	0.114	1.6 (1.4)	0.114
14 Motor	<b>24.7 (1.3)</b>	<0.001	<b>3.8 (1.4)</b>	<0.001	<b>3.5 (1.4)</b>	<0.001	<b>3.5 (1.4)</b>	<0.001
15 Talks	<b>10.4 (1.3)</b>	<0.001	<b>4.7 (1.5)</b>	<0.001	<b>4.1 (1.5)</b>	<0.001	<b>4.1 (1.5)</b>	<0.001
16 Blurts	6.3 (1.2)	<0.001	1.6 (1.4)	0.168	1.6 (1.4)	0.155	1.6 (1.4)	0.155
17 Wait	<b>15.6 (1.2)</b>	<0.001	<b>4.2 (1.4)</b>	<0.001	<b>4.8 (1.4)</b>	<0.001	<b>4.8 (1.4)</b>	<0.001
18 Interrupts	<b>27.2 (1.6)</b>	<0.001	<b>13.3 (2.1)</b>	<b>0.001</b>	<b>12.6 (2.1)</b>	<b>0.001</b>	<b>12.6 (2.1)</b>	<b>0.001</b>

**Table 3**

Odds ratios (standard error) for impairment (yes/no-response variable), in the presence of each item/symptom

- Impairment present: 960 (males: 817)/impairment absent: 807 (females: 662)
- Nagelkerke's  $R^2$  was 0.7 for the total sample models, 0.6 for males and 0.7 for females
- Significant ORs are denoted with bold
- The largest values are denoted with bold italics

**a** All models are adjusted for age and gender **b** All models are adjusted for age, IA and HI scores

	Total sample <sup>a</sup>				Stratified by gender <sup>b</sup>			
	Unadjusted for IA and HI scores		Adjusted for IA and HI scores		males		females	
	OR (SE)	<i>p</i> value	OR (SE)	<i>p</i> value	OR (SE)	<i>p</i> value	OR (SE)	<i>p</i> value
01 Careless	<b>4.5 (1.3)</b>	<0.001	0.8 (1.4)	0.497	1.2 (1.6)	0.677	0.4 (1.8)	0.099
02 Attention	<b>3.3 (1.2)</b>	<0.001	1.1 (1.2)	0.605	1.0 (1.3)	0.940	1.4 (1.5)	0.438
03 Listen	<b>3.6 (1.2)</b>	<0.001	<b>1.6 (1.3)</b>	<b>0.049</b>	1.6 (1.3)	0.093	1.6 (1.6)	0.331
04 Instructions	<b>3.4 (1.2)</b>	<0.001	1.3 (1.3)	0.216	1.3 (1.3)	0.258	1.3 (1.5)	0.549
05 Disorganised	<b>3.8 (1.2)</b>	<0.001	1.1 (1.3)	0.849	1.1 (1.4)	0.794	1.0 (1.6)	0.982
06 Unmotivated	3.3 ( <b>1.2</b> )	<0.001	1.0 (1.3)	0.964	1.2 (1.4)	0.589	0.7 (1.6)	0.542
07 Loses	2.5 ( <b>1.2</b> )	<0.001	0.7 (1.3)	0.203	<b>0.6 (1.3)</b>	<b>0.038</b>	1.6 (1.5)	0.285
08 Distracted	4.4 ( <b>1.7</b> )	<b>0.005</b>	0.6 (1.8)	0.472	1.6 (2.0)	0.500	0.1 (3.5)	0.057
09 Forgetful	2.0 ( <b>1.2</b> )	<0.001	0.7 (1.3)	0.089	0.6 (1.3)	0.094	0.8 (1.6)	0.553
10 Fidgets	<b>4.6 (1.2)</b>	<0.001	1.4 (1.3)	0.268	1.3 (1.4)	0.437	1.5 (1.7)	0.456
11 Seat	3.7 ( <b>1.2</b> )	<0.001	<b>1.7 (1.3)</b>	<b>0.028</b>	1.4 (1.3)	0.209	2.6 (1.6)	0.032
12 Runs/climbs	<b>4.2 (1.2)</b>	<0.001	1.5 (1.3)	0.116	1.4 (1.4)	0.345	1.8 (1.6)	0.213
13 Quiet	2.1 ( <b>1.2</b> )	<0.001	<b>0.6 (1.3)</b>	<b>0.045</b>	0.8 (1.3)	0.379	<b>0.3 (1.7)</b>	<b>0.011</b>
14 Motor	<b>3.7 (1.2)</b>	<0.001	1.1 (1.3)	0.783	0.8 (1.4)	0.614	1.9 (1.7)	0.219
15 Talks	<b>2.4 (1.2)</b>	<0.001	0.9 (1.3)	0.584	1.1 (1.3)	0.748	0.5 (1.6)	0.161
16 Blurts	1.9 ( <b>1.2</b> )	<0.001	0.7 (1.2)	0.079	0.6 (1.3)	0.066	0.9 (1.5)	0.798
17 Wait	<b>3.0 (1.2)</b>	<0.001	0.9 (1.3)	0.695	1.0 (1.3)	0.965	0.7 (1.6)	0.428
18 Interrupts	<b>4.0 (1.3)</b>	<0.001	1.2 (1.4)	0.620	1.5 (1.5)	0.382	0.9 (1.8)	0.809

