The Journal of Child Psychology and Psychiatry

Journal of Child Psychology and Psychiatry **:* (2023), pp **–** doi:10.1111/jcpp.13901

Practitioner Review: Clinical utility of the QbTest for the assessment and diagnosis of attention-deficit/hyperactivity disorder – a systematic review and meta-analysis

Alessio Bellato,[1,2†] Charlotte L. Hall,[3,4,5†] Madeleine J. Groom,[3,4,5] Emily Simonoff,[6] Anita Thapar,[7] Chris Hollis,[3,4,5] and Samuele Cortese[5,8,9,10,11]

[1]School of Psychology, University of Nottingham, Nottingham, Malaysia; [2]Mind & Neurodevelopment (MiND) Research Cluster, University of Nottingham, Nottingham, Malaysia; [3]NIHR MindTech MedTech Co-operative, Institute of Mental Health, School of Medicine, University of Nottingham, Nottingham, UK; [4]NIHR Nottingham Biomedical Research Centre, Institute of Mental Health, University of Nottingham, Nottingham, UK; [5]Mental Health and Clinical Neurosciences, School of Medicine, University of Nottingham, Nottingham, UK; [6]Department of Child and Adolescent Psychiatry, King's College London, Institute of Psychiatry, Psychology and Neuroscience, London, UK; [7]Division of Psychological Medicine and Clinical Neurosciences, Wolfson Centre for Young People's Mental Health, Cardiff University School of Medicine, Cardiff, UK; [8]Centre for Innovation in Mental Health, School of Psychology, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, UK; [9]Solent NHS Trust, Southampton, UK; [10]Clinical and Experimental Sciences (CNS and Psychiatry), Faculty of Medicine, University of Southampton, Southampton, UK; [11]Hassenfeld Children's Hospital at NYU Langone, New York University Child Study Center, New York, NY, USA

Background: Several computerised cognitive tests (e.g. continuous performance test) have been developed to support the clinical assessment of attention-deficit/hyperactivity disorder (ADHD). Here, we appraised the evidence-base underpinning the use of one of these tests – the QbTest – in clinical practice, by conducting a systematic review and meta-analysis investigating its accuracy and clinical utility. **Methods:** Based on a preregistered protocol (CRD42022377671), we searched PubMed, Medline, Ovid Embase, APA PsycINFO and Web of Science on 15th August 2022, with no language/type of document restrictions. We included studies reporting accuracy measures (e.g. sensitivity, specificity, or Area under the Receiver Operating Characteristics Curve, AUC) for QbTest in discriminating between people with and without DSM/ICD ADHD diagnosis. Risk of bias was assessed with the Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2). A generic inverse variance meta-analysis was conducted on AUC scores. Pooled sensitivity and specificity were calculated using a random-effects bivariate model in R. **Results:** We included 15 studies (2,058 participants; 48.6% with ADHD). QbTest Total scores showed acceptable, rather than good, sensitivity (0.78 [95% confidence interval: 0.69; 0.85]) and specificity (0.70 [0.57; 0.81]), while subscales showed low-to-moderate sensitivity (ranging from 0.48 [0.35; 0.61] to 0.65 [0.52; 0.75]) and moderate-to-good specificity (from 0.65 [0.48; 0.78] to 0.83 [0.60; 0.94]). Pooled AUC scores suggested moderate-to-acceptable discriminative ability (*Q*-Total: 0.72 [0.57; 0.87]; *Q*-Activity: 0.67 [0.58; 0.77]); *Q*-Inattention: 0.66 [0.59; 0.72]; *Q*-Impulsivity: 0.59 [0.53; 0.64]). **Conclusions:** When used on their own, QbTest scores available to clinicians are not sufficiently accurate in discriminating between ADHD and non-ADHD clinical cases. Therefore, the QbTest should not be used as stand-alone screening or diagnostic tool, or as a triage system for accepting individuals on the waiting-list for clinical services. However, when used as an adjunct to support a full clinical assessment, QbTest can produce efficiencies in the assessment pathway and reduce the time to diagnosis. **Keywords:** QbTest; ADHD; validity; sensitivity; specificity; AUC.

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental condition characterised by developmentally inappropriate inattention and/or hyperactivity/impulsivity (American Psychiatric Association, 2022) that often persists into adulthood (Faraone, Biederman, & Mick, 2006; Fayyad et al., 2007). If left untreated, it can lead to significant educational, social and occupational impairment (Faraone et al., 2021; Sayal, Prasad, Daley, Ford, & Coghill, 2018). To facilitate access to treatment, it is essential that those who are suspected to have ADHD receive a timely assessment.

Currently, there is no single test to diagnose ADHD, and assessment practices tend to vary across countries. Typically, clinicians make a diagnostic decision based on their judgement, informed by interviews with the patient and their caregivers, and supported by standardised questionnaires completed by multiple informants, including teachers, and/or direct observations. Clinical guidelines (see Coghill et al., 2021; National Institute of Health and Care Excellence (NICE), 2018) support assessment and decision-making but often lack detail on how

[†]Shared first authorship.

Conflict of interest statement: See Acknowledgements for full disclosures.

assessment measures should be combined to reach a diagnostic decision. As for other neurodevelopmental and psychiatric disorders, the lack of accurate diagnostic tests and the complexity of the clinical picture of ADHD (e.g. in relation to comorbidities; Jensen & Steinhausen, 2015) contribute to delays, uncertainty and inconsistencies in the diagnostic process (Hall et al., 2016, 2016), especially for adult cases (Culpepper & Mattingly, 2010; Young et al., 2021) and when corroborative information is missing or contradictory.

Several indices can be used to investigate the validity of a diagnostic test. While *sensitivity* reflects the accuracy (based on a predefined cut-off and compared to a 'gold-standard' tool) in correctly identifying as positive cases those with the condition ('true cases'), *specificity* reflects the accuracy of the test in correctly identifying those without the condition ('non-cases'; Trevethan, 2017). Sensitivity and specificity are expressed on a 0-1 scale, with 1 suggesting perfect accuracy in detecting cases (sensitivity) and ruling out noncases (specificity). A general rule of thumb is that scores between 0.8 and 0.9 reflect 'good' sensitivity/specificity, while scores >0.9 indicate excellent discriminative validity of a test. Conversely, scores less than 0.6 are usually considered 'poor', while scores 0.6–0.7 are 'moderate' and scores 0.7–0.8 indicate 'acceptable' sensitivity/specificity (Cortese et al., 2023). However, these are just arbitrary thresholds that should be interpreted based on the specific purpose of the test (e.g. screening for rare severe conditions versus support decision-making about expensive/dangerous treatments). Moreover, sensitivity and specificity are usually calculated based on a specific cut-off, which is often arbitrary and may not fully discriminate between cases and noncases. The use of the Area under the Receiver Operating Characteristics curve (*AUC*) can overcome these limitations. AUC reflects the performance of a test to separate between true positive and true negative cases for *all* possible cut-off points, with scores ranging between 0 and 1, and indicating perfect discriminative ability at AUC = 1 or suggesting that test performs at a chance level when AUC = 0.5. Discriminative validity is usually considered 'acceptable' when AUC scores are between 0.7–0.8, 'good' between 0.8–0.9, and 'excellent' when >0.9 (Safari, Baratloo, Elfil, & Negida, 2016).

Implementing tests that rely on diagnostic biomarkers (i.e. objective indices that discriminate between those with and without a disorder) may improve diagnostic accuracy of ADHD (Loh et al., 2022), but only when the validity of such tests is confirmed. Indeed, several candidate biomarkers for ADHD have been proposed, but none have shown sufficient sensitivity or specificity to be used in clinical practice (Adamou, Fullen, & Jones, 2020; Buitelaar et al., 2022; Lenartowicz & Loo, 2014; Loo & Barkley, 2005; Loo & Makeig, 2012; Zhang-James, Razavi, Hoogman, Franke, & Faraone, 2023). This conclusion is supported by a recent large-scale systematic review (Cortese et al., 2023) that found no biochemical (e.g. iron levels), neuroimaging (e.g. structural or functional MRI measures), neurophysiological (e.g. event-related potentials and oscillatory patterns) or neuropsychological (e.g. tests of attention) biomarkers of ADHD with good sensitivity/specificity and replicated in at least two independent studies.

Nevertheless, three devices/tools (the neuropsychiatric EEG-Based ADHD Assessment Aid, NEBA; the Test of Variables of Attention, TOVA; and the QbTest) received approval by the Food and Drug Administration (FDA) to be used as a support, rather than as standalone tool, for the clinical assessment of ADHD. The NEBA is based on the theta-beta ratio (TBR), which is a candidate biomarker of ADHD, although the existing literature has shown inconsistent findings and there is evidence that increased TBR is not specific to ADHD (Gloss, Varma, Pringsheim, & Nuwer, 2016; Saad, Kohn, Clarke, Lagopoulos, & Hermens, 2018). The TOVA is a computerised test that primarily assess sustained attention and response inhibition paradigm. As for the NEBA, the evidence of clinical utility of the TOVA is mixed (see Hall et al., 2016, for a systematic review). In this systematic review, we will focus on QbTest. In fact, a meta-analysis on the clinical utility of QbTest has not yet been performed – while the same has already been done for CPT (Arrondo et al., 2023; Hall, Valentine, et al., 2016). The recent NICE Medtech Innovation Briefing on QbTest (NICE, 2023) highlighted the need to conduct this study, which will be informative for policy makers, health commissioners and practitioners at a time of widespread use of QbTest in clinical practice. We believe that the main findings of this study will offer important support for the development of any future NICE health technology assessment and/or ADHD guideline.

QbTest is a commercially available computerised test that is based on the continuous performance test (CPT), in which individuals with ADHD usually show worse performance compared to nonclinical and clinical non-ADHD participants (Epstein et al., 2003), alongside actigraphy measures. In the CPT, participants are instructed to press a response button when certain stimuli (targets) appear on the screen but withhold responding to all other stimuli (nontargets). Hall, Valentine, et al. (2016) and Arrondo et al. (2023) systematically analysed the clinical utility of CPTs for the assessment and diagnosis of ADHD and concluded that CPT, as a stand-alone tool, only has a modest-to-moderate ability to differentiate ADHD from non-ADHD samples. However, combining CPTs with an objective measure of motor activity could be particularly promising as a clinical tool, and worthy of further research.

