NIHR | National Institute for Health and Care Research

Journals Library

Check for updates

Health Technology Assessment

Volume 29 • Issue 58 • November 2025

ISSN 2046-4924

Clinical and cost-effectiveness of technologies for the assessment of attention deficit hyperactivity disorder: a systematic review and economic model

Eve Tomlinson, Mary Ward, Josephine Walker, Melissa Benevente, Hanyu Wang, Chris Cooper, Hayley E Jones, Amanda Owen-Smith, Catalina Lopez Manzano, Sara James, Dietmar Hank, Richard Lee-Kelland, Nicky J Welton and Penny Whiting

Corrigendum

This article was corrected from the original publication in November 2025. Details of the correction can be found in *Health Technol Assess* 2025;**29**(58):i. https://doi.org/10.3310/THGD8525

DOI 10.3310/DRDR7171

NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

Clinical and cost-effectiveness of technologies for the assessment of attention deficit hyperactivity disorder: a systematic review and economic model

Eve Tomlinson[1†] Mary Ward[1†] Josephine Walker[1] Melissa Benevente[1] Hanyu Wang[1] Chris Cooper[1] Hayley E Jones[1] Amanda Owen-Smith[1] Catalina Lopez Manzano[2] Sara James,[2] Dietmar Hank[3] Richard Lee-Kelland[1] Nicky J Welton[1††] and Penny Whiting[1*††]

[1]Bristol TAG, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[2]Patient representative
[3]Avon and Wiltshire Mental Health Partnership NHS Trust, UK
[†]Joint first author
[††]Joint last author

[*]Corresponding author penny.whiting@bristol.ac.uk

Published November 2025
DOI: 10.3310/DRDR7171

This report should be referenced as follows:

Tomlinson E, Ward M, Walker J, Benevente M, Wang H, Cooper C, *et al.* Clinical and cost-effectiveness of technologies for the assessment of attention deficit hyperactivity disorder: a systematic review and economic model. *Health Technol Assess* 2025;**29**(58). https://doi.org/10.3310/DRDR7171

Health Technology Assessment

NICE TAR-DAR

ISSN 2046-4924 (Online)

Impact factor: 4

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal
Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This article

The research reported in this issue of the journal was funded by the Evidence Synthesis programme as award number NIHR136009. The contractual start date was in October 2023. The draft manuscript began editorial review in May 2024 and was accepted for publication in February 2025. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

Abstract

Background: Attention deficit hyperactivity disorder is characterised by inattention, impulsivity and hyperactivity. Diagnosis is complex and time-consuming. Medication requires careful selection and dose titration. Technologies for objective measures of attention deficit hyperactivity disorder that use motion sensors to measure hyperactivity ('sensor continuous performance tests') may help improve the diagnostic process and medication management when used in addition to clinical assessment.

Objective: To determine whether sensor continuous performance tests are clinically effective and cost-effective to the National Health Service. Specific objectives were to determine the effectiveness of sensor continuous performance tests for:

1. diagnosis of attention deficit hyperactivity disorder in people referred with suspected attention deficit hyperactivity disorder
2. diagnosis of attention deficit hyperactivity disorder in people referred with suspected attention deficit hyperactivity disorder for whom current assessment cannot reach a diagnosis
3. during initial dose titration and treatment decisions for people with attention deficit hyperactivity disorder
4. evaluating treatment effectiveness during long-term treatment monitoring for people with attention deficit hyperactivity disorder.

Design: Systematic review and economic model (searches completed 17 November 2023).

Results: Objective 1 [29 studies – 25 QbTest (QbTech Ltd., Stockholm, Sweden), 2 EF Sim (Peili Vision, Oulu, Finland) and 2 Nesplora Kids (Giunti Psychometrics, Florence, Italy)]: most evidence was in children. The AQUA trial was the only study to evaluate the QbTest in combination with clinical assessment and included a comparison with clinical assessment alone. Accuracy was similar and there was no statistical evidence of a difference between groups ($p = 0.14$), but the study was at high risk of bias. The AQUA trial reported that adding QbTest to the diagnostic process resulted in fewer appointments to reach a diagnosis, reduced consultation time, greater clinician confidence and exclusion of the diagnosis in a more children. Findings were supported by limited data from uncontrolled before–after studies. Qualitative and survey data reported increased clinician confidence in clinical decision-making, reduced time to diagnostic decision and improved communication. Barriers to implementation included staffing, training, technology requirements and length and repetitive content of the test. We found that using QbTest in addition to clinical assessment was likely cost-effective due to the reduced time waiting for assessment, reduced appointments until diagnosis and a higher proportion receiving treatment benefits.

Objective 3 (six studies): All evaluated QbTest and most had concerns with risk of bias. Qualitative and survey data suggested that healthcare staff and families valued the QbTest for dose titration, checking medication utility and improving medication adherence. Some data suggested that results may not increase patient understanding and some clinicians highlighted logistical challenges.

No studies were identified for objectives 2 and 4.

Conclusions: Our results suggest that QbTesting as part of the diagnostic workup for attention deficit hyperactivity disorder in children (age < 18 years), when used in combination with clinical assessment, may be cost-effective. This finding was robust to nearly all assumptions made in the model. There are insufficient data on other sensor continuous performance tests in adults or on medication management.

Future work:

- Diagnostic accuracy study evaluating comparing each of the sensor continuous performance tests plus clinical assessment. This should consider accuracy across different patient subgroups.
- Trial comparing patient outcomes and process measures in adults and children tested with and without sensor continuous performance tests with separate analyses for difficult-to-diagnose patients.
- Trial evaluating the role of sensor continuous performance tests in medication management, including long-term follow-up.

Limitations: Lack of good-quality data on all tests, both for diagnosis and medication management, particularly when evaluated in combination with clinical information.

Study registration: This study is registered as PROSPERO CRD42023482963.

Funding: This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR136009) and is published in full in *Health Technology Assessment*; Vol. 29, No. 58. See the NIHR Funding and Awards website for further award information.

Contents

List of tables

List of figures

List of abbreviations

ADHD	attention deficit hyperactivity disorder
AE	adverse event
AHSN	Academic Health Science Networks
ASD	autistic spectrum disorder
ATX	atomoxetine
AUC	area under the curve
AUC ROC	area under the receiver operating characteristics curve
BNF	*British National Formulary*
CAMHS	Children and Adolescent Mental Health Services
CASP	Critical Appraisal Skills Programme
CE	Conformité Européenne
CI	confidence interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
COVID-19	coronavirus disease 2019
CPT	continuous performance test
CrI	credible/credibility interval
DAWBA	Development and Wellbeing Assessment
DCD	developmental co-ordination disorder
DEX	dexamfetaminesulphate
DIVA	Diagnostic Interview for ADHD in Adults
DSM	*Diagnostic and Statistical Manual of Mental Disorders*
DSM-IV	*Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition
DSM-IV-TR	*Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, Text Revision
DSM-V	*Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition
DTA	diagnostic test accuracy
EQ-5D	EuroQol-5 Dimensions
FACT	Functional Assessment of Cancer Therapy
FN	false negative
FP	false positive
FU	follow-up
GP	general practitioner
HCP	healthcare professional
HR	hazard ratio
HRQoL	health-related quality of life
HTA	Health Technology Assessment
ICD-10	International Statistical Classification of Diseases and Related Health Problems, Tenth Revision
ICD-11	*International Classification of Diseases*, 11th Revision
ICER	incremental cost-effectiveness ratio
ICTRP	International Clinical Trials Registry Platform
INB	incremental net benefit
IQ	intelligence quotient
IR	immediate release
JBI	Joanna Briggs Institute
K-SADS-PL	Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime Version
LDX	lisdexamfetamine
MPH	methylphenidate
N/A	not applicable
NCT	National Clinical Trial
NICE	National Institute for Health and Care Excellence
OR	odds ratio
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life-year
Q-SSP	Quality Assessment Checklist for Survey Studies in Psychology

RCT	randomised controlled trial
RD	risk difference
RoB 2	Cochrane Risk of Bias Tool
ROBINS–I	Risk Of Bias In Non-randomized Studies – of Interventions
RRR	relative risk reduction
SCID-I	Structured Clinical Interview for DSM-IV v1
SCID-II	Structured Clinical Interview for DSM-IV v2
SDQ	Strengths and Difficulties Questionnaire
SLI	specific language impairment
SNAP-IV	Swanson Nolan and Pelham Questionnaire
TN	true negative
TP	true positive
TR	Time ratio
VR	virtual reality
VR-CPT	virtual reality continuous performance test
WHO	World Health Organization
WTP	willingness to pay
YOI	Young Offenders Institution

Note

This article is based on the Diagnostic Assessment Report produced for NICE. The full report contained a considerable number of data that were deemed confidential. The full report was used by the Diagnostic Advisory Committee at NICE in their deliberations. The full report with each piece of confidential data removed and replaced by the statement 'confidential information (or data) removed' is available on the NICE website: www.nice.org.uk.

The present article presents as full a version of the report as is possible while retaining readability, but some sections, sentences, tables and figures have been removed. Readers should bear in mind that the discussion, conclusions and implications for practice and research are based on all the data considered in the original full NICE report.

Plain language summary

What is the problem?

Attention deficit hyperactivity disorder is a common condition that affects behaviour in both children and adults. People with attention deficit hyperactivity disorder may find it hard to concentrate, act without thinking and be unable to sit still. This can get in the way of daily life.

Attention deficit hyperactivity disorder is usually diagnosed by a specialist (an expert in attention deficit hyperactivity disorder) based on the person's history, behaviour and symptoms. The expert will typically observe the person and interview the person and others in their life (e.g. partners, parents or teachers).

It can take a long time to be diagnosed with attention deficit hyperactivity disorder and the person may have to go to lots of appointments. Attention deficit hyperactivity disorder is also sometimes confused with mental health conditions that have similar symptoms, making it harder to diagnose.

Tests have been developed that may improve how attention deficit hyperactivity disorder is diagnosed and followed up. They are intended to be used in addition to assessment by an expert. These tests involve the person doing a computer-based task that measures behaviours associated with attention deficit hyperactivity disorder (e.g. ability to concentrate and to control movement) and include the use of sensors to track movement. These tests may reduce the number of appointments needed and could increase the likelihood of diagnosing attention deficit hyperactivity disorder correctly. They might also be able to help work out if treatments are working properly.

What did we do?

We wanted to know whether using these new tests to help diagnose attention deficit hyperactivity disorder will mean that more people are correctly told whether or not they have attention deficit hyperactivity disorder, whether these tests help diagnose attention deficit hyperactivity disorder faster and whether the tests can be used to correctly tell us how well attention deficit hyperactivity disorder treatments work. We also wanted to know whether these tests are a good use of National Health Service money. We looked at existing research and developed cost models to answer these questions.

What did we find?

We found very limited good-quality data. Our findings suggest that using QbTest may help to diagnose attention deficit hyperactivity disorder more quickly, possibly using fewer appointments, and may allow a diagnosis to be made in more people. It is likely to represent a good use of National Health Service money.

Scientific summary

Background

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder that is characterised by persistent patterns of inattention, impulsivity and hyperactivity that can significantly impact daily functioning.

Diagnosis of ADHD is complex and relies on a clinician's judgement combined with information such as questionnaires, third-party reports, patient history and behavioural observations. ADHD is frequently associated with other neurodevelopmental and psychiatric conditions, which can complicate the diagnosis and management of ADHD. It usually takes an average of two to three appointments and around 2.5 hours of clinic time to reach a diagnosis of ADHD. NHS waiting times for ADHD assessment are long, with patients often waiting more than 2 years. One treatment option for ADHD is medication. Identifying the most suitable medication and dose for a particular patient can be challenging.

A number of rating scales and tests are available to help diagnose ADHD, but none have sufficient accuracy to be used as a stand-alone diagnostic tool. There are a number of technologies for objective measures of ADHD, which use motion sensors to measure hyperactivity [referred to as 'sensor continuous performance test (CPT)']. These may help to improve the diagnostic process for people with ADHD and to improve medication management when used in addition to standard clinical assessment.

Objectives

The overall aim of this project was to determine whether sensor CPTs are clinically effective and cost-effective to the NHS.

Objective 1: What are the diagnostic accuracy and clinical effectiveness and cost-effectiveness of sensor CPT for the diagnosis of ADHD in people referred with suspected ADHD?
Objective 2: What are the diagnostic accuracy and clinical effectiveness and cost-effectiveness of sensor CPT for the diagnosis of ADHD in people referred with suspected ADHD for whom current assessment cannot reach a diagnosis?
Objective 3: What are the clinical effectiveness and cost-effectiveness of sensor CPT in evaluating medication effectiveness during initial dose titration and treatment decisions for people with a diagnosis of ADHD?
Objective 4: What are the clinical effectiveness and cost-effectiveness of sensor-based CPT for evaluating treatment effectiveness during long-term treatment monitoring for people with a diagnosis of ADHD?

Methods

Clinical effectiveness review

A systematic review was conducted. Studies that evaluated the QbMini (QbTech Ltd., Stockholm, Sweden), QbTest (6–12 and 12–60) (QbTech Ltd., Stockholm, Sweden), QbCheck (QbTech Ltd., Stockholm, Sweden), EF Sim (Peili Vision, Oulu, Finland), EF Sim Web Version (Peili Vision, Oulu, Finland), Nesplora Kids (Giunti Psychometrics, Florence, Italy) and Nesplora Adults (Giunti Psychometrics, Florence, Italy), alone or in combination with clinical assessment for ADHD, were eligible for inclusion. We included randomised controlled trials (RCTs), non-randomised studies of interventions, including before–after studies [non-randomised study of interventions (NRSI)], diagnostic test accuracy (DTA) studies, surveys and qualitative evaluations that reported on eligible outcomes.

Four databases and two trial registries were searched (inception – 17 November 2023). We screened trial registries, reference lists of reviews and study reports, relevant websites and information submitted by test manufacturers.

Title and abstract screening were conducted by two reviewers independently. Inclusion assessment, data extraction and risk-of-bias assessment were performed by one reviewer and checked by a second. Risk of bias was assessed using the following tools: Cochrane Risk of Bias Tool (RCTs), Risk Of Bias In Non-randomized Studies – of Interventions, QUADAS-2 (DTA studies), Critical Appraisal Skills Programme checklist (qualitative studies), Quality Assessment Checklist for Survey Studies in Psychology (survey studies).

For each objective, we provided a narrative summary of study details, risk of bias and results. Random and fixed-effects meta-analyses were performed to generate summary effect estimates. Forest plots were produced to show individual and summary effect estimates with 95% confidence intervals (CIs). Fisher's exact test was used to compare the estimates of accuracy where studies evaluated multiple index tests. Qualitative evidence was synthesised based on guidance from Joanna Briggs Institute.

Cost-effectiveness model

We developed a de novo model for sensor CPTs for the diagnosis of ADHD in people referred with suspected ADHD. We only evaluated the QbTest in addition to clinical assessment versus clinical assessment alone for children and adolescents due to lack of evidence on the inputs needed for our model for other sensor CPTs and populations. A Markov model structure was used to capture the process of waiting for assessment, assessment, diagnosis and treatment. We populated the model using evidence identified in the clinical effectiveness review, a review of cost-effectiveness studies of diagnostic tests and models of treatment for ADHD and further targeted searches as required.

Results

Objective 1

We included 29 studies (38 reports) for objective 1: 2 RCTs (1 of these also provided data on accuracy; both included a survey and qualitative substudy); 20 DTA studies (2 included a survey of patient views); 5 uncontrolled before–after implementation studies (2 also provided information on patient/clinician views – 1 survey and qualitative evaluation, 1 survey) and 2 studies that only reported on patient's and clinician's acceptability of sensor CPTs. Most studies evaluated the QbTest, two evaluated EF Sim and two evaluated Nesplora Kids; there were no studies of EF Sim web or of Nesplora Adults. The majority of the evidence was in children.

Five studies evaluated the accuracy of the QbTest in combination with clinical information; only one of these (the AQUA trial) evaluated the accuracy in combination with clinical judgement, as would be used in practice. However, data from the AQUA trial were limited due to inclusion of only those who had a diagnostic decision at 6 months and limitations with the reference standard. There are therefore no reliable data on the accuracy of any of the sensor CPTs when used in combination with clinical judgement.

Estimates of the accuracy of the sensor CPTs alone were heterogeneous, and so results should be interpreted with caution. Summary estimates of the accuracy of the QbTest suggested that the sensitivity was highest when the subcomponents were combined into an overall measure (summary sensitivity 79%, 95% CI 69% to 86%), but specificity was lower (summary specificity 59%, 95% CI 42% to 74%) than when the subcategories were assessed individually. There was little evidence of a difference between the accuracy of the three subcategories of activity, impulsivity and inattention. One study of Nesplora Kids and two studies of EF Sim reported similar estimates of accuracy to studies of the QbTest, but this was based on very limited information from studies at a high risk of bias.

Three studies provided a direct comparison between sensor CPT and non-sensor CPT, one study (the AQUA trial) provided a direct comparison between clinical diagnosis combined with QbTest with the accuracy of clinical diagnosis alone and one compared the accuracy of the QbTest alone to the accuracy of QbTest plus clinical information. One study reported that an overall measure from EF Sim was more sensitive than the non-sensor CPT omission errors measure ($p = 0.03$) but was less specific ($p = 0.07$). There was no difference between the overall EF Sim measure and the other two CPT measures. Two studies provided a direct comparison between the Conners' CPT II and the QbTest (12–60). One reported that Qb measures were more sensitive ($p \leq 0.01$) but less specific than the two Conners' CPT measures, while the other reported that the QbTest was less sensitive ($p < 0.01$) with no difference in specificity. The AQUA trial

compared QbTest plus clinical judgement to a control group using the standard diagnostic process. The two groups had very similar specificity, but sensitivity was slightly higher in the clinical diagnosis alone group (96%, 95% CI 87 to 100) compared to the group where diagnosis incorporated the QbTest (86%, 95% CI 72 to 95), but there was no statistical evidence of a difference between groups ($p = 0.14$). One study in older adults presented a comparison between models based on the QbTest alone and a model that incorporated a clinical measure of ADHD symptoms. The model that incorporated the clinical information was much more sensitive (91%, 95% CI 83 to 96) than the QbTest alone (56%, 95% CI 45 to 66; $p < 0.01$). There was no evidence for a difference in specificity ($p = 0.11$).

Five studies evaluated the impact of the QbTest on process measures. All were conducted in the UK and were restricted to children and adolescents. The AQUA trial randomised children to be assessed for ADHD with or without the QbTest as part of the diagnostic process. This study was judged at high risk of bias for time-to-event outcomes, as a large proportion of participants (80/250) were uninformatively censored from the analysis as they dropped out or were discharged from the clinic and so did not have a diagnosis at 6 months. It was at low risk of bias for other outcomes, except cost of clinic appointments which was judged at unclear risk. The other four studies were retrospective record reviews, where data for those evaluated for ADHD prior to implementation of the QbTest were compared to data for those evaluated after the implementation of the QbTest. The largest of these studies, Focus ADHD, was affected by the coronavirus disease 2019 pandemic as the QbTest was implemented over the same period as the pandemic. All four studies were judged at serious risk of bias; none were adjusted for potential confounding factors. The AQUA trial reported a number of benefits associated with adding QbTest to the diagnostic process, including fewer appointments to reach a diagnosis, reduced consultation time, increased proportion of patients with a diagnosis, greater clinician confidence in the diagnostic decision and exclusion of the diagnosis in a greater proportion of children. They also reported that the costs of clinic appointments were less in the QbTest arm compared to the control arm. Limited data from the before–after studies found that following implementation of the QbTest, fewer consultations were required to reach a diagnosis. These studies also reported other benefits included reduced time to reach a diagnosis (two studies) and reduced costs of testing.

Eight studies provided data on the clinician and/or patient and carer views of sensor CPTs for the diagnosis of ADHD. Most of the studies were judged to have some methodological concerns due to a lack of detail reported on the methodology used. Five evaluated the QbTest through interviews, surveys or focus groups. These reported that clinicians felt the test increased confidence in the clinical decision-making, and both clinicians and families felt it may reduce the time to diagnostic decision. Clinicians and families also felt that the test helped to improve communication. However, some families felt that the test results were not properly explained to them and did not help them to understand symptoms or how diagnoses were made. Barriers to implementation included staffing, training and technology requirements. Patients and caregivers highlighted concerns with the length and repetitive content of the test, and staff in one study reported that patients struggled with sensory discomfort and stress during the test. One study of QbCheck reported that participants found it easy to use; however, this was from a brief three-question survey conducted as part of a DTA study. Two survey studies evaluated EF Sim. One of these, funded by the test manufacturer, reported positive findings concerning acceptability for teachers (confidential information has been removed) who had implemented the test. The other study also reported positive acceptability from a short survey to children who had used the test in a DTA study.

We found that QbTest in addition to clinical assessment is likely to be cost-effective, with incremental costs of £238.35 and incremental quality-adjusted life-years (QALYs) of 0.0385 per person evaluated for ADHD. The resulting incremental cost-effectiveness ratio is £6183 per QALY gained, which is cost-effective at a willingness-to-pay (WTP) threshold of £20,000 per QALY. The mean incremental net benefit (probability of being cost-effective) is £532.55 (92%) and £918 (84%) at WTP of £20,000 and £30,000 per QALY, respectively. These findings were driven by reduced time waiting for assessment, reduced appointments until diagnosis and a higher proportion receiving a diagnosis so that more patients with ADHD receive treatment benefits.

We found that our overall conclusions were robust to most of our modelling assumptions. However, if the state costs for responders/non-responders on treatment were assumed to be higher, then QbTest in addition to clinical assessment would not be cost-effective at £20,000/QALY due to the higher proportion who initiate treatment and incur the higher costs. Also, if the proportion of patients with a diagnosis within 6 months for QbTest in addition to clinical assessment is

lower (closer to that for clinical assessment alone), then QbTest in addition to clinical assessment becomes cost-saving but also incurs lower or even less QALYs than clinical assessment alone. In this scenario, the cost savings do not justify the quality of life reductions.

Objective 2

We did not identify studies that met inclusion for objective 2. We ran some exploratory analyses which demonstrated that if there are no consequences in terms of diagnostic accuracy, then using sensor CPTs on the subset of those where a diagnosis is not reached after one or two appointments would be more cost-effective than using sensor CPTs on all patients, because the test cost is incurred for only some patients.

Objective 3

Six studies were included for objective 3; all evaluated the QbTest. One DTA study evaluated the accuracy of QbTest as part of dose titration against the reference standard of 'good outcome' at 1-year follow-up. However, the QbTest formed part of the reference standard which is likely to overestimate the accuracy of the test and so it is not possible to draw strong conclusions from this study.

One study (the QUOTA trial) provided data on process measures; however, it was a small feasibility trial that was not designed and powered to formally evaluate the impact on outcomes. Three RCTs (the AQUA trial and two feasibility RCTs: FACT and QUOTA) and two implementation studies provided interview or survey data on patient and clinician views of the QbTest for medication management and dose titration. Most of the studies had concerns regarding quality due to lack of information on study design. Findings suggested that healthcare staff and families mostly valued the role of the test for dose titration, checking medication utility and improving medication adherence. However, two surveys of patients suggested that the results of the QbTest may not have helped them to understand medication decisions, and some clinicians highlighted that using the QbTest for medication management can present logistical challenges due to having to schedule more appointments.

Objective 4

We did not identify any studies that met inclusion for objective 4.

There was insufficient evidence on model inputs to be able to evaluate cost-effectiveness for objectives 3 or 4.

Conclusions

There was a lack of good-quality data on all tests, both for diagnosis and medication management, particularly when evaluated in combination with clinical information. Our results suggest that QbTesting as part of the diagnostic workup for ADHD in children (age < 18 years), when used in combination with clinical assessment, is cost-effective. We found this finding was robust to nearly all assumptions made in the model. There are insufficient data on other sensor CPTs, in adults or on medication management.

There are a number of areas where further work is required:

- Diagnostic accuracy study evaluating comparing each of the sensor CPT plus clinical assessment. This should consider accuracy across different patient subgroups.
- Trial comparing patient outcomes and process measures in adults and children tested with and without sensor CPT with separate analyses for difficult-to-diagnose patients.
- Trial evaluating the role of sensor CPT in medication management, including long-term follow-up.

Study registration

This study is registered as PROSPERO CRD42023482963.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis programme (NIHR award ref: NIHR136009) and is published in full in *Health Technology Assessment*; Vol. 29, No. 58. See the NIHR Funding and Awards website for further award information.

Chapter 1 Background and definition of decision problem

Sections of this chapter have been reproduced from the review protocol, available at the NICE website.[1]

Epidemiology and burden of attention deficit hyperactivity disorder

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder that is characterised by persistent patterns of inattention, impulsivity and hyperactivity which can significantly impact daily functioning.[2] Different subtypes can be defined based on these key features:

- inattentive subtype
- hyperactive–impulsive subtype
- combined subtype (both inattentive and hyperactive-impulsive).

The exact cause of ADHD is unknown, but it is generally considered to involve multiple genetic and environmental factors that lead to altered brain neurochemistry and structure. ADHD is estimated to affect around 2–7% of school-aged children and young people, with an average estimate of around 5%.[3] There has been a substantial increase in the proportion of children diagnosed with ADHD over the past 30 years, with rates doubling between 2003 and 2018.[4] Increasing awareness of ADHD among healthcare professionals (HCPs), educators and the general public has contributed to higher rates of diagnosis.[3] ADHD often persists into adulthood – studies suggest that around 15% of adults will continue to meet full diagnostic criteria for ADHD, 65% will continue to show symptoms which impact on their life, whereas around 20% will have no symptoms or impairment in adulthood.[5] Certain population may be more likely to have ADHD – a 2018 meta-analysis estimated that up to one in four prisoners had a diagnosis of ADHD,[6] although a more recent reanalysis of these data reported that, after accounting for an outlier and restricting to studies that used random sampling of adults in prison, prevalence was much lower at around 4.5% in men.[7]

Attention deficit hyperactivity disorder can have a significant impact on individuals' academic, social and occupational functioning. Children with ADHD may struggle in school, have difficulty forming and maintaining relationships and experience low self-esteem.[8,9] In adulthood, untreated ADHD can lead to challenges in employment, relationships and mental health.[10] ADHD is often accompanied by substantial comorbidity, including substance use, depression, anxiety and accidents.[11] Symptoms of inattention can make even basic tasks such as reading, watching television and multitasking challenging.[12] Among adults, there is an expectation of being able to function independently, but difficulty in maintaining attention can make this very challenging.[12] However, there are also positive effects of ADHD, with a recent qualitative study highlighting that, sometimes, acting on impulse can have positive effects, leading perhaps to a fulfilled and exciting life.[12] The burden of ADHD extends beyond the affected individuals to their families, schools and the healthcare system – a UK-based study highlighted the impact of ADHD on the quality of life of children with ADHD and of their siblings.[8] The economic burden includes healthcare costs, educational support services and lost productivity for individuals and caregivers.

Attention deficit hyperactivity disorder is usually diagnosed in childhood, with symptoms often becoming noticeable when a child starts school.[13] Boys are more commonly diagnosed with ADHD than girls, with a male-to-female ratio estimated at around 3 : 1.[3,14] People with ADHD may seem restless, have trouble concentrating and may act on impulse.[13] Boys present differently from girls – they often display disruptive behaviour prompting referral, whereas girls are more likely to have the inattentive subtype, making it less likely for girls to be referred for evaluation of ADHD. Symptoms of ADHD may change with age, with symptoms relating to hyperactivity becoming harder to detect with age, while those relating to inattentiveness persist.[5,15]

Current diagnostic and care pathway

Referral

The National Institute for Health and Care Excellence (NICE) guideline on ADHD diagnosis and management (NG87) provides guidance on the diagnostic pathway for ADHD.[16] However, this can be seen as the best practice and is not always reflected in reality in the NHS. The guidance suggests that children and young people with suspected ADHD should be referred from community settings to secondary care for further investigation – often to a paediatrician, with those with significant mental health comorbidities and adolescents often referred to Child and Adolescent Mental Health Services (CAMHS). Community referral is usually made by a health, education or social care professional, for example, the general practitioner (GP), educational psychologist or school special educational needs coordinator. Exact referral and care pathways vary locally.[16]

National Institute for Health and Care Excellence guidelines recommend that adults presenting with symptoms suggestive of ADHD, who do not have a childhood diagnosis of ADHD, should be referred to secondary care for further assessment by a mental health specialist with training in the diagnosis and treatment of ADHD. Referral is usually made from primary care or general adult psychiatric services. Adults who were diagnosed and treated for ADHD as children, or people who present with symptoms suggestive of continuing ADHD, should be referred for further assessment.[16]

The NICE guidelines highlight that the following groups have a higher likelihood of having ADHD than the general population and so a lower threshold for referral may be appropriate in these groups:[16]

- people born pre-term
- looked-after children and young people
- children and young people diagnosed with oppositional defiant disorder or conduct disorder
- children and young people with mood disorders
- people with a close family member diagnosed with ADHD
- people with epilepsy
- people with other neurodevelopmental disorders [e.g. autistic spectrum disorder (ASD), tic disorders and learning difficulties]
- adults with a mental health condition
- people with a history of substance misuse
- people known to the Youth Justice System or Adult Criminal Justice System
- people with acquired brain injury.

The guidelines also highlight that ADHD is likely to be under-recognised in girls and women who may be less likely to be referred for ADHD assessment, may be less likely to be diagnosed with ADHD and may be more likely to receive an incorrect diagnosis of another mental health or neurodevelopmental condition.[16]

Diagnosis

Assessment and diagnosis of ADHD are complex processes that typically rely on a clinician's judgement and involve gathering information from multiple sources, such as assessment questionnaires, third-party reports, patient history and behavioural observations. This approach is largely subjective and can lead to concerns regarding the reliability and consistency of the diagnosis.[17] It is also resource-intensive – it usually takes an average of two to three appointments and around 2.5 hours of clinic time to reach a diagnosis of ADHD.[18] Guidelines from The Royal College of Psychiatrists in Scotland suggest that, in most cases, the assessment and diagnosis of ADHD in adults will require two to three 1-hour sessions.[19] While children are usually assessed face to face in clinic, assessment for adults is often done remotely. This avoids the need to travel long distances to centralised assessment centres and also means that family members can join the consultation from different locations. Waiting times for a diagnosis through the NHS can also be lengthy – a recent survey based on people who had signed a petition to ask for improved ADHD assessment suggested that 10% of respondents had been waiting between 2 and 3 years for an ADHD assessment and 24% had waited between 1 and 2 years.[20] Proportions were slightly higher for children, with 14% waiting between 2 and 3 years for an ADHD assessment and 30% waiting between 1 and 2 years. A recent paper suggests that a realistic estimate for time to

diagnosis for adults newly referred for assessment is likely to be 5–10 years.[21] The average time to diagnosis in children is reported to be 18 months.[22]

The NICE guideline on ADHD diagnosis and management (NG87) recommends diagnosis based on a combination of psychosocial assessment, patient history, symptoms and behaviour.[16] To make a diagnosis of ADHD, symptoms of hyperactivity/impulsivity and/or inattention should meet the diagnostic criteria of the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (DSM-V) or the *International Classification of Diseases*, 11th Revision (ICD-11)[23,24] and should cause at least moderate psychological, social and/or educational impairments. This should be based on interview and/or direct observation in multiple settings. Impairment should be pervasive occurring in at least two important settings, including social, familial, educational and/or occupational settings.[16] The guidance highlights that the diagnosis should only be made by a specialist psychiatrist, paediatrician or other appropriately qualified HCP with training and expertise in the diagnosis of ADHD.[16]

Attention deficit hyperactivity disorder is frequently associated with other neurodevelopmental and psychiatric conditions. Common co-occurring conditions include ASDs, personality disorders, learning disabilities, anxiety disorders, mood disorders, conduct disorders and developmental trauma.[2] The presence of these comorbidities can complicate the diagnosis and management of ADHD.[5] Diagnosis can also be more challenging among those in the criminal justice system.

A number of rating scales are available to help diagnose ADHD. The most commonly evaluated rating scales include Achenbach System of Empirically Based Assessment, Conners Scales, *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV)-based ratings scales (e.g. the ADHD Rating Scale IV) and the Strengths and Difficulties Questionnaire (SDQ). A recent systematic review of these tools concluded that, although most tools have excellent overall diagnostic accuracy [area under the curve (AUC), ranged from 0.76 to 1.00], a single measure completed by a single reporter is unlikely to have sufficient accuracy for clinical use.[25] This finding is reflected in the NICE guidelines, which state that a diagnosis should not be made solely on the basis of such scales.[16]

Other tests that can help with the diagnosis include continuous performance tests (CPTs). These are computer-based tests that assess an individual's sustained attention and impulse control. Examples of these tests include: Test of variables of attention, Gordon's diagnostic system and Conners' CPT. These tests are designed to be used alongside clinical assessment as part of the diagnostic pathway for ADHD. A systematic review found mixed evidence on the clinical utility of CPT as an assessment tool. They highlighted that such tests should not be used as a stand-alone diagnostic tool and suggested that combining CPTs and an objective measure of activity may be particularly useful as a clinical tool and worthy of further pursuit.[26] These tests are not explicitly mentioned in the NICE guidelines.