## Discussion

There are four key findings in this study. First, there was no clear evidence of item redundancy amongst the 18 DSM-IV items using CTT analyses, in regard to KR-20 and item-total correlations. The IRT models, however, indicated that the items *Listen* (IA) and *Blurts* (HI) with flat slopes did not provide satisfactory information. These items had the lowest discrimination parameters within their scales and similar severity with other items (see Fig. 1b).

Second, not all the DSM-IV items (now DSM-5) carry equal weighting. *Loses*, *Forget* and *Attention* mark severity for IA items and *Quiet*, *Motor*, *Seat* and *Blurts* for HI items. In contrast, *Distracted*,

*Careless*, *Interrupts* and *Fidgets* are associated with milder presentations.

Most interestingly, item information curves (see Fig. 1a) showed that, in the IA domain, *Distracted* yielded most information in the low-severity range of the latent trait, *Careless* in the mid-severity range and *Loses* in the high severity range. In the HI domains, *Interrupts* yielded most information in the low-severity range; but *Motor* and *Quiet* in the high-severity range. These three IA and two HI items, therefore, have distinctive characteristics. If replicated, these findings can potentially isolate the unique properties of these specific items in capturing different ranges of ADHD phenotype, analogous to different maths

examination questions capturing students with low, average and high maths abilities. These findings would stimulate more research.

Third, all 18 items predicted impairment. However, after adjusting for the total number of symptoms, only a few items remained significant, with *Listen* and *Seat* predicting increased impairment.

Fourth, in terms of comorbidity with CD, specific items, in particular *Unmotivated* and *Loses*, are associated with 'ADHD + CD' status.

#### (1) Item redundancy and caseness

Our findings provide support for the diagnostic utility of the DSM-IV ADHD 18 items, which have been kept essentially unchanged in DSM-5. Purpura et al. [24] applied IRT analysis to ADHD symptoms in preschool children, using polytomous items. They found also the *Instructions* item in the IA trait and *Runs/climbs* and *Quiet* items in the HI trait to be the most informative. The symptom *Listen* for the IA trait as well as *Blurts* for the HI were also identified as the items with the lowest information in their study. Evidence against those items is reported also by Willcutt et al. [6]. Our results point towards the same direction, that is, these items could be considered for removal, prior to any other item. It should be noted, however, that the item *Listen* (unlike the item *Blurts*) significantly predicted ADHD caseness, impairment and comorbid CD in our data, even after adjusting for the number of symptoms present.

Frick et al. [5] conducted the DSM-IV field trial to select diagnostic symptom criteria for ADHD caseness. They found *Disorganised*, *Forgetful*, *Loses* and *Unmotivated* to carry high positive predictive power amongst the selected IA symptoms; for HI symptoms, they identified *Motor*, *Runs/climbs*, *Fidgets* and *Quiet*. In our sample, after adjusting for

age, gender and the two trait scores, the *Disorganised* item is replicated for IA; Attention, Careless, Instructions and Listen items were significantly associated with caseness. For HI items, *Motor*, *Runs/climbs* and *Fidgets* were replicated; significant predictions for *Wait* and *Talks*, but not *Quiet*, were found.

#### (2) Discriminatory items

Young et al. [25] applied the Partial Credit model to identify the most discriminative ADHD items, in an Australian community twin sample, using the SWAN rating scale; and found *Careless*, *Instructions*, *Loses* in IA trait and *Fidgets*, *Motor* and *Interrupts* in HI trait, as the "first rank" symptoms in discriminating the combined subtype of ADHD cases. Even though the authors studied polytomous items, rather than binary, our results concur with this study for *Careless*, *Motor* and *Fidgets*.