The QbTest (Qbtech Ltd; www.qbtech.com) has been developed as a commercially available CPT

combined with an infrared motion tracker, aimed at providing an objective estimation of the three ADHD core symptoms, that is, attention, impulsivity and hyperactivity (Hall, Valentine, et al., 2016; Valentine et al., 2020; Williams et al., 2021). QbTest is also supported by the National Institute of Health and Care Excellence (NICE, 2023) as a decision-aid to augment, but not replace, clinical assessment of ADHD, and it is now implemented in the USA, the UK, and other European countries.

During the QbTest, the participant sits on a chair in front of a computer screen and performs a CPT, while an infrared camera tracks the motion of a reflective marker attached to a headband placed on the participant's forehead. There are four versions of the test: one for preschoolers (QbMini, <6 years) one for children up to 12 years of age (QbTest 6–12), one for adolescents and adults aged over 12 years (QbTest+; in this version, the target stimuli are more infrequent and challenging to identify), and a more recent version (for people aged 6–60 years old) conducted remotely with the use of a webcam (QbCheck). For a comprehensive description of QbTest, see Hall, Bellato, Kirk, & Hollis (2023).

The QbTest should be conducted in an appropriate environment. In the clinic, a quiet room should be available for about 30 min per patient or, if the QbCheck is conducted at the participant's home, it should be made sure that the environment is adequately comfortable and free from distractions. The administration of QbTest and the preparation of reports can be done by nonspecialist healthcare professionals, but the child's behaviour should be observed during the QbTest, and clinicians experienced in the diagnosis of ADHD must be involved in the interpretation of all information from QbTest (including behavioural observation during the test) and other assessment tools. Notably, the youngest or those with a physical disability or with learning difficulties, may not be able to complete QbTest. Similarly, autistic children and children with sensory hypersensitivity may not tolerate the marker on their forehead. Further information the QbTest is available here: https://www.youtube.com/watch?v=MjsU5-OIs2s; https://www.youtube.com/watch?v=Sfly423zD_E.

The software-generated report compares the participant's CPT performance and motor activity against an age-and-gender normed non-ADHD nonclinical group, providing three cardinal scores (*Q*-Activity, *Q*-Impulsivity and *Q*-Inattention). *Q*-scores reflect the deviation of the participant's performance (in standardised units) from the mean score of the normative group. Notably, there is no fixed rule on how to determine that a certain cardinal *Q*-score (or multiple scores) is considered *positive*, that is, indicative of ADHD, since the scores are only meant to inform the diagnosis made based on clinical information. Similarly, in relation to the interpretation of the cut-offs for AUC or sensitivity/specificity scores for each cardinal score, there is not a fixed rule. In fact, although *z*-scores ≥1.5 are often interpreted as indicative of ADHD, this cut-off is not included in any official document by QbTech or FDA, and there is no evidence that this is the cut-off that best discriminates between ADHD and non-ADHD cases (or between ADHD presentations), nor did the FDA indicate any *Q*-score diagnostic cut-off. However, previous studies interpreted scores above the 93rd percentile (equivalent to *z*-scores ≥1.5) as suggestive of ADHD.

In relation to the clinical utility of QbTest, a randomised controlled trial in child mental health and paediatric services in the UK demonstrated that adding a QbTest report to standard routine clinical assessment for ADHD reduced the number of consultations needed to make a diagnostic decision, compared to when clinicians did not have a QbTest report (Hollis et al., 2018). Additionally, adding a QbTest report reduced the time from assessment to final decision, and increased the number of decisions made as well as the clinicians' confidence in their decision-making, without compromising diagnostic accuracy. There is evidence that QbTest is generally positively accepted by people with ADHD and their families, and clinicians (Hall et al., 2017).

Given the potential efficiencies and cost-savings to clinical pathways, it is not surprising that QbTest has been readily adopted in many clinical services internationally. However, there are anecdotal reports that some clinicians diverge from the intended use of QbTest, applying it as a stand-alone tool to 'screen' for people with suspected ADHD to be prioritised for clinical assessment (Vogt, 2021). Moreover, there are concerns about the diagnostic accuracy of the test. Whereas some studies found good performance of QbTest in discriminating ADHD cases from non-ADHD cases (Emser et al., 2018), others have not (Adamou, Jones, Marks, & Lowe, 2022; Johansson et al., 2021).

Obtaining a quantitative evidence synthesis on the accuracy of QbTest (appraised via sensitivity, specificity, or AUC) is therefore crucial to guide decisions around how and when to integrate QbTest into clinical assessment of ADHD. We therefore aimed to quantitatively summarise and appraise the evidence on the accuracy of QbTest in discriminating people with and without ADHD (non-ADHD nonclinical controls or non-ADHD clinical cases), when used as stand-alone tool *or* as a decision-aid tool to support clinical judgement/assessment as usual (in line with FDA and NICE approval). We also reviewed its clinical utility.

Materials and methods

Search strategy and selection criteria

We followed the 2020 PRISMA guidelines (Page et al., 2021) and the protocol of the current systematic review and meta-analysis was registered in PROSPERO (CRD42022377671). We

searched PubMed, Medline, Embase, PsycINFO, Web of Science on 15th August 2022 with no limits on time, type of document and language. The PRISMA Checklist and the search strategy are reported in Appendices S1 and S2, respectively. We included studies with any design, investigating the clinical utility of the QbTest (either as a stand-alone tool or in support of clinical judgement) for discriminating between ADHD and non-ADHD cases or improving the efficiency of diagnostic decision-making in individuals referred for ADHD assessment or with an established diagnosis of ADHD. Reference lists of relevant previous systematic or narrative reviews were searched to identify any additional eligible studies missed during the electronic search.

Data selection, extraction and quality rating

Titles and abstracts of retrieved studies were independently screened by two authors (AB, CHa) to identify those meeting inclusion criteria. Full texts of potentially eligible studies were also assessed by the same authors. Data extraction and assessment of data quality were performed by one author (AB) and cross-checked by a second author (CHa). Disagreements were settled through discussion. Study quality and risk of bias were assessed with the Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2; Whiting et al., 2011). The risk of bias was rated in relation to each of the following domains: patient selection; index test; reference standard and target conditions; flow and timing (more detailed information is reported in Appendix S3).

Outcomes

We extracted (a) AUC, (b) sensitivity and specificity, (c) positive predictive value (PPV) and negative predictive value (NPV). However, although PPVs and NPVs are helpful to determine the clinical utility of QbTest, they are known to depend on the prevalence of ADHD in the population assessed (e.g. population samples vs. clinically referred groups). Therefore, as their interpretation in meta-analyses may be misleading, we did not pool PPV and NPV.

Data synthesis and analysis

All analyses were conducted in *R 4.1.2* (R Core Team, 2020). For AUC, we used the R package *meta* (Schwarzer, Carpenter, & Rücker 2015) to conduct a generic inverse variance meta-analysis. We used the R package *auctestr* (Gardner, 2017) to estimate the standard error of the AUC, when this was not reported in the publication, by using the standard normal distribution (Hanley & McNeil, 1982). We did not estimate AUC for studies that only reported sensitivity or specificity, since specificity and sensitivity are calculated based on a specific cut-off, while the calculation of AUC scores is not based on a specific cut-off score. Based on 2×2 contingency tables, which were created for each study, we then calculated – via the R package *mada* (Doebler, 2022) – sensitivity and specificity scores, and relative 95% confidence intervals. Pooled sensitivity and specificity were calculated by using a random-effects bivariate model that accounts for any possible correlations between these two metrics (Reitsma et al., 2005). Cross-study heterogeneity was tested with the Q and the I^2 indices, the latter representing the percentage of variation across studies that is due to true heterogeneity rather than chance. As per protocol, we planned to conduct prespecified subgroup analyses to account for potential heterogeneity between studies using different versions of QbTest (QbTest+ for adolescents/adults, QbMini & QbTest 6–12 for children) and studies including different control samples (non-ADHD clinical and non-ADHD non-clinical).

Results

Characteristics of studies included in the systematic review

Our initial search retrieved 240 references, of which 51 were duplicates and 163 were deemed ineligible based on title/abstract. After full-text screening of the remaining 29 references (including three references manually retrieved from reference lists of previously published reviews), 16 references (15 studies/samples) were eligible and were included in the systematic review and meta-analysis (2,058 participants; 48.6% with ADHD; Figure 1; Table 1; see Appendix S4 for list of excluded studies, with reason for exclusion). Potential risk of bias was particularly detected in relation to patient selection, in seven (46%) studies (see Appendix S3).

Table 1 provides an overview of characteristics of the studies included in our systematic review. All studies were conducted in high-income countries, with no study from Africa, South America, Asia, or Oceania. Nine studies used QbTest+ (for adolescents >12 years old, and adults), one used QbTest 6–12 (children ≤12 years old), three used both QbTest+ and QbTest 6–12 (since they had mixed samples of children, adolescents and adults), one used QbCheck (conducted at participants' home, based on both QbTest+ and QbTest 6–12 due to wide age range of the sample), and one used QbMini (for preschoolers). Seven studies (47%) reported sensitivity/specificity, two (13%) reported AUC, while six studies (40%) reported both. Eight studies reported the cut-off used to calculate sensitivity/specificity of QbTest; five used *z*-scores ≥1.5, two used ≥1.3 and one used ≥1.25, while three did not report cut-offs. In nine studies, the sample included people referred for ADHD (who were formally diagnosed with ADHD, or not, at the end of study), while in six studies, people with an established diagnosis of ADHD were compared to a sample of either non-ADHD nonclinical controls (five studies) or autistic participants (one study). All studies reported using DSM criteria for diagnosing ADHD – four used DSM-5, eight used DSM-IV and two used both DSM and ICD, while in one study, DSM criteria were used but the version was not specified (see Table 1 for full details). Although this indicates consistency across studies in using DSM/ICD criteria, we noticed some differences in how diagnostic assessment was described. Most studies (n = 13) relied on clinical assessment conducted by a healthcare professional (e.g. psychiatrist or psychologist), including semistructured interviews and standardised rating scales. However, in two studies (Edebol, Holmberg, Helldin, Gustafsson, & Norlander, 2011; Ulberstad et al., 2020), only 'clinical assessment' was mentioned, without providing further details.