Management and treatment of attention deficit hyperactivity disorder

Managing ADHD requires a multidisciplinary approach, with NICE guidance recommending that individuals with ADHD should have a comprehensive, holistic shared treatment plan that addresses psychological, behavioural and occupational or educational needs.[16] The treatment plan should be developed through discussion with those affected by ADHD and their families – this should be an ongoing process and should undergo regular review. Recommendations on treating ADHD vary according to age, with slightly different recommendations for those under 5 years, children and young people aged over 5 years and adults. Treatment plans will be tailored to the individual but are likely to encompass some or all of the following:[27]

Behavioural interventions

Behavioural therapies are used to improve organisational skills, impulse control and self-regulation. Parent training and classroom management strategies are often included.

Educational support

For children and young people, schools are encouraged to provide support, such as individual education plans and accommodations to address academic challenges.

Psychosocial support

Individual or family counselling may be recommended to address emotional and psychological issues.

Lifestyle and self-care
Encouraging a healthy lifestyle with regular exercise, a balanced diet and adequate sleep is important. Developing structured routines and organisation skills can also be beneficial.

Awareness and education
Parents, caregivers and individuals with ADHD are provided with education and support to help them understand the condition and learn strategies for managing symptoms.

Medication
Medications, such as stimulants [e.g. methylphenidate (MPH) or amphetamine-based drugs] or non-stimulants [e.g. atomoxetine (ATX), guanfacine and clonidine], may be prescribed based on the severity of symptoms and individual response.[28]

Medication should only be given to those with ADHD if their symptoms are 'still causing a significant impairment in at least one domain after environmental modifications have been implemented and reviewed'.[16] However, due to the length of time that it currently takes to receive a diagnosis, by which time most people will have pursued a range of techniques and strategies to manage their difficulties, medication is often started soon after diagnosis. Medication is not recommended in under-fives without a second specialist opinion, ideally from a tertiary centre.[16] Before starting medication, a detailed baseline assessment is required. Medication is usually started at a low dose that is gradually increased as needed.[27] The optimal dose will balance treatment effectiveness against severity of any adverse effects. Potential adverse effects vary according to which medication is prescribed but include: small increases in blood pressure, decreased appetite, trouble sleeping, headaches, stomach aches, drowsiness, dizziness, diarrhoea, nausea and vomiting and mood changes, including feeling aggressive, irritable, depressed, anxious or tense.[27] Treatment is considered optimal when patients demonstrate reduced symptoms, positive behaviour change, improvement in education, employment and relationships, with tolerable adverse effects. Achieving optimal treatment requires regular review, assessment and adjustment of medication.

Once a patient has started treatment, NICE guidelines recommend regular monitoring to assess effectiveness and adverse effects. They recommend that those taking medication should record adverse events (AEs), ideally using an adverse effect checklist. Treatment effectiveness should be monitored using standard symptom and adverse effect rating scales.[16] There are two stages to monitoring treatment effectiveness. The initial stage is during the dose titration phase when patients are reviewed frequently until they are on a stable dose of medication. After this, they are monitored at least annually, mainly to assess whether the treatment remains effective and to assess side effects.

Technologies of interest

Technologies of interest for this appraisal include technologies that combine a CPT with an objective and standardised measure of motor activity for the assessment of ADHD. We use the term 'sensor CPT' to refer to these tests. CPTs that do not incorporate the objective and standardised measures of motor activity are referred to as 'non-sensor CPTs'.

QbTest
The QbTest (QbTech Ltd., Stockholm, Sweden) is a Conformité Européenne (CE)-marked, class I medical device designed for use to aid in the assessment of ADHD and in the evaluation of treatment interventions in those with ADHD aged 6–60 years. It combines computerised assessments with a high-resolution motion tracking system to evaluate three core symptoms of ADHD: attention, impulsivity and hyperactivity.

The QbTest involves a computer-based task that typically takes 15–20 minutes to complete. There are three versions of the test for different age groups to control for developmental differences in cognitive abilities: QbMini for those aged 4–5 years, QbTest (6-–12) for children aged 6–12 years and QbTest (12–60) for those aged 12–60 years (*Table 1*). This version is also referred to as the QbTest Plus. During the test, the individual is required to respond to specific stimuli by pressing a button – they are required to distinguish between 'targets' and 'non-targets'. To monitor motor activity during

TABLE 1 Overview of differences between different versions of the QbTest

Feature	QbMini[29,30]	QbTest (6–12)	QbTest (12–60)
Age group	4–5 years	6–12 years	12–60 years
Stimulus	Yellow smiley face and yellow circle without smiley face	Grey circle and grey circle with a cross	Red circle, blue circle, red square and blue square
Target	Yellow smiley face	Grey circle	Matching pair – identical in shape and colour to the stimulus immediately preceding it
Stimulus rate	One stimulus every 2 seconds	One stimulus every 2 seconds (0.5 Hz)	One stimulus every 2 seconds (0.5 Hz)
Time stimulus is visible	2 seconds	100 milliseconds	200 milliseconds
Total number of stimulus presented	300	450	600
Target–to-non-target ratio	50 : 50	50 : 50	25 : 75

the test, the individual wears a headband. This motion tracking system records and measures hyperactivity and other motor-related behaviours.

To administer the QbTest, a private and quiet room with a computer, desk and chair is needed. Trained healthcare assistants or nurses can oversee the test, and a trained clinician interprets the results. Test results are compared to a normative group of individuals of the same sex and age who do not have ADHD. Outputs of the test are visually reported, detailing the performance in each of the three symptom domains of ADHD (activity, attention and impulsivity) and the level of deviation from non-ADHD score, and are sent directly to the clinician. Results are expressed as the Q-score for subcategories of activity, impulsivity and inattention. Q-scores reflect the deviation of the participant's performance (in standardised units) from the mean score of the normative group. There is no standard threshold for defining a positive Q-score as the scores are only meant to inform the diagnosis – the clinician combines the QbTest data with questionnaire responses and observational information for a comprehensive assessment.

The QbTest was implemented across 69 NHS trusts between 2020 and 2023 as part of an Academic Health Science Networks (AHSN) initiative known as 'Focus ADHD' that aimed to improve the diagnosis of ADHD in children and young people.[22,31] A recent NICE Medical Innovation Briefing highlighted that the QbTest should be used as an addition to routine clinical assessment, not as a stand-alone test. It also highlighted uncertainties in that the evidence reviewed included potentially inappropriate populations and did not use a parallel clinical assessment.[32]

QbCheck

QbCheck (QbTech Ltd., Stockholm, Sweden) is the same as the QbTest, but it is designed for remote testing and can be used without a HCP present. Like the QbTest, it is a CE-marked class I medical device, indicated for use as an online tool to aid in the clinical assessment of ADHD and in the evaluation of treatment interventions in those with ADHD aged 6–60 years. It combines an online CPT with a webcam motion tracking system and, like the QbTest, results are compared to a normative group without ADHD with results reported in the same way as for the QbTest. In addition to the QbTest, the test-taker performs an ability test that gives important information of the test-taker's ability to manage the test situation.

The QbCheck requires a laptop or computer with a stable internet connection in an appropriate location. The test uses the built-in web camera rather than the advanced motion tracking system used for the QbTest. As with the QbTest, there are two different versions targeted at the different age groups – the test stimulus are the same for the QbCheck as for the QbTest. The test can be administered remotely and observed by trained healthcare assistants or nurses and interpreted by a trained clinician alongside questionnaire responses and observational data.

EF Sim Test (previously known as ARVO and EPELI)

The EF Sim (Peili Vision, Oulu, Finland) is a virtual reality (VR) game designed for children and young people aged 8–13 years. It is CE marked as a class I medical device. It involves completing everyday tasks within a simulated home environment and is intended to be used alongside existing clinical assessments for ADHD.

The game consists of a 25-minute in-game session played on an Oculus Go head-mounted display and its hand controller. During gameplay, motion tracking sensors in the goggles and controller capture the participant's movements. An updated version of the EF Sim Test that includes eye movement (saccades) tracking is due to be available in early 2024. The test assesses various performance indicators related to ADHD, including attention, hyperactivity, impulsivity, memory, time management, planning, behaviour regulation, task efficiency and efficiency of information processing.

A web-based, remote version of the EF Sim Test is also in development. This is due to be available in early 2024.

Nesplora Attention Adults Aquarium

The Nesplora Attention Adults Aquarium (Giunti Psychometrics, Florence, Italy) is a Class I CE-marked, virtual reality continuous performance test (VR-CPT) suitable for people aged 16–90 years. It measures symptoms of ADHD, including auditory and visual attention, impulsivity, motor activity and reaction time. It is intended to be used alongside current ADHD clinical assessment.

The test involves an 18–22-minute computerised task that is conducted while wearing a VR headset and headphones. It requires a VR device, computer, stable internet connection and headband headphones. The person undertaking the test uses a handheld button to respond to both visual and auditory stimuli. Results are available immediately and are visually reported, detailing a score for the following categories: attention, inhibitory control (impulsivity), motor activity, processing speed, distractibility and vigilance. This score is calculated by comparing to a normative data set of people without ADHD of the same sex and age. All measures for sustained attention and inhibition are obtained separately for auditory and visual modalities and for the two modalities combined.

Nesplora Attention Kids AULA

The Nesplora Attention Kids AULA (Giunti Psychometrics, Florence, Italy) is a Class I CE-marked VR-CPT. It is very similar to Nesplora Attention Adults Aquarium but is aimed at young people aged 6–16 years – the test also involves a computerised task, measures the same ADHD symptoms as the adult version and is performed and interpreted in the same way as the adult version.

Place of the technology in the diagnostic and treatment pathway

There are four potential roles for the new technologies in the diagnostic and treatment pathway. In all cases, the tests should be used alongside HCP assessment:

1. as part of the initial diagnostic assessment for all people referred with suspected ADHD
2. as part of the initial diagnostic assessment for people where a diagnostic decision cannot be reached using current assessment methods
3. to assess medication effectiveness during initial dose titration and treatment decisions in people with a diagnosis of ADHD
4. to assess treatment (pharmacological or non-pharmacological) effectiveness for long-term treatment monitoring for people with a diagnosis of ADHD.

Chapter 2 Objectives

Sections of this chapter have been reproduced from the review protocol, available at the NICE website.[1]

The overall aim of this project was to determine whether technologies for objective measures of ADHD that use motion sensors to measure hyperactivity are clinically effective and cost-effective to the NHS. We defined the following objectives to address this aim:

1. What are the diagnostic accuracy and clinical effectiveness and cost-effectiveness of technologies that combine measures of cognition and motor activities for the diagnosis of ADHD in people referred with suspected ADHD?
2. What are the diagnostic accuracy and clinical effectiveness and cost-effectiveness of technologies that combine measures of cognition and motor activities for the diagnosis of ADHD in people referred with suspected ADHD for whom current assessment cannot reach a diagnosis?
3. What are the clinical effectiveness and cost-effectiveness of technologies that combine measures of cognition and motor activities in evaluating medication effectiveness during initial dose titration and treatment decisions for people with a diagnosis of ADHD?
4. What are the clinical effectiveness and cost-effectiveness of technologies that combine measures of cognition and motor activities for evaluating treatment effectiveness during long-term treatment monitoring for people with a diagnosis of ADHD?

Chapter 3 Assessment of clinical effectiveness

This manuscript contains reference to confidential information provided as part of the NICE appraisal process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Sections of this chapter have been reproduced from the review protocol, available at the NICE website.[1]

We conducted a systematic review to summarise the evidence on the clinical effectiveness and diagnostic accuracy of technologies that combine measures of cognition and motor activities for diagnosis and management of ADHD. The systematic review followed the principles outlined in the Centre for Reviews and Dissemination guidance for undertaking reviews in health care, the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy and the NICE Health Technology Evaluations Manual.[33–35] The review is reported according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020, PRISMA extension for diagnostic test accuracy (DTA) and PRISMA-Equity (PRISMA-E) guidelines.[36–38] The review was registered on the PROSPERO database (CRD42023482963).

Inclusion and exclusion criteria

Studies that fulfilled the following criteria were eligible for inclusion:

Technology (intervention/index test)

Technologies that combine a CPT with an objective and standardised measure motor activity for the assessment of ADHD. We use the term 'sensor CPT' to refer to these tests. Eligible tests are: QbMini, QbTest (6–12 and 12–60), QbCheck, EF Sim, EF Sim Web Version, Nesplora Kids and Nesplora adults alone or in combination with clinical assessment for ADHD by a HCP. These sensor CPTs were selected by NICE for this review as they are available to the NHS and are likely to have regulatory approval by the time of guidance publication.

Population

Objective 1: Adults and children referred for evaluation of suspected ADHD.

Objective 2: Adults and children referred for evaluation of suspected ADHD in whom a diagnosis had not been made through standard assessment processes.

Objective 3: Adults and children with a diagnosis of ADHD undergoing initial dose titration and treatment decisions.

Objective 4: Adults and children with a diagnosis of ADHD being monitored for treatment effectiveness.

Setting

Secondary care or remote assessment settings were accepted. Studies in which some participants (e.g. control groups) were enrolled in other settings were also eligible.

Comparator

Any diagnostic assessment for ADHD that did not include the technology of interest was included. Studies that compared two or more technologies of interest were also eligible for inclusion. For evaluation of DTA, studies that reported a direct comparison of the accuracy of one of the technologies of interest and another CPT (e.g. Connor's CPT) were also included. These are referred to as 'non-sensor CPTs'.

Reference standard (diagnostic accuracy studies only)

This included any reported diagnostic assessment for ADHD.

Study designs

For assessment of *clinical effectiveness*, we included randomised controlled trials (RCTs) or non-randomised study of interventions. For evaluation of *DTA*, we included DTA studies of any design, including one gate (also known as diagnostic cohort or cross-sectional studies) and multigate (also known as diagnostic case-control studies) designs. Qualitative studies were eligible if they provided data on any of the specified outcomes. Where data were not available on any of the specified outcomes from the designs listed, we also considered UK-based observational studies that included a control group (e.g. before–after study).

Outcomes

Studies were required to report at least one of the following outcomes of interest for this appraisal:

- test performance (diagnostic accuracy) for example, sensitivity, specificity, area under the receiver operating characteristics (ROC) curve (AUC ROC)
- test failure
- time to assessment or to reach a diagnostic decision
- use of NHS and Personal Social Services (PSS) services (such as the number and length of clinical appointments prior to diagnosis)
- impact on clinical decision-making
- confidence of HCPs in assessment
- ease of use/acceptability for clinicians
- use of interventions (such as ADHD medication)
- morbidity
- mortality
- health-related quality of life (HRQoL)
- ease of use/acceptability for patients or carers
- patient and carer experience
- costs related to using the technologies
- cost of training staff to operate technology and interpret results
- costs of resources associated with diagnosing and reviewing ADHD
- cost of interventions to help manage ADHD HRQoL.

If studies met the above inclusion criteria, they were included regardless of their language of publication, country of publication or the date they were reported.

Existing systematic reviews were included if they fulfilled inclusion criteria, were judged as low risk of bias based on the ROBIS tool,[39] had searches conducted within the past year and stratified the synthesis as described in our synthesis section (see *Synthesis methods*); otherwise, they were used a source of potentially relevant studies.

Study identification

Studies were identified using bibliographic and non-bibliographic search methods following guidance from the NICE Health Technology manual.[34] The search was not limited by date, language or country of publication.

Bibliographic searching

The following databases were searched:

- MEDLINE (Ovid)
- EMBASE (Ovid)
- PsycInfo® (American Psychological Association, Washington, DC, USA) (Ovid)
- Cumulative Index to Nursing and Allied Health Literature (CINAHL) [EBSCO*host* (Elton B. Stephens Company)].

We used sensitive search strategy based on terms for each of the technologies eligible for inclusion and for the manufacturers of these technologies. Full search strategies are reported in *Appendix 1*.

Non-bibliographic search methods

Completed and ongoing trials were identified through searches of the following trial registries:

- ClinicalTrials.gov via www.clinicaltrials.gov
- World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) via www.who.int/clinical-trials-registry-platform.

Additional relevant studies were identified by:

- screening reference lists of any reviews (systematic or non-systematic) identified by our searches
- reviewing the reference lists of any primary study report included at full text
- hand searching the websites of the manufacturer/or licence holders for each test
- information submitted by test manufacturers.

Managing the searches

Search results were exported to EndNote 20 [Clarivate Analytics (formerly Thomson Reuters), Philadelphia, PA, USA] for deduplication using the default deduplication settings and manual review of records. Search results were then exported from EndNote to Microsoft Access® (Microsoft Corporation, Redmond, WA, USA) for screening.

Review strategy

Two reviewers independently screened titles and abstracts identified by the searches. Full copies of all reports considered potentially relevant were obtained and two reviewers independently assessed these for inclusion. Disagreements were resolved through discussion.

The three test manufacturers (Peili Vision, Nesplora and QbTech) submitted reports containing information about the tests and citations to potentially relevant reports. One reviewer extracted all relevant information and citations from the test manufacturer's submissions into a separate document for each manufacturer in Microsoft Word (Microsoft Corporation, Redmond, WA, USA). One reviewer screened each citation as follows: (1) checked our review searches to see if it had been identified already; (2) if it had not been identified by our searches, or identified by our searches but only screened at title and abstract stages, we located the full-text report, saved it and assessed it for inclusion. Any queries were discussed with a second reviewer.

Data were extracted using standardised data extraction forms developed in Microsoft Access or Microsoft Word depending on the quantity of data available. Data extraction forms were piloted on a small sample of papers and adapted as necessary. Data were extracted by one reviewer and checked in detail by a second reviewer. Any disagreements were resolved through discussion.

Data were extracted on the following: study design (RCTs, DTA studies, before–after implementation study, qualitative and survey), objective that study addresses, funding sources (public, industry and mixed), country, setting, inclusion criteria, ADHD subtype, test details (test and threshold), comparator or reference standard test(s), sample size and outcomes specified in inclusion criteria (see *Inclusion and exclusion criteria*).

We considered the PROGRESS-Plus population factors, where reported.[40] PROGRESS-Plus is an acronym that describes characteristics that contribute to health inequity. PROGRESS stands for: place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status and social capital. 'Plus' stands for any additional factors considered important for the specific topic under review. We extracted the following 'Plus' factors:

- personal characteristics associated with discrimination: characteristics of relevance to the current review include age, sex, ethnicity, learning disability, neurodevelopmental disorders (including ASDs and personality disorders), developmental trauma
- looked-after children
- features of relationships for example, exclusion from school
- time-dependent relationships for example, instances where a person may be temporarily at disadvantage
- people in the Youth Justice System or Adult Criminal Justice System.

We extracted whether each PROGRESS-Plus factor was reported at baseline (yes/no), the baseline data concerning the factor as reported by the authors, and whether the study reports results data stratified by the factor. Where stratified data were reported, these were extracted.

Dichotomous clinical effectiveness data were extracted as number of patients with events and/or number of events and total number of patients in each treatment arm. For categorical data, we extracted details on the categories assessed, the total number of patients in each treatment arm and the number of patients in each outcome category. For continuous clinical effectiveness data, we extracted means/medians together with ranges, standard deviations (SDs), standard errors (SEs) and/or confidence intervals (CIs) for the outcome at baseline, follow-up (FU) and for change from baseline in each treatment group. For all types of clinical effectiveness data, summary effect estimates together with 95% CIs and *p*-values for comparisons between groups together with details on the methods of analysis, any variables controlled for in the analysis and the test statistic were extracted.

Accuracy data were extracted as 2 × 2 tables comparing the ADHD test against the reference standard, where available. The AUC ROC was also extracted, with 95% CI or SE. Where 2 × 2 tables were not reported in the paper, these were calculated from estimates of sensitivity and specificity together with the total number of patients with and without ADHD. For one study,[41] 2 × 2 tables were approximated from reported point estimates for sensitivity, specificity, positive predictive value and negative predictive value, the total sample size and an assumption that the proportion of individuals excluded from the test accuracy evaluation was the same in both the QbTest (6–12) and the QbTest (12–60) groups. Where SEs or CIs were not reported for an AUC estimate, these were estimated from the AUC and number of patients with and without ADHD, using the R (The R Foundation for Statistical Computing, Vienna, Austria) package auctestr.[42,43] If a measure of accuracy (e.g. sensitivity, specificity and AUC) was reported without providing the information needed to calculate 2 × 2 tables, then these data were extracted.

Where multiple sets of 2 × 2 data were reported in a single study, for example, for different tests, test components, target conditions, ADHD subtypes, thresholds or subgroups of interest, all data were extracted. For studies comparing two or more index tests (at least one of which was a sensor CPT) and a reference standard, if full cross-classifications of test results (2 × 2 × 2 data) were reported, these were also extracted.

For studies that reported data on qualitative interviews or survey data, data were extracted on the following: author (year), study name, country, language, setting, study design, funding and sensor CPT. For each relevant study component (e.g. interview with young people and survey with HCPs), we extracted information about participants, sampling strategy, data collection and analysis.

Where studies were only available as abstracts, or where insufficient data were reported in a study to extract the required information, study authors were contacted for additional information.

Quality assessment strategy

The methodological quality of included RCTs was assessed using the updated Cochrane Risk of Bias Tool (RoB 2).[44] DTA studies were assessed for methodological quality using QUADAS-2.[45] Before-and-after implementation studies were assessed using the Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS–I) tool.[46] Studies that contributed qualitative data were assessed with an amended version of the Critical Appraisal Skills Programme (CASP) checklist for qualitative studies (we excluded question 10 'how valuable is the research?').[47] Studies that contributed

survey data were assessed with the Quality Assessment Checklist for Survey Studies in Psychology (Q-SSP).[48] One reviewer assessed the quality of included studies and this was checked by a second reviewer. Any disagreements were resolved by consensus or discussion with a third reviewer.

Synthesis methods

For each of the four objectives, a narrative summary of included studies is presented. This includes a summary of study characteristics (e.g. study designs, sample size, geographical location, year, age group and test evaluated), outcomes reported and study quality. We also narratively summarised whether studies reported baseline data for PROGRESS-Plus characteristics and whether the studies report results data stratified by these characteristics.

We stratified the synthesis on whether the tests were evaluated in isolation or in combination with clinical assessments and on specific sensor CPT tests evaluated. For each test, the analysis was further stratified on the test subcategory evaluated. We had intended to conduct subgroup analyses based on the following subgroups; however, there were only sufficient data available to stratify on age:

- age (children, young people and adults)
- sex
- ethnicity
- people with mental health, behavioural and neurodevelopmental conditions
- people with developmental trauma
- people in the Youth Justice System or Adult Criminal Justice System
- looked-after children.

Where sufficient data were available, meta-analysis was carried out to generate summary effect estimates. We only had sufficient data on test accuracy outcomes (sensitivity, specificity and AUC) to perform meta-analysis. If a single study reported multiple estimates of 2 × 2 data that could have been included in a single meta-analysis, we selected one set of data for each analysis based on the following hierarchy:

- If multiple control groups were available, we selected the control group most similar to the group in which the test will be used in practice:
    - control group of participants who had been evaluated for suspected ADHD and in whom the condition was ruled out selected in preference to other groups.
    - diseased controls selected in preference to healthy controls.
- Where results were reported for multiple thresholds, we selected the threshold most similar to that evaluated in other studies.
- If data were reported for the whole population and separately for specific population subgroups, we selected data for the full population.

Where at least two sets of 2 × 2 data were available, meta-analysis of sensitivity and specificity was performed using the metadta command[49] in the Stata® (StataCorp LP, College Station, TX, USA) statistical software package (StataCorp. 2023. Software: Release 18. College Station, TX: StataCorp LLC). For analyses based on at least three sets of 2 × 2 data, bivariate random-effects meta-analyses of sensitivity and specificity were performed, with binomial likelihoods.[50,51] Where only two sets of 2 × 2 data contributed to a meta-analysis, we used univariate fixed-effects meta-analysis. Study-level and pooled results were plotted as coupled forest plots and in ROC space. In ROC space, uncertainty around summary results from bivariate and univariate analyses are represented with 95% confidence ellipses or 95% CIs, respectively. Subgroup analysis was performed by QbTest (6–12) and QbTest (12–60). We did not have sufficient studies for formal investigation of other sources of heterogeneity. We also produced summary estimates of the AUC using inverse-variance random-effects models. These were fitted using the metagen command[52] within the 'meta'

package of the R statistical software package [R Core Team. *R: A Language and Environment for Statistical ## Computing*. R Foundation for Statistical Computing, Vienna, Austria; 2021. URL: www.R-project.org/ (accessed December 2023)].

Where studies compared the accuracy of two index tests, we produced plots showing estimates and 95% CI for the two tests in the same population. We tested for differences between estimates of sensitivity or specificity using Fisher's exact test.[53]

If two or more qualitative studies were identified that reported data on the same outcomes, we used the meta-aggregative approach to qualitative synthesis based on guidance from the Joanna Briggs Institute (JBI).[54] One reviewer (Eve Tomlinson) extracted themes from the included studies and then organised them into conceptual categories. This was checked by a second reviewer (Amanda Owen-Smith). We extracted direct quotes to evidence what the synthesised themes presented. Where conflicted information, or negative cases, were identified, these were pursued further to enhance methodological rigour. Where available, data from survey studies were also used to evidence the themes presented, clearly marked in the full synthesis in *Appendix 5* as additional 'findings from quantitative data'.

Protocol changes

The following changes were made to the methods specified in the review protocol:[1]

- We clarified the eligibility criteria for study setting to make it clear that studies with control groups recruited in other settings were eligible: 'Studies in which some participants (e.g. control groups) were enrolled in other settings (e.g. community setting) were also eligible'.
- We broadened our inclusion criteria for comparative studies to also include data from studies that compared the accuracy of sensor CPTs (alone or in combination with clinical diagnosis) with the accuracy of clinical diagnosis alone.
- We identified one study of the QbMini. Although the original protocol did not specify that this test would be eligible, as it very similar to the QbTest, just aimed at younger children, this was also included.

Chapter 4 Results of clinical effectiveness review

Results of the searches

The searches of bibliographic databases and trials registries identified 507 unique reports. Additional methods of study identification (website checking, reference checking of included studies, checking studies included in systematic reviews and checking manufacturer's submissions) identified 1200 unique reports. In total, 30 studies in 43 reports were included in the review (*Figure 1*). We identified nine systematic reviews.[26,30,55–61] None of these fulfilled the criteria specified for inclusion of systematic reviews and so they were screened to identify potentially relevant studies.

Most studies evaluated the QbTest, there was one study of the QbCheck, one of QbMini, two of Nesplora AULA and two of the EF Sim test. Three studies were only reported as conference abstracts – two DTA studies,[62,63] and one implementation study.[64] The authors of these studies did not respond to our request for a full publication. All included studies were reported in English, except for one conference abstract of a DTA study, which was reported in Spanish and we translated it using Google Translate (Google Inc., Mountain View, CA, USA).[63] The translation was checked by a native Spanish speaker to ensure it was an accurate summary of the abstract. We could not locate the full text for one unpublished potentially relevant study for the QbTest, which we identified by checking references of the included studies.[65]

We contacted the authors of nine studies to request additional data or to clarify information presented in the study reports. Five responded to our requests, including the authors of four DTA studies,[18,66–68] and one implementation study.[69] Four did not respond to our requests, including three DTA studies[41,62,70] and one implementation study.[64]

We identified two ongoing studies. One ongoing study was identified by our searches.[71] This study is evaluating the EF Sim test in children aged 8–13 years and includes a group with diagnosed ADHD and a control group of children without ADHD. One ongoing study was highlighted in the submissions from the manufacturers, with limited detail [no National Clinical Trial (NCT) number or reference to study provided]. Peili Vision reported that several pilots using the EF Sim test are being set up in spring 2024 in the UK to implement it as part of an early triage tool (no further information provided). *Appendix 2* provides an overview of included (see *Table 33*), ongoing (see *Table 34*) and excluded (see *Tables 35–40*) studies.

Objective 1: diagnostic accuracy and clinical effectiveness of sensor continuous performance tests for the diagnosis of attention deficit hyperactivity disorder in people referred with suspected attention deficit hyperactivity disorder

We included 29 studies (38 reports) for objective 1: 2 RCTs (1 of these also provided data on accuracy,[18] both also included a survey and qualitative substudy),[18,72] 20 DTA studies[29,41,62,63,66–68,70,73–84] (2 included a survey of patient views on the acceptability of the test),[75,77] 5 uncontrolled before–after implementation studies[31,64,69,85,86] (2 also provided information on patient/clinician views – 1 was survey and qualitative evaluation),[31] 1 was survey[69] and 2 studies that only reported on patient and clinicians acceptability of sensor CPTs.[87,88]

Impact of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder on patient outcomes

Only one study, the Functional Assessment of Cancer Therapy (FACT) UK-based feasibility RCT, considered the impact of sensor CPTs on patient outcomes.[72] As this study was a feasibility trial, the primary objective was to determine the feasibility of conducting a full trial rather than to compare the outcomes between intervention groups. This study was conducted in the very specific population of boys with symptoms of possible ADHD, aged 15–18 years, in Young Offenders Institutions (YOIs) in England. It compared usual care combined with the QbTest (12–60) to usual care alone in 60 boys (30 in each treatment group). FU was poor, with only 32% of participants followed up at 6 months, although the authors report that this was affected by coronavirus disease 2019 (COVID-19) restrictions. As shown

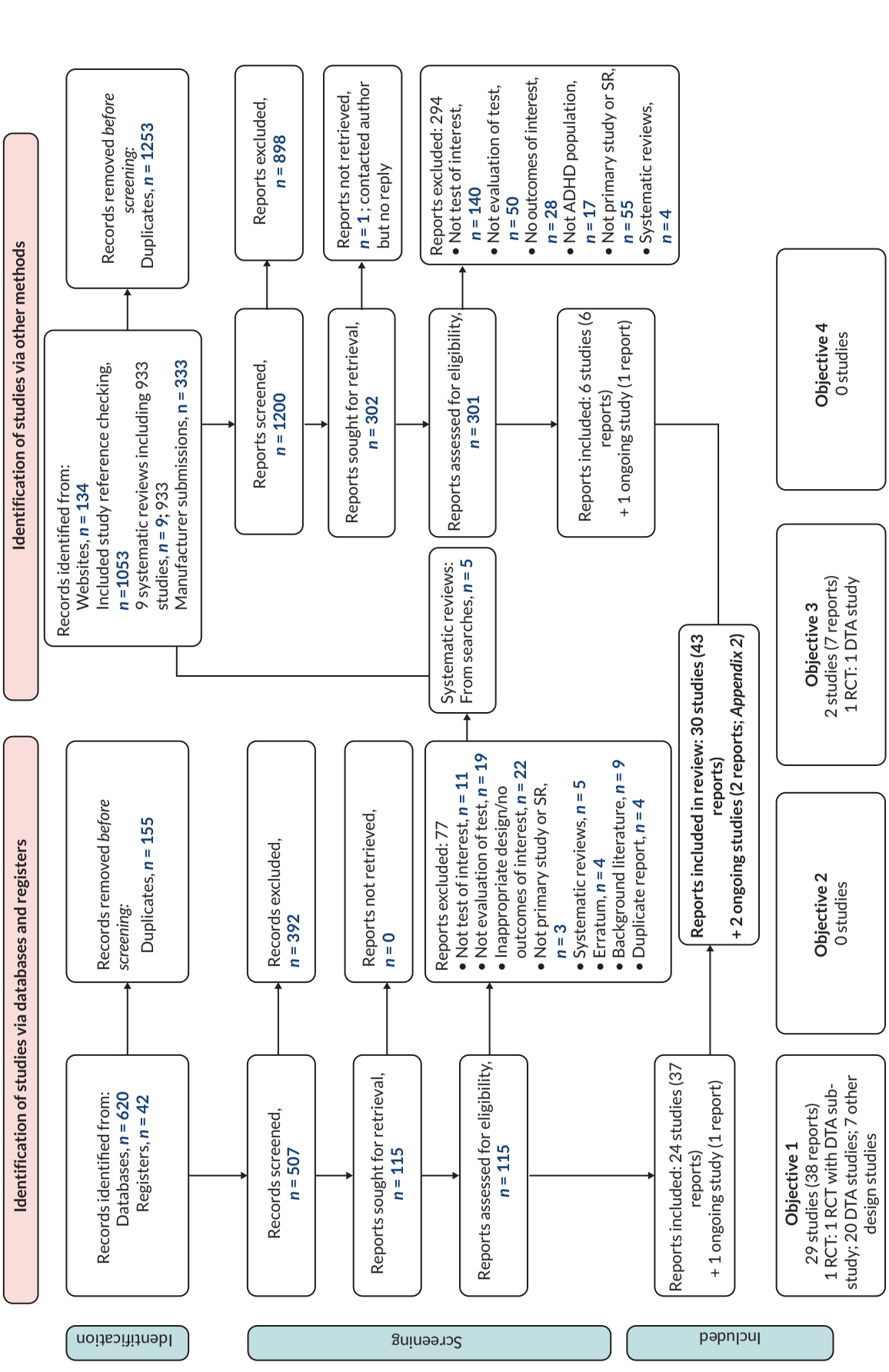

FIGURE 1 The PRISMA flow chart. SR, systematic review.

in *Appendix 3*, *Table 46*, this study reported baseline data on four PROGRESS-Plus characteristics (sex, ethnicity, education and time-dependent relationships). Due to the feasibility design, small sample size and low FU rates, it was not possible to draw conclusions regarding clinical effectiveness from this study.