Gomez et al. [26] applied IRT to a sample of primary school-aged children using parent-rated DBRS scale and found *Instructions*, *Careless*, *Attention*, *Disorganised* and *Unmotivated* items in the IA trait, and *Runs/climbs* item in the HI trait, as the best discriminatory symptoms. Apart from *Instructions*, these items overlap partially with our detected discriminatory items.

Overall, the identified discriminatory items vary across studies and could be susceptible to sample and measurement differences. Our sample is unique in ascertaining ADHD probands and their siblings and therefore probably isolating items that best discriminate clinical cases from sub-threshold symptomatic siblings. The items identified in our study may therefore have survived a rigorous test: that is, discriminating between individuals with high levels of the traits.

According to our results, a notable finding is that not all the DSM-IV items (now DSM-5) carry equal weighting. This is a striking finding from the clinical point of

view. *Loses*, *Forgetful* and *Attention* mark severity for IA items and *Quiet*, *Motor*, *Seat* and *Blurts* for HI items. Interestingly, three severe HI items (*Quiet*, *Motor* and *Seat*) are significantly associated with comorbid CD. The clinical relevance and treatment implications remain unexplored and warrant further research. However, if replicated, these symptoms could index a need for more vigorous clinical management, to reduce the risk of developing conduct problems and potentially associated antisocial personality disorder.

Polanczyk et al. [27] have used the Rasch model of IRT to compare adolescents and adults with ADHD in a representative Brazilian population sample. The authors report that the adolescent data provide a better fit with the Rasch model than adults. Amongst the six items examined, the decreasing order of severity of the paraphrased items is as follows: *Careless*, *Disorganised*, *Instructions*, *Forgetful*, *Motor* and *Fidgets*. The authors suggest viewing the phenomenology of ADHD within a developmental framework. These findings are not directly comparable to our results, since screening items were measured and all fitted via the one parameter Rasch model and therefore equal discrimination was assumed for all items, using data from a community sample.

### (3) Prediction of impairment

There are items that predict decreased impairment: *Quiet* in the overall sample and in females and *Loses* in males. Our measure of impairment reflects the parental perception. A possible explanation could be that noisy children are more closely monitored by their parents, thereby reducing the likelihood of risky behaviour. Likewise, children who lose things might be more scaffolded by their parents.

### (4) Differences of items in ADHD-only and ADHD + conduct disorder cases

Some items are significantly more common in 'ADHD + CD' cases. There could be three possible interpretations. First, these items could be more pathogenic leading to the developmental expression of CD complication; however, this hypothesis has been refuted by Lahey et al. [14]. Second, 'ADHD + CD' may delineate a distinct disorder, as suggested by the ICD-10 category of Hyperkinetic Conduct Disorder [28] by its differential response profile to treatment [29], also indexed by ADHD severity and as a distinct familial subtype [15], and by possible genetic heterogeneity with one study showing DAT1 genotypes as having a significantly greater genetic influence on ADHD without CD [30]. Our 'ADHD + CD' specific items could potentially alert to comorbidity with CD. Third, the ADHD items common to both conditions are more specific to 'ADHD per se', i.e. representing a purer ADHD phenotype.

If our findings are replicated by prospective longitudinal data, these ADHD + CD specific items could potentially alert the clinician to the possibility of comorbid CD in a child or young person diagnosed with ADHD.

Several limitations need to be considered. Our analysis relied mostly on information provided by the carers and is therefore subject to reporter bias (information from teachers as independent sources of information did, however, contribute towards diagnostic status of case-ness). Our sample consists of ADHD cases ascertained from psychiatric and paediatric clinics and their non-ADHD siblings form the comparison group, so the comparison group was not independently ascertained. Nevertheless, the findings confirming the robust utility of the ADHD 18 items, which have been kept in DSM-5, are striking, as this limitation would have biased the results towards the null hypothesis. As highlighted by Willcutt et

al. [6], the diagnoses of ADHD subtypes are not stable over time. Our study is cross-sectional comparing a population with ADHD and non-ADHD siblings. Given the design of the study, we may not be considering subjects with varying developmental trajectories.