Only one study (Hollis et al., 2018) investigated QbTest in line with its FDA-approved and NICE-

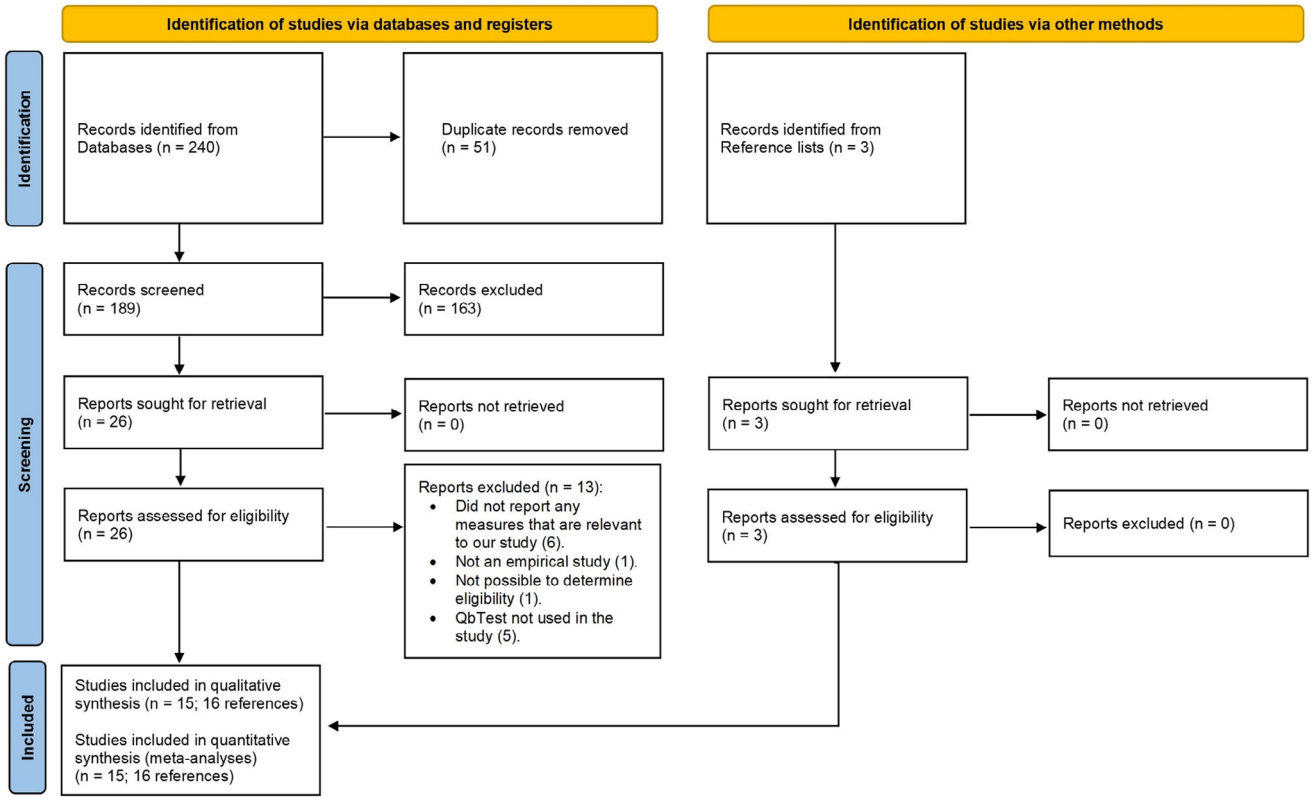

Figure 1 PRISMA flowchart

recommended intended use, that is, in combination with other assessment tools as part of clinical assessment. Most of the studies (n = 12) included in the systematic review investigated the validity of QbTest as a stand-alone tool, that is, they calculated specificity, sensitivity and/or AUC of QbTest in comparison to clinical assessment, while two studies investigated the validity of QbTest in combination with other measures, compared to clinical assessment alone: Emser et al. (2018) investigated the accuracy of QbTest – combined with a battery of attentional tasks – in discriminating between ADHD cases and non-ADHD nonclinical controls in a sample of children and adults; Groom, Young, Hall, Gillott, and Hollis (2016) analysed the capacity of QbTest – in addition to Conners Adult ADHD Rating Scale (CAARS; Conners et al., 1999) and Autism Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) to discriminate between ADHD and autism in a sample of adults.

Pooled estimate of AUC

Seven studies from eight publications reported AUC values for QbTest (Brunkhorst-Kanaan et al., 2020; Groom et al., 2016; Hamadache, Hoberg, Zaplana Labarga, & Günther, 2021; Hult, Kadesjö, Kadesjö, Gillberg, & Billstedt, 2018; Johansson et al., 2021; Labarga, Hoberg, Hamadache, & Günther, 2019; Pettersson, Söderström, & Nilsson, 2018; Söderström, Pettersson, & Nilsson, 2014). Three used *Q*-Total Score (Groom et al., 2016; Hamadache et al., 2021; Johansson et al., 2021), derived from averaging the three cardinal *Q*-scores (*Q*-Inattention, *Q*-Activity and *Q*-Impulsivity). The remaining five studies reported cardinal *Q*-scores (*Q*-Inattention, *Q*-Activity or *Q*-Impulsivity; Brunkhorst-Kanaan et al., 2020; Hult et al., 2018; Labarga et al., 2019; Pettersson et al., 2018; Söderström et al., 2014). Six studies used QbTest as a stand-alone tool, while one study (Groom et al., 2016) investigated the validity of QbTest in combination with CAARS (Conners et al., 1999) and AQ (Baron-Cohen et al., 2001).

AUC for *Q*-Total Score, across studies, ranged between 0.58 and 0.87, while for cardinal *Q*-Scores it was in the range 0.48–0.82 (see Table 2 and Table 3). Pooled AUC estimates were all higher than 0.5 (and their 95% CI did not cross 0.5, Figure 2), suggesting that QbTest was overall performing better than chance in discriminating between ADHD and non-ADHD cases. However, only AUC for *Q*-Total Score was considered acceptable, rather than good (0.72 [95% CI = 0.57; 0.87]), while the other cardinal *Q*-scores showed AUC <0.7. We conducted a sensitivity analysis and excluded Groom et al. (2016) from the meta-analysis on AUC for *Q*-Total Score, since this study investigated the validity of QbTest, CAARS and AQ combined (i.e. not QbTest alone, as all other studies did). The sensitivity analysis showed that pooled AUC was moderate (0.67 [95% CI = 0.51; 0.82]) when only considering studies that investigated QbTest as a

Table 1 Overview of studies included in the systematic review

Study	Country	QbTest version	*Q*-scores, cut-offs & validity outcomes	ADHD diagnostic assessment & criteria	Sample demographics & characteristics, and main findings
Adamou, 2022	UK	QbTest+ (adults)	Sensitivity and Specificity *Q*-Total Score (cut-off: 1.5)	DSM-5. Psychiatric assessment by a doctor with expertise in ADHD and General Psychiatry, which also included DIVA	Referred sample. 69 adults (mean age: 33 ± 9.9 years; 34.8% females; IQ >70; no information about ethnicity) referred for ADHD assessment. 38 participants (55%) received ADHD diagnosis by clinical consensus No differences between those who did and did not receive ADHD diagnosis on QbTest+ measures, but those who received the diagnosis scored higher on symptoms of inattention and hyperactivity/impulsivity (DIVA)
Bijlenga, 2019	Germany, Sweden, The Netherlands	QbTest+ (adults)	Sensitivity and Specificity *Q*-Total Score (cut-off: 1.5)	DSM-IV-TR. Diagnoses were established using DIVA, CAARS, WURS and/or BADDS. Self-reported instruments were also used: ADHD-RS and ASRS (based on location)	Research sample. 97 adults with ADHD (mean age: 63.2 ± 4.8 years; 54% females; no cognitive decline/impairment; no information about ethnicity), recruited from adult outpatient and psychiatric clinics 112 non-ADHD non-clinical controls (mean age: 64.4 ± 5.4 years; 55% females) recruited from the community The ADHD group scored worse, compared with the control group, on all QbTest parameters except commission errors. They also scored higher on number of alcohol units/week, smoking rate and self-reported ADHD symptom severity
Brunkhorst-Kanaan, 2020	Germany	QbTest+ (adults)	Sensitivity, Specificity, AUC *Q*-Inattention, *Q*-Impulsivity and *Q*-Activity (cut-off: 1.5)	DSM-5. Clinical interview including DIVA, WURS and ASRS	Referred sample. 114 adults (mean age: 35.4 years, 45.6% females; no information about ethnicity) undergoing diagnostic assessment for ADHD 94 participants (82.5%) received ADHD diagnosis. ADHD group scored worse on *Q*-Activity. Higher rate of depression was found in the non-ADHD group, while a higher percentage of patients with substance abuse disorder were found in the ADHD group
Edebol, 2011	Sweden	QbTest+ (adults)	Sensitivity and Specificity *Q*-Total Score (cut-off: 1.3)	DSM-IV. Clinical assessment	Referred sample. 19 adults (mean age 31.7 years, 47.3% females; no information about ethnicity) undergoing diagnostic assessment for ADHD QbTest was helpful for identifying most of ADHD cases (n = 10), but specificity was just above chance
Edebol,2013[a]	Sweden	QbTest+ (adults)	Sensitivity and Specificity Composite score, based on three cardinal *Q*-scores (different cut-offs analysed)	DSM-IV. Psychiatric assessment and neuropsychological tests, including ASRS	Research sample. 55 adults with ADHD (mean age: 33.35 ± 8.84 years; 55% females; no information about ethnicity), 202 non-ADHD nonclinical controls, part of a normative group (mean age: 31.06 ± 10.27 years; 44% females; no information about ethnicity) The composite score, based on cardinal *Q*-scores, correctly identified most ADHD and non-ADHD cases

(continues)

Table 1 (continued)