Diagnostic accuracy of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder

Twenty-one studies (28 reports) evaluated the accuracy of sensor CPTs for the diagnosis of ADHD (*Tables 2* and *3*). One of these studies was a RCT (AQUA trial) included in *Impact of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder on process measures*, which also reported a DTA substudy;[18] all others were DTA

TABLE 2 Overview of studies that provide information on the diagnostic accuracy of sensor CPTs for the diagnosis of ADHD

Feature	Category	Number of studies
Design	One-gate (diagnostic cohort/cross-sectional)	10
	Multigate (diagnostic case-control)	10
	Unclear	1
Test evaluated	QbTestPlus	10
	QbTest or QbTestPlus	4
	QbTest	2
	QbMini	1
	QbCheck	1
	EPELI (EF Sim)	1
	Nesplora AULA	2
Combination with clinical information	Test evaluated alone	17
	Test evaluated in combination with clinical information	3
	Both	1
Comparison with other tests	Accuracy of other CPT compared with accuracy of sensor CPT	3
	Comparison with clinical diagnosis alone	1
	No comparison	19
Reference standard	DSM-IV diagnostic criteria	9
	DSM-V diagnostic criteria	5
	DSM (version not specified)	1
	Independent consensus diagnosis using DAWBA (based on DSM-V and ICD-10)	1
	K-SADS-PL interview	1
	ICD-10	1
	Diagnostic process according to clinic's standard diagnostic procedure – no further information	1
	Assessment of disruptive behaviour pathway used locally as the standard	1
	Not reported	1
Country	Sweden	8
	Germany	3
	UK	3
	The Netherlands, Germany and Sweden	1

TABLE 2 Overview of studies that provide information on the diagnostic accuracy of sensor CPTs for the diagnosis of ADHD (*continued*)

Feature	Category	Number of studies
	Sweden and Germany	1
	Finland	1
	Germany, Sweden and the USA	1
	Spain	1
	Not reported	2
Setting	Secondary care	19
	Population-based	1
	Not reported	1
Funding	Non-industry	6
	Non-industry (but the authors developed the test)	1
	Mixed (non-industry and industry)	4
	Industry (authors employed by QbTech)	1
	Not reported but one author employed by QbTech	1
	Unfunded	3
	Not reported	4
	'N/A' (no further information)	1
Sample size (number analysed)	< 50	1
	50–100	4
	100–200	7
	200–500	6
	> 500	1
Age group	Children aged 5 years	1
	Children (6–12 years)	5
	Children (5–15 years)	1
	Children (6–12 years) and adolescents (12–18 years)	2
	Children (age not reported)	1
	Adolescents (12–18 years)	1
	Adolescents (12–18 years) and adults	1
	Adults	8
	Older adults	1
% Male	< 25%	0
	25–50%	5
	50–75%	11
	> 75%	2
	Unclear	3

DAWBA, Development and Wellbeing Assessment; DSM, *Diagnostic and Statistical Manual of Mental Disorders*; ICD-10, *International Statistical Classification of Diseases and Related Health Problems*, Tenth Revision; K-SADS-PL, Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime Version; N/A, not applicable.

TABLE 3 Details of studies that provided information on the diagnostic accuracy of sensor CPTs for the diagnosis of ADHD

Author, design and location	Test	Population and reference standard
QbTest combined with clinical assessment		
Bijlenga (2019)[78] The Netherlands, Germany and Sweden; two-gate design (healthy controls)	QbTest (12–60); QbTest (12–60) + clinical judgement (symptom severity self-report scale)	ADHD group: adults (age 55+); DSM-IV-TR ADHD diagnosis (*n* = 97) Healthy controls: adults (age 55+) with score below cut-off on symptom severity measures (*n* = 112) matched on age and gender
Emser (2018)[83] Germany; two-gate design (healthy controls)	QbTest (6–12) or QbTest (12–60) + objective clinical assessment (KiTAP and TAP)	Children and adults ADHD: DSM-IV-oriented clinical interview by experienced clinician including KSADS and rating scales (*n* = 68). Controls: No established or suspected ADHD diagnosis or family history of ADHD, unclear how assessed. Age/gender matched at group level (*n* = 68)
Groom (2016)[80] UK; two-gate design (ADHD controls)	QbTest (12–60) + clinical judgment (Conners Adult Rating Scale and Autism Quotient-10)	Adults (aged 18–60 years) ADHD group (*n* = 32): DSM-V diagnosis of ADHD Autism (ASD) group (*n* = 25): ICD-10 diagnosis of Asperger's syndrome
Hollis (2018)[18] UK; one-gate design	QbTest (6–12 or 12–60) + clinical judgement Clinical judgement alone	Children and adolescents (aged 6–17 years) enrolled in AQUA trial Consensus diagnosis using DAWBA[89] QbTest group: ADHD confirmed (*n* = 43); no-ADHD (*n* = 43) Control group: ADHD confirmed (*n* = 51); no-ADHD (*n* = 25)
QbTest alone		
Adamou (2022)[81] UK; one-gate design	QbTest (12–60)	Adults referred to Specialist Adult ADHD and Autism service DSM-V – ADHD confirmed (*n* = 38) vs. no ADHD (*n* = 31)
Bijlenga (2019)[78] The Netherlands, Germany and Sweden; two-gate design (healthy controls)	QbTest (12–60); QbTest (12–60) + clinical judgement (symptom severity self-report scale)	ADHD group: adults (age 55 +); DSM-IV-TR ADHD diagnosis (*n* = 97) Healthy controls: adults (age 55 +) with score below cut-off on symptom severity measures (*n* = 112) matched on age and gender
Brunkhorst-Kanaan (2020)[68] Germany; one-gate	QbTest (12–60)	Adults referred to specialist outpatient clinic for suspected ADHD diagnosis DSM-V: DIVA interview. ADHD confirmed (*n* = 94); no ADHD (*n* = 20)
Edebol (2013)[79] Sweden and Germany; two-gate design (healthy controls)	QbTest (12–60)	ADHD group: adults diagnosed with ADHD following clinical assessment adhering to DSM-IV Non-ADHD control group: Healthy adults with no known psychiatric diagnoses
Edebol (2012)[76] Sweden; four-gate design (diseased and healthy controls)	QbTest (12–60)	ADHD group: DSM diagnosis (version not specified) (*n* = 53) B/B group: diagnosed with B/B (*n* = 45) Disconfirmed ADHD (*n* = 29) (retained for analysis) Healthy controls (*n* = 179)
Edebol (2011)[84] Sweden; one-gate design	QbTest (12–60)	Adults awaiting clinical assessment of ADHD. DSM-IV – clinical assessments. ADHD confirmed (*n* = 12) and no ADHD group (*n* = 7)
Hult (2018)[67] Sweden; one-gate	QbTest (6–12)	Children (aged 6–12 years) with suspected ADHD, autism or another neurodevelopmental disorder. Diagnosis based on DSM-IV; assessed by multiprofessional team ADHD confirmed (*n* = 124); no-ADHD (*n* = 58)
Johansson (2021)[70] Sweden; one-gate	QbTest (12–60)	Adolescent (age 15) population with high occurrence of neurodevelopmental disorders, including ADHD K-SADS-PL interview confirmed ADHD (*n* = 89) and no ADHD (*n* = 248)
Pettersson (2018)[82] Sweden; one-gate design	QbTest (12–60)	Adults referred for ADHD assessment ADHD diagnosed based on expert clinical assessment (DSM-IV), SCID-I, SCID-II. ADHD confirmed (*n* = 60) and no ADHD group (*n* = 48)

TABLE 3 Details of studies that provided information on the diagnostic accuracy of sensor CPTs for the diagnosis of ADHD (*continued*)

Author, design and location	Test	Population and reference standard
Sharma (2009)[62] UK; unclear design	QbTest (6–12) or QbTest (12–60)	Children and adolescents (aged 5–15 years, *n* = 50) selected from QbTest database, which were evaluated for ADHD as per local protocol or as diagnosed by child/family guidance Assessment of disruptive behaviour pathway used locally as standard; no. with/without ADHD not reported
Söderström (2014)[74] Sweden; one-gate	QbTest (12–60)	Adults referred to neuropsychological clinic for ADHD assessment DSM-IV: Clinical assessment confirmed ADHD (*n* = 41) and no ADHD (*n* = 20)
Stevanovic (2023)[41] Sweden; one-gate	QbTest (6–12) and QbTest (12–60)	Children and adults referred for evaluation of suspected neurodevelopmental/psychiatric disorder. Diagnosis based on clinic's standard diagnostic procedure (no further information). ADHD confirmed (*n* = 708); no-ADHD (*n* = 220)
Tallberg (2019)[66] Sweden; one-gate	QbTest (6–12)	Children who screened positive for ADHD and were referred for further assessments in CAP clinic. Diagnosis based on DSM-IV ADHD confirmed (*n* = 80); no-ADHD (*n* = 38)
QbMini		
Hamadache (2021)[29] Germany; three-gate design (healthy and diseased controls)	QbMini	Children (age 5): ADHD-based DSM-IV (*n* = 37) SLI (*n* = 27) Healthy controls: tested at pre-schools and found to be normally developing (*n* = 55)
QbCheck		
Ulberstad (2020)[77] Germany, Sweden, USA Two-gate (healthy controls)	QbCheck	Adolescents and adults (12–59 years) Cases: DSM-V diagnostic criteria (*n* = 69) Controls: healthy controls; those with high levels of inattention/hyperactivity/impulsivity according to DSM-V excluded (*n* = 73)
Nesplora Kids (AULA)		
Rufo-Campos (2012)[63] Not reported; two-gate	Nesplora Kids (AULA)	Children (age not reported) ADHD group: children diagnosed with ADHD – no further information reported (*n* = 62) Non-ADHD group: children without ADHD diagnosis – no further information reported (*n* = 62)
Zulueta (2019)[73] Spain; two-gate (healthy controls)	Nesplora Kids (AULA)	Children (aged 6–16 years) ADHD group: fulfilled DSM-V criteria; recruited from outpatient department (*n* = 213) Healthy control group: from schools and neurology clinics minimal ADHD symptoms and no other behavioural disorder (*n* = 194 included)
EPELI		
Seesjärvi (2022)[75] Finland; two-gate (healthy controls)	EPELI	Children (aged 9–12 years) ADHD group (*n* = 38): ADHD diagnosis by licensed physician using ICD-10 Non-ADHD group (*n* = 38): no mental or behavioural disorder; matched to cases

CAP, Child and Adolescent Psychiatry; DAWBA, Development and Wellbeing Assessment; DIVA, Diagnostic Interview for ADHD in Adults; DSM-IV-TR, *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, Text Revision; ICD-10, *International Statistical Classification of Diseases and Related Health Problems*, Tenth Revision; KiTAP, Test of Attentional Performance for Children (child version); K-SADS-PL, Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime Version; SCID-I, Structured Clinical Interview for DSM-IV v1; SCID-II, Structured Clinical Interview for DSM-IV v2; SLI, specific language impairment; TAP, Test of Attentional Performance for Children.

studies. *Table 2* provides a summary of study characteristics for these studies. *Appendix 3*, *Table 41* provides further baseline details.

The majority of studies evaluated the QbTest (6–12 or 12–60 depending on age), with single studies evaluating the QbMini and QbCheck (online) versions of this test. There was only one study of EF sim (reported as Epeli test) and two of Nesplora Kids; there were no studies of EF Sim web or of Nesplora Adults. Most studies evaluated the accuracy of the tests in isolation, three evaluated the accuracy of the QbTest in combination with some form of clinical assessment[18,80,83] and one evaluated the test both in isolation and combined with clinical assessment.[78] Three studies provided a direct comparison of the accuracy of the sensor CPT with that of a non-sensor CPT,[66,75,82] and one compared the accuracy of QbTest combined with clinical information with QbTest alone.[18] Fifteen studies used the DSM-IV or DSM-V criteria for the diagnosis of ADHD as the reference standard, with single studies using *International Statistical Classification of Diseases and Related Health Problems*, Tenth Revision (ICD-10),[75] Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime Version (K-SADS-PL) interview,[83] independent consensus diagnosis using Development and Wellbeing Assessment (DAWBA).[18,89] One reported that the diagnostic process was according to the clinic's standard diagnostic procedure without providing any further details,[41] one used an assessment of disruptive behaviour pathway used locally as the standard[62] and one did not report any details about a reference standard (conference abstract).[63] *Table 4* provides an overview of the reference standards used in the included studies.

TABLE 4 Overview of the reference standards used in the studies that contribute accuracy data to objective 1

Reference standards	Details
DSM-IV diagnostic criteria[90]	Fourth edition of the DSM. The DSM contains standardised diagnostic criteria for mental disorders, which are used by HCPs to guide diagnosis. For ADHD, it includes 18 symptoms divided into two domains: inattention and hyperactivity/impulsivity. At least six symptoms in one domain are required for diagnosis[91]
DSM-V diagnostic criteria[92]	Fifth and most recent version of DSM. The same 18 symptoms and domains are included as in DSM-IV, but there were also several changes to the handbook including (but not limited to): only 5 symptoms are required in one domain for adult diagnosis (still six for younger persons); examples have been added to facilitate application across the lifespan; comorbid diagnosis with ASD is now allowed; ADHD moved to 'neurodevelopmental disorders' chapter[91]
DSM (version not specified)	The DSM (as above)
ICD-10[93]	The ICD-10 is the tenth revision of the classification system (the current version is ICD-11) created by the WHO to provide a standardised way to report and code mortality and morbidity data. The classification contains codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, external causes of injury or diseases. The ICD-10 calls ADHD 'hyperkinetic disorder' and requires hyperactivity, inattention and impulsivity to be present. The ICD-10 diagnostic criteria for ADHD are more restrictive than DSM criteria[94]
Independent consensus diagnosis using the DAWBA[89]	The DAWBA consists of interviews and rating scales to generate an ICD-10 or DSM-V psychiatric diagnoses in 5- to 16-year-olds. It involves a parent interview, an interview for young people aged 11+, a teacher questionnaire and a computer-assisted clinical diagnostic rating based on the information. Clinical raters use the computer-generated rating to decide whether to accept or overturn the computer diagnosis (or lack of diagnosis) after reviewing all the information
K-SADS-PL[95]	The Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version interview (K-SADS-PL) is a semistructured diagnostic interview to assess mental disorders, including, but not limited to, ADHD, schizophrenia and major depressive disorder. The schedule has six components [developmental history, diagnostic screening interview, completion checklist supplement to screen for additional disorders, appropriate diagnostic supplements (review presence/absence of symptoms for other disorders), supplementary lifetime diagnosis checklist (summarises which disorders have been present from first episode to now), CGAS (level of functioning)]. It generates DSM-III-R and DSM-IV diagnoses[95]
Diagnostic process according to clinic's standard diagnostic procedure – no further information	N/A
Assessment of disruptive behaviour pathway used locally as the standard	N/A
Not reported	N/A

CGAS, Children's Global Assessment Scale; ICD-10, *International Statistical Classification of Diseases and Related Health Problems*, Tenth Revision; N/A, not applicable.

Ten studies used the more reliable one-gate design (also known as diagnostic cohort or cross-sectional study) where a single group of participants was enrolled and all then received both the index test and reference standard. Five of these single-gate studies enrolled adults with suspected ADHD referred for ADHD assessment in secondary care.[68,74,81,82,84] Two studies enrolled children only – of these, one recruited children with suspected ADHD, autism or another neurodevelopmental disorder,[67] and one recruited children who had screened positive for ADHD and were referred for further ADHD assessment.[66] One study enrolled adolescents with a high occurrence of neurodevelopmental disorders, including ADHD.[70] The remaining two single-gate studies included mixed populations: one enrolled children and adolescents who had been referred for their first ADHD assessment (and enrolled in the AQUA trial QbOpen arm),[18] and one enrolled children and adults referred for evaluation of suspected neurodevelopmental or psychiatric disorder.[41] Ten studies used a multigate design (also known as diagnostic case-control study) where two or more separate groups of participants were enrolled – one with known ADHD and one or more without ADHD, and the participants then received the sensor CPT. Eight studies had a two-gate design, in which they enrolled an ADHD group (cases) and one control group. Seven of these studies enrolled healthy controls[63,73,75,77–79,83] and one enrolled controls with autism.[80] One study enrolled four groups: an ADHD group (cases) and three different control groups (a group who had been assessed for ADHD and in whom this had been ruled out, a group with bipolar disease and healthy controls)[76] and another enrolled three groups [an ADHD group, a group with specific language impairment (SLI) and healthy controls].[29] For the four-gate study, we selected the group that had been assessed for ADHD as the control group to use for the analysis; for the three-gate study, we used the group with SLI. One study had an unclear study design, with limited study details reported in a conference abstract.[62]

Studies were conducted almost exclusively in Europe, with eight studies conducted in Sweden; one study was a multinational study that included sites in the USA in addition to Germany and Sweden. One study was conducted in a population-based setting recruiting participants from a twins registry, one study (reported in a conference abstract only) did not report setting[63] and all other studies were conducted in secondary care (e.g. recruiting participants from specialised ADHD outpatient clinics, neuropsychiatric centres, university ADHD outpatient clinics or CAMHS), although some included controls recruited from community settings (e.g. university, waiting areas, schools and workplaces). Five studies were at least partly funded by industry, and in further two studies, the authors either worked for the test manufacturer or developed the test.

Eight studies were conducted in adults, five in children (aged 6–12 years), two in children and adolescents (aged 12–18 years), with single studies in children aged 5 years, children aged 5–15 years, children (age not specified), adolescents, adolescents and adults and older adults. The majority of studies included more male participants than female participants, particularly in the ADHD groups, although five studies included slightly higher proportions of female participants.

Twenty out of 21 DTA studies reported baseline data on at least one PROGRESS-Plus characteristic. The one study that did not report on PROGRESS-Plus was a conference abstract with limited detail on the population.[63] Data on place of residence were reported by 2 studies (10%),[76,83] ethnicity by 1 study (5%),[18] occupation by 4 studies (19%),[76,79,82,84] sex by 18 studies (86%), religion by 0 studies (0%), education by 8 studies (38%);[70,75,76,78,79,82–84] socioeconomic status by 3 studies (14%)[75,79,80] and social capital by 0 studies (0%). Baseline data on neurodevelopmental/learning disorders were reported by 13 studies (62%), and data on mental health disorders were reported by 11 studies (52%). Features of relationships (e.g. marital status, household set-up and major school problems) were reported by three studies (14%).[70,76,79] None of the studies reported data stratified by PROGRESS-Plus characteristics. *Appendix 3*, *Table 42* presents the PROGRESS-Plus data extracted from each study.

Risk of bias

Only 3 of the 21 studies were judged at low risk of bias across all QUADAS-2 domains, 3 were judged at unclear risk of bias and 15 were judged at high risk of bias (*Table 5* and *Appendix 3*, *Table 44*).

Eleven studies were judged at high risk of bias for the patient spectrum domain. Ten studies were judged as high risk because they used a two-gate design where studies recruited a group of patients with known ADHD and a group without ADHD, either a healthy control group or a group of patients with an alternative diagnosis. One other study (the

TABLE 5 Results of the QUADAS-2 assessment of risk of bias in DTA studies included for objective 1

	Patient selection	Index test	Ref stand	Patient flow	Overall bias	Rationale
Adamou (2022)[81]	☺	☺	?	☺	?	Unclear whether reference standard interpreted blind to QbTest results
Bijlenga (2019)[78]	☹	?	☺	☹	☹	Two-gate design. No information on threshold. High proportion of dropouts (25/234)
QbTest alone						
Brunkhorst-Kanaan (2020)[68]	☺	☺	☺	☺	☺	No concerns
Edebol (2013)[79]	☹	☺	☺	☹	☹	Two-gate design. 4/55 ADHD group excluded from analysis
Edebol (2011)[84]	☺	☺	☺	☺	☺	No concerns
Edebol (2012)[76]	☹	☺	?	☺	☹	Four-gate design. Limited details on reference standard
Emser (2018)[83]	☹	?	☺	☺	☹	Two-gate design. No information on threshold for QbTest + clinical assessment or on blinding of reference standard.
Groom (2016)[80]	☹	?	☺	☹	☹	Two-gate design. No information on blinding of QbTest to case/control status. No detail on threshold. High proportion of dropouts (5/37 in ADHD group)
Hamadache (2021)[29]	☹	?	☺	☺	☹	Mutligate design. Limited details on QbMini
Hollis (2018)[18]	☹	☺	☹	☺	☹	Participants eligible for DTA substudy if diagnostic decision had been made at 6 months (QbOpen eligible sample *n* = 94/123; QbBlind *n* = 86/127). Reference standard diagnosis made using limited data for around 50% participants as either parent or teacher assessment was missing
Hult (2018)[67]	☺	☺	☺	☺	☺	No concerns
Johansson (2018)[70]	☺	?	☹	☹	☹	Reference standard K-SADS-PL – not ADHD-specific and so may not correctly diagnose ADHD. High proportion of participants excluded from 2 × 2 table
Pettersson (2018)[82]	☺	☺	?	☺	?	Unclear if reference standard is blind to QbTest result
Rufo-Campos (2012)[63]	☹	?	?	?	☹	Two-gate design; no details about conduct/interpretation of index test, reference standard, or flow and timing
Seesjärvi (2022)[75]	☹	?	☺	☹	☹	Two-gate design; patients with other listed comorbidities excluded from cases and controls. No information on whether Epeli test interpreters were blinded to diagnosis; high proportion excluded from 2 × 2 table
Sharma (2009)[62]	?	?	?	?	?	Very limited information available from conference abstract
Söderström (2014)[74]	☺	☺	☹	☺	☹	Clinicians aware of QbTest results when interpreting reference standard
Stevanovic (2023)[41]	☺	☺	☹	☹	☹	Unlikely that reference standard interpreted was blind to index test; insufficient details on reference standard but was based on clinic records, not DSM criteria. High proportion of dropouts
Tallberg (2019)[66] – accuracy	☺	?	?	☹	☹	High proportion of missing data. Unclear if reference standard was blinded to QbTest; was not blinded to other tests evaluated
Ulberstad (2020)[77]	☹	?	☺	☹	☹	Two-gate design. Unclear who interpreted the test and if blinded to ADHD status. 7/149 patients were not included in 2 × 2 table
Zulueta (2019)[73]	☹	?	☺	☺	☹	Two-gate design. No information on test interpretation or threshold

AQUA trial) was judged as high risk for this domain because participants were only eligible for the DTA substudy if they had a diagnostic decision at 6 months (94/123 participants in QbTest group and 76/127 in the control group).

None of the studies were judged at high risk of bias for the index test domain, although nine were judged at unclear risk of bias as they did not provide sufficient information on how the sensor CPT was evaluated or on the threshold to determine a 'positive' test result. While the QbTest does not specify a threshold for positivity so there is no standard threshold that can be applied, it is important that study authors pre-specify any threshold that is used to dichotomise results.

Four studies were judged at high risk of bias for the reference standard domain – one study used the K-SADS-PL criteria (see *Table 4*) which is not specific for ADHD and so may not be as accurate as DSM or ICD criteria; and in two studies, information on the sensor CPT was available to the person interpreting the reference standard results; in one of these, the ADHD diagnosis was made based on criteria used within the clinic rather than on accepted criteria such as the DSM-V criteria. The AQUA trial used independent consensus diagnosis by two independent child psychiatrists based on the DAWBA criteria, which is considered as an accepted reference standard. However, it was judged at high risk of bias for the reference standard domain, as in 123/241 participants, DAWBAs were missing from 1 informant (i.e. either parent or teacher), meaning the independent assessors did not have access to this information when making a diagnosis. A further seven studies were judged at unclear risk of bias – five did not provide sufficient information to judge whether the reference standard was interpreted blind to the index test results, and in three studies, it was unclear whether the reference standard was likely to correctly classify participants as having ADHD.

Eight studies were judged at high risk of bias for the flow and timing domain due to a large number of enrolled participants not being included in the analysis.

Concerns regarding applicability

Six studies were judged at low concerns regarding applicability, 3 at unclear concerns and 12 at high concerns (*Table 6* and *Appendix 3*, *Table 44*). All 10 studies that used a two-gate design were considered to have concerns regarding applicability as they did not enrol a group of participants with suspected ADHD. Two of the one-gate studies were also considered to have concerns regarding applicability as they enrolled a selected subgroup to assess for ADHD – both enrolled participants with a high level of neurodevelopmental/neuropsychological disorders. Concerns regarding applicability were high for the index test for one study – in this study, the conduct of the QbTest did not follow the manufacturer's submission instructions. In a further 11 studies, the applicability was judged as unclear for the index test as there were insufficient details on how the sensor CPT was performed. Five studies were judged at unclear concerns regarding applicability for the reference standard domain as there were insufficient details on the reference standard to determine how this was classifying ADHD.

Accuracy of QbTest plus clinical information

Four studies[18,78,80,83] provided information on the accuracy of QbTest in combination with clinical information; one of these studies reported results separately for QbTest (12–60) and for QbTest (6–12) (*Figure 2* and *Appendix 3*, *Table 43*). We did not identify any studies of any of the other sensor CPTs in combination with clinical information.

The Hollis (2018) AQUA trial was the only study to combine the QbTest information with clinical assessment in the same way that it would be used in practice. Other studies constructed prediction models that combined information from specific clinical scales with results from the QbTest. The Hollis (2019) and Groom (2016) studies used an overall combined output from the QbTest. Bijlenga (2019) used information from the hyperactivity and inattention domains, and Emser used individual QbTest outputs. *Table 7* provides a summary of the clinical information used and how studies combined this with QbTest results. As the type of clinical information and QbTest data used varied across studies, it was not considered appropriate to pool data.

The AQUA trial used the more reliable one-gate design, and all others used a two-gate design. Risk of bias was high for all studies that used a two-gate design. The AQUA trial was also judged at high risk of bias due to limitations with the reference standard and restriction to those with a diagnosis at 6 months.

TABLE 6 Results of the QUADAS-2 assessment of concerns regarding applicability of DTA studies included for objective 1

	Patients	Index test	Reference stand	Overall	Rationale
Adamou (2022)[81]	☺	☺	☺	☺	No concerns
Bijlenga (2019)[78]	☹	☺	☺	☹	Two-gate design
Brunkhorst-Kanaan (2020)[68]	☺	?	☺	?	Limited details on test conduct
Edebol (2013)[79]	☹	☺	☺	☹	Two-gate design
Edebol (2011)[84]	☺	☺	☺	☺	No concerns
Edebol (2012)[76]	☹	☺	?	☹	Four-gate design; limited details on reference standard
Emser (2018)[83]	☹	?	☺	☹	Two-gate design; limited details on test conduct
Groom (2016)[80]	☹	?	☺	☹	Two-gate design; limited details on test conduct
Hamadache (2021)[29]	☹	?	☺	☹	Three-gate design; limited details on test conduct
Hollis (2018)[18]	☺	☺	☺	☺	No concerns
Hult (2015)[67]	☺	☺	☺	☺	No concerns
Johansson (2018)[70]	☹	?	?	☹	High proportion of neurodevelopmental disorders – unlikely to be reflective of population with symptoms of ADHD
Pettersson (2018)[82]	☺	☺	☺	☺	No concerns
Rufo-Campos (2012)[63]	☹	?	?	☹	Two-gate design. Limited details on index test conduct and interpretation; no details about reference standard
Seesjärvi (2022)[75]	☹	?	☺	☹	Two-gate deign; limited details on test conduct
Sharma (2009)[62]	?	?	?	?	Very limited information available from conference abstract
Söderström (2014)[74]	☺	☺	☺	☺	No concerns
Stevanovic (2023)[41]	☹	☹	?	☹	Children referred for evaluation of various neuropsychological conditions (not just ADHD). Test conduct did not follow manufacturer's instructions
Tallberg (2019)[66] – accuracy	?	?	☺	?	Children had screened positive for ADHD and so were referred for further evaluation – unclear if representative of review population
Ulberstad (2020)[77]	☹	?	☺	☹	Two-gate design. Limited details on test conduct
Zulueta (2019)[73]	☹	?	☺	☹	Two-gate design. Limited details on test conduct

Estimates of sensitivity ranged from 80% (95% CI 61% to 92%) to 94% (95% CI 79% to 99%). Estimates of specificity ranged from 40% (95% CI 25% to 56%) to 91% (95% CI 84% to 96%), but these were above 76% for all but the AQUA trial. It is likely that the limited information available to those making the reference standard diagnosis may have resulted in the diagnosis being too stringent – this would have resulted in more FP results, leading to an underestimate of specificity. Restriction to those with a diagnosis at 6 months is likely to have overestimated the accuracy of the test, as those without a diagnosis are more likely to be a difficult-to-diagnose group.

TABLE 7 Overview of how studies combined clinical information with QbTest results

Study author (date)	Details of 'QbTest + clinical information'
Bijlenga (2019)[78]	QbTest + self-reported ADHD symptom severity: Several self-report questionnaires were used to assess symptom severity, ADHD-RS was used in the Netherlands, which assesses the DSM-IV-TR ADHD symptoms.[96] In Sweden, the Swedish version of the ADHD Symptom Rating Scale was used, which also assesses DSM-IV-TR ADHD criteria.[97] In Germany, the German version of the Conners' Adult ADHD Rating Scale (self-report long version) was used, which assesses DSM-IV ADHD criteria.[98] In order to establish a unified symptom severity outcome, the total scores per patient were transformed into a 0–100% score, taking into account the score range of each measure. This unified outcome was called the 'ADHD symptom severity score' The authors conducted two binary logistic regressions – the first model included only QbTest factors (QbHyperactivity and QbInattention) and the second model included both QbTest factor scores and self-reported ADHD symptom severity. Estimates of sensitivity and specificity were derived from the models; details on how this was done were not reported
Emser (2018)[83]	QbTest + objective clinical assessment (KiTAP and TAP): Three subtests from the TAP (test battery of attention)[99] and KiTAP (child version of the test battery of attention)[100] were used: Go/NoGo task, divided attention and sustained attention. The authors provided accuracy of ADHD diagnosis using the output from the QbTest and TAP tasks The authors developed prediction models that combined the QbTest components and TAP assessment variables. Estimates of sensitivity and specificity were derived from the models; details on how this was done were not reported
Groom (2016)[80]	QbTest + Conners Adult Rating Scale and Autism Quotient-10: Self- and observer-reported symptom ratings were collected from all participants using the E-ADHD subscale of the CAARS-E,[101] which measures ADHD symptoms, and the Autism Quotient-10 (AQ-10), which screens for ASDs[102] The authors conducted binary logistic regression to combine data from the QbTest composite score with data from the CAARS-E and AQ-10. Sensitivity and specificity were calculated based on the % of participants correctly assigned to the ADHD and ASD control groups
Hollis (2018)[18]	Usual diagnostic workup (typically, this included an interview with the child and their family and the completion of at least one standardised informant-based behavioural assessment measure) with QbTest results being available to clinician

AQ-10, Autism Spectrum Quotient-10; CAARS-E, Conners Adult ADHD Rating Scale; KiTAP, Test of Attentional Performance for Children (child version); TAP, Test of Attentional Performance for Children.

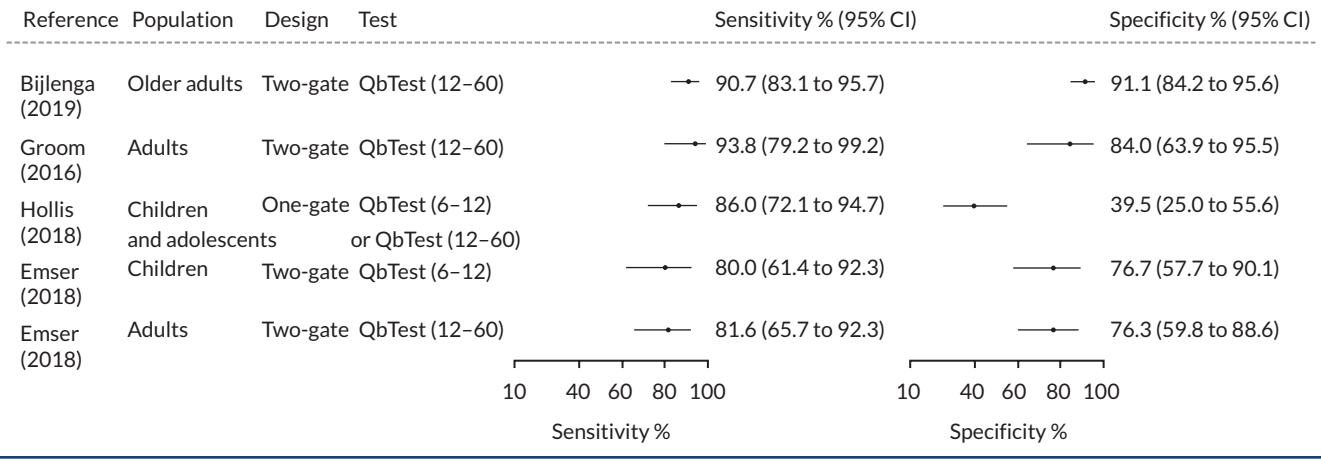

FIGURE 2 Forest plot showing estimates of sensitivity and specificity with 95% CIs for studies that evaluated sensor CPTs in combination with clinical assessment.