In summary, first, all 18 DSM-IV items contributed significantly and independently to the clinical diagnosis of ADHD. Second, not all the DSM-IV items carry equal weighting. *Loses*, *Forgetful* and *Attention* mark severity for Inattentiveness (IA) items and *Quiet*, *Motor*, *Seat* and *Blurts* for Hyperactivity/Impulsivity (HI) items. *Distracted*, *Careless* *Interrupts* and

*Fidgets* are associated with milder presentations. In the IA domain, *Distracted* yields most information in the low-severity range of the latent trait, *Careless* in the mid-severity range and *Loses* in the high severity range. In the HI domains, *Interrupts* yields most information in the low-severity range and *Motor* in the high-severity range. Third, all 18 items predicted impairment. Fourth, specific ADHD items are associated with ADHD + CD status. To conclude, the DSM-IV diagnostic items were valid and not redundant; however, some carried more weight than others. All items were associated with impairment.

**Table 4**

Odds ratios (standard error) for comorbid CD (present/absent-response variable), in the presence of each item/symptom

- CD present: 343 (females: 302) / CD absent: 1078 (females: 904)
- Nagelkerke's  $R^2$  was 0.04–0.05 for the total sample models, 0.03–0.04 for males and 0.16–0.21 for females
- Significant ORs are denoted with bold
- The largest values are denoted with bold italics

\* Not enough cases to estimate the OR

**a** All models are adjusted for age and gender    **b** All models are adjusted for age, IA and HI scores

	Bivariate Associations				Total sample <sup>a</sup>				Stratified by gender <sup>b</sup>			
	CD (%)		Chi square (df =1)		Unadjusted for IA & HI scores		Adjusted for IA & HI scores		males		females	
	No	Yes	$\chi^2$	<i>p</i> value	OR (SE)	<i>p</i> value	OR (SE)	<i>p</i> value	OR (SE)	<i>p</i> value	OR (SE)	<i>p</i> value
01 Careless	98	99	0.387	0.534	1.5 (1.7)	0.442	0.7 (1.8)	0.542	1.2 (2.2)	0.831	0.2 (3.0)	0.115
02 Attention	<b>85</b>	<b>91</b>	<b>6.692</b>	<b>0.010</b>	1.7 (1.2)	0.111	1.2 (1.3)	0.435	1.0 (1.3)	0.873	3.1 (2.1)	0.125
03 Listen	92	91	0.157	0.692	0.9 (1.2)	0.568	<b>0.6 (1.3)</b>	<b>0.022</b>	<b>0.6 (1.3)</b>	<b>0.046</b>	0.6 (1.8)	0.354
04 Instructions	86	89	1.069	0.301	1.2 (1.2)	0.385	0.8 (1.2)	0.463	0.8 (1.3)	0.388	1.0 (1.7)	0.943
05 Disorganised	96	96	0.001	0.999	1.0 (1.4)	0.937	0.6 (1.4)	0.177	0.6 (1.5)	0.241	0.6 (2.3)	0.607
06 Unmotivated	<b>91</b>	<b>97</b>	<b>11.616</b>	<b>0.001</b>	<b>2.9 (1.4)</b>	<b>0.001</b>	<b>2.2 (1.4)</b>	<b>0.023</b>	2.0 (1.5)	0.063	3.3 (2.4)	0.167
07 Loses	<b>79</b>	<b>87</b>	<b>12.654</b>	<b>&lt;0.001</b>	<b>1.9 (1.2)</b>	<b>0.000</b>	<b>1.5 (1.2)</b>	<b>0.048</b>	<b>1.7 (1.3)</b>	<b>0.026</b>	0.9 (1.7)	0.915
08 Distracted	99	99	0.094	0.759	0.4 (2.5)	0.361	0.3 (2.7)	0.183	0.6 (3.4)	0.711	0.1 (7.3)	0.149
09 Forgetful	79	83	2.252	0.133	1.2 (1.2)	0.197	0.8 (1.2)	0.430	0.8 (1.3)	0.417	0.8 (1.8)	0.653
10 Fidgets	94	97	2.573	0.109	1.8 (1.4)	0.888	0.8 (1.5)	0.515	0.8 (1.5)	0.615	0.4 (3.5)	0.428
11 Seat	<b>84</b>	<b>89</b>	<b>4.633</b>	<b>0.031</b>	<b>1.6 (1.2)</b>	<b>0.012</b>	1.0 (1.3)	0.904	1.0 (1.3)	0.971	1.1 (2.1)	0.932
12 Runs/climbs	<b>89</b>	<b>95</b>	<b>11.029</b>	<b>0.001</b>	<b>2.6 (1.3)</b>	<b>0.001</b>	1.5 (1.4)	0.203	1.6 (1.4)	0.152	0.7 (2.3)	0.684
13 Quiet	<b>72</b>	<b>84</b>	<b>19.036</b>	<b>&lt;0.001</b>	<b>2.1 (1.2)</b>	<b>0.000</b>	1.4 (1.2)	0.069	1.3 (1.2)	0.181	2.8 (2.0)	0.127
14 Motor	<b>86</b>	<b>93</b>	<b>11.982</b>	<b>0.001</b>	<b>2.1 (1.3)</b>	<b>0.001</b>	1.1 (1.3)	0.787	1.2 (1.4)	0.525	0.4 (2.5)	0.270
15 Talks	89	89	0.001	0.999	1.0 (1.2)	0.878	<b>0.5 (1.3)</b>	<b>0.003</b>	<b>0.6 (1.3)</b>	<b>0.045</b>	<b>0.1 (2.4)</b>	<b>0.002</b>
16 Blurts	79	83	2.361	0.124	1.3 (1.2)	0.107	0.8 (1.2)	0.213	0.7 (1.2)	0.107	2.7 (2.3)	0.228
17 Wait	<b>87</b>	<b>94</b>	<b>9.995</b>	<b>0.002</b>	<b>2.1 (1.3)</b>	<b>0.002</b>	1.3 (1.3)	0.331	1.2 (1.3)	0.556	1.8 (2.4)	0.499
18 Interrupts	98	99	1.217	0.270	3.5 (2.1)	0.089	1.9 (2.2)	0.419	1.7 (2.2)	0.517	*	