Study	Country	QbTest version	*Q*-scores, cut-offs & validity outcomes	ADHD diagnostic assessment & criteria	Sample demographics & characteristics, and main findings
Emser, 2018[a]	Germany	QbTest 6–12 (children) and QbTest+ (adults)	Sensitivity and Specificity *Q*-Total Score (in combination with TAP; cut-off: unspecified)	DSM-IV. Clinical interview including K-SADS-PL (children), WRI (adults) and rating scales (CAARS, CRS)	Research sample. 30 children with ADHD (mean age: 8.9 ± 1.4; 30% females; IQ > 80; no information about ethnicity) and 38 adults with ADHD (mean age: 35.1 ± 11.7; 34% females; no information about ethnicity), recruited from outpatient university clinic and local schools 30 non-ADHD nonclinical children (mean age: 8.7 ± 1.2; 37% females; no information about ethnicity) and 38 non-ADHD nonclinical adults (mean age: 32.2 ± 9.6; 34% females; no information about ethnicity) recruited from the community CAARS scores were significantly more predictive than QbTest in predicting ADHD diagnosis
Groom, 2016	UK	QbTest+ (adults)	Sensitivity, Specificity, AUC *Q*-Total Score (in addition to CAARS and AQ; cut-off: unspecified since ROC analysis was conducted)	DSM-IV. Clinical interview including DIVA, CAARS, AQ and ASRS	Research sample. 37 adults with ADHD (mean age: 8.7 ± 1.2; 37% females; no information about ethnicity) and 25 autistic adults (mean age: 8.7 ± 1.2; 37% females; no information about ethnicity) QbTest had more specificity than CAARS in identifying ADHD cases, and it improved classification accuracy when combined with CAARS and AQ, compared with using the rating scales alone
Hamadache, 2021; Labarga, 2019	Germany	QbMini (preschoolers)	Sensitivity, Specificity, AUC *Q*-Total Score and cardinal *Q*-scores (cut-off: unspecified)	DSM-IV/ ICD-10. Clinical interview and observation of child playing with others. CBCL and FBB-ADHS were used, but not for diagnosis	Research sample. 40 children with ADHD (mean age: 5.54 ± 0.27; 19% females; no information about ethnicity), 26 children with specific language impairment (mean age: 5.45 ± 0.25; 33% females; no information about ethnicity) and 55 non-ADHD nonclinical children (mean age: 5.45 ± 0.27; 44% females; no information about ethnicity) Children with ADHD performed worse than non-ADHD non-clinical controls on QbMini indices, but not differently than those with specific language impairment. The CBCL and FBB-ADHS were better in predicting ADHD versus non-ADHD nonclinical diagnosis but were as poor as QbMini in differentiating between ADHD and children with specific language impairment

(continues)

Table 1 (continued)

Study	Country	QbTest version	*Q*-scores, cut-offs & validity outcomes	ADHD diagnostic assessment & criteria	Sample demographics & characteristics, and main findings
Hollis, 2018	UK	QbTest 6–12 (children) and QbTest+ (adolescents)	Sensitivity and Specificity *Q*-Total Score (cut-off: unspecified).	DSM-5/ICD-10. 'Gold standard' DAWBA, CGAS, SNAP, compared with routine clinical assessment including QbTest (QbOpen) versus without using QbTest (QbBlind)	Referred sample. 250 children (mean age: 9.5 ± 2.8 years; 22% females; 90% White, 10% Mixed and other) referred for ADHD assessment Speed to diagnostic decision increased by 44% with access to QbTest. No differences in diagnostic accuracy for clinicians who had or did not have access to QbTest
Hult, 2018	Sweden	QbTest 6–12 (children)	Sensitivity, Specificity, AUC *Q*-Inattention, *Q*-Impulsivity and *Q*-Activity (cut-off: 1.25)	DSM-IV. Gold-standard clinical assessment by a multi-professional team using LEAD procedure	Referred sample. 182 children (mean age: 10.6 years, 18% females; no information about ethnicity) undergoing diagnostic assessment for ADHD 124 children were diagnosed with ADHD, while 58 were considered as part of a non-ADHD clinical comparison group. Worse QbTest performance was found in those diagnosed with ADHD, versus non-ADHD clinical cases
Johansson, 2021	Sweden	QbTest+ (adolescents)	Sensitivity, Specificity, AUC *Q*-Total Score (cut-off: unspecified)	DSM (version not reported). Clinical assessment (blind) including the A-TAC inventory, K-SADS, CGAS and SDQ	Referred sample. 340 adolescents (median Age: 15 years; 45% females; no information about ethnicity) from Swedish twin registry, screened for ADHD and clinically assessed for ADHD Adolescents with ADHD performed worse than non-ADHD on QbTest, but the ability of QbTest to correctly classify ADHD was poor
Pettersson, 2018[a]	Sweden	QbTest+ (adults)	Sensitivity, Specificity, AUC *Q*-Activity and *Q*-Inattention (cut-off: 1.5)	DSM-IV. Clinical assessment that included interview and DIVA	Referred sample. 60 adults with ADHD (mean age: 28.18 ± 9.09 years, 47% females; no information about ethnicity) and 48 non-ADHD adults (clinical comparison group; mean age: 32.75 ± 10.61 years, 48% females; no information about ethnicity) Both the ADHD group and the non-ADHD clinical comparison group performed above the normative mean on all QbTest domains. However, DIVA was superior than QbTest (Inattention and Activity indices) in predicting ADHD diagnosis
Söderström, 2014	Sweden	QbTest+ (adults)	Sensitivity, Specificity, AUC *Q*-Inattention, *Q*-Impulsivity, *Q*-Activity (cut-off: 1.5)	DSM-IV. Clinical and psychiatric interview/ assessment, ASRS and CSS	Referred sample. 61 adults (mean age: 31.7 years; 57% females; no information about ethnicity) undergoing diagnostic assessment for ADHD 41 participants (67%) met diagnostic criteria for ADHD. Worse QbTest performance in the ADHD-diagnosed group. *Q*-Inattention and *Q*-Activity added significant predictive power to ASRS and CSS

(continues)

Table 1 (continued)

Study	Country	QbTest version	*Q*-scores, cut-offs & validity outcomes	ADHD diagnostic assessment & criteria	Sample demographics & characteristics, and main findings
Tallberg, 2018	Sweden	QbTest 6–12 (children) and QbTest+ (adolescents)	AUC *Q*-Inattention, *Q*-Impulsivity, *Q*-Activity (cut-off: unspecified since ROC analysis was conducted)	DSM-IV. Multidisciplinary assessment and consensus diagnosis, SNAP, and neuropsychological assessment	Referred sample. 118 children (median age: 12 years. 31% females; no information about ethnicity) screened for ADHD and referred for assessment 80 children (68%) met diagnostic criteria for ADHD. ADHD group performed worse only on *Q*-Impulsivity, but not *Q*-Activity or *Q*-Inattention. QbTest measures were not statistically significant in predicting ADHD diagnosis
Ulberstad, 2020	Germany, Sweden, USA	QbCheck (based on QbTest 6–12 and QbTest+)	Sensitivity and Specificity *Q*-Total Score (cut-off: unspecified)	DSM-5. Clinical assessment	Research sample. 69 adults with ADHD (mean age: 27.58 ± 12.12 years; 48% females; no information about ethnicity) and 73 non-ADHD nonclinical controls (mean age: 26.16 ± 9.55 years; 56% females; no information about ethnicity) recruited from the community Adults with ADHD showed worse performance than non-ADHD non-clinical controls on QbCheck, which had good accuracy in identifying ADHD cases

ADHD, Attention-Deficit/Hyperactivity Disorder; ADHD-RS, ADHD Rating Scale; AQ, Autism Quotient; ASRS, Adult ADHD Self-Report Scale; A-TAC, Autism-Tics, ADHD, and Other Comorbidities; AUC, Area Under the Receiver Operating Characteristics Curve; BADDS, Brown Attention Deficit Disorder Scales for Adults; CAARS, Conners' Adult ADHD Rating Scales; CGAS, Children's Global Assessment Scale; CRS, Conner's Rating Scales; CSS, Current Symptoms Scale; DAWBA, Development and Well-being Assessment; DIVA, Diagnostic interview for ADHD in adults; DSM, Diagnostic and Statistical Manual of Mental Disorders; FBB-ADHS-V, Fremdbeurteilungsbogen für Vorschüler mit Aufmerksamkeits- und Hyperaktivitätsstörungen; ICD, International Classification of Diseases; IQ, Intellectual Quotient; K-SADS, Schedule for Affective Disorders and Schizophrenia for School-Age Children interview; LEAD, Longitudinal, Experts, All, Data; ROC, Receiver Operating Characteristics Curve; SDQ, Strengths and Difficulties Questionnaire; SNAP, Swanson, Nolan and Pelham; TAP, Test battery of attention; WRI, Wender Reimherr Interview; WURS, Wender-Utah Rating Scale.

[a]Study retrieved from reference lists of previous systematic or narrative reviews.

Table 2 Pooled AUC and sensitivity/specificity scores for *Q*-Total Score and cardinal *Q*-scores

	N of studies	*N* of participants	Pooled AUC [95% CI]	Heterogeneity
Q-Total score	4	554	0.72 [0.57; 0.87]	*Q*: 33.37[#]. I^2: 91.0%
Q-Inattention	6	682	0.66 [0.59; 0.72]	*Q*: 12.16[#]. I^2: 58.9%
Q-Impulsivity	5	574	0.59 [0.53; 0.64]	*Q*: 5.51. I^2: 27.4%
Q-Activity	6	682	0.67 [0.58; 0.77]	*Q*: 23.81[#]. I^2: 79.0%

	N of studies	*N* of participants	Pooled Sensitivity [95% CI]	Pooled Specificity [95% CI]	Heterogeneity
Q-Total score	9	1,354	0.78 [0.69; 0.85]	0.70 [0.57; 0.81]	I^2: 30.9%
Q-Inattention	3	354	0.48 [0.35; 0.61]	0.83 [0.60; 0.94]	n/a[a]
Q-Impulsivity	2	243	0.49 [0.33; 0.65]	0.76 [0.63; 0.86]	n/a[a]
Q-Activity	4	469	0.65 [0.52; 0.75]	0.65 [0.48; 0.78]	n/a[a]

[#]$p < .05$.
[a]Not possible to calculate, due to limited number of included studies.

stand-alone tool and not in combination with other measures.

There was significant heterogeneity for all measures except *Q*-Impulsivity (see Table 2). Due to the scarcity of studies, it was not possible to conduct subgroup analyses as planned. We could therefore only consider the data descriptively. We observed an overlap in the 95% CI of AUC scores reported by studies that used QbMini or QbTest 6–12 in preschoolers and children (0.69 [95% CI = 0.61; 0.77]) versus QbTest+ in adolescents and adults (0.66 [95% CI = 0.60; 0.72]), and AUC scores reported by studies including non-ADHD non-clinical (0.63; [95% CI = 0.58; 0.68]) versus non-ADHD clinical cases (0.64; [95% CI = 0.60; 0.68]) as comparison groups.