Accuracy of QbTest

Thirteen studies evaluated the accuracy of the QbTest alone (*Figures 3* and *4* and *Appendix 4*, *Figure 19*). Three studies evaluated the version for children aged 6–12 years,[41,66,67] 10 studies evaluated the version for older children and adults aged 12–60 years,[41,68,70,74,76,78,79,81,82,84] 1 evaluated both versions[62] and 1 evaluated both versions, reporting data

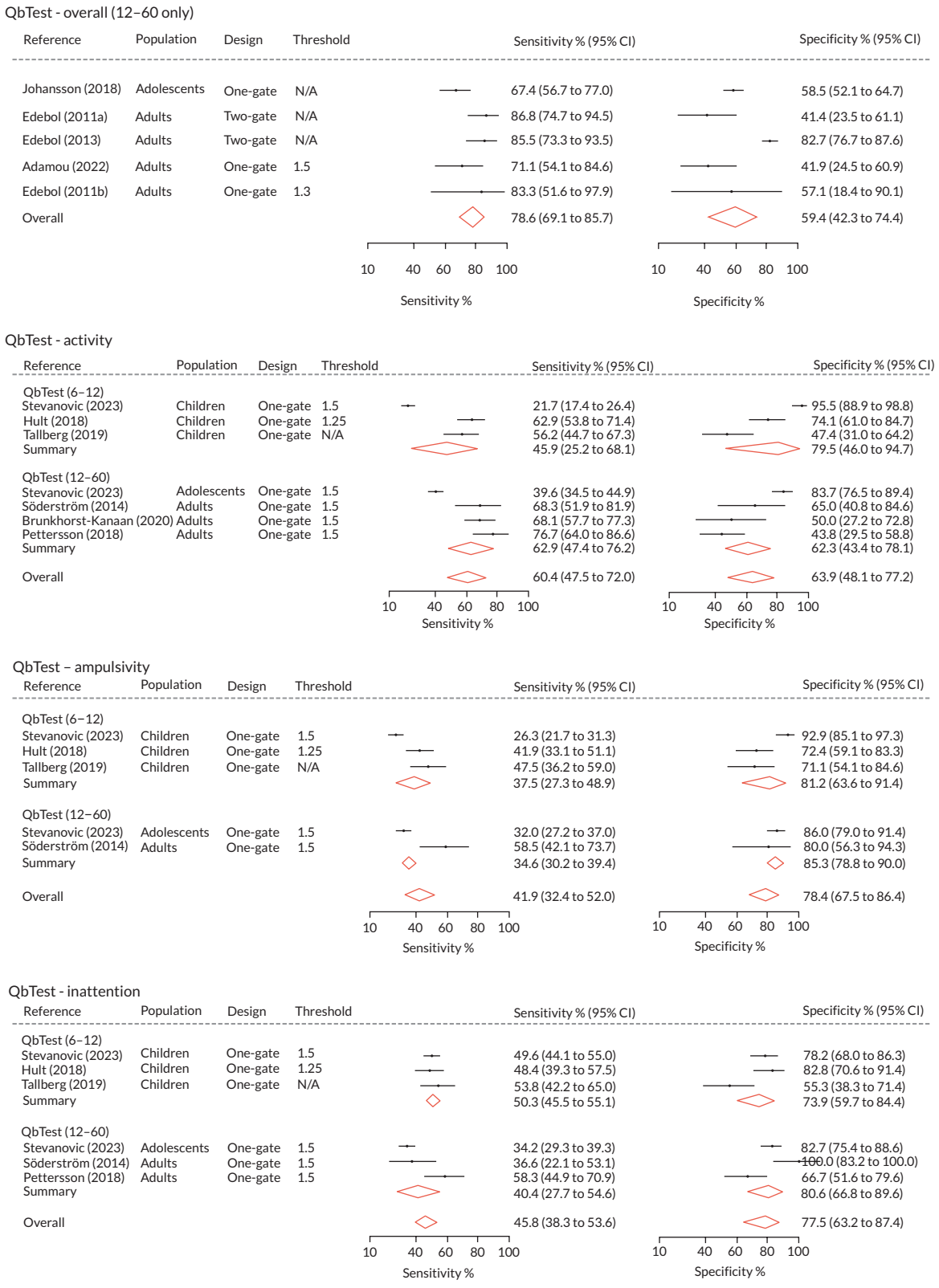

FIGURE 3 Forest plot showing individual study and summary estimates of sensitivity and specificity, with 95% CIs for studies that evaluated the QbTest stratified according to QbTest domain. N/A, not applicable.

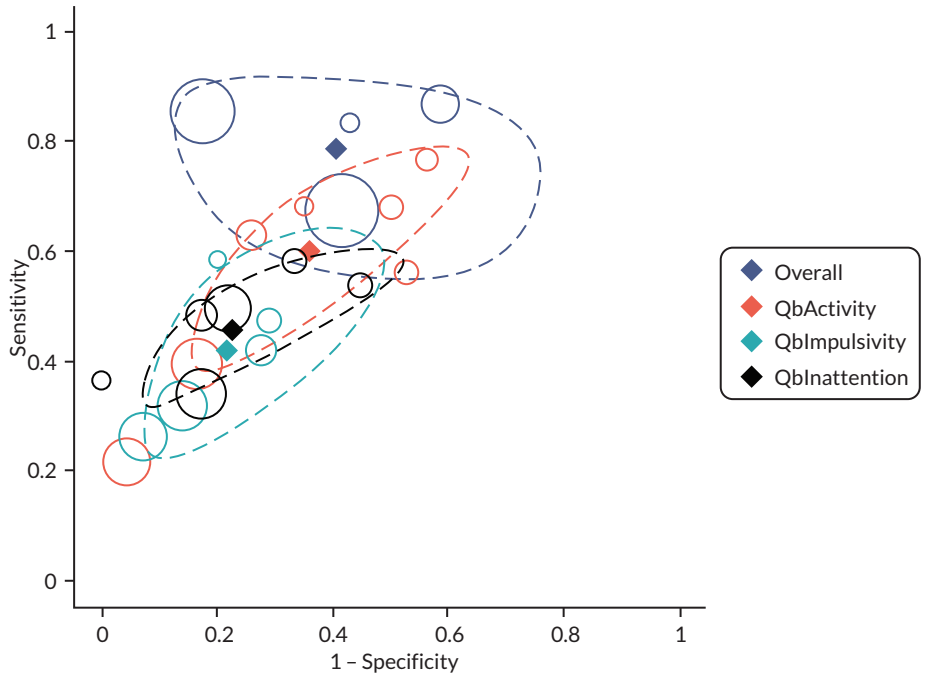

FIGURE 4 Individual study and summary estimates of sensitivity and specificity, with 95% CIs plotted in summary receiver operating characteristics space for studies that evaluated the QbTest, stratified according to QbTest domain.

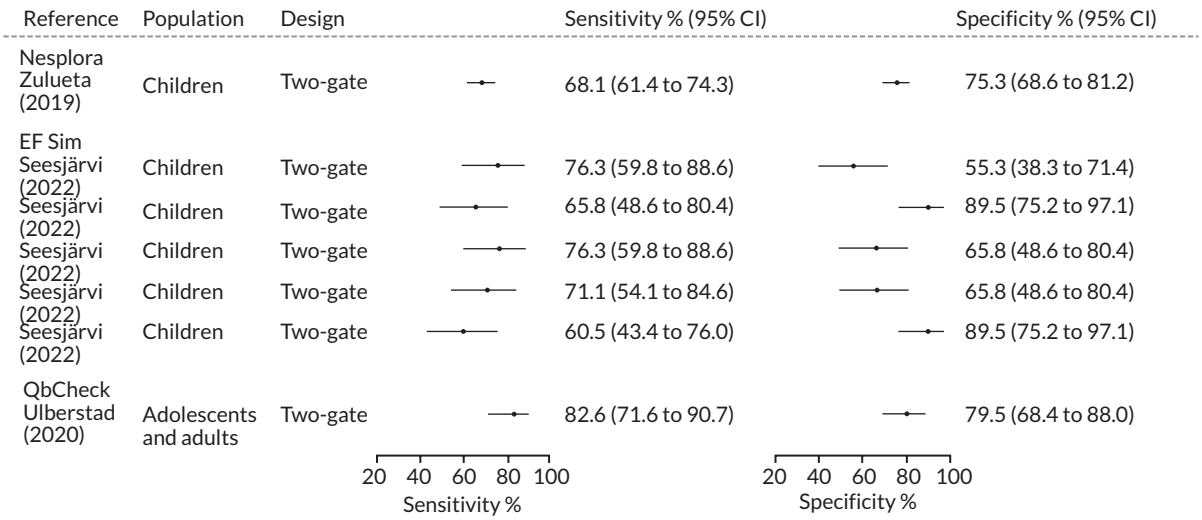

FIGURE 5 Forest plot showing estimates of sensitivity and specificity, with 95% CIs for sensor CPTs that were evaluated in single studies.

separately for the different age groups.[41] Where reported, thresholds ranged from 1.25 to 1.5, with most studies using a threshold of 1.5. Estimates of the accuracy of QbTest evaluated in isolation were generally lower than when evaluated in combination with clinical judgement

QbTest: overall

Six studies reported an overall measure of QbTest based on the three subcategories – all evaluated the version in adolescents or adults. Three studies, all by Edebol,[76,79,84] evaluated a measure that they called 'prediction of ADHD'. This was based on qualitative analyses of raw scores from the different QbTest subcategories. The studies by Johansson and Adamou based the total score on the mean of the three subcategory scores.[70,81] Two studies were judged at high risk of bias as they used a two-gate design,[81,84] and the others were judged at low risk of bias.

Estimates of sensitivity ranged from 67% (95% CI 57% to 77%) to 87% (95% CI 75% to 95%), with a summary estimate of 79% (95% CI 69% to 86%). Estimates of specificity were slightly lower and ranged from 41% (95% CI 24% to 61%) to 83% (95% CI 77% to 88%), with a summary estimate of 60% (41% to 76%). There was some suggestion that sensitivity was higher in two-gate studies, and the highest estimate of specificity was from a two-gate study that enrolled a healthy control group. None of the studies reported AUC data for the overall combined measure, although one provided AUC data for the QbTest subcategories.[70] None reported data on sensitivity and specificity for the QbTest subcategories.

One study (not shown on the plots) conducted in older adults and judged at low risk of bias only provided data for a combination of scores across the QbActivity and QbInattention subcategories.[78] Estimated sensitivity was 56% (95% CI 45% to 66%) and specificity was 83% (75% to 0.89%). Another study (not shown on plots), available only as an abstract, did not provide any information on what QbTest outputs were used for the analysis.[62] This study reported a sensitivity of 96% (95% CI 82% to 100%) and specificity of 81% (95% CI 58% to 95%).

QbTest: subcategories

Six studies evaluated the accuracy of subcategories of the QbTest – QbActivity, QbImpulsivity or QbInattention. One of the studies provided data separately for the QbTest (6–12) and QbTest (12–60) versions of the test. Two studies were judged at high risk of bias as they used a two-gate design, and the others were at low risk of bias. All studies provided data on the AUC – all provided data on the QbActivity and QbInattention scores and five provided data on the QbImpulsivity scores. Summary estimates of AUC were similar across the three domains ranging from 0.58 (95% CI 0.55 to 0.61) to 0.63 (95% CI 0.58 to 0.68). The summary estimate of sensitivity was lowest for QbImpulsivity (42%, 95% CI 32% to 52%), followed by QbInattention (46%, 95% CI 38% to 54%) and was highest for QbActivity (60%, 95% CI 47% to 72%), although CIs overlapped for all estimates. Summary estimates of specificity were similar for QbImpulsivity (78%, 95% CI 67% to 86%) and QbInattention (77%, 95% CI 63% to 87%) and was lower for QbActivity (64%, 95% CI 78% to 77%), although CIs also overlapped for these estimates. There was little evidence of a difference in accuracy of the tests between adults and children for all accuracy measures across all domains. Note that summary estimates that combined data from the different age groups are more different than summary estimates stratified based on age. This is because the combined data are summarised using random-effects models, whereas stratified data are summarised using fixed-effects models due to the small number of studies.

QbCheck

One study, Ulberstad (2020)[77] evaluated the accuracy of the QbCheck test – the remote version of the QbTest. This study used a two-gate design with healthy controls and so was considered at high risk of bias. Estimated sensitivity for the overall results (unclear how this was calculated) was 83% (95% CI 72% to 91%) and specificity was 79% (95% CI, 68% to 88%) (*Figure 5*). Estimates of sensitivity and specificity were not reported for the individual components of the QbCheck test, but AUC data were reported (see *Appendix 4*, *Figure 19*). Estimates ranged from 0.73 (95% CI 0.65 to 0.81) to 0.81 (95% CI 0.74 to 0.88), with CIs overlapping for all estimates.

QBMini

One study, Hamadache (2021)[29] evaluated the QbMini test – the version of the QbTest designed for children aged 4–5 years. This study was judged at high risk of bias as it used a two-gate design with two control groups – healthy controls and those with SLI. We selected the group with language impairment for analysis, as healthy controls are more likely to overestimate specificity. The study only reported AUC data for the three subcategories of the test – QbActivity, QbInattention and QbImpulsivity. The AUC were close to 0.5, suggesting no discriminative ability of the test (see *Appendix 4*, *Figure 20*).

Accuracy of EF Sim Test (previously known as ARVO and EPELI)

Only one study provided data on the accuracy of the EF Sim test – referred to in this paper as the EPELI test. This study was judged at high risk of bias as it used a two-gate design with healthy controls in which controls were matched to cases – this is not appropriate for the evaluation of test accuracy. There was also a high proportion of missing data from the 2 × 2 table. It reported estimates of sensitivity, specificity and AUC for various subcategories of the tests as well as for a single overall measure. AUC estimates ranged from 0.70 (95% CI 0.58 to 0.82) for the overall measure to 0.83 (95% CI 0.74 to 0.92) for the Task Efficacy measure (see *Appendix 4*, *Figure 20*). Estimates of sensitivity ranged

from 61% (95% CI 43 to 76%) for the Actions measure to 76% (60% to 89%) for the Navigation Efficacy and Overall measures. Estimates of specificity ranged from 55% (95% CI 38% to 71%) for the overall measure to 89% (95% CI 75% to 97%) for the Task Efficacy and Actions measures.

Accuracy of Nesplora Attention Kids AULA

Two studies evaluated the accuracy of the Nesplora Attention Kids AULA test; there were no studies of the adult version of this test. Both studies were judged at high risk of bias as they used a two-gate design with healthy controls.[73] One study reported an overall estimate of sensitivity of 68% (95% CI 61% to 74%) and specificity of 75% (95% CI 68% to 81%). The other study, available only as an abstract, reported that the test had an overall accuracy of 93.5% but did not provide any further information or report data separately for sensitivity and specificity.

Comparison of sensor continuous performance tests with non-sensor continuous performance tests or clinical diagnosis alone

Three studies provided a direct comparison between non-sensor CPT and sensor CPTs,[66,75,82] and one study compared QbTest alone to QbTest combined with clinical symptoms,[78] and the AQUA trial compared QbTest combined with clinical diagnosis to clinical diagnosis alone.[18] Results are summarised in *Figure 6* and *Appendix 4*, *Figure 21*. There were insufficient data to allow full cross-classification of results. Formal comparisons between estimated sensitivity and specificity were performed for each measure reported in each study (*Table 8*).

Four studies provided a paired comparison of tests, that is, all participants received both tests; the AQUA trial randomised participants to diagnosis incorporating the QbTest or to clinical diagnosis alone. Both designs are considered appropriate to compare the accuracy of multiple index tests. Four studies were judged at high risk of bias and one at unclear risk of bias (see *Appendix 3*, *Table 45*). The only limitations in the studies identified by the QUADAS-C assessments, in addition to those identified by the standard QUADAS-2 assessment, were that the only study in which information was provided on whether each test was interpreted blind to the other was the AQUA trial, as participants were randomised to testing groups.

Seesjärvi (2022)[75] compared three measures from a non-sensor CPT[103] with the EF Sim test. The overall EF Sim measure was more sensitive than the non-sensor CPT omission errors measure ($p = 0.03$), but it was less specific ($p = 0.07$). There was no difference between the overall EF Sim measure and the other two CPT measures.

Pettersson (2018)[82] and Tallberg (2019)[66] provided a direct comparison between the Connors' CPT II[104] and the QbTest (12–60). The Pettersson study reported that all three of the Qb measures (QbActivity, QbInattention and QbOmmision errors) were more sensitive ($p \leq 0.01$) but less specific than CPT II commission errors and CPT II reaction time variability. There was no difference for QbTest reaction time variance. In contrast, Tallberg reported that the QbTest was less sensitive ($p < 0.01$) than the CPT II with no difference in specificity.

The AQUA trial[18] compared QbTest (6–12) or QbTest (12–60) plus clinical judgement ('QbOpen') to a control group using the standard diagnostic process ('QbBlind'; in this group, the QbTest was also conducted, but the results were not shared with the clinician or used to guide diagnosis). Both groups were evaluated against independent consensus diagnosis using DAWBA, and the limitations with this reference standard are highlighted above. The two groups had very similar specificity: 40% (95% CI 25 to 56) for QbOpen and 36% (95% CI 1 to 58) for QbBlind [odds ratio (OR) 1.16 (95% CI 0.38 to 3.71)], p-value = 0.80). Sensitivity was slightly higher in the QbBlind group (96%, 95% CI 87 to 100) compared to the QbOpen group (86%, 95% CI 72 to 95), but there was no statistical evidence of a difference between groups [OR 0.26 (95% CI 0.02 to 1.53); p-value = 0.14].

The study by Bijlenga (2019)[78] in older adults presented a comparison between models based on the QbTest alone and a model that incorporated a clinical measure of ADHD symptoms (*Table 7*). The model that incorporated the clinical information was much more sensitive (91%, 95% CI 83 to 96) than the QbTest alone (56%, 95% CI 45 to 66; $p < 0.01$). There was no evidence for a difference in specificity ($p = 0.11$).

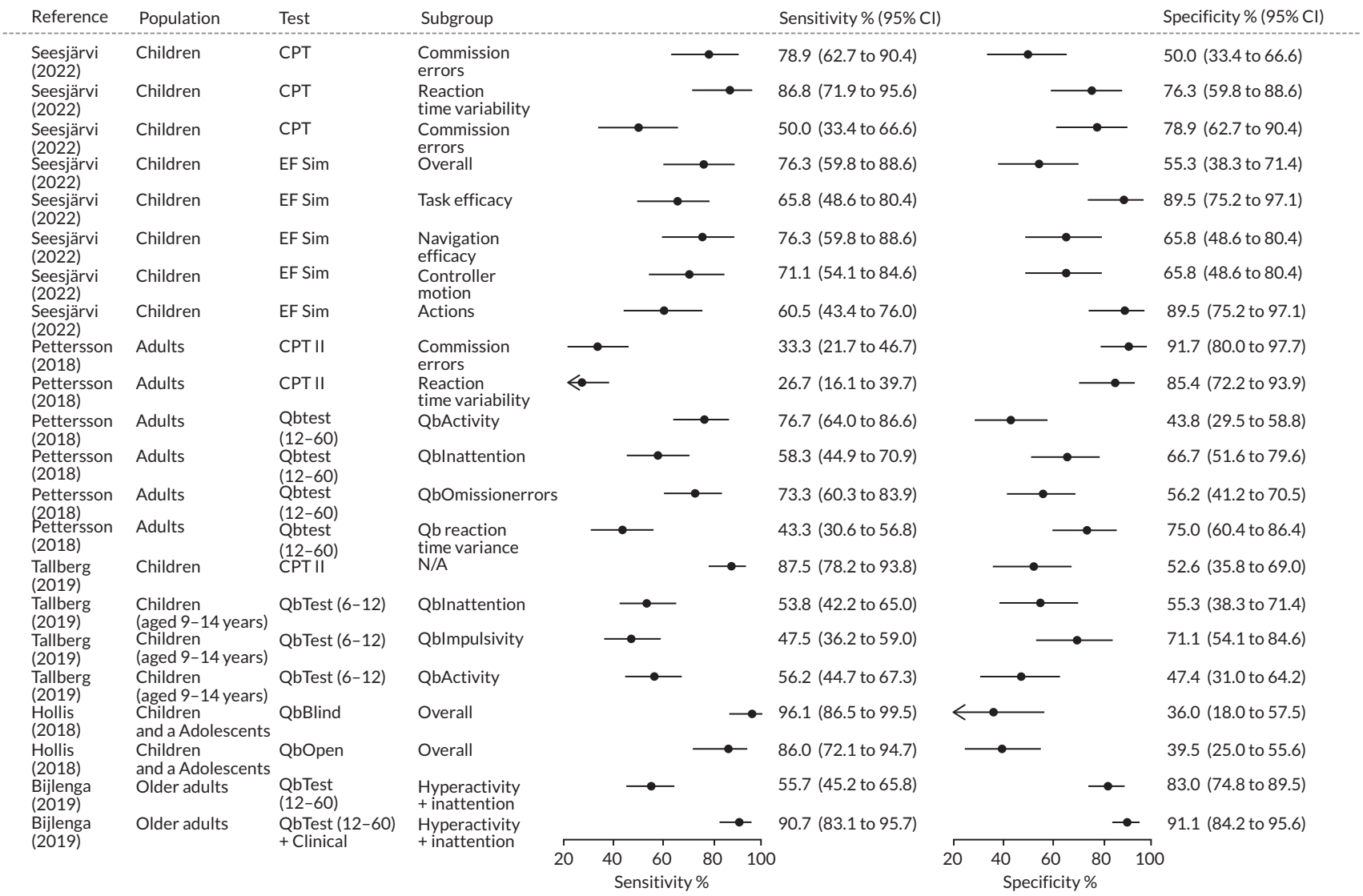

FIGURE 6 Forest plot showing estimates of sensitivity and specificity, with 95% CIs for studies that compared multiple index tests.

TABLE 8 Formal statistical comparisons of sensitivity and specificity within studies that compared multiple index tests

Test 1	Test 2	OR (95% CI) sensitivity	*p*-value sensitivity	OR (95% CI) specificity	*p*-value specificity
Seesjärvi (2022) in children					
CPT – commission errors	EF Sim – overall	1.16 (0.34 to 3.99)	1	0.81 (0.30 to 2.19)	0.82
CPT – reaction time variability	EF Sim – overall	2.03 (0.54 to 8.64)	0.38	2.57 (0.88 to 7.95)	0.09
CPT – omission errors	EF Sim – overall	0.32 (0.10 to 0.91)	0.03	2.99 (1.00 to 9.61)	0.05
Pettersson (2018) in adults					
CPT – commission errors	QbActivity	0.15 (0.06 to 0.36)	< 0.01	13.71 (4.07 to 60.83)	< 0.01
CPT – commission errors	QbInattention	0.36 (0.16 to 0.80)	0.01	5.41 (1.55 to 24.34)	< 0.01
CPT – commission errors	Omission errors	0.18 (0.08 to 0.43)	< 0.01	8.36 (2.46 to 37.13)	< 0.01
CPT – commission errors	Qb reaction time variance	0.66 (0.29 to 1.46)	0.35	3.62 (0.99 to 16.73)	0.05
CPT – reaction time variability	QbActivity	0.11 (0.04 to 0.27)	< 0.01	7.36 (2.59 to 23.50)	< 0.01
CPT – reaction time variability	QbInattention	0.26 (0.11 to 0.60)	< 0.01	7.36 (2.59 to 23.50)	< 0.01
CPT – reaction time variability	Omission errors	0.13 (0.05 to 0.32)	< 0.01	2.90 (0.98 to 9.38)	0.05
CPT – reaction time variability	Qb reaction time variance	0.48 (0.20 to 1.09)	0.08	1.94 (0.62 to 6.48)	0.31
Tallberg (2019) in children					
CPT II	QbInattention	5.95 (2.58 to 14.86)	< 0.01	0.90 (0.33 to 2.44)	1
CPT II	QbImpulsivity	7.63 (3.32 to 19.04)	< 0.01	0.46 (0.16 to 1.29)	0.16
CPT II	QbActivity	5.39 (2.33 to 13.46)	< 0.01	1.23 (0.46 to 3.34)	0.82
Hollis (2018) in children *and* ***adolescents***					
QbOpen	QbBlind	0.26 (0.02 to 1.53)	0.14	1.16 (0.38 to 3.71)	0.8
Bijlenga (2019) in older adults					
QbTest + clinical	QbTest	7.70 (3.37 to 19.43)	< 0.01	2.08 (0.87 to 5.27)	0.11

Impact of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder on process measures

Ten studies provided data on process measures (*Table 9*). This included the AQUA trial[18] and five studies conducted in England that compared results before and after implementation of QbTest (referred to as 'before–after implementation studies').[31,64,69,85,86] Although our inclusion criteria specified that we would only consider before–after studies conducted in the UK, we did not find any studies conducted outside of the UK. Four of the studies that evaluated accuracy (see *Diagnostic accuracy of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder*) also provided additional data on test failure rates.[75,77,78,80] All studies were conducted in children and adolescents (age < 18 years). Baseline data for studies that contributed process measures data are provided in *Appendix 3*, *Tables 47* and *49*.

The AQUA trial compared usual care with QbTest (6–12 and 12–60, depending on age), with test results available to clinician ('QbOpen'), to a control group where diagnosis was based on the usual diagnostic pathway. The QbTest was also performed in the control group, but test results were withheld from the clinician (and so this arm was described as 'QbBlind') and so did not form part of the diagnostic workup of patients. Participants received the QbTest during one of their first three appointments, with 98.4% having received the test by their second appointment. The primary outcome was the number of consultations until a diagnostic decision confirming or excluding the diagnosis of ADHD.

The five before–after implementation studies explored the impact of implementing the QbTest in addition to standard diagnostic assessment by comparing data from clinical records, pre- and post-QbTest implementation in England. One study was restricted to cases with a diagnosis of ADHD, selecting 40 cases diagnosed without QbTest and 40 cases diagnosed with QbTest.[85] The other four studies all selected a group of patients that had been evaluated for suspected ADHD prior to the introduction of the QbTest and a group of patients evaluated for suspected ADHD, who had received the QbTest as part of their diagnostic workup. Sample size ranged from 20 to 549 patients in each group; in one study, the sample size was unclear, and the authors only state that 20–30 children per site across three sites (so 60–90 total) were enrolled.

Risk of bias

The AQUA trial was judged as being at high risk of bias for outcomes involving time to event data (number of consultations to diagnostic decision, minutes spent at clinic appointments, number of clinic appointments and number of days to diagnostic decision) based on the RoB 2 assessment (see *Appendix 3*, *Table 48*).[18] This was due to a large proportion of participants being censored from the analysis as they dropped out or were discharged from the clinic and so did not have a diagnosis at 6 months – 29/123 in the QbTest group and 51/127 in the control group. Reasons and numbers for dropouts and discharge from clinic were not reported. The analysis for these outcomes assumed that participants were uninformatively censored and so had equivalent outcomes to those for whom full FU data were available. It was unclear how cost data were calculated, and how censored participants contributed to these data, and so the trial was judged at unclear risk of bias for this outcome. The trial was judged at low risk of bias for other outcomes (proportion of participants with a diagnostic decision, diagnostic status, diagnostic confidence and stability of diagnosis). HRQoL was pre-specified as an outcome in the study protocol and the data were not reported; therefore, there is potential for selective reporting in the trial.

All five implementation studies that reported on process measures were judged as being at serious risk of bias based on the ROBINS–I tool assessment (see *Appendix 3*, *Table 51*). Four were rated as serious risk of bias due to confounding.[31,69,85,86] This was because important confounders (age at the point of seeking ADHD referral, sex, comorbidities, nature and severity of symptoms at presentation, socioeconomic status and ethnicity) were not controlled for and there was potential for confounding of the effect of intervention. Additionally, one of these studies (Focus ADHD) was confounded by the COVID-19 pandemic, which coincided with the 'post-Qb Implementation' group in the trial. The confounding domain was judged as 'no information' for the other study due to being a conference abstract with very limited detail.[64] This study was, however, rated at serious risk of bias due to the selection of participants, as participants were excluded if their assessment resulted in an inconclusive diagnosis or they did not have a diagnosis in the time frame.[64]

TABLE 9 Overview of studies that evaluated the impact of sensor CPTs for diagnosis of ADHD on process measures

Author, design and location	Group 1	Group 2	Population
Hall (2016)[85] UK; uncontrolled before–after implementation study	QbTest + standard ADHD assessment (*n* = 40)	Standard ADHD assessment (SDQ) and school information form to parents/teachers; Conners' parent and teacher rating scales; child developmental history taken by clinician) (*n* = 40)	**Children and adolescents** (4.5–14.6 years) with ADHD diagnosis confirmed in community paediatric clinic
Hollis (2018)[18] UK; RCT with embedded qualitative evaluation and accuracy data (DTA substudy) AQUA trial	Usual care + QbTest (6–12 and 12–60), with test results available to clinician ('QbOpen') (*n* = 123)	Usual care + (6–12 and 12–60), with test results withheld from clinician ('QbBlind') (*n* = 127)	**Children and adolescents** (6–17 years) referred for first ADHD assessment in CAMHS or community paediatric clinics in England
Vogt (2011)[86] UK; uncontrolled before–after implementation study	QbTest + standard ADHD assessment (*n* = 62)	Standard ADHD assessment (clinical interview by psychiatrists, medical examination, rating scales (e.g. SDQ; Conners) to parents/teachers) (*n* = 46)	**Children and adolescents** (Qb group mean age 9 years; control mean age 10.5 years) referred for ADHD assessment in CAMHS
Sharma (2022)[64] UK; uncontrolled before–after implementation study	QbTest + standard ADHD assessment (*n* = 20)	Standard ADHD assessment (no detail provided) (*n* = 20)	**Children** (mean age 11.7 years, SD 2.4) referred for ADHD/non-specific behavioural problems/ASD who completed ADHD assessment in hospital paediatric clinic
Humphreys (2018)[69] UK; uncontrolled before–after implementation study (East Midlands AHSN) + survey	QbTest + standard assessment (unclear)	Standard assessment (no detail provided) (*n* = unclear)	**Children and adolescents** (5–16 years) referred for ADHD assessment in eight community paediatric mental health settings in three NHS trusts
McKenzie (2022)[31] UK; uncontrolled before–after implementation study ('Focus ADHD') plus survey and qualitative study	QbTest + standard assessment (*n* = 549)	Standard assessment (no detail provided) (*n* = 549)	**Children** referred for ADHD assessment in 20 CAMHS and paediatric sites
Bijlenga (2019)[78] The Netherlands; Germany; Sweden; two-gate DTA study	QbTest (12–60) (*n* = 234)	N/A – only process measure data the study reported is test failure rate for the sensor CPT	**Adults** ADHD group: adults (aged 55+ years); DSM-IV-TR ADHD diagnosis Healthy controls: adults (aged 55+ years) with score below cut-off on symptom severity measures, matched on age and gender
Groom (2016)[80] UK; two-gate DTA study	QbTest (12–60) (*n* = 84)		**Adults** ADHD group: DSM-V diagnosis of ADHD. Autism (ASD) group: ICD-10 diagnosis of Asperger syndrome

continued

TABLE 9 Overview of studies that evaluated the impact of sensor CPTs for diagnosis of ADHD on process measures (*continued*)

Author, design and location	Group 1	Group 2	Population
Seesjärvi (2022)[75] Finland; two-gate DTA study	EPELI (*n* = 115)		**Children (aged 9–12 years)** ADHD group: ADHD diagnosis by licensed physician using ICD-10 Non-ADHD group: no mental or behavioural disorder; matched to cases
Ulberstad (2020)[77] Germany; Sweden; USA; two-gate DTA study	QbCheck (*n* = 149)		**Adolescents and adults** (12–59 years) Cases: DSM-V diagnostic criteria Controls: healthy controls; those with high levels of inattention/hyperactivity/impulsivity according to DSM-V excluded

CAMHS, Child and Adolescent Mental Health Services; ICD-10, *International Statistical Classification of Diseases and Related Health Problems*, Tenth Revision.

Of the other four studies, one was rated at low risk due to random selection of cases,[85] and three were rated as no information.[85,105] Three studies were rated at low risk for bias in deviations due to intended interventions,[31,85,86] one study was rated as no information due to being a conference abstract with limited detail[64] and the other study was rated as moderate risk of bias due to there having been a full pathway redesign of the service in two/three sites after the introduction of the QbTest.[69] One study was rated as moderate risk of bias for missing data (people with a final diagnosis were selected, so we do not know the number of individuals referred who never received a diagnosis),[85] one as no information[69] and three as low risk of bias.[31,64,106] All studies were rated at low risk for bias in the classification of interventions, as intervention groups were clearly defined. All studies were rated as moderate risk of bias for measurement of the outcomes (measurement of the outcome may have been influenced by knowledge of the intervention received) and for selection of the reported result (no protocol).

The four DTA studies that also reported process measures were judged as being high risk of bias based on the QUADAS-2 assessment.[75,77,78,80]

Results

Table 10 provides a summary of results from studies that evaluated the impact of introducing the QbTest as part of the diagnostic process for ADHD on process outcomes. Very few studies provided a formal statistical comparison of results between intervention groups. All process measures data are reported in *Appendix 3*, *Tables 50* and *52*.