**Acknowledgments:** The IMAGE project is a multi-site, international effort supported by NIH grants R01MH62873 and R01MH081803 to S. V. Faraone. The IMAGE site principal investigators are Philip Asherson, Tobias Banaschewski, Jan Buitelaar, Richard P. Ebstein, Stephen V. Faraone, Michael Gill, Ana Miranda, Fernando Mulas, Robert D. Oades, Herbert Roeyers, Aribert Rothenberger, Joseph Sergeant, Edmund Sonuga-Barke and Hans-Christoph Steinhausen. Chief investigators at each site are Rafaela Marco, Nanda Rommelse, Wai Chen, Henrik Uebel, Hanna Christiansen, Ueli C. Mueller, Cathelijne Buschgens, Barbara Franke and Lamprini Psychogiou. We thank all. Silia Vitoratou was funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The authors are grateful to Rand R. Wilcox, for his statistical advice; Professor Eric Taylor and Professor Andrew Pickles for advice and support; and the Alicia Koplowitz Foundation, which sponsored Dr Alexandra Garcia Rosales. We also thank the children and families participating in the IMAGE project.

**Conflict of interest:** AGR: AGR reports no biomedical financial interests or potential conflict of interests. SV: SV reports no biomedical financial interests or potential conflict of interests. TB: TB served in an advisory or consultancy role for Hexal Pharma, Lilly, Medice, Novartis, Otsuka, Oxford outcomes, PCM scientific, Shire and Viforpharma. He received conference attendance support and conference support or received speaker's fee from Lilly, Medice, Novartis and Shire. He is/has been involved in clinical trials conducted by Lilly, Shire and Viforpharma. The present work is unrelated to the above

grants and relationships. PA: King's College London has received funds for work completed by PA: grants from Janssen, Shire, Lilly, QBTech and Vifor; sponsored talks from Shire, Novartis, Janssen and Lilly; consultancy for Shire, Lilly and Novartis. JB: JB has been in the past 3 years a consultant to/member of advisory board of/and/or speaker for Janssen Cilag BV, Eli Lilly, Bristol-Myer Squibb, Schering Plough, UCB, Shire, Novartis and Servier. He is not an employee of any of these companies and not a stock shareholder of these companies. He has no other financial or material support, including expert testimony, patents or royalties. RO: In the last 24 months, RO reports no financial interests or potential conflicts of interest. AR: Advisory Board and Speakers Bureau: Lilly, Shire, Medice, Novartis; research support: Shire, German Research Society, Schwaabe; travel support: Shire; educational grant: Shire; consultant: UCB/Shire; Lilly. H-C S: In the last 36 months, H-C S has been a member of the advisory board and/or received payment for lectures including service on speakers bureaus by Shire and Medice. The present work is unrelated to the above grants and relationships. SVF: In the past year, SVF received consulting income, travel expenses and/or research support from Akili Interactive Labs, Alcobra, VAYA Pharma and SynapDx and research support from the National Institutes of Health (NIH). His institution is seeking a patent for the use of sodium-hydrogen exchange inhibitors in the treatment of ADHD. In previous years, he received consulting fees or was on advisory boards or participated in continuing medical education programs sponsored by: Shire, Alcobra, Otsuka, McNeil, Janssen, Novartis, Pfizer and Eli Lilly. Dr. Faraone receives royalties from books published by Guilford Press: Straight Talk about Your Child's Mental Health and Oxford