Pooled estimates of sensitivity and specificity

Thirteen studies reported sensitivity as well as specificity (Adamou et al., 2022; Bijlenga et al., 2019; Brunkhorst-Kanaan et al., 2020; Edebol et al., 2011; Edebol, Helldin, & Norlander, 2013; Emser et al., 2018; Groom et al., 2016; Hollis et al., 2018; Hult et al., 2018; Johansson et al., 2021; Pettersson et al., 2018; Söderström et al., 2014; Ulberstad et al., 2020). Of these, nine reported sensitivity and specificity for *Q*-Total Score (Adamou et al., 2022; Bijlenga et al., 2019; Edebol et al., 2011, 2013; Emser et al., 2018; Groom et al., 2016; Hollis et al., 2018; Johansson et al., 2021; Ulberstad et al., 2020), and four for cardinal *Q*-scores (Brunkhorst-Kanaan et al., 2020; Hult et al., 2018; Pettersson et al., 2018; Söderström et al., 2014).

Sensitivity ranged between 0.56 and 0.95 for *Q*-Total Score, and between 0.37 and 0.77 for cardinal *Q*-scores. Specificity ranged between 0.36 and 0.84 for *Q*-Total Score, and between 0.44 and 0.96 for cardinal *Q*-scores. Pooled sensitivity and specificity for *Q*-Total Score were 0.78 [95% CI = 0.69; 0.85] and 0.70 [95% CI = 0.57; 0.81], respectively, suggesting acceptable, rather than good, sensitivity and specificity of the *Q*-Total Score in discriminating between ADHD and non-ADHD cases (Table 2; Figure 3). We conducted a sensitivity analysis and excluded Emser et al. (2018) and Groom et al. (2016) from this meta-analysis, as these studies did not investigate QbTest as a stand-alone tool. The results showed acceptable sensitivity for *Q*-Total Score but only moderate specificity (sensitivity: 0.76, [95% CI = 0.66; 0.84]; specificity: 0.66, [95% CI = 0.49; 0.79]). In Hollis et al. (2018), where QbTest was used in support of clinical assessment, no significant differences in diagnostic accuracy for clinicians who had or did not have access to QbTest during clinical assessment were found. However, QbTest reduced the time needed for an ADHD diagnosis and improved clinicians' confidence in confirming or ruling out ADHD. In Emser et al. (2018), QbTest did not prove helpful in discriminating between ADHD and non-ADHD cases in addition to CAARS, while Groom et al. (2016) found that adding QbTest to CAARS and AQ improved the accuracy of discrimination between ADHD and non-ADHD cases (classification accuracy improved from 81% to 90%).

Pooled sensitivity was moderate for *Q*-Activity (0.65 [95% CI = 0.52; 0.75]) and low for *Q*-Inattention (0.48 [95% CI = 0.35; 0.61]) and *Q*-Impulsivity (0.49 [95% CI = 0.33; 0.65]). Pooled specificity was good for *Q*-Inattention (0.83 [95% CI = 0.60; 0.94]), acceptable for *Q*-Impulsivity (0.76 [95% CI = 0.63; 0.86]) and moderate for *Q*-Activity (0.65 [95% CI = 0.48; 0.78]). Cross-study heterogeneity could only be computed for *Q*-Total Score and was nonsignificant (I^2 = 30.9%; Q not available). As for AUC measures, due to the scarcity of studies, it was not possible to conduct subgroup analyses. However, we observed overlapping 95% CI for sensitivity and specificity (for *Q*-total and cardinal scores, combined) reported by studies that used QbTest+ in adolescents and adults (sensitivity = 0.67 [95% CI = 0.57; 0.76]; specificity = 0.70 [95% CI = 0.60; 0.79]) and those using QbMini or QbTest 6–12 in preschoolers and children (sensitivity = 0.62 [95% CI = 0.39; 0.80];

Table 3 Sensitivity, specificity and AUC, and relative 95% CI, for each study included in the review

Study	Developmental stage	Non-ADHD comparison group	Test(s)	Total sample size	Sensitivity [95% CI]	Specificity [95% CI]	AUC [95% CI]
Q-total score							
Adamou, 2022	Adults	Referred but not diagnosed	QbTest	69	0.71 [0.55; 0.83]	0.42 [0.26; 0.60]	–
Bijlenga, 2019	Older Adults	Non-ADHD nonclinical controls	QbTest	209	0.56 [0.46; 0.65]	0.83 [0.75; 0.89]	–
Edebol, 2011	Adults	Referred but not diagnosed	QbTest	19	0.83 [0.52; 0.96]	0.57 [0.23; 0.86]	–
Edebol, 2013[a]	Adults	Non-ADHD nonclinical controls	QbTest	257	0.85 [0.74; 0.93]	0.83 [0.77; 0.87]	–
Emser, 2018[a]	Children and Adults	Non-ADHD nonclinical controls	QbTest + TAP	136	0.77 [0.65; 0.85]	0.82 [0.71; 0.90]	–
Groom, 2016	Adults	Autistic group	QbTest, CAARS and AQ	62	0.95 [0.81; 0.99]	0.84 [0.64; 0.94]	0.87 [0.78; 0.96]
Hamadache, 2021; Labarga, 2019	Children	Non-ADHD nonclinical controls	QbMini	92	–	–	0.82 [0.72; 0.91]
Hamadache, 2021; Labarga, 2019	Children	SLI	QbMini	63	–	–	0.60 [0.45; 0.74]
Hollis, 2018	Children	Referred but not diagnosed	QbTest in support of clinical assessment	123	0.86 [0.77; 0.91]	0.36 [0.20; 0.56]	–
Johansson, 2021	Adolescents	Referred but not diagnosed	QbTest	337	0.67 [0.57; 0.76]	0.59 [0.52; 0.64]	0.58 [0.51; 0.65]
Ulberstad, 2020	Children and Adults	Non-ADHD nonclinical controls	QbCheck	142	0.83 [0.72; 0.90]	0.80 [0.69; 0.87]	–
Q-Inattention							
Brunkhorst-Kanaan, 2020	Adults	Referred but not diagnosed	QbTest	118	–	–	0.56 [0.43; 0.68]
Hamadache, 2021; Labarga, 2019	Children	Non-ADHD nonclinical controls	QbMini	95	–	–	0.63 [0.52; 0.75]
Hult, 2018	Children	Referred but not diagnosed	QbTest	182	0.48 [0.39; 0.57]	0.83 [0.71; 0.91]	0.76 [0.69; 0.83]
Pettersson, 2018[a]	Adults	Referred but not diagnosed	QbTest	108	0.58 [0.46; 0.70]	0.67 [0.52; 0.78]	0.67 [0.57; 0.77]
Söderström, 2014	Adults	Referred but not diagnosed	QbTest	61	0.37 [0.23; 0.52]	0.96 [0.74; 0.99]	0.69 [0.56; 0.83]
Tallberg, 2018	Children	Referred but not diagnosed	QbTest alone	118	–	–	0.59 [0.48; 0.70]
Q-Impulsivity							
Brunkhorst-Kanaan, 2020	Adults	Referred but not diagnosed	QbTest	118	–	–	0.54 [0.41; 0.67]
Hamadache, 2021; Labarga, 2019	Children	Non-ADHD nonclinical controls	QbMini	95	–	–	0.49 [0.38; 0.61]
Hult, 2018	Children	Referred but not diagnosed	QbTest	182	0.42 [0.34; 0.51]	0.72 [0.60; 0.82]	0.62 [0.54; 0.70]
Söderström, 2014	Adults	Referred but not diagnosed	QbTest	61	0.59 [0.43; 0.72]	0.80 [0.57; 0.92]	0.68 [0.55; 0.82]
Tallberg, 2018	Children	Referred but not diagnosed	QbTest alone	118	–	–	0.60 [0.49; 0.71]
Q-Activity							
Brunkhorst-Kanaan, 2020	Adults	Referred but not diagnosed	QbTest	118	0.50 [0.40; 0.60]	0.71 [0.50; 0.85]	0.65 [0.54; 0.77]
Hamadache, 2021; Labarga, 2019	Children	Non-ADHD nonclinical controls	QbMini	95	–	–	0.82 [0.73; 0.91]
Hult, 2018	Children	Referred but not diagnosed	QbTest	182	0.63 [0.54; 0.71]	0.74 [0.61; 0.84]	0.74 [0.67; 0.81]
Pettersson, 2018[a]	Adults	Referred but not diagnosed	QbTest	108	0.77 [0.64; 0.86]	0.44 [0.31; 0.58]	0.66 [0.56; 0.77]
Söderström, 2014	Adults	Referred but not diagnosed	QbTest	61	0.68 [0.53; 0.81]	0.65 [0.43; 0.82]	0.67 [0.53; 0.81]
Tallberg, 2018	Children	Referred but not diagnosed	QbTest alone	118	–	–	0.48 [0.37; 0.59]

–: Not reported in the study.
[a]Study retrieved from reference lists of previous systematic or narrative reviews.