Time to diagnostic decision

Five studies reported the data on time to diagnostic decision. The AQUA trial reported that the number of appointments required to reach a consultation was less in the QbTest group compared to control [hazard ratio (HR) 1.44, 95% CI 1.04 to 2.01; $p = 0.029$)]. When results were stratified by QbTest version, only those using the QbTest (6–12) version were found to have fewer appointments (HR 1.84, 95% CI 1.23 to 2.68; $p = 0.001$), and this was not seen in the QbTest (12–60) group (HR 0.82, 95% CI 0.37 to 1.80; $p = 0.618$). The AQUA trial also reported that the mean number of appointments to a diagnosis was slightly less in the QbTest arm compared to control (2.69 vs. 2.72). (Confidential information has been removed.) The time spent at clinic appointments until diagnosis was less in the QbTest group compared to the control group [median 150 minutes vs. 165 minutes; time ratio (TR) 0.85; 95% CI 0.77 to 0.93; $p = 0.001$]. There was also a suggestion that the number of days to diagnosis was less in the QbTest group, but the evidence for this was weak [median 96 vs. 108; TR 0.90 (95% CI 0.73 to 1.10; $p = 0.285$)]. However, the HR and TR estimates should be interpreted with some caution due to the large proportion of participants who were censored (i.e. dropped out of the study or were discharged from clinic). Estimates are based on an analysis of the full data set, where those without a diagnosis are censored after their last appointment under the assumption that they would have similar HRs or TRs as those who had a diagnosis.

Four of the before–after studies also reported on the number of consultation to reach a diagnosis – in all studies, this was reported to be less following implementation of the QbTest, although only one study reported strong evidence for a difference between groups ($p = 0.02$); another study reported no difference between groups ($p > 0.05$) and the other two studies did not make a formal comparison between groups. Two of the before–after studies also reported that the time to diagnosis was reduced following implementation of the QbTest, but they did not provide a statistical comparison of results. The Focus ADHD reported that time from referral to diagnosis ($p < 0.01$) and time to reach a diagnostic decision (p-value not reported) were increased in the period following implementation of the QbTest, but these data are likely to have been confounded by the COVID-19 pandemic.

Impact on clinical decision-making

The AQUA trial reported an improved diagnostic decision-making [diagnostic decision was made for 76.4% (95% CI 68.9% to 83.9%) in the QbTest group compared to 59.8% (95% CI 51.3% to 68.4%) in the control group at 6 months], OR 2.43 (95% CI 1.34 to 4.39) and greater confidence in the diagnostic decision ($p = 0.022$). Clinician confidence in the diagnostic decision was greater in the QbTest group compared to control (OR 1.77, 95% CI 1.09 to 2.89). There was no difference in the stability of the diagnosis over time (change from when the diagnosis was first confirmed) ($p = 0.32$). They also reported that ADHD could be ruled out in more cases within the QbTest group [relative risk reduction (RRR)

TABLE 10 Overview of results from studies that evaluated the impact of sensor CPTs for diagnosis of ADHD on process measures

Outcome category	Outcome details	Hollis AQUA trial[18]	Hall (2016)[85]	Vogt (2011)[86]	Sharma (2022)[64]	Humphreys (2018)[69]	McKenzie (2022)[31] Focus ADHD
Time to diagnostic decision	No. consultation to ADHD diagnosis	Diagnosis rate (appointment number units): HR 1.44, 95% CI 1.04 to 2.01 (p = 0.03); 1.84 (1.23 to 2.68) 6–12 years; 0.82 (0.37 to 1.82) 12–17 years Mean number of appointments to diagnosis: QbTest: 2.69 (SD = 0.85) Control: 2.72 (SD = 0.91)	IRR 0.71 (95% CI 0.54 to 0.94); p = 0.02		QbTest: mean 2.4 (SD 0.8) Control: mean 2.7 (SD 0.7); p > 0.05	QbTest: 0.24–1.04 less per child Control: range 3–8 appointments	QbTest: mean 2.85 (range 1–32) Control: mean 3.22 (range 1–50)
	Time from referral to diagnosis				QbTest: 5.5 (SD 1.8) months Control: mean 6.5 (SD 3) months	*QbTest:* average ranged from 15 to 252 days *Control:* average ranged from 161 to 453 days	QbTest: mean 507 (range 43–1281) days Control: mean 452 (range 15–3276) days; p < 0.01*
	Total consultation time	Median time to diagnosis: QbTest: 150 (95% CI 140 to 155) Control: 165 (95% CI 150 to 180) minutes TR 0.85 (95% CI 0.77 to 0.93)					
	Days to reach diagnostic decision	QbTest: median 96 (95% CI 85 to 99) Control: median 108 (95% CI 91 to 140) TR 0.90 (95% CI 0.73 to 1.10)					QbTest: mean 129 (range 0–1378) Control: mean 117 (range 0–1570)
Impact on clinical decision-making	Proportion of patients with a diagnosis	OR 2.43 (95% CI 1.34 to 4.39)					

TABLE 10 Overview of results from studies that evaluated the impact of sensor CPTs for diagnosis of ADHD on process measures (*continued*)

Outcome category	Outcome details	Hollis AQUA trial[18]	Hall (2016)[85]	Vogt (2011)[86]	Sharma (2022)[64]	Humphreys (2018)[69]	McKenzie (2022)[31] Focus ADHD
	Stability of diagnosis	No difference, $p = 0.032$					
	Number with ADHD diagnosis						QbTest: 418/549 (76%) Control: 445/549 (81%)
	Confidence in diagnostic decision	OR 1.77 (95% CI 1.09 to 2.89)					
	Number of individuals in whom ADHD diagnosis was excluded	RRR 2.14 (95% CI, 1 to 4.59)					
	Number of children in whom school observations were utilised						QbTest: 49/549 (9%) Control: 120/549 (22%)
Outcomes at 1-year FU	Outcomes for those with ADHD			No difference between groups ($p = 0.24$)			
	Diagnosis of ADHD in those with diagnosis rejected at initial assessment			QbTest: 0/19 Control: 7/19 (37%) $p < 0.0035$			
Cost	Cost of clinic appointments (*unclear how much it costed*)	QbTest: £87.62 Control: £90.06					
	Cost per patient to diagnosis		QbTest: £265.90 Control: £329.40				

IRR, incidence rate ratio.

2.14, 95% CI 1.00 to 4.59)]. As highlighted above, these data should be interpreted with some caution due to the exclusion of those who dropped out or who were discharged from clinic. The Focus ADHD study reported that fewer children were diagnosed with ADHD after the QbTest was implemented (76%) compared to the control period (81%). They also reported that fewer in-school observations were used to help make the ADHD diagnosis in the post-QbTest group (9%) compared to the control group (22%); however, these data are likely to have been influenced by the COVID-19 pandemic.

Outcomes at 1-year follow-up

The Vogt (2011) study reported outcomes of patients at 1-year FU and found no difference between groups in the proportion of children in each of the following categories (p = 0.24): ADHD diagnosis changed, medication trial, continuing on medication, discontinued medication and lost to FU. It reported that a higher proportion of children who had initially been diagnosed as not having ADHD received a revised diagnosis of ADHD at 1-year follow up in the control group (37%) compared to none in the QbTest group.

Cost

The AQUA trial reported that the cost of clinic appointments was slightly less in the QbTest group (£87.62) compared to control (£90.06). The study by Hall also reported that costs were lower following QbTest with an average cost per patient for a diagnosis of £265.90, following the introduction of the QbTest, and £329.40 prior to introduction of the QbTest. Neither study provided a formal statistical comparison between groups.

Test failure rate

Four DTA studies, all two-gate designs, provided data on the test failure rate.[75,77,78,80] Two studies reported test failure rate for the QbTest (12–60). One reported that 25/234 (11%; 9 ADHD, 16 controls) participants had an unavailable test result. Reasons for missing results included: not understanding the task, being an extreme outlier, not following instructions, technical errors and aborted tests.[78] The other study reported that 4/84 (5%) had an unavailable test result, described as non-completion of the test (no further information provided).

The study that evaluated QbCheck reported that 7/149 (5%; 6 ADHD, 1 control) of participants had an unavailable test result. Reasons included failure to complete the test due to technical problems with the camera (2), participant ending test in the middle of the session for unknown reasons (4) and intentionally discontinuing the test (1).[77] The study that evaluated the EF Sim (EPELI version) test reported that 22/115 (19%; 5 ADHD, 17 controls) had an unavailable test result due to technical failures or human error (no further information provided).[75]

Clinician and patient views of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder

Eight studies evaluated the clinician, patient or carer views of sensor CPTs for the diagnosis of ADHD, which were collected through surveys, qualitative interviews or focus groups.[31,69,72,75,77,87,88,107] Five evaluated the QbTest,[31,69,72,87,107] one assessed the QbCheck[77] and two assessed the EF Sim test.[75,88] An overview of these studies is provided in *Table 11* and further details are outlined in *Appendix 3*, *Table 53*.

Of the five studies that evaluated the QbTest, two combined qualitative interviews and a survey. One was conducted as part of the FACT feasibility RCT (in the very specific population of young boys in a YOI)[72] and the other as part of the AQUA trial.[107] Two studies were implementation studies included for *Impact of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder on process measures*.[31,69] One reported survey data from patients, families and clinical staff who had used QbTest on their experience of using the test,[69] and one (Focus ADHD) reported both qualitative interview data from staff and survey data from patients, families and staff on their experiences of using the QbTest.[31] All four of these studies were conducted in England. The remaining study was a mixed-methods study that reported focus group data and survey data concerning clinicians, young service users and their families' experiences of using QbTest in addition to standard ADHD assessment in CAMHS.[87] This study, which was conducted in Ireland, only provided data on patient and/or clinicians views and so was only included for this section of the review.

TABLE 11 Overview of studies that evaluated clinician and/or patient views of sensor CPTs for diagnosis of ADHD

Author, location, design and test	Study components
Studies with interview and survey data	
Chitsabesan (2022)[72,108] England; interview and survey components of FACT feasibility RCT; QbTest + standard assessment	1. Semistructured interviews with 6 adolescent boys from the QbTest group of the FACT trial 2. Semistructured interviews with 1 research assistant and 5 staff members who used QbTest in the FACT trial 3. Survey completed by 10 adolescent boys from the QbTest group of the FACT trial
Hollis (2018)[18] England; qualitative substudy of AQUA trial; usual care + QbTest (6–12 and 12–60), with test results available to clinician ('QbOpen')	1. Semistructured interviews with the 10 clinical leads of sites involved in the AQUA trial 2. Semistructured interviews with 20 families from the AQUA trial 'QbOpen' Group 3. Survey completed by the 10 clinical leads and 76 families involved in AQUA trial
McKenzie (2022)[31] England; qualitative interview and survey components of an uncontrolled before–after implementation study (Focus ADHD); QbTest (6–12) or QbTest (12–60) + standard ADHD assessment	1. Interviews with 21 healthcare staff involved in implementation of QbTest at their site, or conducting the test/interpreting test results, in the Focus ADHD study 2. Survey completed by 65 healthcare staff involved in the Focus ADHD study 3. Survey completed by 22 patients who had been assessed with the QbTest in the Focus ADHD study
Pellegrini (2020)[87] Ireland; mixed-methods study of real-world impact of test implementation; QbTest + standard ADHD assessment	1. Focus groups with 19 clinicians who were using the QbTest in 1 of the 3 CAMHS teams selected for this study in Ireland 2. Survey to 17 clinicians, 15 young people and their parents/guardians (n = 18) who had used QbTest in 1 of the 3 CAMHS teams involved in this study
Studies with survey data only	
Humphreys (2018)[69] England; survey component of an uncontrolled before–after implementation study; QbTest (6–12) or QbTest (12–60) + standard ADHD assessment	1. Survey completed by 48 patients (children who had ADHD assessment using QbTest in CAMHS in the before–after study) and their families 2. Survey to staff who had used QbTest in the study (n = unknown)
Peili Vision (n.r.)[88] Finland; pilot cohort study; EF Sim test + psychologist evaluation	1. Survey completed by 21 teachers of participating schools that used EF Sim for students in the Health Service Pilot (Confidential information has been removed)
Seesjärvi (2022)[75] Finland; survey data from two-gate DTA study; EF Sim (EPELI version)	1. Survey completed by children (some with ADHD; some healthy controls – n = not reported) who took part in the DTA study using EF Sim (EPELI version) test and completed the survey component
Ulberstad (2020)[77] Germany, Sweden, USA; Survey data from two-gate DTA study; QbCheck	1. Survey completed by patients who used QbCheck in the DTA study and who completed the survey (n = 125; 59 ADHD and 69 healthy controls)

The study that evaluated the QbCheck test (in Germany, Sweden and the USA)[77] and one of the studies that evaluated the EF Sim test (in Finland)[75] were DTA studies included in *Diagnostic accuracy of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder* that reported survey data from patients on the ease of use/acceptability of the tests. The other study of the EF Sim test, included in the manufacturer's submission from Peili Vision, only reported survey data on views of the test and therefore was only included for this section of the review. This study was a pilot project in which 50 students in Finland completed the EF Sim test, and survey data were gathered from teachers (confidential information has been removed) about their experience of using the test.[88]

Risk of bias

Qualitative study components

Two of the four studies that provided qualitative data on patient and carer views had no concerns regarding study quality based on the CASP checklist assessment (see *Appendix 3*, *Table 54*). This was the qualitative component of the study by Pellegrini (2020),[87] which involved focus groups with 19 clinicians who had used the QbTest in CAMHS in Ireland, and the qualitative substudy of the AQUA trial, which involved interviews with clinicians and families who had used the QbTest in the trial.[107]

The other two studies appeared to use appropriate methodology, but they reported limited detail which made it difficult to judge certain items in the CASP checklist. In Chitsabesan (2022; reports interview data as a secondary outcome of the FACT feasibility RCT) and McKenzie (2022; reports on the interview component of the Focus ADHD study), there were limited details on the relationship between researcher and participant and data analysis and so it was not possible to fully assess the quality of the approach taken.[31,72]

Survey study components

Two of the eight studies that provided survey data on patient and carer views had very few concerns regarding study quality based on the Q-SSP assessment (see *Appendix 3*, *Table 55*): the AQUA trial substudy,[107] and Pellegrini (2020).[87] The other six studies were judged to have some concerns due to a lack of information about participants, methodology and analysis.[31,69,72,75,77,88]

Results

Below, we summarise our synthesis of findings from these studies. The full synthesis is presented in *Appendix 5*.

QbTest

We identified two broad themes from the findings concerning the QbTest: views around the helpfulness of the test and barriers to the implementation of the test. Conceptual categories that pertained to views around the helpfulness of the QbTest included contribution to ADHD diagnosis and communication with caregivers.

Findings from qualitative data suggested that healthcare staff felt that the QbTest increased their confidence in decision-making,[31,87,107] helped to differentiate ADHD subtypes (particularly subtle presentation, common in girls)[31,107] and supported diagnosis in the presence of comorbidities.[72,107] Healthcare staff also felt that the test could decrease the time to diagnostic decision.[31,72,87,107] For example, some sites in the Focus ADHD study commented that the QbTest implementation had resulted in fewer appointments by replacing the school observation and that the faster assessment pathway supported the young person in getting educational support quickly.[31] Families also appeared to feel that the QbTest could have a positive impact on the diagnostic process. They recognised the role that the QbTest could have in shortening the emotionally overwhelming diagnostic procedure and they emphasised the need for a quick diagnostic decision.[107] However, they also felt that the process should not be rushed and that their child should not be 'labelled' quickly.[107]

Clinicians valued the perceived objectivity of the test, which they felt added important information to clinical assessments and, in some cases, increased confidence in decision-making and reduced the burden on clinician time.[31,72,87,107] However, clinicians also reported a need to establish where the QbTest falls on the ADHD assessment pathway[31,107] and expressed uncertainty about whether the clinical setting of the test is representative of what happens in other settings (e.g. school).[107]

Findings suggested that the QbTest helped to improve communication between clinicians and patients and their families,[31,87,107] between clinicians and schools,[107] between clinical colleagues[87] and between patients and families.[107] In the AQUA trial, clinicians reported that being able to show a comparison of the child's performance to a normative sample helped them to communicate the diagnostic decision to families, and they thought that this helped families to accept the decision.[107] However, some clinicians in the Focus ADHD study commented that families could still struggle to accept a diagnostic decision.[31] Some families interviewed in the AQUA trial were unclear about how the QbTest report was being used to inform decision-making.[107] This was also reflected in survey responses, which suggested that some families did not think the QbTest helped them to understand how diagnoses were made.[31,107] Furthermore, some

families and young people felt that the results of the QbTest were not properly explained to them[31] and did not help them to understand symptoms.[72,107]

Barriers to the implementation of the QbTest

Conceptual categories that pertained to views around barriers to the implementation of the QbTest included practical barriers and acceptability to patients and caregivers. Interviews and focus groups with healthcare staff highlighted that staffing (i.e. the need for someone trained to administer the task), room requirements and technology were barriers to QbTest implementation.[31,72,87,107]

Regarding patient views on the acceptability of the test, some patients found the test to be boring, long and repetitive.[31,72] In the Focus ADHD study,[31] interviews with healthcare staff highlighted that some individuals (particularly, young people and people with ASD) experienced sensory discomfort and struggled with wearing the tight headband. Staff commented that other young people struggled to follow the instructions, and felt anxious during the test, due to the test itself and/or being without their caregivers. Additionally, concerns were raised about the lack of representation of different ethnicities in the test explanation video, the requirement to choose biological sex before conducting the test and the use of the word 'test', which staff felt induced stress in participants.[31] The study of QbCheck reported that participants found it easy to use; however, this was from a brief three-question survey conducted as part of a DTA study.[77]

EF Sim Test

Two studies evaluated the EF Sim Test.[75,88] One study, run by the test manufacturer, surveyed 21 teachers of participating schools that had implemented the EF Sim test for students in a pilot study. On average, the majority of the teachers found the test results usable and reported that they can support communication with guardians and that they are helpful to identify executive functioning challenges in students that may otherwise go unnoticed.[88] (Confidential information has been removed.) The other study was a DTA study of the EF Sim test (previous version named EPELI) in children (some with ADHD, some healthy controls, *n* = not reported). Answers to a short survey suggested that, on average, children appeared to feel enthusiastic about the tasks, found them interesting and they put effort into their performance on the test.[75]

Objective 2: diagnostic accuracy and clinical effectiveness of sensor continuous performance tests for the diagnosis of attention deficit hyperactivity disorder in people referred with suspected attention deficit hyperactivity disorder for whom current assessment cannot reach a diagnosis

We did not identify any studies that met inclusion criteria for this objective.

Objective 3: clinical effectiveness of sensor continuous performance tests in evaluating medication effectiveness during initial dose titration and treatment decisions for people with a diagnosis of attention deficit hyperactivity disorder

Six studies were included for objective 3 – three RCTs (FACT, QUOTA and AQUA),[72,107,110] one DTA study[66] and two implementation studies.[31,69] All studies evaluated the QbTest. One RCT (the QUOTA trial), that included a qualitative substudy, was only included for objective 3.[110] The other five studies also contributed to objective 1; one reported data on the accuracy of QbTest for medication dose titration,[68] and the other four reported qualitative and survey data on the use of QbTest for medication management.[31,66,69,72,107]

Diagnostic accuracy of sensor continuous performance tests during initial dose titration and treatment decisions for people with a diagnosis of attention deficit hyperactivity disorder

The DTA study by Tallberg (2019)[66] evaluated the accuracy of the QbTest for medication dose titration in children with ADHD (*Table 12*). The study enrolled a single group of patients with ADHD. They were assessed with the QbTest and

a behaviour rating scale for ADHD [Swanson Nolan and Pelham Questionnaire (SNAP-IV)] before starting treatment with MPH.[111] Dose titration started at a low dose of 18 or 20 mg and the dose was titrated in steps of 10 or 18 mg, depending on the drug brand, to a maximal dose of 60 mg (less in case of side effects). At each dose titration, children were tested with both the SNAP-IV behavioural test and the QbTest. To determine the accuracy of the QbTest for medical titration, QbTest results at 1-year FU were cross-tabulated with 'good' or 'poor' outcome. A 'good' outcome was defined as being on the optimal dose 1 year after titration as defined by EITHER a SNAP-IV score increase of at least 0.2 (equivalent to 0.4 SDs) OR a QbTest score decrease of at least 0.4 SD. This is problematic as the QbTest formed part of the reference standard, which is likely to overestimate the accuracy of the test. The study was therefore judged at high risk of bias (see *Appendix 3*, *Table 61*). Accuracy was estimated separately for the QbInattention and QbActivity subcategories. Sensitivity was estimated at 82% (95% CI 69% to 91%) and specificity at 60% (95% CI 26% to 88%) for the QbInattention domain, and sensitivity was 76% (95% CI 62% to 87%) and specificity was 40% (95% CI 12% to 74%) for the QbActivity domains. All baseline data extracted from Tallberg (2019) is provided in *Table 59* (see *Appendix 3*) and all extracted results are given in *Table 60* (see *Appendix 3*).

Impact of sensor continuous performance tests during initial dose titration and treatment decisions for people with a diagnosis of attention deficit hyperactivity disorder on patient and process outcomes

The QUOTA trial[110] was a feasibility trial conducted in England which explored the feasibility of conducting a RCT to evaluate the efficacy of the QbTest as part of medication management for children with ADHD (*Table 12*; *Appendix 3*, *Table 56*). It compared the QbTest protocol in which participants completed the QbTest at baseline and two FU points on medication (2–4 weeks and 8–10 weeks) and control where participants received treatment as usual, which included at least two FU consultations. Outcomes evaluated included: use of interventions, impact on clinical decision-making, ease of use/acceptability and confidence in HCP assessment. However, as this was a feasibility study, it was designed and powered to assess the feasibility of conducting a full trial and not to formally evaluate the impact on outcomes. For this reason, a formal risk-of-bias assessment was also not undertaken for this study. The number of participants was very small with 44 children randomised – 21 to the intervention arm and 23 to the control.

Results suggested that those in the QbTest arm were more likely to have had their medication changed (type of dose of ADHD medication) at the first FU point (10/18 in intervention vs. 7/21 in control), but figures were more similar at FU 2 (7/17 in intervention vs. 9/19 in control) (see *Appendix 3*, *Table 57*). These findings should be interpreted with caution; due to the feasibility design and small sample size, it was not possible to draw conclusions regarding clinical effectiveness from this study.

TABLE 12 Details of studies that provide information on sensor CPTs in evaluating medication effectiveness during initial dose titration and treatment decisions for people with a diagnosis of ADHD

Study	Williams (2021)[110] (QUOTA trial)	Tallberg (2019)[66]
Design	Feasibility RCT	DTA study (one-gate)
Sample size	44 (44 analysed); 21 in intervention arm and 23 in control group	186 (56 analysed)
Population	Children aged 6–15 years, diagnosed with ADHD and referred to CAMHS/community paediatric clinic in the UK to commence ADHD medication	Children and adolescents aged 7–18 years, with ADHD, from a Child and Adolescent Psychiatry clinic in Sweden
Group or test	Intervention: QbTest (6–12 or 12–60) + usual care Control: usual care	Index test: QbTest (6–12 or 12–60) + SNAP-IV behaviour rating scale Reference standard: SNAP-IV or QbTest score
Funding	Non-industry	Non-industry

Clinician and patient views of sensor continuous performance tests during initial dose titration and treatment decisions for people with a diagnosis of attention deficit hyperactivity disorder

Five studies provided data on clinician and patient's views of the QbTest for dose titration and treatment decision-making (see *Appendix 3*, *Table 58* and *Appendix 5*). Three RCTs (FACT, QUOTA and AQUA trials) reported interview and survey data concerning patient and clinician views of the QbTest for medication management and dose titration;[72,107,110] one implementation study reported patient and carer views of the test from survey data,[69] and one implementation study reported qualitative interview and survey data (the Focus ADHD study).[31]

Risk of bias

Qualitative study components

The AQUA trial had no concerns regarding study quality based on the CASP checklist assessment (see *Appendix 3*, *Table 54*).[107] The QUOTA trial had very few concerns (see *Appendix 3*, *Table 62*).[110] The FACT trial[72] and the Focus ADHD study[31] reported limited details on the relationship between researcher and participant and data analysis and so it was not possible to fully assess the quality of the approach taken (see *Appendix 3*, *Table 54*).

Survey study components

The AQUA trial had very few concerns regarding study quality based on the Q-SSP assessment (see *Appendix 3*, *Table 55*).[107] The other four studies that contributed survey data were judged to have some concerns due to a lack of information about the participants, methodology and analysis (see *Appendix 3*, *Tables 55* and *63*).[31,69,72,107,110]

Results

Across the five studies that reported on the clinician and patient views of the QbTest for dose titration and treatment decision-making,[31,69,72,107,110] healthcare staff and families mostly appeared to value the role of the test for dose titration, checking medication utility and improving medication adherence.

Clinicians interviewed in the AQUA trial qualitative substudy reported greater support from parents on initiating and continuing medication, and greater adherence to medication, as a result of being able to directly observe the effect of medication with a QbTest.[107] Additionally, families interviewed in the AQUA trial reported that seeing the QbTest results made them more confident that the medication would help their child.[107] This objectivity was also highlighted as a positive point in interviews with clinicians in the QUOTA trial, who valued the objectivity of the QbTest in comparison to informant measures traditionally used to monitor medication.[110] Interviews with healthcare staff in the Focus ADHD study also identified that the QbTest could be helpful in managing dose titration and checking medication utility, and the staff felt that the QbTest helped young people/caregivers to understand medication decisions and the effects of the medication. This study only involved interviews with staff, not patients/carers.[31]

Survey data from two studies suggested that patients/caregivers were not convinced that the results of the QbTest helped them to understand medication decisions.[72,107] Less than half (20/52) of families surveyed in the AQUA trial felt that it helped them to understand the decisions made about medication, although it is notable that most participants did not commence medication, so the results are difficult to interpret.[107] Likewise, in the FACT RCT, there was no consensus among 10 adolescent boys assessed for ADHD as to whether the QbTest results helped them to understand how the decisions about medication had been made (the majority voted 'neither agree/disagree').[72] By contrast, interviews with parents (six in the intervention and two in the control group) in the QUOTA trial provided mainly positive feedback. The QbTest was found to increase parents' confidence in their child's treatment and ongoing medication decisions. Parents described how a visual representation of the child's symptoms helped them to better understand the treatment impact, though the test was noted to be boring by some and it required taking time out of school to have multiple appointments to monitor medication.

Some HCPs in the Focus ADHD also felt that the QbTest helped them to decide how effective the medication is and had increased their confidence in decision-making about treatment.[31] By contrast, in a survey to clinicians in CAMHS (*n* = not reported), only half of the respondents agreed that the QbTest results aided treatment decisions (30% of respondents remained neutral).[69] In the QUOTA trial, the survey of clinicians showed that, across both FUs, 73% (24/33 responses) of clinicians reported that the QbTest was useful in determining treatment, 18% (6) were neutral and 9% (3) stated it was not helpful. More clinicians found the QbTest to be helpful at FU 1 (76.5%; 13/17) than FU 2 (68.8%;

11/16). In interviews, clinicians also highlighted the potential role of the suggestion that the QbTest appears to help parents to be more accepting of treatment recommendations, and they reported that it increased their confidence in treatment and helped to communication around treatment impact. However, they did also note that having more appointments for medication management can present logistical issues in scheduling appointments, and they reported a preference to only add additional QbTest appointments when it was perceived to add value.

Objective 4: clinical effectiveness of sensor continuous performance tests for evaluating treatment effectiveness during long-term treatment monitoring for people with a diagnosis of attention deficit hyperactivity disorder

We did not identify any studies that met inclusion criteria for this objective.

Chapter 5 Assessment of cost-effectiveness

Sections of this chapter have been reproduced from the review protocol, available at the NICE website.[1]

Review of cost-effectiveness models of diagnostic testing and treatment of attention deficit hyperactivity disorder

Review methods

We conducted a systematic review to identify previous cost-effectiveness studies of diagnostic tests for the assessment of ADHD and previous cost-effectiveness models of treatment for ADHD.

We searched the following databases:

- MEDLINE (MEDALL) via Ovid
- EMBASE via Ovid
- PsycInfo via Ovid
- CINAHL via EBSCO*host*.

The full search strategies are reported in Appendix 1.2.

We also included any relevant papers on the cost-effectiveness of sensor CPTs for the assessment of ADHD, which were identified in the clinical effectiveness review, searched citations in relevant publications and asked experts in the field. We also ran additional targeted searches to identify specific inputs required in the economic model.

We assessed the quality of cost-effectiveness studies of diagnostic tests for the assessment of ADHD using the Drummond checklist.[112]

Results of the cost-effectiveness review

Appendix 6, *Figure 22* shows the PRISMA flowchart showing the studies identified from the systematic review of cost-effectiveness models for diagnosis or treatment of ADHD, and *Appendix 6*, *Figure 23* shows the PRISMA flowchart for economic evaluations of sensor CPTs for the assessment of ADHD.[38]

Cost-effectiveness models of diagnosing attention deficit hyperactivity disorder

We did not find any studies reporting cost-effectiveness models of diagnostic tests for the assessment of ADHD. To keep the review manageable, we did not search for models of diagnostic tests in other neurodevelopmental conditions or conditions with similar symptoms to ADHD, which could have been informative, and acknowledge this as a limitation.

Economic evaluations of sensor continuous performance tests for diagnosing attention deficit hyperactivity disorder

We found one RCT that assessed the cost-effectiveness of diagnostic tests for the assessment of ADHD[18] and one implementation study (two reports[69,113]). The quality assessment of the two economic evaluations using the Drummond checklist is given in *Appendix 7*, *Table 64*.

AQUA trial

Hollis *et al.*[18] presents the results of the AQUA trial of ADHD diagnosis in children and adolescents, including a cost-effectiveness analysis. They use an NHS perspective, and the cost analysis focuses on the staff time (number and length of appointments) required to reach a diagnosis confirming or excluding ADHD. The analysis compares QbTest plus usual care (QbOpen) to usual care, with the usual care arm including QbTest, but the results were not provided to the diagnosing clinicians (QbBlind). In the costing analysis, they incorrectly exclude the cost of QbTest from both arms rather

than including it for the QbOpen arm only. While QbTest was used in both arms of their trial, the QbBlind arm reflects the situation where QbTest is not used and so cost of the test should be applied for QbOpen and not for QbBlind.

The EQ-5D-Y was used to calculate the quality-adjusted life-year (QALY) weights for participants in each intervention group, relying on multiple imputation as only 43% of study participants completed the questionnaire. The EQ-5D-Y questionnaire is stated in the analysis plan to be measured at baseline, 4–8 weeks after medication titration and 6 month FU; however, it is not stated how the repeated measures were combined within the multiple imputation analysis, or what value set was used to convert EQ-5D-Y to QALY weights, and the results are not given; only the incremental QALYs for the two arms are reported.

Cost-effectiveness is also reported in terms of incremental cost per incremental time to diagnosis, in which the time to diagnosis was reduced in the QbOpen arm.

For the purposes of this analysis, data presented on resource use, time to diagnosis and the proportions with a diagnosis (ADHD or no ADHD) will be used to model the impact of QbTest.

East Midlands Academic Health Science Networks study, Kent Surrey Sussex Academic Health Science Networks report

Humphreys *et al.* report a study by the East Midlands AHSN which collected data from three East Midland trusts (Derbyshire, Leicestershire and Lincolnshire).[69] The Kent Surrey Sussex AHSN conducted a cost–benefit analysis using the data collected by the East Midlands AHSN.[113]

The assessment uses a return on investment calculation that accounts for the costs of implementing QbTest and benefits to the NHS in terms of reduced number of appointments for clinical assessment and school nurses and social benefit in terms of improved quality of life while on the waiting list.

Cost calculations are not shown explicitly and resource use units are not clearly described. All input costs are increased based on a bias scale and then total costs are increased by 15% and benefits are decreased by 15% to give a more conservative result.

Three scenarios are presented; one based on data collected by the study authors at three East Midlands trusts, the second based on data provided by the QbTest manufacturer and the third using scenario 1 data and additional assumptions to estimate the return on investment of a national scale up of QbTest.

The cost–benefit analysis is presented in a report which does not appear to be peer reviewed. There is some lack of clarity in presentation of the methods; some of these are more clearly described in the *Results* and *Discussion* sections.

No decision model is used, rather, the net benefit is calculated within each scenario. In all cases, there is a positive return on investment result, which is primarily driven by the cost of implementing QbTest being lower than the NHS cost savings due to two fewer appointments being needed for each patient when QbTest is implemented.

Review results for cost-effectiveness models of treatment for attention deficit hyperactivity disorder

We found 24 studies describing cost-effectiveness models for treatment of ADHD, which are summarised in *Table 13*. All studies described either Markov models or decision tree models, and one study[122] also described a trajectory analysis model as an alternative to their Markov model. One report, the NICE guideline NG87,[16] included two separate studies; these were of parent training (see appendix 1 of NG87) and combination treatment (see appendix 2 of NG87).