University Press: Schizophrenia: The Facts. WC: Wai Chen has received honorarium for lecturing and consultancy for Shire, Janssen and Flynn Pharma, research support from Shire, and support for attending meetings from Shire and Janssen. The present work is unrelated to the above grants and relationships.

## References

1. Taylor E (2009) Developing ADHD. *J Child Psychol Psychiatry* 50(1–2):126–132
2. Taylor E (2011) Antecedents of ADHD: a historical account of diagnostic concepts. *Atten Defic Hyperact Disord* 3(2):69–75
3. Bussing R, Mason DM, Bell L, Porter P, Garvan C (2010) Adolescent outcomes of childhood attention-deficit/hyperactivity disorder in a diverse community sample. *J Am Acad Child Adolesc Psychiatry* 49:595–605
4. Ramtekkar UP, Reiersen AM, Todorov AA, Todd RD (2010) Sex and age differences in attention-deficit/hyperactivity disorder symptoms and diagnoses: implications for DSM-V and ICD-11. *J Am Acad Child Adolesc Psychiatry* 49:217–228
5. Frick PJ, Lahey BB, Applegate B, Kerdyck L, Ollendick T, Hynd GW, Garfinkel B, Greenhill L, Biederman J, Barkley RA et al (1994) DSM-IV field trials for the disruptive behavior disorders: symptom utility estimates. *J Am Acad Child Adolesc Psychiatry* 33(4):529–539
6. Willcutt EG, Nigg JT, Pennington BF, Solanto MV, Rohde LA, Tannock R, Loo SK, Carlson CL, McBurnett K, Lahey BB (2012) Validity of DSM-IV attention-deficit/hyperactivity disorder symptom dimensions and subtypes. *J Abnorm Psychol* 121(4):991–1010
7. Wolraich ML, Bard DE, Neas B, Doffing M, Beck L (2013) The psychometric properties of the Vanderbilt attention-deficit hyperactivity disorder diagnostic teacher rating scale in a community population. *J Dev Behav Pediatr* 34(2):83–89
8. Sung M, Erkanli A, Angold A, Costello EJ (2004) Effects of age at first substance use and psychiatric comorbidity on the development of substance use disorders. *Drug Alcohol Depend* 75:287–299
9. Caron C, Rutter M (1991) Comorbidity in child psychopathology: concepts, issues and research strategies. *J Child Psychol Psychiatry* 32(7):1063–1080
10. Newcorn JH, Halperin JM, Jensen PS, Abikoff HB, Arnold LE, Cantwell DP, Conners CK, Elliott GR, Epstein JN, Greenhill LL, Hechtman L, Hinshaw SP, Hoza B, Kraemer HC, Pelham WE, Severe JB, Swanson JM, Wells KC, Wigal T, Vitiello B (2001) Symptom profiles in children with ADHD: effects of comorbidity and gender. *J Am Acad Child Adolesc Psychiatry* 40(2):137–146
11. Oosterlaan J, Sergeant JA (1998) Response inhibition and response re-engagement in attention-deficit/hyperactivity disorder, disruptive, anxious and normal children. *Behav Brain Res* 94(1):33–43
12. Oosterlaan J, Logan GD, Sergeant JA (1998) Response inhibition in AD/HD, CD, comorbid AD/HD + CD, anxious, and control children: a meta-analysis of studies with the stop task. *J Child Psychol Psychiatry* 39(3):411–425
13. Schachar R, Tannock R (1995) Test of four hypotheses for the comorbidity of attention-deficit hyperactivity disorder and conduct disorder. *J Am Acad Child Adolesc Psychiatry* 34:639–648
14. Lahey BB, Van Hulle CA, Rathouz PJ, Rodgers JL, D’Onofrio BM, Waldman ID (2009) Are oppositional-defiant and hyperactive-inattentive symptoms developmental precursors to conduct problems in late childhood?: genetic and environmental links. *J Abnorm Child Psychol* 37(1):45–58
15. Christiansen H, Chen W, Oades RD, Asherson P, Taylor EA, Lasky-Su J et al (2008) Co-transmission of conduct problems with attention deficit hyperactivity disorder: familial evidence for a