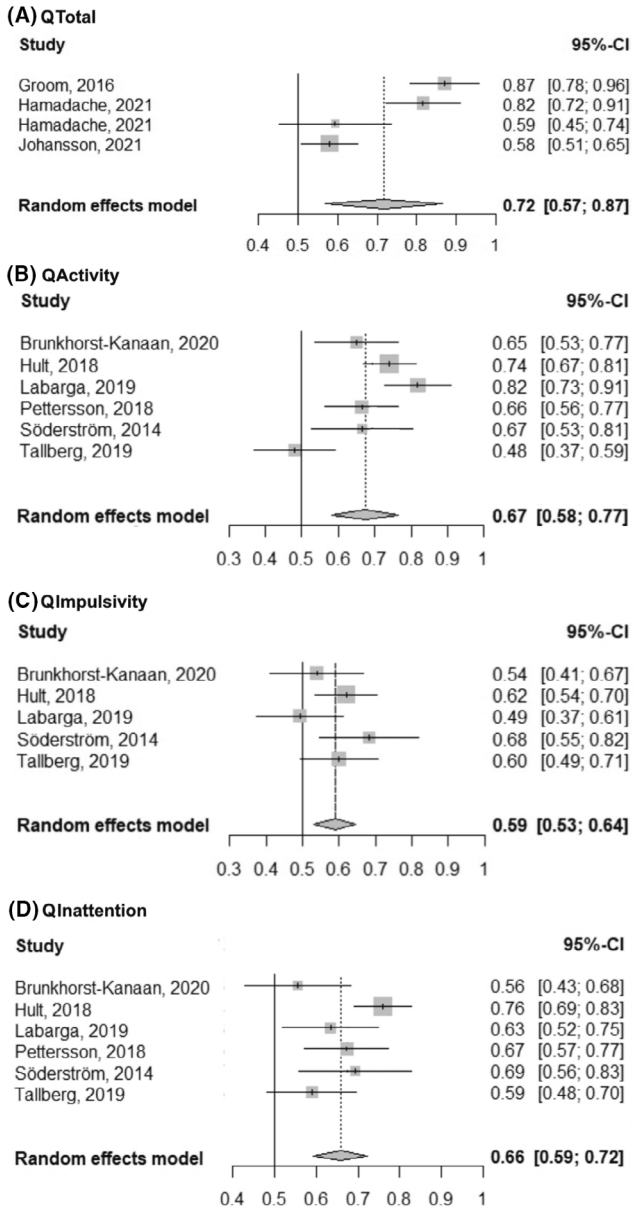

Figure 2 Forest plots of the meta-analysis of studies investigating AUC measures for *Q*-Total Score and cardinal *Q*-scores

specificity = 0.68 [95% CI = 0.47; 0.84]). Moreover, we observed higher specificity (but not sensitivity) reported by studies that involved, as comparison group, non-ADHD nonclinical controls (sensitivity = 0.76 [95% CI = 0.61; 0.86]; specificity = 0.82 [95% CI = 0.78; 0.86]) versus non-ADHD cases (i.e. including clinical participants; sensitivity = 0.63 [95% CI = 0.54; 0.71]; specificity = 0.66 [95% CI = 0.56; 0.74]).

Discussion

We conducted the first systematic review and meta-analysis investigating the clinical utility and validity of QbTest for the assessment/diagnosis of ADHD. Reflecting the current literature, our results are mostly derived from studies that assessed the validity of QbTest as a stand-alone test, which is not in line with its intended and FDA-approved use.

We found that using *Q*-Total Scores – arbitrarily calculated by averaging *Q*-Inattention, *Q*-Activity and *Q*-Impulsivity cardinal scores, but not available to users – showed acceptable, rather than good, sensitivity, specificity and AUC. We also found that using *Q*-Total Scores showed lower accuracy in discriminating ADHD from non-ADHD clinical cases, compared to when non-ADHD nonclinical controls were the control group. This is particularly relevant considering that QbTest is mainly used, in clinical practice, to discriminate between ADHD and non-ADHD clinical cases, and less frequently to discriminate between ADHD cases and non-ADHD nonclinical controls. Focusing on the three cardinal *Q*-scores – the scores available to users of QbTest – separately, led to mixed results. *Q*-Activity had moderate sensitivity, while low sensitivity was found for *Q*-Inattention and *Q*-Impulsivity. Conversely, good specificity was found for *Q*-Inattention, but it was only acceptable for *Q*-Impulsivity and moderate for *Q*-Activity.

Overall, these findings indicate that QbTest, as a stand-alone tool, does not show good discriminative ability between ADHD and non-ADHD cases. Notably, only a few studies used QbTest as part of a diagnostic process, instead as a stand-alone tool. Good discriminative ability (i.e. AUC >0.8) was obtained by studies where QbTest was used in combination with other measures (e.g. Groom et al., 2016) or to discriminate between ADHD cases and non-ADHD nonclinical controls (see Hamadache et al., 2021, for *Q*-Total Score; and Labarga et al., 2019, for *Q*-Activity).

Our findings have important implications for practitioners using QbTest. First, using the combined *Q*-Total Score (compared to using separate cardinal *Q*-scores) could support the discrimination between ADHD and non-ADHD cases, even though accuracy metrics were acceptable rather than good. However, clinicians using QbTest in their daily practice have access only to the three cardinal *Q*-scores. Conversely, focusing on *Q*-Inattention (for which, good specificity but low sensitivity was found) may not only lead to more confidence in confirming ADHD but also run the risk of missing (ruling out) some individuals who have ADHD. Although we could not conduct subgroup analyses (due to the scarcity of studies included in the meta-analyses), we found increased specificity in studies where non-ADHD nonclinical controls were used as control group (compared to when non-ADHD clinical cases were the comparison group). Considering that most (if not all) children and young people who pass the triage system in CAMHS and are formally assessed are clinical cases, clinicians need instruments that help them discriminate ADHD from other clinical conditions, in which circumstances QbTest appears less accurate.

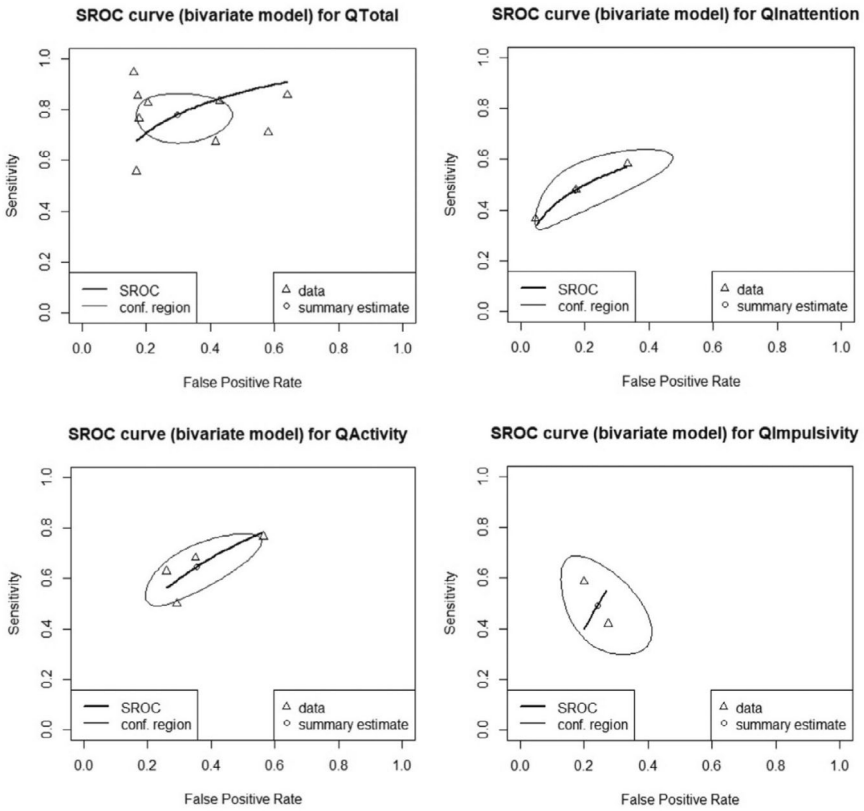

Figure 3 SROC curve plots for studies investigating sensitivity and specificity of *Q*-Total Score and cardinal *Q*-scores. The Summary receiver operating characteristic (SROC) plots represent the trade-off between sensitivity and specificity (lower values of false-positive rate indicate higher specificity). Each triangle in the plots represents sensitivity and specificity data for a single study, while the small circle (*summary estimate*) represents the pooled sensitivity and specificity data, across studies, for a particular QbTest score (e.g. Total Score). The bigger circle around the summary estimate (*conf. region*) indicates the 95% confidence region for the pooled sensitivity/specificity data. SROC curves (bold lines) that occupy the upper left space of the plot indicate high discriminative ability for that specific QbTest score

Second, QbTest is not sufficiently accurate in discriminating between ADHD and non-ADHD cases to be used as a stand-alone tool. Therefore, there is no evidence that QbTest can be used as a screening tool (e.g. to detect those who should be prioritised for full assessment) or to confirm/rule out ADHD, without conducting a healthcare professional-led clinical assessment. Importantly, we observed that most of the research to date has investigated QbTest in a way that is not aligned with its intended and approved use. Further research is therefore needed to establish the added value of QbTest when combined with other measures as part of a multidisciplinary, multicomponent assessment. Third, clinicians should be careful when using specific *Q*-score cut-offs (e.g. 1.5 or 1.25) to guide clinical decisions as it is not clear how well these cut-offs perform.

The present study has some limitations, related to the included studies rather than to the systematic review and meta-analysis per se. A limited number of eligible studies were retrieved, most of which conducted in European countries or the USA. However, considering that QbTest was approved by the FDA only 10 years ago, and that most included studies were published in the last 5 years, we acknowledge the fact that this is a recent area of research. Second, a significant percentage of studies included in our review were on adults (53%), and only a very limited percentage of studies included children whose mean age was less than 10 years old (20%), which is the age of most cases referred for ADHD assessment to child and adolescent services. More research is needed to understand if using the QbTest with preschoolers and younger children will improve accuracy of differential diagnosis in referred samples. Third, potential risk of bias was detected in many studies in relation to patient selection; future studies should recruit people with a formal or suspected diagnosis of ADHD, which should be compared to heterogeneous samples of non-ADHD nonclinical controls and those with other clinical conditions (e.g. to analyse variations in QbTest performance dependent on cognitive abilities and presence of comorbidities). This will allow clarification of those cases in which QbTest has the potential to be more helpful, that is, when used alongside other clinical tools, to discriminate between ADHD and non-ADHD cases.