Thirteen studies described Markov models;[116–120,122–124,126,128–130,133,134] five of these studies used or closely based their models on a previously published model: the Cottrell (2008) Markov model[116] was adapted by Hong (2009)[120] and Prasad (2009);[128] the Faber (2008) Markov model[118] was adapted by Schawo (2015)[129] and van der Schans (2015);[134] and the Sikirica (2012) Markov model[130] was adapted by Lachaine (2016).[123]

TABLE 13 Overview of cost-effectiveness models for treatment of ADHD

Study	Model type	Target population	Treatments studied, including any treatment sequencing	Study perspective	States/events in model	Time horizon/ discounting	Key assumptions
Nice guideline NG87 (2018)[114], appendix 1	Decision tree	Children in UK with ADHD	Parent training vs. no parent training	NHS and PSS	Response or no response to parent training	1 year	A proportion of children will also be on drug treatment
NICE guideline NG87 (2018)[115], appendix 2	Decision tree	Children in UK with ADHD	Combination treatment vs. medication alone or behavioural therapy alone	NHS and PSS	Response or no response to treatment, and stopping treatment due to AEs	1 year	Patients may experience tolerable AEs, which do not lead them to discontinue treatment, but do have associated disutilities
Cottrell (2008)[116]	Markov model. Monthly cycles over period of 1 year	Children with ADHD in UK. Split into subgroups based on stimulant history	ATX, compared against MPH, dexamphetamine and no treatment. Patients either start on ATX or a comparator and then follow same treatment sequence if not successful	NHS	18 health states, based on different combinations of treatment/response/ side effects	1 year	Model assumes that all non-drug healthcare costs and indirect costs are equivalent between the treatment groups
Erder (2012)[117]	Markov model. Weekly cycles over period of 1 year. Split into 4-week drug titration period, and 48-week maintenance period	Children and adolescents with ADHD in USA	Comparing GXR vs. ATX	US third-party payer. Only considered direct costs (drug costs and direct medical costs)	Titration phase: response, non-response (on treatment), discontinuation. Maintenance phase: response, discontinuation, non-response (off treatment)	1 year	Patients who discontinued treatment had the same utility and medical costs as non-responders. AEs reduced the patients' health utilities during the titration period
Faber (2008)[118]	Markov model, with a primary 2-month titration phase, followed by a Markov phase of length of 10 years, with 1-day cycles	Youths with ADHD in the Netherlands, who have suboptimal response to IR MPH	Long-acting MPH OROS vs. IR MPH	Societal/community	Non-response, optimal response, suboptimal response, treatment stopped, functional remission, non-compliance	10-year horizon, discounting at 4% per year	Costs of non-pharmacological interventions were incurred in the first and sixth year of treatment, when the child is aged 8 and 13 years, respectively
Freriks (2019)[119]	Markov model	Children in the Netherlands with ADHD	Medication, behavioural or combination treatment	Includes healthcare costs and criminal justice system costs	No delinquency, minor-to-moderate delinquency, serious delinquency	10-year horizon, discounting at 4% per year	Serious delinquency is an absorbing health state

continued

TABLE 13 Overview of cost-effectiveness models for treatment of ADHD (*continued*)

Study	Model type	Target population	Treatments studied, including any treatment sequencing	Study perspective	States/events in model	Time horizon/ discounting	Key assumptions
Hong (2009)[120]	Cottrell (2008) Markov model adapted to Spain. Monthly cycles and 1-year time horizon	Children and adolescents with ADHD in Spain	Patients start on ATX or MPH, then move to other if drug unsuccessful and finally stop medication if neither drug is successful	NHS in Spain	10 health states, based on different combinations of treatment/response/ side effects	1 year	Model assumes that all non-drug healthcare costs and indirect costs are equivalent between the treatment groups
King (2006)[121]	Decision tree	Children and adolescents with ADHD in UK	Treatment sequences of MPH, ATX, DEX in different orders, followed by fourth line of no treatment	NHS and PSS	Tolerate or intolerable side effects. Response or no response	1 year, with a secondary analysis extrapolating beyond 1 year	Drug titration period lasts 1 month, after which non-responders move to next drug in treatment sequence
Klein (2011)[122]	Two approaches: Markov model and trajectory analysis. Considered 1-month and 1-year cycle lengths	Youth with ADHD in USA	Models different patient groups transitions between treatment modalities (out of treatment, medication only, services only, combination)	N/A (costs not reported)	Out of treatment, medication only, services only, combination	1 year	Time on treatment assumes that medication is taken daily to completion of prescription
Lachaine (2016)[123]	Markov model [similar to Sikirica (2012)] with two stages: weeks 0–8 where all patients remain on treatment, and weeks 9–52 where patients in moderate/ severe state may discontinue as they are considered to be non-responsive. Length 1 year, with weekly cycles	Children aged 6–12 years with ADHD in Canada, with a suboptimal response to GXR	GXR adjunctive to long-acting stimulants	Two perspectives: Canadian Ministry of Health, and societal	Mild, moderate, severe or normal, assigned using clinician-reported CGI-S scores	1 year	Annual medical costs for patients in normal health state are assumed to be the same as median medical costs for non-ADHD patients. AEs are assumed to result in a utility decrement lasting 4 weeks
Maia (2016)[124]	Unclear. Appears to be a Markov model but is also described as a decision tree	Children and adolescents in Brazil with ADHD	MPH vs. natural course	Brazilian Unified Health System	Treatment (not) maintained, (no) spontaneous improvement, (no) improvement maintained	6-year horizon, discounting at 5% per year	Patients who discontinue treatment do not later restart treatment

TABLE 13 Overview of cost-effectiveness models for treatment of ADHD (*continued*)

Study	Model type	Target population	Treatments studied, including any treatment sequencing	Study perspective	States/events in model	Time horizon/ discounting	Key assumptions
Marchetti (2001)[125]	Decision tree, with up to four treatment evaluation periods, each lasting 4 weeks	Children in USA with ADHD	Treatment adjustment and sequencing. MPH (immediate or extended release), Adderall	Payer perspective	Success and failure of treatments. Followed by management by psychologist/psychiatrist if there are four failures	1 year	Once a child responds to medication, they continue on that dose for the remainder of the evaluation period
Nagy (2017)[126]	Four-layer conceptual model, including Markov	Childhood through to adulthood in patients with ADHD	Treatment sequencing of drugs	Includes societal perspective	Drug toleration, response, compliance and persistence	Not stated	Provides an example of three layers of conceptual model, making some strong assumptions on the links between short-term and long-term outcomes
Narayan (2004)[127]	Decision tree	Children in USA with ADHD	Treatment sequencing of MPH or amphetamine/dextroamphetamine, followed by the other treatment, then no treatment	Societal perspective (though some indirect costs not included)	Response, non-response or discontinuation of treatment. Tolerance of side effects	1 year	Side effects are assumed to result in a utility decrement lasting 1 month
Prasad (2009)[128]	Uses Cottrell (2008) Markov model	Children and adolescents in UK with ADHD	ATX, compared against MPH, dexamphetamine and no treatment. Patients either start on ATX or a comparator and then follow same treatment sequence if not successful	NHS	18 health states based on different combinations of treatment/response/side effects	1 year	Model assumes that all costs other than study drug costs are equivalent between treatment groups
Schawo (2015)[129]	Markov model [similar to Faber (2008)], with 1-day cycle length and 12-year horizon	Children and adolescents in the Netherlands with ADHD	MPH OROS vs. IR	Societal perspective	Suboptimal medication intake, optimal medication intake, remission, treatment stopped	12-year horizon. Costs discounted at 4%, effects discounted at 1.5%	Costs of non-pharmacological interventions were incurred at ages 6 and 12 years, around when children change schools

continued

TABLE 13 Overview of cost-effectiveness models for treatment of ADHD (*continued*)

Study	Model type	Target population	Treatments studied, including any treatment sequencing	Study perspective	States/events in model	Time horizon/ discounting	Key assumptions
Sikirica (2012)[130]	Markov model, length 1 year, and cycle length 1 week. The model has two stages: weeks 0–8 and weeks 9–52. Patients are considered to be non-responsive at week 8 and perma-nently discontinue treatment	Children and adolescents in USA with ADHD	GXR vs. stimulant monotherapy	US third-party payer	Mild, moderate, severe, or normal, assigned using CGI-S scores	1-year time horizon	Patients who do not respond to the initial therapy by week 8 discontinue treatment and do not switch to a new treatment
Sohn (2016)[131]	Decision tree, with several arms for AEs	Children and adolescents in USA with ADHD	Atypical antipsychotics vs. other alternatives to stimulants	US third-party payer	Drug effectiveness, and several side effects, including weight gain and high blood pressure	1-year time horizon	Side effects seen within 6 weeks of initial treatment; will persist for the entire year as treatment is continued
Tajik (2023)[132]	Decision tree	Children and adolescents in Iran with ADHD	LDX vs. MPH	Social perspective	Toleration or non-toleration of treatment. Response or no response	1-year time horizon	Patients who discontinue treatment due to intolerance are assumed to have the same utilities and costs as non-responders for the remainder of the 1-year model time horizon
Tockhorn-Heidenreich (2015)[133]	Markov model with 1-month cycle and 1-year horizon	Adults in Spain with ADHD	ATX vs. no treatment	Spanish National Healthcare System	Treatment initiation, response or no response	1-year time horizon	During the first 3 months, patients may only discontinue due to AEs as ATX has a prolonged onset of treatment response
van der Schans (2015)[134]	Markov model, similar to Faber (2008), with a 2-month titration phase followed by a 10-year Markov phase with 1-day cycle	Children and adolescents in the Netherlands with ADHD with a suboptimal response to IR MPH	IR vs. slow-release MPH	Societal perspective	Optimal response, suboptimal response, natural remission, discontinuing treatment	10-year horizon, future costs discounted at 4% per year and future outcomes discounted at 1.5% per year	Patients may restart treatment [unlike the Faber (2008) model]

TABLE 13 Overview of cost-effectiveness models for treatment of ADHD (*continued*)

Study	Model type	Target population	Treatments studied, including any treatment sequencing	Study perspective	States/events in model	Time horizon/ discounting	Key assumptions
Vanoverbeke (2003)[135]	Decision tree	Children and adolescents in UK with ADHD	Behavioural treatment, immediate- or slow-release MPH, followed by an alternative or combination treatment if first treatment fails	NHS and PSS	Success and failure of treatments	1-year time horizon	Assumes medication compliance is the same for slow vs. IR MPH
Zimovetz (2016)[136]	Decision tree	Children and adolescents in UK with ADHD, who have responded inadequately to MPH	LDX dimesylate vs. ATX	NHS	Toleration or non-toleration of treatment over a 28-day titration phase, followed by response or no response to treatment over a 48-week post-titration phase	1-year time horizon	Patients who discontinue treatment due to intolerance are assumed to have the same utilities and costs as non-responders for the remainder of the 1-year model time horizon
Zimovetz (2018)[137]	Decision tree	UK adults with ADHD	LDX dimesylate as a first- or second-line treatment vs. slow-release MPH and ATX	NHS	Toleration or non-toleration of treatment over a 28-day titration phase, followed by response or no response to treatment over a 48-week post-titration phase	1-year time horizon. Also used a 5-year time horizon in a sensitivity analysis, discounting at 3.5% per year	Patients who discontinue treatment due to intolerance have the same utilities and costs as non-responders for the remainder of the 1-year model time horizon. Patients who responded to and tolerated treatment are persistent over the 1-year model time horizon
Miller *et al.* (1998)[138]	Decision tree	Children in Canada with ADHD	MPH vs. dextroamphetamine vs. pemoline vs. non-drug therapy vs. combined therapy vs. no treatment	Third-party payer	Toxicity or no toxicity, compliance or non-compliance	1-year time horizon	Children on no treatment visit their family physician the same number of times per year as children on drug treatment

CGI-S, Clinical Global Impression Severity Scale; DEX, dexamfetaminesulphate; GXR, guanfacine extended-release; IR, immediate release; LDX, lisdexamfetamine; OROS, osmotic release oral system.

Eleven studies described decision tree models.[114,115,121,125,127,131,132,135–138]

The Zimovetz (2018)[137] study used the same model as Zimovetz (2016),[136] but applied it to adults rather than children and adolescents.

Treatments modelled were either drug treatments, behavioural therapy, a combination of the two or no treatment. Only three models[124,133,138] compared directly against no treatment (which is required for our diagnostic strategies models), but most models included treatment discontinuation with consequences of not being on treatment. Switching between treatments in sequence was conducted in nine models, of which five were Markov models[122,116,120,126,128] and four were decision tree models.[121,125,127]

Only two models were on adults,[133,137] and all other models were for children and/or adolescents.

Most studies took either a health system or payer perspective. Seven studies considered a societal perspective as their sole perspective or as an additional perspective.[118,123,126,129,134,127,132]

Most studies used a time horizon of 1 year, stating a lack of long-term data as the reason for this choice. Five studies used a longer time horizon, with the most common choice being a 10-year time horizon.[118,119,124,129,134]

All studies were cost-effectiveness or cost–utility analyses, except for Klein (2011)[122] which modelled treatment trajectories, Nagy (2017)[126] which described a conceptual model, and Vanoverbeke (2003)[135] which was a cost analysis only.

The majority of studies (22 studies) had target populations in Europe, the USA or Canada, with 8 study models in the UK.[114,115,116,121,128,135–137] The other two studies were modelled using children and adolescents in Brazil[124] and Iran.[132]

The most common states or events used to structure the economic models were response or no response to treatment and discontinuation of treatment due to non-tolerance of AEs. Some studies used more detailed stratification to differentiate between patients' symptom levels. The Sikirica (2012) Markov model,[130] also used by Lachaine (2016),[123] consisted of four health states to stratify patients according to the severity of their ADHD symptoms. These four health states are normal, mild, moderate and severe and are based on a clinician-completed ADHD rating scale. The Faber (2008) Markov model,[118] also used by Schawo (2015)[129] and van der Schans (2015),[134] differentiated between optimal and suboptimal responses to treatment.

In most models which included drug titration periods, the each titration period was either around 4 weeks long,[116,117,120,121,125,127,128,132,136,137] or 8 weeks long.[118,123,130,134] Models which included utility decrements from AEs leading to treatment discontinuation assumed that the decrements last for 4 weeks.[117,123,127]

Implications of cost-effectiveness review for this economic evaluation

The two cost-effectiveness evaluations of diagnostic assessment for ADHD provide information, which we used to parameterise our model. The AQUA trial[18] is of direct relevance to objective 1, as it compares QbTest plus clinical assessment to clinical assessment alone, with information on the resource use required to reach a diagnosis in each arm. The East Midlands study and Kent economic evaluation provide some additional information on resource use needed to reach a diagnosis.[69,113]

Neither of these evaluations contain an economic model, and no previous economic models of diagnosis of ADHD were identified, so we needed to develop de novo models for this assessment. However, there have been several previous economic models of treatment of ADHD, which are relevant for modelling the costs and outcomes of ADHD treatment following diagnosis and for the evaluation of sensor CPTs in the assessment of dose titration and long-term monitoring.

Most of the models of treatments for ADHD included treatment response, adverse effects of treatment and treatment discontinuation, all of which are relevant for our models. Some modelled different types of response (optimal or suboptimal),[118,129,134] which is particularly relevant for models of dose titration and long-term treatment monitoring.

Many models capture patients moving through several lines of treatment, and some included remission, both of which are relevant for a model of long-term monitoring. Only three models[124,133,138] compared an active treatment strategy against no treatment, and none of these were UK-based. However, outcomes on 'no treatment' were assumed in many of the models for patients who discontinue treatment, which can be used for patients not on treatment in our model. A limitation of many of the previous models of treatment for ADHD is that they restrict to a 1-year time horizon. This may be appropriate for comparisons of different active treatments, as patients are monitored every 6 months or annually. For a model of diagnostic strategies, however, the time horizon needs to be long enough to capture the time period before a diagnosis is eventually reached in all patients with ADHD, which is likely to be longer than 1 year.

We considered studies which were conducted in the UK to be the most appropriate source of information for health-state costs and utility inputs to the model. Cottrell (2008)[116] and Prasad (2009)[128] only included drug costs, assuming all other costs were the same between their comparators, and therefore were not useful for our model. Studies which reported costs in terms of responders versus non-responders[114,115,121,136,137] were of most relevance to our model.

Model structure and methods of economic evaluation

We aimed to develop decision-analytic models to estimate the incremental costs and QALYs for sensor CPTs in addition to current methods of assessment compared with current methods of assessment alone, for each of the following purposes:

i. assisting diagnosis of ADHD in people referred with suspected ADHD (objective 1)
ii. assisting diagnosis of ADHD in people referred with suspected ADHD for whom current assessment cannot reach a diagnosis (objective 2)
iii. to assist in dose titration and treatment decisions in people with a diagnosis of ADHD (objective 3)
iv. to assess treatment effectiveness for long-term treatment monitoring for people with a diagnosis of ADHD (objective 4).

However, the majority of the evidence on sensor CPTs identified in the clinical review (Section 0) was relevant for objective 1 only. We did not identify any evidence for objective 2, but we present a scenario analysis for objective 1 to give some speculative results relevant to objective 2, albeit with strong assumptions. There was insufficient evidence available to assess the cost-effectiveness of the use of sensor CPTs for dose titration and long-term treatment monitoring (objectives 3 and 4), and so we describe potential model structures only and do not populate the models or report results for these objectives.

Population

For objectives 1 and 2, the population are patients suspected of having ADHD who have been referred for assessment. For our scenario analysis to explore objective 2, we assume that the technology is only used in those where a diagnosis was not reached after two appointments using standard assessment, and the results of the diagnostic test would be available at the third appointment.

For objectives 3 and 4, the population are patients diagnosed with ADHD who initiate pharmacological treatment. We did not identify sufficient evidence on sensor CPTs for this population to be able to conduct an economic evaluation.

Subgroups

The key source of evidence on the effectiveness of sensor CPTs was the AQUA trial,[18] which evaluated QbTest (6–12) (in children aged 7–12 years) and QbTest (12–60) (for adolescents aged 12–17 years). We did not identify any studies in adults that reported information on the time and number of appointments until diagnosis and no studies with diagnostic accuracy data for sensor CPTs in combination with clinical assessment. Our main analyses are therefore only directly applicable for children and adolescents. We conducted scenario analyses using the HR for diagnosis in children and adolescents separately, but note that this is the only the only outcome reported separately for children and adolescents, and all other model inputs are assumed to be the same.

There was insufficient evidence to conduct subgroup analyses for: sex, ethnicity, people with mental health, behavioural and neurodevelopmental conditions, people with developmental trauma, looked-after children or people in the Youth Justice System or Adult Criminal Justice System. There was a feasibility study conducted in the very specific population of boys with symptoms of possible ADHD, aged 15–18 years, in YOIs in England.[72] However, as noted in *Impact of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder on patient outcomes* due to the feasibility design, small sample size, low numbers of appointments (only 14 decisions were made, and all were exclusions of ADHD) and impact of COVID-19, there was insufficient evidence to conduct a subgroup analysis for male young offenders aged 15–18 years.

Attention deficit hyperactivity disorder assessment strategies

We included sensor CPTs identified in the clinical effectiveness review (Section 0) and for which there was sufficient evidence available for the model. This meant that the economic evaluation focussed on QbTest (6–12) and QbTest (12–60) (for adolescents aged 12–17 years), as we did not have sufficient evidence for other sensor CPTs. We refer to these tests collectively as 'QbTest'. We conducted scenario analyses changing the test cost to match that of other tests where we had information on test costs, but note that these assume all other inputs are as for QbTest and have to be interpreted as such.

Assessment strategies for attention deficit hyperactivity disorder diagnosis

Current methods for diagnosing ADHD are assessment by a HCP (without use of the sensor CPTs) using history-taking, third-party observational reports and questionnaires.[16] Children are usually assessed face to face in the clinic, while assessment for adults is often done remotely.

We evaluated the following diagnostic assessment strategies (restricted to QbTest as the only test with sufficient data).

Standard

All patients receive standard clinical assessment using current methods for diagnosis of ADHD.

QbTestAll

All patients are offered QbTest, the results of which are available to the HCP making the assessment at the second appointment along with all other evidence used for standard assessment.

QbTestUnclear

All patients receive standard assessment, and those patients who do not receive a diagnosis after two appointments are offered QbTest, the results of which are made available to the HCP making the assessment at the third appointment.

QbTestUnclear is only evaluated as a scenario analysis to explore objective 2.

Assessment strategies for dose titration

Following a diagnosis of ADHD, symptoms are managed using a combination of non-pharmacological and pharmacological interventions (see *Management and treatment of attention deficit hyperactivity disorder*). For those patients where pharmacological treatment is indicated, medications licensed in the UK include stimulants [MPH, lisdexamfetamine (LDX) and DEX] and non-stimulants (ATX or guanfacine). Patients undergo a 'dose titration' period during which they begin with a low dose of first-line treatment and they are then assessed at 2-week intervals for efficacy and side effects, where decisions to change the dose or treatment are made. NICE guidelines recommend patients to start with MPH for 6 weeks; then, if no response, they recommend switching to LDX for 6 weeks; then, if no response, the patients switch to ATX;[16] although in practice, treatment choice is based on individual circumstances, response, tolerability and adherence.[28] The period of time before the treatment and dose are settled upon varies greatly across patients, but we heard that the majority reach a stable dose by 12 weeks (six appointments).

Williams (2021)[110] conducted a feasibility study to compare the use of QbTest in addition to clinical assessment with clinical assessment alone for dose titration. Patients completed a QbTest prior to initiating medication and two further QbTests while on medication (2–4 weeks and 8–10 weeks after initiating medication). The study found that to fit with

clinical practice, there needed to be flexibility on the timing of the pre-medication QbTest and the HCPs making the assessments should be allowed to determine the number and timing of subsequent QbTests post medication.

For a model to evaluate sensor CPTs for dose titration, we would therefore assume the sensor CPT is performed pre medication (which could be during the diagnostic assessment) and either once or twice more while on medication during the dose titration period. The cost of the pre-medication sensor CPT would only be incurred in the case where this is not part of the routine diagnosis. The sensor CPTs conducted during the titration period would need to be conducted in a dedicated in-person appointment because dose titration assessments are largely conducted remotely, which needs to be reflected in the costs.

Dose titration assessment strategies relevant to be evaluated for objective 3 are explained in the following subsections.

Standard

All patients receive standard assessment using current methods for dose titration with fortnightly appointments until a stable dose/treatment is reached.

Sensor continuous performance test

The sensor CPT is completed pre medication and either once or twice post medication, the results of which are available to the HCP making the assessment at fortnightly appointments.

Assessment strategies for long-term monitoring

Following the dose titration period, patients are monitored regularly (annually for adults and at least every 6 months for children), including an assessment of whether medication needs to be adjusted. Patients may also take a 'drug holiday' to see if they still need to take medication (our clinical advisors consider this every 3–5 years for adults and maybe during school holidays for children).

We did not find any studies of the use of sensor CPT for long-term monitoring of ADHD patients, and it is not clear what format such monitoring would take. For this reason, we were unable to describe the assessment strategies to compare the cost-effectiveness of the use of sensor CPTs to assist treatment decisions in the long-term management of patients (objective 4).

Setting

The AQUA trial, which provided the main source of data for our model, recruited participants who were referred for assessment for ADHD in CAMHS (48%) or community paediatric clinics (52%) in England.[18] The model is therefore applicable for patients referred through these routes based on a similar patient mix as seen in the AQUA trial. The East Midlands AHSN study gathered data from three trusts in Derbyshire, Leicestershire and Lincolnshire.[69] In addition, they used data provided by QbTest manufacturers from undisclosed clinical settings. Details are not given on where assessments take place within the three trusts. QbTest was used for ADHD diagnosis in children, but an age range was not specified.

Model structures

The model structures were developed to capture the short- and long-term costs and benefits of sensor CPTs for the assessment of ADHD, informed by the findings of our review of clinical and cost-effectiveness studies and discussions with our clinical advisors and patient representatives.

Model structure for diagnostic assessment (objectives 1 and 2)

A Markov model structure was used to capture the process of diagnosis of ADHD (*Figure 7*). Patients enter the model after a referral for assessment for ADHD and join a waiting list for assessment. The time spent waiting for assessment is assumed to depend on whether a sensor CPT is used or not, because a potential benefit of the use of sensor CPTs is to reduce the time and resources required to reach a diagnosis and hence reduce the clinician, time which can be used to reduce waiting times for assessment. Patients then undergo diagnostic assessment for ADHD, which consists of a series of appointments until a diagnosis is reached or assessment is discontinued.

The AQUA trial presents the proportion of patients for whom a diagnosis is reached against the number of appointments (figure 2 in Hollis *et al.*[18]) and a corresponding survival analysis that accounts for censoring for the high proportions who were lost to clinic (appendix S6 in Hollis *et al.*[18]). The survival analysis indicates that most diagnoses had been reached by six appointments, but note that this makes the strong assumption of non-informative censoring. This is unlikely to be the case, as those lost to clinic are unlikely to achieve a diagnosis at the same rate as those attending clinic, and we know they are not diagnosed within 6 months. We therefore distinguish between those who attend clinic and diagnosis can be reached within 6 months (for whom the survival analysis results are applicable to) and those who do not receive a diagnosis within 6 months (a proportion of whom may have further assessments and eventual diagnosis beyond 6 months). We treat these as two distinct subgroups of patients, with the proportion in each group depending on the assessment strategy used (as can be seen from the differential proportion of patients for whom a diagnosis is reached within 6 months in figure 2 in Hollis *et al.*[18]). Furthermore, the case-mix of those with a diagnosis within 6 months differs between assessment strategies, with QbTest plus clinical assessment being more likely to make a diagnosis excluding ADHD than clinical assessment alone.[18,31] Therefore, the modelled prevalence of ADHD among those receiving diagnosis within 6 months depends on assessment strategy to reflect the differences in the proportions whose diagnosis is to exclude ADHD based on the AQUA study. Note that the overall prevalence across all patients is equal regardless of assessment strategy. Thus, the modelled prevalence of ADHD in those who do not have a diagnosis within 6 months also depends on assessment strategy, in order that the total of prevalence is fixed, regardless of the time of diagnosis.

Patients who have a diagnosis within 6 months are either diagnosed as having ADHD and will go on to receive treatment for ADHD or are diagnosed as not having ADHD and do not receive further treatment or assessments for ADHD. We heard from our clinical advisors that the main impact of QbTest is likely to be on the time waiting for assessment, number and length of appointments, which makes it easier to exclude ADHD without leading to appeal, rather than on the diagnostic accuracy of the eventual diagnosis. Adding QbTest to clinical assessment was not expected to make clinical assessment any less accurate, and this is assumed in our base-case model, although note that we do include the proportion of diagnoses made within 6 months and the proportion of those diagnoses that are ADHD in the model, both of which depend on the test. We also include DTA in a scenario analysis where those with a positive diagnosis include those who do have ADHD [true positives (TPs)] and those who do not have ADHD [false positives (FPs)], and those with a negative diagnosis include those who do have ADHD [false negatives (FNs)] and those who do not have ADHD [true negatives (TNs)], as illustrated in *Figure 7*. FPs are assumed to incur costs of treatment during the dose titration period but without any benefits in terms of response to treatment. We heard from our clinical advisors that treatment may continue into the long term for many patients who do not have ADHD but initiate treatment, and so we include costs of non-responders beyond the titration period to capture these ongoing costs. FNs do not incur treatment costs, but they do not gain any treatment benefits.

QbTest is administered early in the assessment period in our model (and in the AQUA trial), and the results from the AQUA trial show that there is little additional benefit of QbTest after five appointments. Based on this, we assume that the diagnoses after 6 months are no different than for clinical assessment alone, since the additional appointments beyond 6 months are likely to be based on additional reports other than QbTest (which has already been considered). However, because the prevalence of ADHD in those who receive a diagnosis within 6 months depends on assessment strategy, so too does the prevalence of ADHD in those who receive a diagnosis after 6 months (since the overall prevalence must be the same regardless of assessment strategy).

Patients who have not been diagnosed by 6 months are likely to be a mixture of those who have stopped attending assessments and do not have further assessment (where for those with ADHD their diagnosis will be 'missed') and those who will continue to have assessments and who get an eventual diagnosis. In other words, there are the following four groups of patients:

- Those who undergo further assessment for ADHD and receive a diagnosis of ADHD and go on to receive treatment for ADHD.
- Those who undergo further assessment for ADHD and receive a diagnosis of not having ADHD and receive no further treatment or assessments for ADHD.

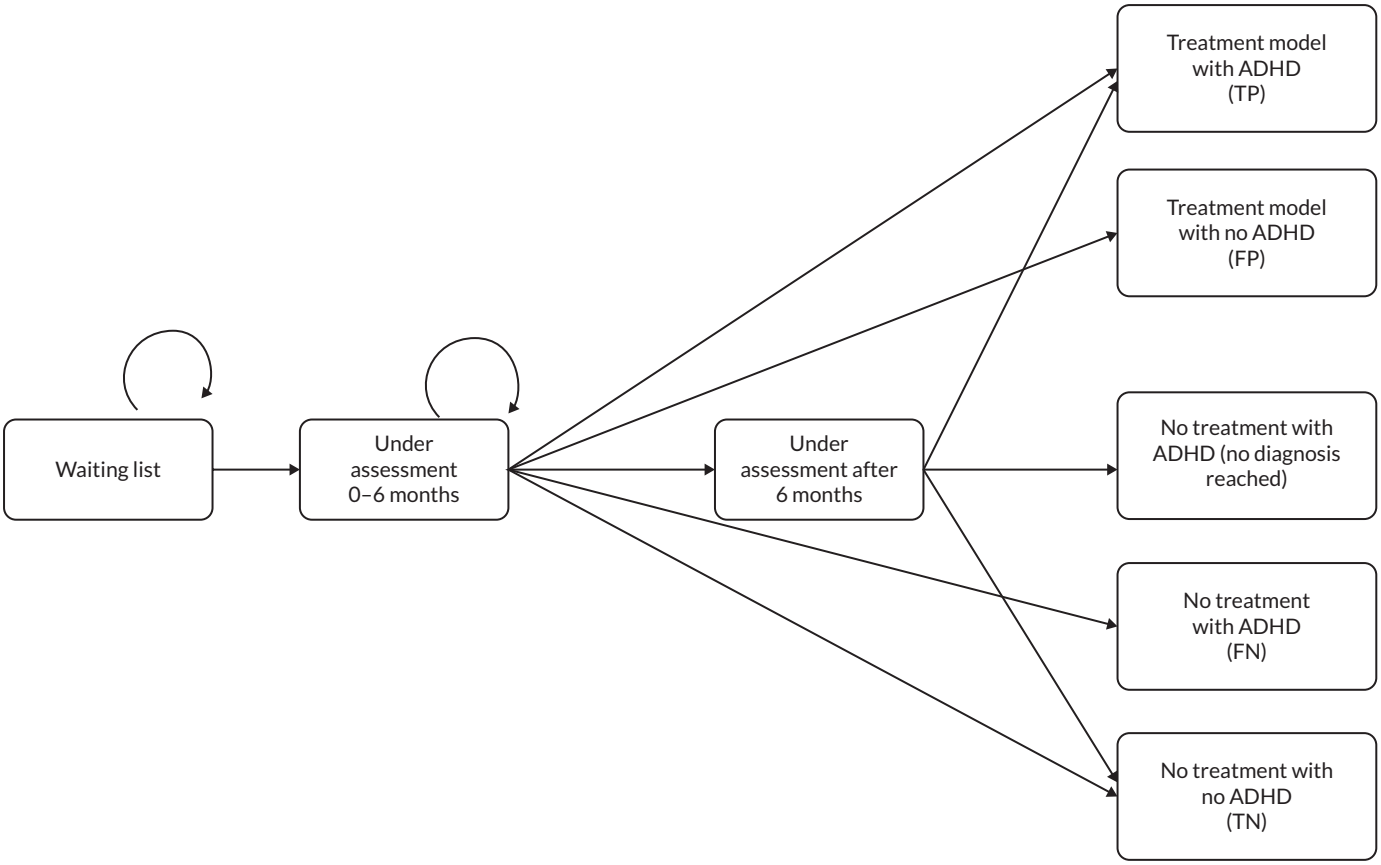

FIGURE 7 Markov model structure for the diagnosis of ADHD.

- Those who have ADHD but do not undergo further assessment and so do not receive appropriate treatment ('missed diagnosis').
- Those who do not have ADHD and do not undergo further assessment for ADHD, so further treatment for ADHD is not received or required. These patients are captured in the 'No Treatment with No ADHD (TNs)' state, even though they do not actually receive a diagnosis, since the health states are equivalent.

To evaluate the diagnosis model, we use an alternative (but equivalent) model structure, which is illustrated in *Figure 8*. Here, we evaluate the model separately for those who do and do not have a diagnosis within 6 months and then form an average over the proportions in each subgroup, which varies depending on whether QbTest is used or not. This makes it possible to have different model parameters for those who do not have a diagnosis within 6 months and to use tunnel states to ensure that the assessment period for those who have further assessments is longer than 6 months. For those who do not have further assessments, we assume they have an average of three assessments before they stop attending assessments, which is based on the data provided to us from the authors of the AQUA trial on the number of appointments for those patients who were censored. These patients will follow the same path in the model as those with further assessments, but they do not incur the assessment costs.