- distinct disorder. *J Neural Transm* 115:163–175
16. Chen W, Zhou K, Sham P, Franke B, Kuntsi J et al (2008) DSMIV combined type ADHD shows familial association with sibling trait scores: a sampling strategy for QTL linkage. *Am J Med Genet Part B Neuropsychiatr Genet* 147B (8):1450–1460
17. Müller UC, Asherson P, Banaschewski T et al (2011) The impact of study design and diagnostic approach in a large multi-centre ADHD study. Part 1: ADHD symptom patterns. *BMC Psychiatry* 11:54
18. Müller UC, Asherson P, Banaschewski T et al (2011) The impact of study design and diagnostic approach in a large multi-centre ADHD study. Part 2: dimensional measures of psychopathology and intelligence. *BMC Psychiatry* 11:54
19. Curran S, Newman S, Taylor E, Asherson P (2000) Hypescheme: an operational criteria checklist and minimum data set for molecular genetic studies of attention deficit and hyperactivity disorders. *Am J Med Genet* 96:244–250
20. Chen W, Taylor EA (2006) Parental account of children's symptoms (PACS), ADHD phenotypes and its application to molecular genetic studies. In: Oades RD (Ed) *Attention-deficit/hyperactivity disorder and the hyperkinetic syndrome*. Nova Science Publishing Inc, Hauppauge, pp 3–20
21. Kuder GF, Richardson MW (1937) The theory of the estimation of test reliability. *Psychometrika* 2(3):151–160
22. Muthén LK, Muthén BO (1998) *M plus User's Guide*, 6th edn. Los Angeles, Muthén & Muthén
23. R Development Core Team (2011) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna
24. Purpura DJ, Wilson SB, Lonigan CJ (2010) Attention-deficit/hyperactivity disorder symptoms in preschool children: examining psychometric properties using item response theory. *Psychol Assess* 22(3):546–558
25. Young DJ, Levy F, Martin NC, Hay DA (2009) Attention deficit hyperactivity disorder: a Rasch analysis of the SWAN Rating Scale. *Child Psychiatry Hum Dev* 40(4):543–559
26. Gomez R, Vance A, Gomez A (2011) Item response theory analyses of parent and teacher ratings of the ADHD symptoms for recoded dichotomous scores. *J Atten Disord* 15(4):269–285
27. Polanczyk G, Laranjeira R, Zaleski M, Pinsky I, Caetano R, Rohde LA (2010) ADHD in a representative sample of the Brazilian population: estimated prevalence and comparative adequacy of criteria between adolescents and adults according to the item response theory. *Int J Methods Psychiatr Res* 19(3):177–184
28. World Health Organization (2010) *International classification of diseases (ICD-10)*. World Health Organization, Switzerland
29. Jensen PS, Hinshaw SP, Kraemer HC, Lenora N, Newcorn JH, Abikoff HB, March JS, Arnold LE, Cantwell DP, Conners CK, Elliott GR, Greenhill LL, Hechtman L, Hoza B, Pelham WE, Severe JB, Swanson JM, Wells KC, Wigal T, Vitiello B (2001) ADHD comorbidity findings from the MTA study: comparing comorbid subgroups. *J Am Acad Child Adolesc Psychiatry* 40(2):147–158
30. Zhou K, Chen W, Buitelaar J, Banaschewski T, Oades RD, Franke B, Sonuga-Barke E, Ebstein R, Eisenberg J, Gill M, Manor I, Miranda A, Mulas F, Roeyers H, Rothenberger A, Sergeant J, Steinhausen HC, Lasky-Su J, Taylor E, Brookes KJ, Xu X, Neale BM, Rijdsdijk F, Thompson M, Asherson P, Faraone SV (2008) Genetic heterogeneity in ADHD: DAT1 gene only affects probands without CD. *Am J Med Genet B Neuropsychiatr Genet* 147B (8):1481–1487