Fourth, significant cross-study heterogeneity was detected in almost every meta-analysis, or it was not possible to calculate it due to the scarcity of studies meta-analysed, indicating that study and sample characteristics may have influenced our findings, which should therefore not be considered conclusive. For example, although there are similarities across international clinical guidelines (Coghill et al., 2021), there was heterogeneity in the protocols and instruments used for 'gold-standard' assessment of ADHD across studies (e.g. using different scales or details not reported). Considering that there is not a truly 'gold-standard' (i.e. 100% accurate) method for ADHD assessment, the clinician's judgement – integrating all different sources of information, including parent, teacher and young-person reports as well as direct observations of the person with suspected ADHD – is considered the best practice (NICE, 2018). There was also heterogeneity in the study setting. In most studies, participants were part of a clinically referred group who underwent diagnostic assessment (similar to what is usually seen in mental health or child and adolescent services), but there were also studies where participants with an already established diagnosis of ADHD were asked to complete QbTest as part of a research project (these samples may be less representative of the ADHD population, e.g. referral samples may present with more co-occurring conditions than research samples). Lastly, although all samples were heterogeneous in relation to ADHD presentations, only two studies specifically investigated differences between ADHD-Combined and ADHD-Inattentive presentations. Hult et al. (2018) found that QbTest performed similarly in people with ADHD-Combined and ADHD-Inattentive presentations (especially when considering *Q*-Inattention and *Q*-Activity). Conversely, Petterson et al. (2018) found that *Q*-Activity was able to discriminate between ADHD-Combined and ADHD-Inattentive presentations. All other studies did not assess possible differences on QbTest performances across ADHD presentations.

Further research aimed at providing clear evidence-based guidelines on how to best integrate different sources of information (including QbTest) is urgently needed. To support a more consistent and appropriate use of QbTest in clinical practice, further rigorous studies should be conducted to evaluate its role and added value as an aid to diagnostic assessment, and its cost-effectiveness, possibly alongside other technologies (e.g. virtual reality (Goharinejad, Goharinejad, Hajesmaeel-Gohari, & Bahaadinbeigy, 2022) or wearable technologies (Welch et al., 2022) that are in their infancy in terms of clinical implementation. As done by Hollis et al. (2018), we encourage researchers to conduct studies to investigate to what extent adding QbTest to 'traditional' clinical assessment increases the diagnostic agreement between clinicians and, therefore, diagnostic reliability. We recommend the development of guidelines to inform clinicians how best to integrate QbTest in their practice, for example, where in the clinical assessment pathway they would benefit more from QbTest results (before or after identifying a preliminary diagnosis?) and how receiving QbTest results in line with (or against) their clinical judgement influences their final diagnostic decision.

Conclusion

In line with FDA-approved and NICE-recommended intended use of QbTest as an adjunctive decision-aid to clinical assessment, we found evidence that QbTest is not sufficiently accurate in discriminating between ADHD and non-ADHD cases to be used as a stand-alone diagnostic tool. By contrast, QbTest should be used in support of clinical assessment, with the aim of augmenting the clinician's judgement, improving the efficiency of the diagnostic process, and leading to earlier intervention. Our findings support the recent NICE Medtech Innovation Briefing (NICE, 2023) which concluded that QbTest should be used as an addition to routine clinical assessment of ADHD and not as a standalone assessment. When used as an adjunctive decision-aid, the potential benefits are quicker assessment and cost savings because of reduced clinician time and greater efficiency of the pathway. Further research is needed to understand how, exactly, the QbTest should be integrated with clinical information. The use of QbTest as stand-alone tool or as a triage system to address long waiting lists in child and adolescent mental services should be discouraged. Until the pathophysiology and underlying mechanisms of ADHD are fully understood, fully valid tests of ADHD cannot be designed. Therefore, the 'gold-standard' method for assessing and diagnosis ADHD remains a clinician-led assessment including semistructured clinical interviews and self-, parent- or teacher-reports aimed at providing an overview of the behavioural characteristics of the assessed case, which should be evaluated against DSM- and/or ICD-based diagnostic criteria for ADHD. Objective tests may supplement, rather than replace, this process.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. PRISMA Checklist.

Appendix S2. Search strategy (last search: 15th August 2022).

Appendix S3. Quality appraisal of studies included in the review.

Appendix S4. List of excluded papers after full-text screening.

Acknowledgements

The authors would like to thank Pia Tallberg for providing additional information about their publication and data for the meta-analysis. A.B. does not have any interests to declare. C.L.H. was a research fellow and trial manager for the NIHR funded AQUA trial of QbTest; and she supported the East Midlands Academic Health Sciences Centre's Focus ADHD programme involving QbTest. M.J.G. received travel expenses from QbTech in 2013–2016. E.S. was Senior Clinical Advisor to the NICE ADHD guideline [NG87]. A.T.'s research is funded by the Wolfson Foundation and Wellcome Trust; she is a trustee (unpaid) of the UK Charity ADHD Foundation. C.H. was a member of the NICE guideline committee for ADHD [NG87], Expert Advisor to the NICE Medtech innovation briefing [MIB318] for QbTest, and chief investigator for the NIHR funded AQUA trial of QbTest. S.C. declares honoraria and reimbursement for travel and accommodation expenses for lectures from the following nonprofit associations: the Association for Child and Adolescent Mental Health (ACAMH), the Canadian ADHD Resource Alliance (CADDRA), the British Association for Pharmacology (BAP); and from Healthcare Convention for educational activity on ADHD. S.C. is currently supported by funding from the National Institute for Health and Care Research, NIHR (grants NIHR203684, NIHR203035, NIHR130077, NIHR128472, RP-PG-0618-20003), Solent NHS Trust (Research Capability Funding, 2022), University of Southampton (Knowledge and Exchange Enterprise Fund 2023, Enterprise Development Fund 2022) and European Research Agency (Horizon: Project 101095568).

Correspondence

Alessio Bellato, School of Psychology, University of Nottingham Malaysia, Jalan Broga, Semenyih, Selangor, 43500, Malaysia; Email: alessio.bellato@nottingham.edu.my

Key points

- We conducted the first systematic review and meta-analysis investigating the clinical utility and accuracy of QbTest [an FDA-approved test used as an adjunctive decision-aid to attention-deficit/hyperactivity disorder (ADHD) clinical assessment] in discriminating ADHD from non-ADHD cases.
- Based on 15 studies included in our meta-analysis, we found evidence that QbTest is not adequately specific or sensitive to be used as a stand-alone diagnostic tool to replace a full clinical assessment or as a triage system to allocate to assessment waiting-list individuals referred to clinical services.
- QbTest should be used as an aid to support clinical judgement where it can reduce the time to diagnosis.

Data availability statement

Data is presented in the manuscript and supplementary materials. The raw data and R codes used for the meta-analyses are available upon request to the corresponding author.

References

Adamou, M., Fullen, T., & Jones, S.L. (2020). EEG for diagnosis of adult ADHD: A systematic review with narrative analysis. *Frontiers in Psychiatry*, *11*, 871.

Adamou, M., Jones, S.L., Marks, L., & Lowe, D. (2022). Efficacy of continuous performance testing in adult ADHD in a clinical sample using QbTest. *Journal of Attention Disorders*, *26*, 1483–1491.

American Psychiatric Association. (2022). Neurodevelopmental disorders. In *Diagnostic and statistical manual of mental disorders* (5th edn, text rev.), Washington, DC: American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425787.x01_Neurodevelopmental_Disorders

Arrondo, G., Mulraney, M., Iturmendi-Sabater, I., Musullulu, H., Gambra, L., Niculcea, T., … Cortese, S. (2023). Systematic review and meta-analysis: Clinical utility of continuous performance tests for the identification of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*. https://doi.org/10.1016/j.jaac.2023.03.011

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17.

Bijlenga, D., Ulberstad, F., Thorell, L.B., Christiansen, H., Hirsch, O., & Kooij, J.J.S. (2019). Objective assessment of attention-deficit/hyperactivity disorder in older adults compared with controls using the QbTest. *International Journal of Geriatric Psychiatry*, *34*, 1526–1533.

Brunkhorst-Kanaan, N., Verdenhalven, M., Kittel-Schneider, S., Vainieri, I., Reif, A., & Grimm, O. (2020). The quantified behavioral test – A confirmatory test in the diagnostic process of adult ADHD? *Frontiers in Psychiatry*, *11*, 216.

Buitelaar, J., Bölte, S., Brandeis, D., Caye, A., Christmann, N., Cortese, S., … Banaschewski, T. (2022). Toward precision medicine in ADHD. *Frontiers in Behavioral Neuroscience*, *16*, 900981.

Coghill, D., Banaschewski, T., Cortese, S., Asherson, P., Brandeis, D., Buitelaar, J., … Simonoff, E. (2021). The management of ADHD in children and adolescents: Bringing evidence to the clinic: Perspective from the European ADHD guidelines group (EAGG). *European Child & Adolescent Psychiatry*, *32*, 1337–1361.

Conners, C.K., Erhardt, D., Epstein, J.N., Parker, J.D.A., Sitarenios, G., & Sparrow, E. (1999). Self-ratings of ADHD symptoms in adults I: Factor structure and normative data. *Journal of Attention Disorders*, *3*, 141–151.

Cortese, S., Solmi, M., Michelini, G., Bellato, A., Blanner, C., Canozzi, A., … Correll, C.U. (2023). Candidate diagnostic biomarkers for neurodevelopmental disorders in children and adolescents: A systematic review. *World Psychiatry*, *22*, 129–149.

Culpepper, L., & Mattingly, G. (2010). Challenges in identifying and managing attention-deficit/hyperactivity disorder in adults in the primary care setting: A review of the literature. *Primary Care Companion to the Journal of Clinical Psychiatry*, *12*, PCC.10r00951.

Doebler, P. (2022). Meta-Analysis of Diagnostic Accuracy [R package mada version 0.5.11]. R-Project.org.r-forge.r-project.org/projects/mada/.

Edebol, H., Helldin, L., & Norlander, T. (2013). Measuring adult attention deficit hyperactivity disorder using the quantified behavior test plus. *PsyCh Journal*, *2*, 48–62.

Edebol, H., Holmberg, E., Helldin, L., Gustafsson, S.A., & Norlander, T. (2011). In search for objective measures of hyperactivity, impulsivity and inattention in adult attention deficit hyperactivity disorder using the quantified behavior test plus. *Europes Journal of Psychology*, *7*, 443–457.

Emser, T.S., Johnston, B.A., Steele, J.D., Kooij, S., Thorell, L., & Christiansen, H. (2018). Assessing ADHD symptoms in children and adults: Evaluating the role of objective measures. *Behavioral and Brain Functions*, *14*, 11.

Epstein, J.N., Erkanli, A., Conners, C.K., Klaric, J., Costello, J.E., & Angold, A. (2003). Relations between continuous performance test performance measures and ADHD behaviors. *Journal of Abnormal Child Psychology*, *31*, 543–554.