We assume patients with an ADHD diagnosis to initiate pharmacological treatment following NICE guidance,[16] starting with MPH for 2 monthly cycles; then, if no response, they switch to LDX for 2 monthly cycles; then, if no response, they switch to ATX. Note that the guidance is to switch treatments for non-responders every 6 weeks, but we have approximated this with 2 months to align with the cycle length of our model. The treatment model is shown for ADHD patients who initiate treatment (TPs) in *Figure 9a*, where costs and utilities depend on the treatment and response status. Patients who discontinue treatment due to adverse effects are modelled as if they are non-responders. For ADHD patients not on treatment (FNs and those who did not receive a diagnosis), we assume they are non-responders and incur costs and utilities for non-responders but without treatment costs (*Figure 9b*), although this may be an overestimate as non-responders are likely to be monitored more closely. For patients who do not have ADHD but receive a diagnosis (FPs), we assume that they initiate treatment, but do not respond, but may continue to incur

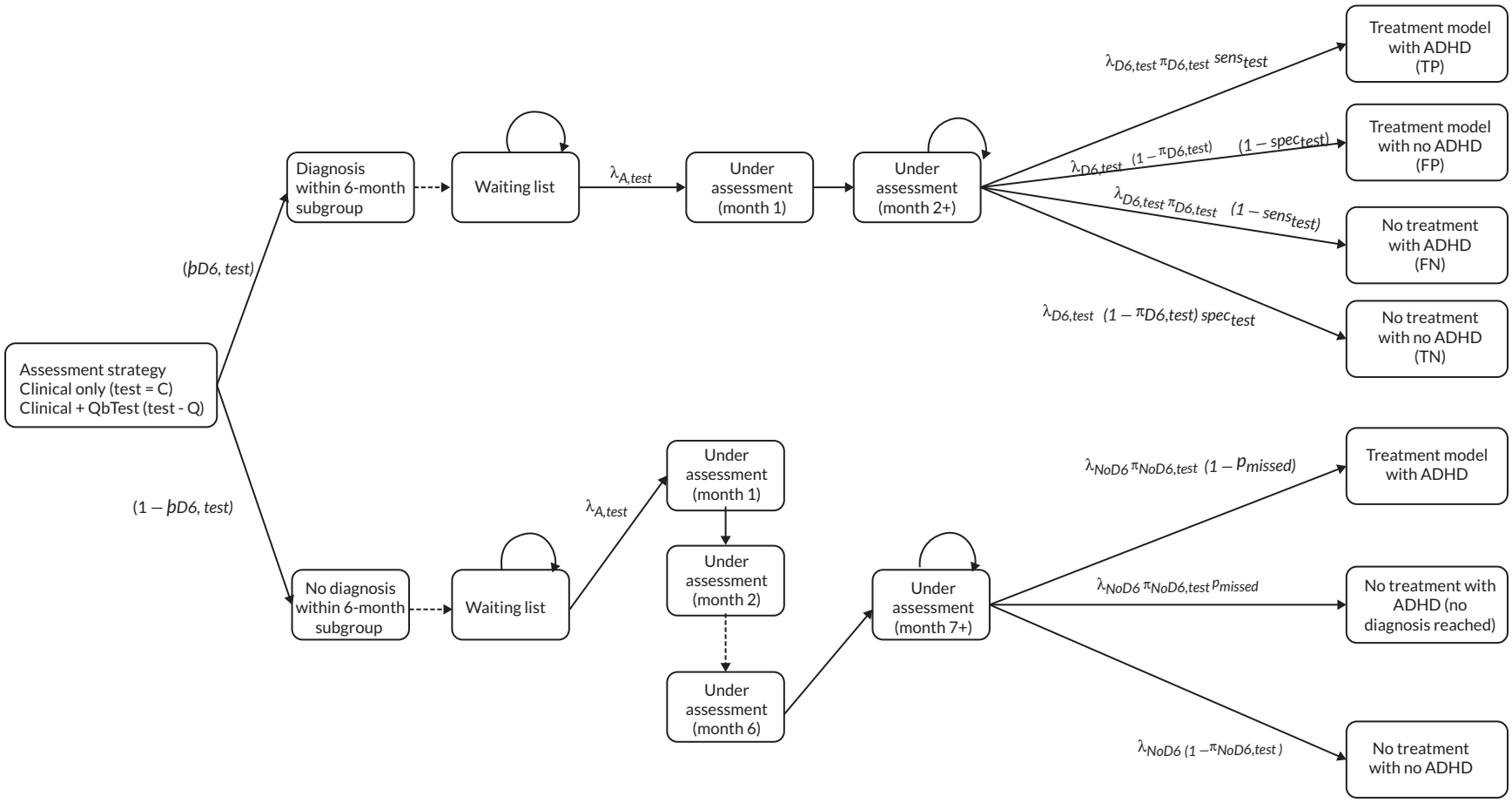

FIGURE 8 Markov model structure for the diagnosis of ADHD, restructured by subgroups with diagnosis before/after 6 months.

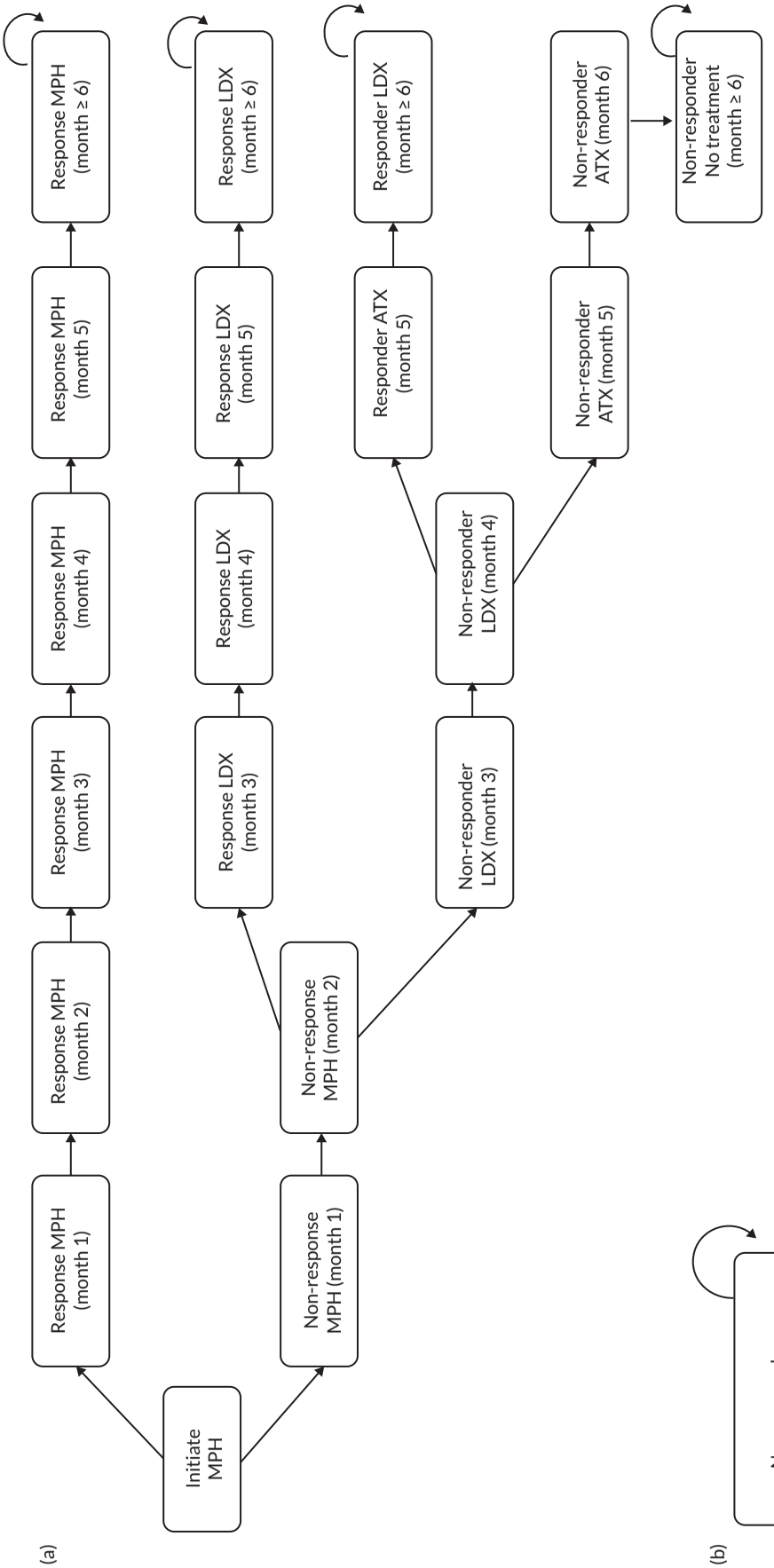

FIGURE 9 Markov model structure following diagnosis for patients with ADHD (a) for those diagnosed with ADHD (TPs) and (b) for those not diagnosed with ADHD (FNs).

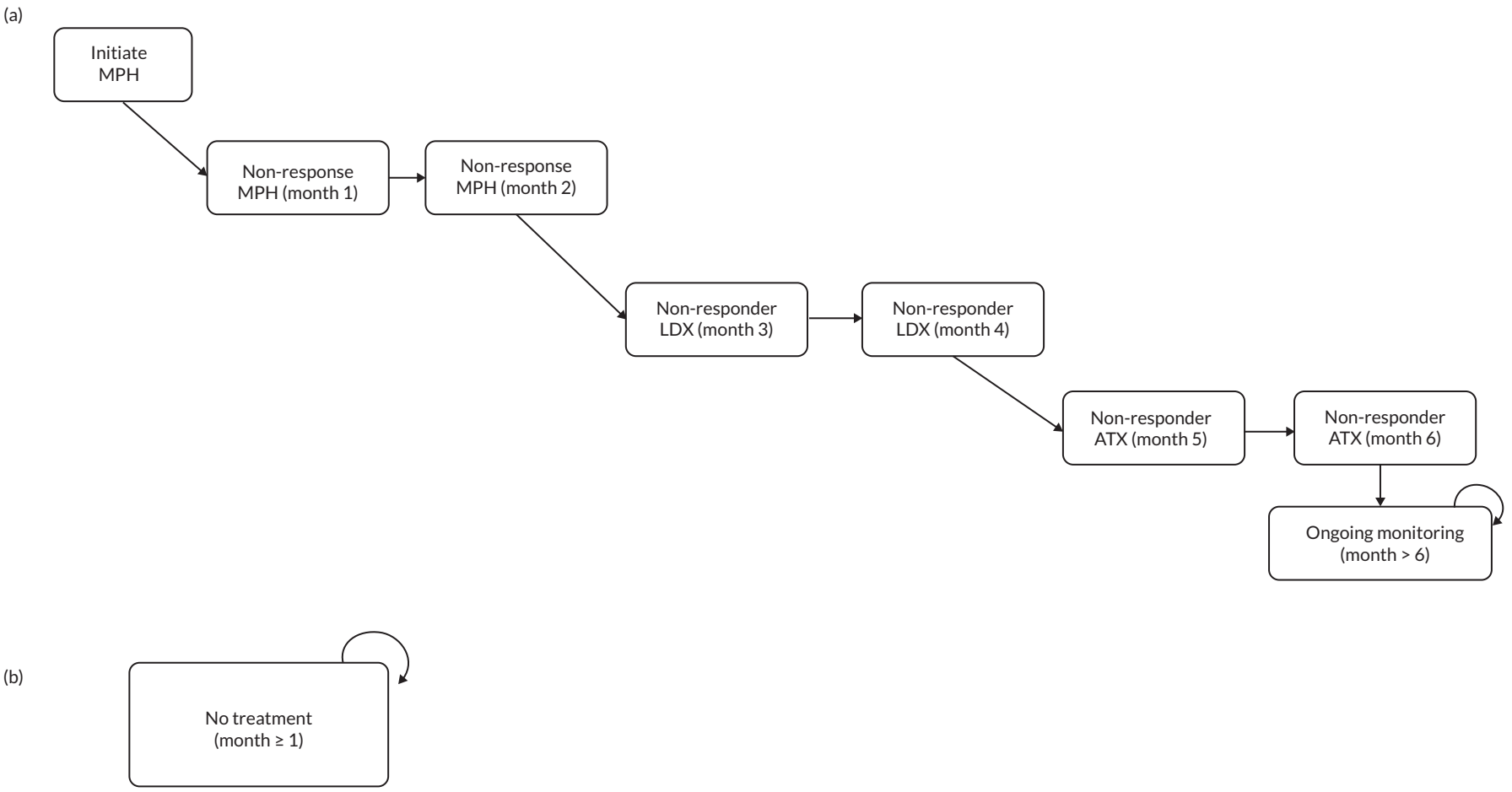

FIGURE 10 Markov model structure following diagnosis for patients without ADHD (a) for those diagnosed with ADHD (FPs) and (b) for those not diagnosed with ADHD (TNs).

monitoring costs in the long term (*Figure 10a*). In a scenario analysis, we assume that the FPs do not incur monitoring costs in the long term. For patients who do not receive a diagnosis and do not have ADHD (TNs), they are not on treatment and do not incur any additional costs (*Figure 10b*).

While waiting for diagnosis (either on waiting list or under assessment), the proportion of patients with ADHD receive QALYs corresponding to those with ADHD but not on treatment. In our base case, we assume that there are no additional costs while waiting, but in a scenario, we explore this being the same as ADHD patients not on treatment. While under assessment, all patients incur appointment costs and QbTest costs as appropriate.

The transition parameters of the model in continuous patient time are indicated in *Figure 8* and are defined below, where *test* indicates whether the assessment is made using standard clinical assessment only (*test* = *C*) or QbTest alongside clinical assessment (*test* = *Q*):

$p_{D6,test}$ is the proportion of patients with a diagnosis within 6 months, and it depends on assessment strategy.

$\lambda_{A,test}$ is the rate at which patients leave the waiting list for assessment, and it depends on assessment strategy.

$\lambda_{D6,test}$ is the rate at which patients receive a diagnosis for the subgroup that receives a diagnosis within 6 months, and it depends on assessment strategy.

λ_{NoD6} is the rate at which patients receive a diagnosis for the subgroup that does not receive a diagnosis within 6 months but go on to have further assessments, and it does not depend on assessment strategy.

$\pi_{D6,test}$ is the proportion of patients with ADHD in the subgroup of patients who receive a diagnosis within 6 months, and it depends on assessment strategy due to the difference in case-mix of those diagnosed by QbTest plus clinical assessment compared to clinical assessment alone.

$\pi_{NoD6,test}$ is the proportion of patients with ADHD in the subgroup of patients who do not receive a diagnosis within 6 months, and it depends on test because $\pi_{D6,test}$ depends on test and the overall prevalence of ADHD must be the same regardless of assessment strategy.

$sens_{test}$ is the proportion with ADHD having a positive diagnosis (sensitivity) for each assessment strategy in those who are diagnosed within 6 months. In our base case, the sensitivity of the assessment is assumed to be perfect ($sens_{test} = 1$), and we run scenario and threshold analyses assuming a lower sensitivity for QbTest plus clinical assessment.

$spec_{test}$ is the proportion of those without ADHD having a negative diagnosis (specificity) for each assessment strategy in those who are diagnosed within 6 months. In our base case, the specificity of the assessment is assumed to be perfect ($spec_{test} = 1$), and we vary this in scenario analyses.

p_{missed} is the proportion of patients without a diagnosis within 6 months who do not undergo further assessment and so do not receive a diagnosis. It is assumed that this does not depend on test, although the proportion without a diagnosis within 6 months does depend on test. This parameter is a key uncertainty which we vary in scenario analysis and threshold analysis.

To obtain transition probabilities from transition rates, we used the following relationships, where t is the cycle length in months.

The probability of moving from the waiting list to assessment in a cycle is the same regardless of patient subgroup:

$$p(wait \rightarrow assessment) = 1(-\lambda_{A,test} t) \quad (1)$$

For those who receive a diagnosis within 6 months, the proportion that receives an ADHD or no ADHD diagnosis in a cycle is:

$$
\begin{aligned}
&p(\textit{assessment} \rightarrow \textit{ADHD diagnosis}(\textit{true} + \textit{ve})) = (1 - \exp(-\lambda_{D6,test}t))\pi_{D6,test}\textit{sens}_{test} \\
&p(\textit{assessment} \rightarrow \textit{ADHD diagnosis}(\textit{false} + \textit{ve})) = (1 - \exp(-\lambda_{D6,test}t))(1 - \pi_{D6,test})(1 - \textit{spec}_{test}) \\
&p(\textit{assessment} \rightarrow \textit{No ADHD diagnosis}(\textit{false} - \textit{ve})) = (1 - \exp(-\lambda_{D6,test}t))\pi_{D6,test}(1 - \textit{sens}_{test}) \\
&p(\textit{assessment} \rightarrow \textit{No ADHD diagnosis}(\textit{true} - \textit{ve})) = (1 - \exp(-\lambda_{D6,test}t))(1 - \pi_{D6,test})\textit{spec}_{test}
\end{aligned} \tag{2}
$$

For those who do not receive a diagnosis within 6 months, the proportion that receives an ADHD diagnosis, no ADHD diagnosis or missed ADHD diagnosis in a cycle (after 6 months) is:

$$
\begin{aligned}
&p(\textit{assessment} \rightarrow \textit{ADHD diagnosis}) = (1 - \exp(-\lambda_{NoD6}t))\pi_{NoD6,test}(1 - p_{missed}) \\
&p(\textit{assessment} \rightarrow \textit{Missed ADHD diagnosis}) = (1 - \exp(-\lambda_{NoD6}t))\pi_{NoD6,test}p_{missed} \\
&p(\textit{assessment} \rightarrow \textit{No ADHD diagnosis}) = (1 - \exp(-\lambda_{NoD6}t))(1 - \pi_{NoD6,test})
\end{aligned} \tag{3}
$$

In the AQUA trial, there were up-to six appointments over a 6-month period, and so we assume that appointments are scheduled approximately every month and so use a monthly cycle for the model.

Evaluating strategy QbTestUnclear (objective 2)

We did not find any evidence on the use of sensor CPTs in those for whom a diagnosis could not be reached using standard assessment (objective 2) and so we need to make some assumptions, and it is important to note these are speculative and only presented as a scenario analysis. The AQUA trial evaluated the use of QbTest in all patients referred for assessment, and not in those with an unclear diagnosis, but we use some of the findings to support our assumptions for objective 2. In the AQUA trial, QbTest was administered after the first appointment and before the second appointment, and it was shown that there was no difference in the proportion of patients receiving a diagnosis after two appointments (figure 2 in Hollis *et al.*[18]) and no difference in the appointment time until diagnosis for the first 120 minutes appointment time, which also corresponds to two appointments (supplementary figure S7 in Hollis *et al.* 2018[18]). There was an increase in the proportions diagnosed and a reduction in appointment time to reach a diagnosis for QbTest from the third appointment onwards (i.e. the second appointment after administering the QbTest). This suggests that it may be reasonable to assume that there is a proportion of patients (approximately, 20% from figure 2 of Hollis *et al.* 2018) for whom diagnosis is relatively straightforward and can be achieved after two appointments (one appointment after QbTest is administered) regardless of whether QbTest results were used. This view agrees with our clinical advisers' experience who use QbTest only if a diagnosis is not reached after one assessment appointment (following the initial appointment).

To assess the QbTestUnclear strategy (objective 2), we ran a scenario analysis where it is assumed that QbTest is not administered until after two appointments and then only in those where a diagnosis has not yet been reached. We assume that 20% of patients reach a diagnosis after two appointments without QbTest, after which QbTest is administered to the remaining 80% of patients whose diagnosis is less clear. We vary this proportion in a scenario and threshold analysis. We assumed that the only difference between strategy QbTestUnclear compared with strategy QbTestAll was the proportion of patients incurring the cost of QbTest under the assumption that the diagnosis for the straightforward diagnoses does not depend on whether QbTest is used or not.

Model structure for dose titration (objective 3)

We developed a conceptual model to capture the impact of sensor CPT compared with standard assessment for dose titration in patients initiating pharmacotherapy for ADHD if sufficient data were available to populate it. The model captures the time period from initiating treatment until the first long-term monitoring assessment (assumed 6 months from the end of the titration period for children and 12 months for adults). *Figure 11* shows a Markov model structure for the titration period, including sequences of treatments following the recommendations in NICE Guidelines NG87. It is assumed that patients undergo a period of dose titration until they reach a stable dose and treatment, which may either be an optimal or suboptimal dose/treatment. This is followed by a period on treatment until their first long-term monitoring assessment when their medication will be reviewed. During this time, it is assumed that patients remain on the stable treatment/dose but that an optimal dose may become suboptimal over time and patients move from the

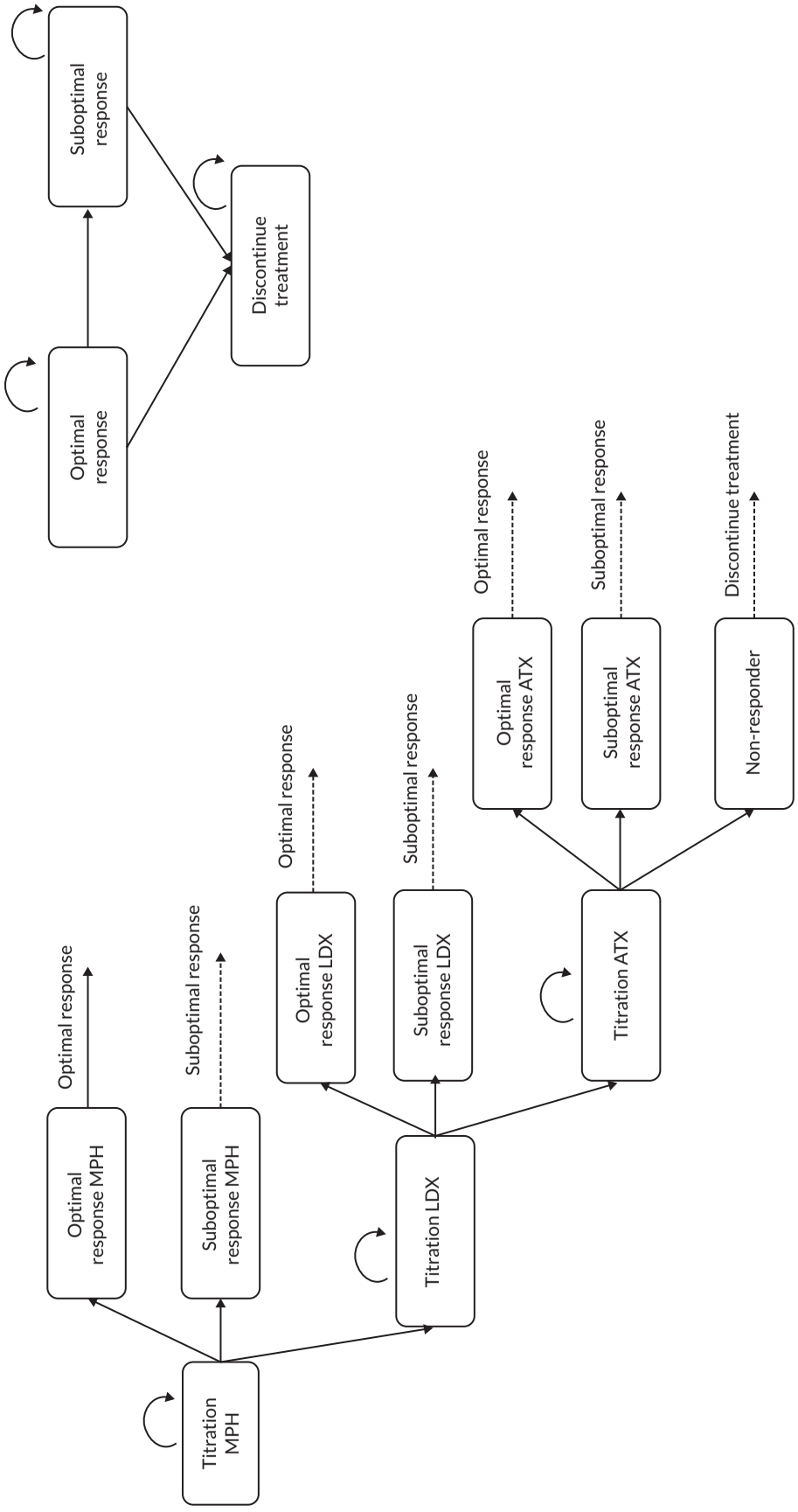

FIGURE 11 Markov model structure for dose titration in the pharmacological treatment of ADHD. Dotted arrows indicate the starting state in maintenance period model on the right-hand side.

'optimal response' state to the 'suboptimal response' state. Also, patients may discontinue treatment due to adverse effects, lack of adherence or lack of response. During the dose titration period, patients are monitored every 2 weeks when they incur appointment costs (likely remote appointments) and depending on assessment strategy the costs of the sensor CPT. Patients accrue costs and QALYs depending on whether they are have optimal response, suboptimal response or have discontinued treatment. The cycle length is 2 weeks to reflect the titration process, and the time horizon reflects the time from the initiation of treatment until the first long-term monitoring appointment.

Model structure for long-term monitoring (objective 4)

We did not identify any studies on the use of sensor CPT for long-term treatment monitoring of patients with ADHD, and so it is unclear how sensor CPT would be used in this context. However, we have developed a conceptual model setting which could potentially be used if sufficient evidence on the use of sensor CPTs in this context were available to populate it. *Figure 12* shows a model which cycles between two phases, with the first phase modelling the routine long-term monitoring assessment where those patients with optimal response continue on medication until the next monitoring appointment, and those with suboptimal response or not on treatment have their medication adjusted with dose titration if required, and a proportion of patients may be deemed to be in remission following a treatment holiday. Patients on treatment then enter into the response model until their next monitoring assessment, which is identical to that used for objective 3 post titration. Patients in remission are assumed to stay in remission until their next monitoring assessment when they may have relapsed. Routine monitoring is assumed to occur annually for adults and between 6 and 12 months for children.

Perspective and time horizon

An NHS and PSS perspective was taken where costs and QALYs were discounted at an annual rate of 3.5%. Because longer waiting times lead to lower test costs under discounting, we also run a scenario where discounting is not applied. For the diagnostic assessment model, we used a 10-year time horizon, which was considered to be long enough to capture the time waiting for assessment, time to reach a diagnosis and consequences of treatment in children before they enter adult services, by which time we assume that all have been appropriately diagnosed and treated. We run sensitivity analyses to the time horizon. The model included health effects for both patients and carers, but run a scenario analyses to include carer disutility.

We did not evaluate the dose titration and long-term monitoring models due to insufficient evidence. For the dose titration model, the time horizon should reflect the time until the first long-term monitoring appointment (6 months for children/adolescents and 12 months for adults) and so discounting is not necessary. The long-term monitoring model should use a life-time horizon, or until the cohort of patients have all stopped treatment.

Uncertainty

To reflect uncertainty in model inputs, we conducted probabilistic sensitivity analysis (PSA), where the parameter uncertainty is captured with probability distributions and simulation is used to estimate expected (mean) costs, expected QALYs, incremental cost-effectiveness ratios (ICERs) and expected incremental net benefit (INB) at willingness to pay (WTP) of £20,000 and £30,000 per QALY. The impact of uncertainty is presented using cost-effectiveness planes and the probability that QbTest is cost-effective at WTP of £20,000 and £30,000 per QALY. One-way sensitivity analyses were performed for all key parameters.

Model implementation and validation

The model is implemented in the R programming language. All files to run the model are provided, including a guide to running the model. The model underwent internal validation by two members of the team not involved in the building of the model, following Büyükkaramikli *et al.*[139] The validation included face validity tests, checks of model calculations and examination of the model outputs.

Model parameters and inputs

Model inputs for the diagnostic assessment model (see *Figure 8*) are described below. These were derived from the clinical and cost-effectiveness reviews where possible, mostly from the AQUA trial, supplemented by targeted literature

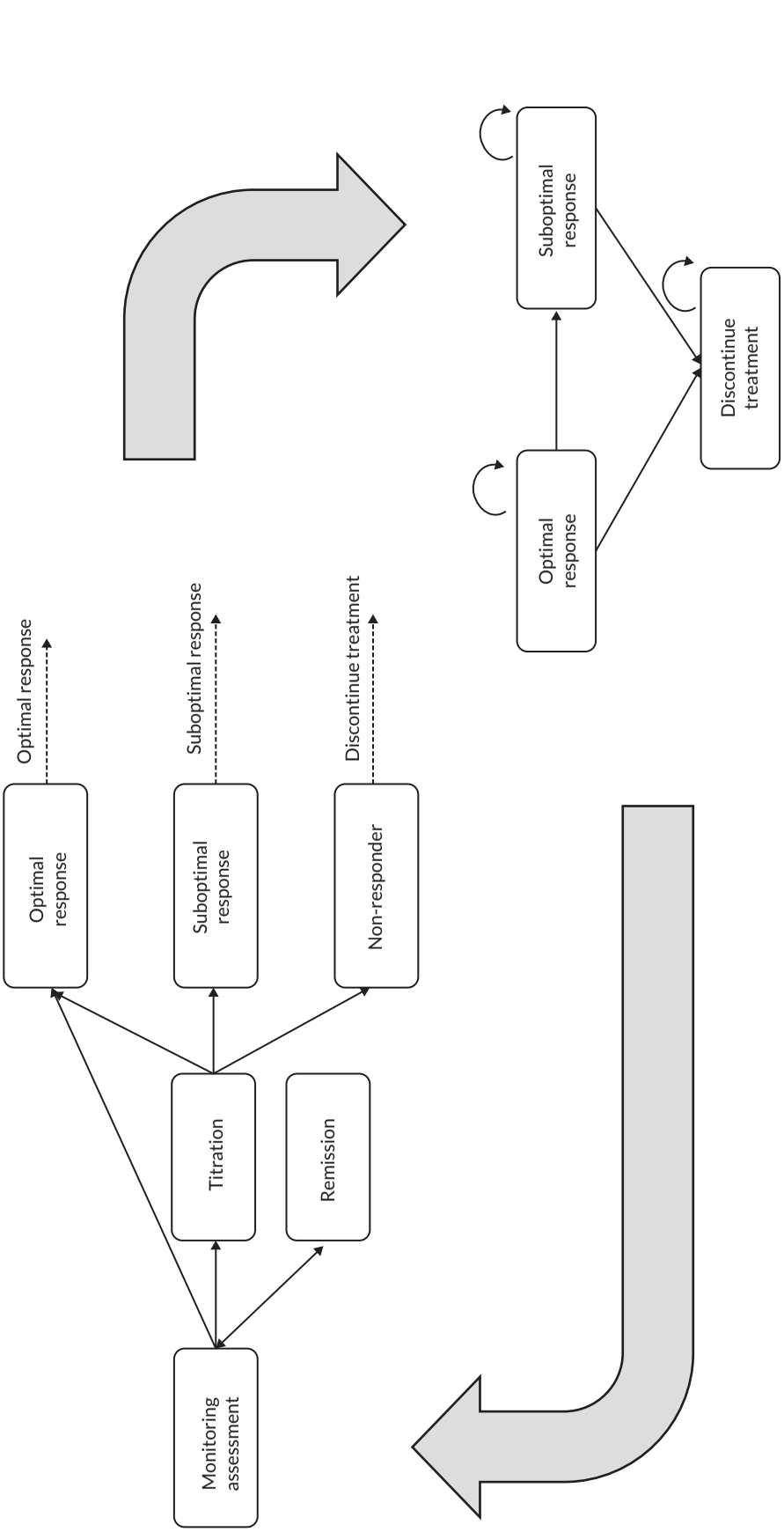

FIGURE 12 Markov model structure for long-term monitoring in the pharmacological treatment of ADHD.

searches. Where there was insufficient evidence available, we consulted with our clinical advisors as to plausible parameter assumptions and conducted scenario analyses to explore the impact of these assumptions on the results. A summary of all model inputs, assumed values, assumed distributions and evidence source is provided in *Table 26*.

Proportion receiving a diagnosis within 6 months after initiating assessment

In the AQUA trial, 94/123 = 76.4% (95% CI 68.9% to 83.9%) of patients received a diagnosis within 6 months after initiating assessment in the QbTest group (which corresponds to our strategy QbTestAll), whereas 76/127 = 59.8% (95% CI 51.3% to 68.4%) received a diagnosis within 6 months after initiating assessment in the control group (which corresponds to our strategy standard).[18] We used these figures to inform the proportion receiving a diagnosis within 6 months, $p_{D6,test}$, in the model.

Waiting time for assessment

Under standard clinical assessment only

Studies providing information on the waiting time for assessment under clinical assessment inform the rate that patients leave the waiting list for assessment for test = C $\lambda_{A,C}$, which is the reciprocal of the mean waiting time (for an exponential waiting time distribution). *Table 14* shows the results from a survey on waiting times for children, which was conducted by the Petitions Committee of those who had signed petitions for improvements to ADHD assessment.[20] Applying the proportions to the range mid-points gives a mean waiting time of 368.65 days, although this may be an overestimate due to selection bias (the sample being those who had signed a petition). The Focus ADHD study reports the mean time from referral to diagnosis and the mean time from assessment to diagnosis, from which we can calculate the mean time from referral to assessment, which was 335 days (with an approximation to the SE of 25.0).[31] We prefer this estimate because it is based on the data from 20 different sites, including a mix of CAMHS and paediatric services. We only use the data from the clinical assessment group (pre-QbTest) from the Focus ADHD study because the data post QbTest were impacted significantly by the COVID-19 pandemic.

Under QbTestAll and QbTestUnclear assessment strategies

The only study that directly provides evidence on the time from referral to assessment when using QbTest is the Focus ADHD study.[31] As noted in the Focus ADHD report, however, the estimates of number of days until assessment and until diagnosis for the post-QbTest group were significantly impacted by the COVID-19 pandemic and as such are not usable in our model. We therefore need a different approach.

TABLE 14 Studies with information on waiting time for assessment

Waiting time for clinical assessment	
Petitions committee survey[20]	
1–6 months	18%
6 months–1 year	22%
1–2 years	30%
2–3 years	14%
Approximate mean time (days)	368.65
Focus ADHD,[31] ***clinical assessment group (pre-QbTest)***	
Referral -> diagnosis (days)	Mean 452; range 15–3276; approx. SE[a] 22.5
Assessment -> diagnosis (days)	Mean 117; range 0–1570; approx. SE[a] 10.8
Referral -> assessment (days)	Mean 335; approx. SE[a] 25.0

a Approximate SE obtained by assuming range represents a 99.9% CI.