Faraone, S.V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M.A., … Wang, Y. (2021). The World Federation of ADHD International Consensus Statement: 208 evidence-based conclusions about the disorder. *Neuroscience and Biobehavioral Reviews*, *128*, 789–818. https://doi.org/10.1016/j.neubiorev.2021.01.022

Faraone, S.V., Biederman, J., & Mick, E. (2006). The age-dependent decline of attention deficit hyperactivity disorder: A meta-analysis of follow-up studies. *Psychological Medicine*, *36*, 159–165.

Fayyad, J., De Graaf, R., Kessler, R., Alonso, J., Angermeyer, M., Demyttenaere, K., … Jin, R. (2007). Cross-national prevalence and correlates of adult attention-deficit hyperactivity disorder. *The British Journal of Psychiatry*, *190*, 402–409.

Gardner, J. (2017). Statistical Testing for AUC Data [R package auctestr version 1.0.0]. R-Project.org. https://cran.r-project.org/package=auctestr

Gloss, D., Varma, J.K., Pringsheim, T., & Nuwer, M.R. (2016). Practice advisory: The utility of EEG theta/beta power ratio in ADHD diagnosis: Report of the guideline development, dissemination, and implementation Subcommittee of the American Academy of neurology. *Neurology*, *87*, 2375–2379.

Goharinejad, S., Goharinejad, S., Hajesmaeel-Gohari, S., & Bahaadinbeigy, K. (2022). The usefulness of virtual, augmented, and mixed reality technologies in the diagnosis and treatment of attention deficit hyperactivity disorder in children: An overview of relevant studies. *BMC Psychiatry*, *22*, 4.

Groom, M.J., Young, Z., Hall, C.L., Gillott, A., & Hollis, C. (2016). The incremental validity of a computerised assessment added to clinical rating scales to differentiate adult ADHD from autism spectrum disorder. *Psychiatry Research*, *243*, 168–173.

Hall, C.L., Bellato, A., Kirk, J.D., & Hollis, C. (2023). The clinical utility of QbTest in supporting the assessment and monitoring of attention-deficit/hyperactivity disorder (ADHD): What do paediatricians need to know? *Paediatrics and Child Health*, *33*, 259–264. https://doi.org/10.1016/j.paed.2023.06.006

Hall, C.L., Selby, K., Guo, B., Valentine, A.Z., Walker, G.M., & Hollis, C. (2016). Innovations in practice: An objective measure of attention, impulsivity and activity reduces time to confirm attention deficit/hyperactivity disorder diagnosis in children - a completed audit cycle. *Child and Adolescent Mental Health*, *21*, 175–178.

Hall, C.L., Taylor, J.A., Newell, K., Baldwin, L., Sayal, K., & Hollis, C. (2016). The challenges of implementing ADHD clinical guidelines and research best evidence in routine clinical care settings: Delphi survey and mixed-methods study. *BJPsych Open*, *2*, 25–31.

Hall, C.L., Valentine, A.Z., Groom, M.J., Walker, G.M., Sayal, K., Daley, D., & Hollis, C. (2016). The clinical utility of the continuous performance test and objective measures of activity for diagnosing and monitoring ADHD in children: A systematic review. *European Child & Adolescent Psychiatry*, *25*, 677–699.

Hall, C.L., Valentine, A.Z., Walker, G.M., Ball, H.M., Cogger, H., Daley, D., … Hollis, C. (2017). Study of user experience of an objective test (QbTest) to aid ADHD assessment and medication management: A multi-methods approach. *BMC Psychiatry*, *17*, 66.

Hamadache, S., Hoberg, K., Zaplana Labarga, S., & Günther, T. (2021). Is the QbMini a valid instrument for ADHD assessment? *Journal of Attention Disorders*, *25*, 1384–1394.

Hanley, J.A., & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hollis, C., Hall, C.L., Guo, B., James, M., Boadu, J., Groom, M.J., … the AQUA Trial Group. (2018). The impact of a computerised test of attention and activity (QbTest) on diagnostic decision-making in children and young people with suspected attention deficit hyperactivity disorder: Single-blind randomised controlled trial. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *59*, 1298–1308.

Hult, N., Kadesjö, J., Kadesjö, B., Gillberg, C., & Billstedt, E. (2018). ADHD and the QbTest: Diagnostic validity of QbTest. *Journal of Attention Disorders*, *22*, 1074–1080.

Jensen, C.M., & Steinhausen, H.C. (2015). Comorbid mental disorders in children and adolescents with attention-deficit/hyperactivity disorder in a large nationwide study. *Attention Deficit and Hyperactivity Disorders*, *7*, 27–38.

Johansson, V., Norén Selinus, E., Kuja-Halkola, R., Lundström, S., Durbeej, N., Anckarsäter, H., … Hellner, C. (2021). The quantified behavioral test failed to differentiate ADHD in adolescents with neurodevelopmental problems. *Journal of Attention Disorders*, *25*, 312–321.

Labarga, S.Z., Hoberg, K., Hamadache, S., & Günther, T. (2019). Validierung des QbMini Tests zur Diagnose der Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) bei fünfjährigen Kindern. *Zeitschrift für Neuropsychologie*, *30*, 149–156.

Lenartowicz, A., & Loo, S.K. (2014). Use of EEG to diagnose ADHD. *Current Psychiatry Reports*, *16*, 498.

Loh, H.W., Ooi, C.P., Barua, P.D., Palmer, E.E., Molinari, F., & Acharya, U.R. (2022). Automated detection of ADHD: Current trends and future perspective. *Computers in Biology and Medicine*, *146*, 105525.

Loo, S.K., & Barkley, R.A. (2005). Clinical utility of EEG in attention deficit hyperactivity disorder. *Applied Neuropsychology*, *12*, 64–76. https://doi.org/10.1207/s15324826an1202_2

Loo, S.K., & Makeig, S. (2012). Clinical utility of EEG in attention-deficit/hyperactivity disorder: A research update. *Neurotherapeutics*, *9*, 569–587.

National Institute for Health and Care Excellence (NICE). (2018). Attention deficit hyperactivity disorder: diagnosis and management NICE guideline [NG87]. https://www.nice.org.uk/guidance/ng87

National Institute for Health and Care Excellence (NICE). (2023). QbTest for the assessment of attention deficit hyperactivity disorder (ADHD). Medtech innovation briefing [MIB318]. www.nice.org.uk/guidance/mib318

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Pettersson, R., Söderström, S., & Nilsson, K.W. (2018). Diagnosing ADHD in adults: An examination of the discriminative validity of neuropsychological tests and diagnostic assessment instruments. *Journal of Attention Disorders*, *22*, 1019–1031.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reitsma, J.B., Glas, A.S., Rutjes, A.W.S., Scholten, R.J.P.M., Bossuyt, P.M., & Zwinderman, A.H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, *58*, 982–990. https://doi.org/10.1016/j.jclinepi.2005.02.022

Saad, J.F., Kohn, M.R., Clarke, S., Lagopoulos, J., & Hermens, D.F. (2018). Is the theta/Beta EEG marker for ADHD inherently flawed? *Journal of Attention Disorders*, *22*, 815–826.

Safari, S., Baratloo, A., Elfil, M., & Negida, A. (2016). Evidence based emergency medicine; part 5 receiver operating curve and area under the curve. *Emergency*, *4*, 111–113.

Sayal, K., Prasad, V., Daley, D., Ford, T., & Coghill, D. (2018). ADHD in children and young people: Prevalence, care pathways, and service provision. *The Lancet. Psychiatry*, *5*, 175–186.

Schwarzer, G., Carpenter, J.R., & Rücker, G. (2015). Meta-analysis with R. Use R! https://doi.org/10.1007/978-3-319-21416-0

Söderström, S., Pettersson, R., & Nilsson, K.W. (2014). Quantitative and subjective behavioural aspects in the assessment of attention-deficit hyperactivity disorder (ADHD) in adults. *Nordic Journal of Psychiatry*, *68*, 30–37.

Tallberg, P., Råstam, M., Wenhov, L., Eliasson, G., & Gustafsson, P. (2018). Incremental clinical utility of continuous performance tests in childhood ADHD—An evidence-based assessment approach. *Scandinavian Journal of Psychology*, *60*, 26–35. https://doi.org/10.1111/sjop.12499

Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, Pliabilities, and pitfalls in research and practice. *Frontiers in Public Health*, *5*, 307.

Ulberstad, F., Bostrom, H., Chavanon, M.-L., Knollmann, M., Wiley, J., Christiansen, H., & Thorell, L.B. (2020). Objective measurement of attention deficit hyperactivity disorder symptoms outside the clinic using the QbCheck: Reliability and validity. *International Journal of Methods in Psychiatric Research*, *29*, e1822.

Valentine, A.Z., Brown, B.J., Groom, M.J., Young, E., Hollis, C., & Hall, C.L. (2020). A systematic review evaluating the implementation of technologies to assess, monitor and treat neurodevelopmental disorders: A map of the current evidence. *Clinical Psychology Review*, *80*, 101870.

Vogt, C. (2021). Clinical conundrums when integrating the QbTest into a standard ADHD assessment of children and Young people. *Neuropediatrics*, *52*, 155–162.

Welch, V., Wy, T.J., Ligezka, A., Hassett, L.C., Croarkin, P.E., Athreya, A.P., & Romanowicz, M. (2022). Use of mobile and wearable artificial intelligence in child and adolescent psychiatry: Scoping review. *Journal of Medical Internet Research*, *24*, e33560.

Whiting, P.F. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*, 529. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

Williams, L., Hall, C.L., Brown, S., Guo, B., James, M., Franceschini, M., … Groom, M.J. (2021). Optimising medication management in children and young people with ADHD using a computerised test (QbTest): A feasibility randomised controlled trial. *Pilot and Feasibility Studies*, *7*, 68.

Young, S., Asherson, P., Lloyd, T., Absoud, M., Arif, M., Colley, W.A., … Skirrow, C. (2021). Failure of healthcare provision for attention-deficit/hyperactivity disorder in the United Kingdom: A consensus statement. *Frontiers in Psychiatry*, *12*, 649399.

Zhang-James, Y., Razavi, A.S., Hoogman, M., Franke, B., & Faraone, S.V. (2023). Machine learning and MRI-based diagnostic models for ADHD: Are we there yet? *Journal of Attention Disorders*, *27*, 335–353.

Accepted for publication: 4 August 2023