The AQUA trial provides an estimated TR (*TR* 0.85, 95% CI 0.77 to 0.93) for clinical appointment time for QbTest in addition to clinical assessment versus clinical assessment alone.[18] This estimate is based on an analysis of the full data set where those without a diagnosis are censored after their last appointment under the assumption that they would have similar TRs as those that had a diagnosis. We therefore consider this TR to be most applicable to those with a diagnosis within 6 months. The TR can be interpreted as a proportional reduction in number of months (appointments) to reach a diagnosis. Assuming that those appointments could be offered to those on the waiting list, it may be reasonable to assume a similar proportional reduction in the number of months waiting for an appointment (assuming that there are no changes in the referral rate). So, for a mean waiting time of 12 × 335/365 = 11.01 months under clinical assessment alone (*Table 14*), this would imply an adjusted mean waiting time of 0.85 × 11.01 = 9.36 months for QbTest in those having a diagnosis within 6 months of initiating assessment. We use a weighted average of an adjusted and non-adjusted transition rate according to the proportions of who have a diagnosis within 6 months of initiating assessment:

$$\lambda_{A,Q} = \frac{p_{D6,Q}\lambda_{A,C}}{TR} + (1 - p_{D6,Q})\lambda_{A,C} \quad (4)$$

Based on our assumed point estimates, this gives a rate of 0.103, which corresponds to a mean waiting time of 9.70 months with QbTest compared with 11.01 months without it, that is, a reduction in the mean waiting time of 1.31 months.

The transition rate from waiting to assessment is applied to all patients regardless of whether they are in the subgroup of patients with diagnosis within 6 months or not, because that is unknown while the patient is on the waiting list. We vary both the mean waiting time under standard assessment and the TR in sensitivity analyses to see the impact of assumptions around waiting list reduction on model results.

For the scenario where we explore the QbTestUnclear strategy, the waiting time is the same as for the QbTestAll strategy because the number of consultations for the patients with straightforward diagnoses is assumed to be unaffected by using QbTest.

Time from initiating assessment until a diagnosis is reached

Studies identified in the clinical review that report information on time from initial assessment to diagnosis are summarised in *Table 10*. The mean number of appointments roughly corresponds to the mean time if it is assumed that appointments are scheduled at monthly intervals until a diagnosis is reached. We use data on the number of appointments, assumed to be monthly, to obtain estimates of the mean time until diagnosis.

The pre-QbTest group in the Focus ADHD study[31] provides the largest and most representative evidence on the number of appointments until diagnosis under standard clinical assessment. We scanned and digitised data from the histogram for the number of appointments until diagnosis for pre-QbTest (see appendix 2, figure 6 of Focus ADHD study[31]), which enabled us to estimate the mean number of appointments separately in those who have a diagnosis within 6 months (appointments) and those who have further assessments after six appointments, as presented in *Table 15*.

In our model, we assume everyone has an initial appointment, after which the QbTest is administered and results are made available in time for the next clinical appointment. The rate at which a diagnosis is reached for those who receive a diagnosis within 6 months under standard clinical assessment is estimated as the reciprocal of the mean (minus 1 for initial appointment), giving an estimate of the rate $\lambda_{D6,\,C}$ of 0.76 95% CI (0.706 to 0.831), which we use in the model.

We then apply the HR reported in the AQUA trial to obtain the diagnosis rate (after the first clinical appointment) under the QbTest assessment strategy:

$$\lambda_{D6,\,Q} = \lambda_{D6,\,C} \times HR \quad (5)$$

In a scenario analysis, we use the HR for children and adolescents separately (*Table 10*).

We assume the diagnosis rate for those without a diagnosis within 6 months who continue to undergo further assessment is the same regardless of the test, and estimated that from the mean number of additional appointments (above six) from the Focus ADHD study (*Table 15*), which gives a monthly rate of 0.12, 95% CI (0.080 to 0.189), which we use in the model.

Prevalence of attention deficit hyperactivity disorder in those referred for assessment

Estimates of prevalence of ADHD in children range from 2% to 7%.[3] However, our model requires the prevalence of ADHD in those who have been referred for assessment for ADHD, which will be much higher due to the reasons for referral. We model prevalence of ADHD separately for those who have a diagnosis within 6 months, $\pi_{D6,test}$, and those who do not have a diagnosis within 6 months,$\pi_{No\ D6,test}$.

Prevalence in those who have a diagnosis within 6 months

Studies providing information on the proportion whose diagnosis was ADHD in those who obtained a diagnosis (within 6 months in AQUA) are shown in *Table 16*. In the AQUA trial and Focus ADHD before–after study, there was a higher proportion of ADHD diagnoses (in those with a diagnosis) in the clinical assessment group compared to QbTest plus clinical assessment. We prefer to use the results from the AQUA trial which was a RCT and not influenced by the COVID-19 pandemic (which unfortunately impacted on the results from Focus ADHD). However, we note that the patterns seen are similar to those seen in Focus ADHD.

In our base case, we assume that there is perfect sensitivity and specificity, and so the estimates from the AQUA trial from *Table 16* can be used directly to inform $\pi_{D6,C}$ and $\pi_{D6,Q}$.

In a scenario analyses, we explore alternative values for sensitivity and specificity. To do this, we assume that the results from the AQUA trial inform the prevalence of a positive diagnosis of ADHD from *Table 16* (which includes both TPs and FPs),$\pi_{AQUA,test}$.

We can then rearrange to write the prevalence of ADHD in those with a diagnosis within 6 months as a function of $\pi_{AQUA,test}$, $sens_{test}$, and $spec_{test}$:

$$\pi_{D6,test} = \frac{\pi_{AQUA,test} - (1 - spec_{test})}{sens_{test} - (1 - spec_{test})} \tag{6}$$

TABLE 15 Mean number of appointments until diagnosis in the Focus ADHD study in those who have a diagnosis

Assessment -> diagnosis	Clinical assessment	QbTest plus clinical assessment
Focus ADHD*[31] *(n = 549 per group)		
Mean (range) appointments	3.22 (1–50)	2.85 (1–32)[a]
Mean (SE) appointments in those with ≤ 6 appointments (*n* = 508)	2.3 (0.054)[b]	
Mean (SE) additional (above 6) appointments in those with > 6 appointments (*n* = 40)	8.9 (1.858)[b]	

a Note these figures were impacted by the COVID-19 pandemic.
b Computed using reconstructed data from scanning the histogram in appendix 2 in McKenzie *et al.*[31]

TABLE 16 Studies with the prevalence of ADHD diagnosis conditional on those with a diagnosis (within 6 months in the AQUA trial)

Prevalence of ADHD diagnosis (95% CI)	Clinical assessment	QbTest plus clinical assessment
AQUA[18]	65/76 = 85.5% (77.6% to 93.4%)	69/94 = 73.4% (64.5% to 82.3%)
Focus ADHD[31]	445/549 = 81.1% (77.8% to 84.3%)	418/549 = 76.1% (72.6% to 79.7%)

Prevalence in those who do not have a diagnosis within 6 months

For those that did not receive a diagnosis within 6 months, the proportion of patients with ADHD is expected to be lower than in those who had a diagnosis within 6 months. We did not identify any studies proving information on this directly. Vogt *et al.* 2011[86] reported that 7/19 = 36.8% 95% CI (15.2% to 58.5%) of patients who did not receive an ADHD under clinical assessment alone subsequently received an ADHD diagnosis after 1-year FU. We use this estimate for $\pi_{No\ D6,\ C}$ in the model and vary it in a sensitivity analysis.

To estimate the prevalence of ADHD in those who did not get a diagnosis by 6 months for the QbTest strategy, we note that the total prevalence of ADHD must be the same regardless of test. The means that:

$$\pi_{D6,C}p_{D6,C} + \pi_{No\ D6,\ C}(1 - p_{D6,C}) = \pi_{D6,Q}p_{D6,Q} + \pi_{No\ D6,\ Q}(1 - p_{D6,Q}) \quad (7)$$

Rearranging we obtain:

$$\pi_{No\ D6,\ Q} = \frac{\pi_{D6,C}p_{D6,C} + \pi_{No\ D6,\ C}(1 - p_{D6,C}) - \pi_{D6,Q}p_{D6,Q}}{(1 - p_{D6,Q})} \quad (8)$$

Sensitivity and specificity

We did not find any suitable evidence to estimate the sensitivity and specificity of QbTest in addition to clinical assessment in our clinical review due to issues with the reference standard used in the AQUA trial (see *Diagnostic accuracy of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder*). In our base case, we make an assumption that the sensitivity and specificity of QbTest plus clinical assessment are the same as that for clinical assessment alone (which is assumed a gold standard). The rationale for this is that adding additional information on which to base the assessment is not expected to lead to a less accurate diagnosis in those where a diagnosis is reached. This is to some extent supported by the ROC analysis conducted by Hollis *et al.*,[18] which found there was no evidence of a difference in the diagnostic accuracy between QbTest plus clinical assessment and clinical assessment alone.

We conduct scenario and threshold analyses to explore the impact of changing sensitivity and specificity. The ratio (QbTest plus clinical assessment vs. clinical assessment alone) of sensitivity against the imperfect reference standard from the AQUA trial was 0.86/0.96 = 0.895. The corresponding ratio for specificity was 39.5/36.0 = 1.097; that is, QbTest was more specific than clinical assessment alone, so specificity of standard relative to QbTest is 0.9116. We therefore conduct a range of scenarios with alternative sensitivity and specificity assumptions.

Missed diagnosis (attention deficit hyperactivity disorder or no attention deficit hyperactivity disorder)

There was a high proportion of patients who did not receive a diagnosis in the AQUA trial, and this proportion was higher under standard clinical assessment. The median number of appointments for those who did not receive a diagnosis in the AQUA trial was three (calculated from data provided by the AQUA authors); however, we do not know if they attended further assessment after the trial and eventually received a diagnosis (whether that was for ADHD or to exclude ADHD). To get an understanding of the proportion of patients who have assessments beyond six appointments, the Focus ADHD study provides a histogram of the number of appointments until diagnosis in those who received a diagnosis (see appendix 2, figure 6 of Focus ADHD study[31]). Based on this, we estimated that 7.25% of patients who eventually receive diagnoses (whether for ADHD or excluding ADHD) had more than six appointments. Applying this to the 40.2% of cases that did not receive a diagnosis in the standard clinical assessment arm of the AQUA trial suggests that (100–7.25/40.2) = 82% of those who do not have a diagnosis within six appointments will not attend for further assessment and their diagnosis is missed. If their diagnosis would have been for ADHD, then they will not receive treatment benefits or costs as for FNs. If their diagnosis would have been to exclude ADHD, then they will appropriately not receive treatment as for the TNs. In our base case, we assume the proportion who do not have further assessment, $p_{missed} = 0.82$. This is an important assumption in the model due to the large and different proportions who do not receive a diagnosis in the AQUA trial, and so we vary this in scenario and threshold analyses.

Proportion of patients with a less clear diagnosis

We conducted a scenario analysis to evaluate the QbTestUnclear strategy (objective 2), as explained in *Model structures*. We assume that everyone has two appointments, after which 20% of patients are diagnosed. QbTest is administered to the remaining 80% prior to the third appointment. We base this estimate on the proportion without a diagnosis after two appointments in the AQUA trial, 80% 95% CI (75.0% to 85.0%), noting that the AQUA trial was not designed to evaluate the QbTestUnclear strategy, and so these assumptions are speculative. We vary this proportion in the sensitivity analysis.

Resource use and costs

Costs were obtained from routine NHS sources to represent costs in the 2023–4 financial year values. For staff and unit costs related to administration of the sensor CPT and ADHD treatment, we use the latest NHS cost collection available from 2021 to 2022,[141] and for costs from Personal Social Services Research Unit (PSSRU), we used the Unit Costs of Health and Social Care 2023 Manual for 2022–3 costs.[140] For 2023–4 NHS reference costs, we also referred to the NHS payment scheme for 2023–4 published in August 2023,[146] however, none of the required unit costs were reported there. We inflated NHS cost collection 2021–2 and PSSRU 2022–3 to 2023–4 costs using the Consumer Price Inflation index 06.2.1/3 (medical services and paramedical services) using the ratio March 2024 : March 2023 (122.9/118.8 = 3.45% inflation)[141,147] or March 2022 (122.9/114.4 = 7.43% inflation). For drug costs, we use the *British National Formulary* (BNF) updated on 26 March 2024.[148] Resource use was estimated from our reviews of previous cost-effectiveness models, targeted literature searches and through discussions with the manufacturers and clinical advisors. Unit costs of the sensor CPT were provided by the manufacturers. We did not include costs that are incurred regardless of assessment strategy, such as long-term treatment costs incurred for patients without ADHD.

Staff costs

Nurse time to administer QbTest

QbTest takes 15–20 minutes to complete, but the appointment to administer the test will need to be longer to conduct administrative tasks and set the test up. Hall *et al.* found that a 30-minute nurse-led appointment was required to administer the test,[85] whereas a previous economic evaluation assumed that a 1-hour appointment was required (based on assumption).[113] The economic evaluation[113] of the East Midlands AHSN study[69] noted that band 4 nurses were used in two trusts and band 2 in the other, whereas the manufacturer's submission suggests a band 3 healthcare assistant. We assume a one-off 30-minute band 4 nurse-led appointment to administer the test (*Table 17*) based on an hourly cost of £38 inflated to £39.31.[140]

Consultant paediatrician time for assessment

We assume that each assessment appointment (with or without QbTest) is at a community paediatric service or CAMHS service. No costs for these services are available in the 2023 PSSRU, so we take the mean of the costs of CAMHS Outpatient Attendance (£383.46) and Community Paediatric Service (Outpatient Attendance – code 290) (£350) from 2021 to 2022 NHS reference costs.[141] Each appointment cost is therefore £366.73,[141] inflated to £393.98.

Costs related to using the technologies

A laptop computer, camera, tripod and headband with reflective spot are required to conduct the QbTest. A plastic sleeve is replaced on the headband each time the test is conducted, and this is the only consumable used. When the test is completed, the results are automatically uploaded to QbTest's central server in order to generate the report

TABLE 17 Cost of QbTest administration

Item	Cost
Band 4 nurse for 30 minutes (£39.31 per hour[a])	£19.66
QbTest unit cost per test	£31.20 (range £23–96)
Total	£50.86

a Band 4 nurse per hour (excluding qualifications) (PSSRU 2023).[140]

comparing the patient's results to the normative data. The device equipment is all provided as part of QbTest, as well as clinical advisor support, and training material, and this is included in the cost. Manufacturer advised that the cost per test ranges from £23 to £96 per test depending on volume used, and most NHS trusts pay £31.20 per test. We vary this in a sensitivity analysis.

To administer the QbTest, a private and quiet room with a computer, desk and chair is needed. A hard stool with no back or arms is required for ages 6–12 years and a hard chair with a back but no arms is required for ages 12–60 years. The room must be free of visual distractions for the patient or reflective areas, so windows must be able to be darkened. As staff time estimates account for overhead/space costs in PSSRU, we do not include additional costs for space, but note that appropriate space has be available, which may be an issue for implementation.

Trained healthcare assistants or nurses can oversee the test, and a trained clinician interprets the results. According to training material available on the QbTest website, there are three training modules: administration (2–3 hours), interpretation (2–3 hours) and intermediate interpretation (2–3 hours). The healthcare assistant or nurse (band 4) administering the test would only need to complete the administration portion of the training, while clinicians interpreting the results would complete all three portions. We assume the clinicians are medical or psychiatric consultants with an hourly cost of £109/hour,[140] inflated to £112.76. The cost of training is likely to be approximately £118 per (band 4) nurse trained, and £1015 per consultant trained, but we do not account for this in our model as it is a start-up cost that is not allocated per patient treated.

The clinical review found that some patients (between 5% and 11%, *Impact of sensor continuous performance tests for diagnosis of attention deficit hyperactivity disorder on process measures* on test failure) were unable to complete the QbTest assessment; however, the test administration costs will still be incurred, and so test costs are incurred for all patients in the model. For patients for whom the test is not appropriate [e.g. those with intelligence quotient (IQ) < 70] (manufacturer's submission), QbTest would not be used under any assessment strategy, and so those patients are not included in our model. We run scenario analyses where 5% or 11% incur the test administration cost, but the outcomes are as for standard assessment rather than the QbTest.

QbTest is the only sensor CPT for which we found effectiveness data; however, we do have cost information for Nesplora AULA (suitable for paediatrics) and EF Sim.

Nesplora AULA costs £21.03 for a single use (plus a one-off registration fee of £84.12), £75.70 for 7 uses (monthly), £227.11 for 22 uses (quarterly) or £1345.85 per year for unlimited use on a single VR device. The actual price paid will therefore depend on the volume of tests required and the plan chosen. For the purposes of illustration, we run a scenario with all inputs as for QbTest, but with the test costs of £21.03 (for a single use), a scenario with the test cost of £10.32 (based on 22 uses per quarter) and a scenario with test cost of £2.80 (based on the annual professional plan with 40 assessments per month as estimated by Nesplora in their response to the Evidence Assessment Group report). The cost of the nurse time to administer the test is as given in *Table 17* and is added to the test cost.

Peili Vision Oy (ARVO) proposes a different delivery model where a dedicated healthcare assistant travels to each practice 1 day per month to provide EF Sim assessments to all patients with suspected ADHD based on initial screening. They estimate a cost per practice 7.5 hour working day of £197.05. Based on an assumed 30-minute slot for each test, 15 tests would be conducted per day at a cost of 197.03/15 = £13.14 per test. This includes the healthcare assistant cost. We include an illustrative scenario using this cost with all other inputs as for QbTest.

We stress that the scenarios using the costs for Nesplora AULA and EF Sim should not be interpreted as cost-effectiveness analyses of those technologies, since there are no effectiveness data for these tests, and in the case of EF Sim, the delivery model is quite different.

Health-state costs for attention deficit hyperactivity disorder patients who do and do not respond to treatment

We identified health-state costs from analyses within our review of cost-effectiveness of ADHD treatment which were conducted in the UK, of which we considered the King Health Technology Assessment (HTA),[121,136] and NICE guideline

NG87,[16] in particular *Appendix 2*,[115] to be the most appropriate sources for health-state costs in paediatrics (see *Results of the cost-effectiveness review*).

Zimovetz (2016)[136] was the most recent UK-based study and it updates the health-state costs for the items from the King *et al.*'s (2006) HTA report[121] using a survey of 21 UK specialists. However, the NICE NG87 Guideline highlights concerns about potential bias in Zimovetz (2016) due to industry funding. Using the resource use and unit costs presented in table 2 in Zimovetz (2016), and in table 88 in King (2006), we updated the costs for responders and non-responders using PSSRU (2023)[140] and National Schedule of Reference Costs 2021–2.[141] inflated to 2024. The resulting costs are shown in *Tables 18* and *19*.

The appendix 2 of the NICE NG87 guideline[115] presents the resource use during dose titration and maintenance and for non-responders to other treatments. Unit costs of psychiatrist time and band 7 nurse are updated to costs from PSSRU (2023). We assume a ratio of 1 : 0.95 for contact hours for consultants, while the ratio for band 7 nurse is 1 : 0.33.[140] The hourly cost for a consultant psychiatrist is £109 and for a band 7 nurse is £68 (excluding qualifications),[140] with inflated unit costs accounting for TRs of £228.34 and £97.16 per contact hour, respectively. The resource use and costs per month on treatment are shown in *Table 20*.

We used resource use values for dose titration and responders and non-responders to apply to the states in the treatment models (*Figures 9* and *10*). FPs are assumed to incur the cost of a non-responder post titration, reflecting that patients are likely to continue to be monitored and treated, but we run a scenario analysis where no further costs are incurred post titration for FP cases. We use appendix 2 values in the updated NICE NG87 in the base case (*Table 20* £38.06 and £76.11 for responder and non-responder costs per month after dose titration). In a scenario analysis, we used the higher values for responder and non-responder costs after dose titration from Zimovetz (2016) (£170.52 and £325.90) and King (2006) (£398.86 and £573.13).

TABLE 18 Annual health-state costs of paediatric responder vs. non-responder to ADHD treatment updated from table 2 in Zimovetz (2016)[136]

Item	Responder resource use	Non-responder resource use	2022 unit cost	Responder cost	Non-responder cost
Psychiatrist[a]	2.48	5.19	411.95	1021.64	2138.03
Paediatrician[b]	2.33	4.1	306.18	713.39	1255.32
GP[c]	2.62	4.24	50.69	132.81	214.93
Nurse[d]	2.71	4.48	60.00	162.60	268.81
Blood test[e]	0.42	0.72	3.18	1.34	2.29
ECG[f]	0.18	0.39	80.48	14.49	31.39
Total annual				2046.27	3910.76
Monthly				**170.52**	**325.90**

ECG, electrocardiogram.
a CAMHS outpatient attendances.[141]
b Paediatric outpatient attendance, 420.[141]
c Table 9.4.2. Each consultation lasting 10 minutes, including direct care staff costs, excluding qualification costs [PSSRU (2023) unit cost manual].[140]
d Table 9.2.1. Band 6 cost per hour excluding qualifications [PSSRU (2023) unit cost manual].[140]
e DAPS05 – Haematology.[141]
f DADS EY51Z Electrocardiogram Monitoring or Stress Testing.[141]

Note
Unit costs updated to 2023–4 values with sources matched as closely as possible with unit costs used in Zimovetz (2016).[136]

TABLE 19 Annual health-state costs of paediatric responder vs. non-responder to ADHD treatment updated from table 88 in King *et al.*[121]

Item	Responder resource use	Non-responder resource use	Unit cost	Responder cost	Non-responder cost
Psychiatrist[a]	3.5	5.75	411.95	1441.83	2368.72
Paediatrician[b]	2.25	2.5	306.18	688.90	765.44
GP[c]	3	2.75	50.69	152.07	139.40
Blood test[d]	0.05	0.35	3.18	0.16	1.11
ECG[e]	0.18	0.33	80.48	14.49	26.56
EEG[f]	0	0.43	286.53	0.00	123.21
Allergy test[g]	0	0.5	8.18	0.00	4.09
Total annual				2297.44	3428.52
Monthly				191.45	285.71

ECG, electrocardiogram; EEG, electroencephalogram
a CAMHS outpatient attendances.[141]
b Paediatric outpatient attendance, 420.[141]
c Table 9.4.2. Each consultation lasting 10 minutes, including direct care staff costs, excluding qualification costs (PSSRU (2023) unit cost manual).[140]
d DAPS05 – Haematology.[141]
e DADS EY51Z Electrocardiogram Monitoring or Stress Testing.[141]
f DADS AA33D Conventional EEG, EMG or Nerve Conduction Studies, 18 years and under.[141]
g DAPS06 Immunology.[141]

Note
Unit costs updated to 2023–4 values.

TABLE 20 Monthly health-state costs of paediatric responder vs. non-responder to ADHD treatment updated from appendix 2 in NG87 and applied to model structure shown in *Figures 9* and *10*

State	Month	Resource use psychiatrist (minutes)	Resource use nurse (minutes)	Total cost
Response MPH	1	60	20	260.73
	2	100	0	380.57
	3	10	0	38.06
	4	10	0	38.06
	5	10	0	38.06
	6+	10	0	38.06
Non-response MPH	1	60	40	293.12
	2	0	0	0.00
Response LDX	3	60	20	260.73
	4	100	0	380.57
	5	10	0	38.06
	6+	10	0	38.06
Non-response LDX	3	60	40	293.12
	4	0	0	0.00
Response ATX	5	60	20	260.73
	6	100	0	380.57

continued

TABLE 20 Monthly health-state costs of paediatric responder vs. non-responder to ADHD treatment updated from appendix 2 in NG87 and applied to model structure shown in *Figures 9* and *10* (*continued*)

State	Month	Resource use psychiatrist (minutes)	Resource use nurse (minutes)	Total cost
	7+	10	0	38.06
Non-response ATX	5	60	40	293.12
	6	0	0	0.00
	7+	20	0	76.11
No treatment with ADHD	1+	0	0	0.00

Note
Unit cost for psychiatrist is £228.34 per hour and unit cost for nurse is £97.16 per hour, accounting for ratios of contact time. Following 2 months of dose titration, responders are assumed to have 2 hours of psychiatrist contact per year (averaged to 10 minutes per month), and non-responders are assumed to have 4 hours of psychiatrist contact per year (averaged to 20 minutes per month). All resource uses are adapted from NICE[115] and unit costs are adapted from PSSRU 2022 unit cost manual.[140]

Drug costs

Methylphenidate is available in modified-release (12-hour tablets or 8-hour capsules) and immediate release (IR) formulations. The NHS Specialist Pharmacy Service notes that modified release may be preferred in general (unless flexible dosing is required).[149] We therefore identified costs for modified-release formulations based on the average doses (*Table 21*) during titration and after titration used in the King HTA[121] using the nearest actual dose available. There is a variation in monthly costs across the different formulations available (*Table 22*), and in the absence of information on the market share of the different formulations, we used an average cost across formulations in our model.

Lisdexamfetamine is available as an oral capsule. Dittmann *et al.* (2013)[142] found that, starting with a 30-mg daily dose, the mean dose after optimisation was 52.5 mg (*Table 21*), which is close to the 50-mg capsule. We assume that the average dose during titration is 40 mg (the mid-point between starting and optimised dose). The estimated monthly costs are shown in *Table 22*.

Mean dose for ATX after titration in the King HTA was 45 mg,[121] whereas it was 40.2 mg in the Dittman trial.[142] The Dittmann trial was specifically on patients who have not responded after a trial of MPH, which is most relevant to our model, and so we use a dose of 40 mg for ATX after titration. We use the estimate from the King HTA which is close to the 25-mg available dose. There were 12 different products listed on the BNF, all with similar costs, and so we present an average cost for ATX in *Table 22*.

Treatment effects

Adverse events

Adverse events rates were estimated using the NG87 NICE guideline review (summary forest plots for children aged 5–18 years displayed in section E2 of document D).[16] There was no evidence of differences between the treatments in the total number of AEs with risk ratio for MPH versus ATX relative risk = 0.99 95% CI (0.87 to 1.13) and risk difference (RD) for ATX versus LDX RD = −0.01 95% CI (−0.12 to 0.10). For the purpose of our model, which focuses on diagnosis decisions, rather than treatment decisions, we consider it to be reasonable to assume that the overall AE rate is the same for each treatment. To estimate the AE rate attributable to treatment, we estimate the RD compared with placebo. We pool the results from the studies of ATX, LDX or MPH versus placebo to get a pooled RD of 0.1435 95% CI (0.0734 to 0.2186) (*Table 23*). We assume that this proportion of patients will experience AEs while on treatment and there will be a disutility associated with this.

Some patients will discontinue treatment due to adverse effects, which we obtained from studies in the NG87 NICE guideline review of pharmacological studies (document C). For MPH as a first-line treatment, there was no evidence of heterogeneity and so we used a fixed-effect meta-analysis to give a pooled estimate for discontinuation due to adverse effects of 0.0244, 95% credible/credibility interval (CrI) (0.0127 to 0.0396) (*Table 24*).

TABLE 21 Average drug dosage

Drug	Average dose during titration	Average dose after titration	Source
MPH modified-release 12	27 mg	35 mg	King HTA[121]
MPH modified-release 8	25 mg	41 mg	King HTA[121]
LDX	–	52.5 mg	Dittmann trial[142]
ATX	28 mg	45 mg	King HTA[121]
	–	40.2 mg	Dittmann trial[142]

TABLE 22 Drug costs

Item	Pack price/size	Monthly cost	Source
MPH hydrochloride			
Dose titration average dose			
Concerta XL 27 mg	£36.81/30	£36.81	BNF
Affenid XL 27 mg[a]	£12.87/30	£12.87	BNF
Delmosart 27 mg[a]	£15.57/30	£15.57	BNF
Matoride XL 27 mg[a]	£15.58/30	£15.58	BNF
Xaggitin XL 27 mg[a]	£15.58/30	£15.58	BNF
Xenidate XL 27 mg[a]	£15.57/30	£15.57	BNF
Equasym XL 20 mg	£30.00/30	£30.00	BNF
Medikinet XL 20 mg	£28.86/30	£28.86	BNF
Metyrol XL 20 mg	£20.43/30	£20.43	BNF
	Average	**£21.25**	
Average dose after titration			
Concerta XL 36 mg	£42.45/30	£42.45	BNF
Affenid XL 36 mg[a]	£14.85/30	£14.85	BNF
Delmosart 36 mg[a]	£21.21/30	£21.21	BNF
Matoride XL 36 mg[a]	£21.22/30	£21.22	BNF
Xaggitin XL 36 mg[a]	£21.22/30	£21.22	BNF
Xenidate XL 36 mg[a]	£21.21/30	£21.21	BNF
Equasym XL 40 mg	£60.00/30	£60.00	BNF
Medikinet XL 40 mg	£57.72/30	£57.72	BNF
Metyrol XL 40 mg	£39.88/30	£39.88	BNF
	Average	**£33.31**	
LDX mesilate			
Average dose during titration			
Elvanse 40 mg	£62.82/28	£67.31	BNF
Average dose after titration			
Elvanse 50 mg	£68.60/28	£73.50	BNF

continued

TABLE 22 Drug costs (*continued*)

Item	Pack price/size	Monthly cost	Source
ATX			
Dose titration average dose			
ATX 25 mg (average)[b]	£49.43/28	£52.96	BNF
Average dose after titration			
ATX 40 mg (average)[b]	£50.79/28	£54.42	BNF

a Bio-similar to Concerta.
b Average over 12 available products.

TABLE 23 Proportions with AEs and pooled RD for active treatment compared with placebo

AE/*n* (prop.) Study	Placebo	ATX	LDX	MPH
Hervas (2014)[150]	73/111 (0.658)	76/112 (0.679)		
Martenyi (2010)[151]	11/33 (0.333)	44/72 (0.611)		
Newcorn (2008)[152]	40/74 (0.541)	149/221 (0.674)		146/219 (0.667)
Takahashi (2009)[153]	43/62 (0.694)	144/183 (0.789)		
Wehmeier (2012)[154]	27/62 (0.435)	32/63 (0.508)		
Montoya (2009)[155]	19/51 (0.373)	65/100 (0.650)		
Childress (2014)[156]	34/72 (0.472)		162/218 (0.743)	
Findling (2011)[157]	45/77 (0.584)		160/233 (0.687)	

Note
Random-effects meta-analysis: pooled RD treatment vs. placebo: 0.1435, 95% CI (0.0734 to 0.2186).

The LDX and ATX are used for patients who have not responded to MPH. The NG87 NICE guideline review of pharmacological sequencing (figure 274 of document C)[16] identified one study with information for LDX and ATX in the population of those who have not responded to MPH, which we use as inputs to our model (*Table 24*).

Response to treatment

Response rate for MPH as a first-line treatment was based on the studies with modified-release MPH in the NG87 NICE guideline review of pharmacological studies (see figure 42 of document C).[16] We pooled these in a fixed-effect meta-analysis that gave a pooled estimate of 0.502, 95% CI (0.434 to 0.571) (*Table 25*).

The LDX and ATX are used for patients who have not responded to MPH. The NG87 NICE guideline review of pharmacological sequencing (see figures 268 and 273 of document C)[16] identified two studies of LDX and one study of ATX in the population of those who have not responded to MPH. We pool the two studies of LDX in a fixed-effect meta-analysis and use the single study for ATX[142] as inputs to our model (*Table 25*).

Health-state utilities

Utilities on waiting list and under assessment

While patients are waiting for assessment and diagnosis, we assume that the proportion with ADHD have the same HRQoL as an ADHD patient who is not on treatment or not responding to treatment. For the proportion of patients

TABLE 24 Treatment discontinuation due to adverse effects

Study	Discontinued	Total	Proportion that discontinued	SE
MPH				
Coghill (2013)[158]	2	112	0.017857	0.012514
Findling (2008)[159]	2	91	0.021978	0.015369
Wolraich (2001)[160]	1	94	0.010638	0.010582
Palumbo (2008)[161]	1	29	0.034483	0.033883
Wang (2007)[162]	6	166	0.036145	0.014487
Fixed-effect meta-analysis: pooled proportion discontinuing 0.0244, 95% CrI (0.0127 to 0.0396)				
LDX (in those who have not responded to MPH)				
Dittmann (2013)[142]	8	128	0.0625	0.0214
ATX (in those who have not responded to MPH)				
Dittmann (2013)[142]	10	134	0.0746	0.0227

TABLE 25 Proportion responding for MPH, LDX and ATX

Study	Responders	Total	Proportion responders	SE
MPH				
Coghill (2013)[158]	57	107	0.53271	0.048233
Wolraich (2001)[160]	44	94	0.468085	0.051466
Fixed-effects meta-analysis: pooled proportion responders 0.502, 95% CI (0.434 to 0.571)				
LDX (in those who have not responded to MPH)				
Dittmann (2013)[142]	103	126	0.81746	0.034413
Jain (2011)[163]	15	19	0.789474	0.093529
Fixed-effects meta-analysis: pooled proportion responders 0.672, 95% CI (0.409 to 0.872)				
ATX (in those who have not responded to MPH)				
Dittmann (2013)[142]	84	132	0.636	0.04187

without ADHD, we assume they have the same HRQoL as an ADHD patient who is responding to treatment, which we consider to be more appropriate than using values from the general population, since they have been referred for ADHD assessment and likely have another condition which affects their quality of life. So, the average utility for a patient waiting for assessment and diagnosis is

$$utility = (u_{non-responder} - u_{carer-dis})\ prev + u_{responder}(1 - prev) \quad (9)$$

where $u_{carer-dis}$ is the disutility (utility decrement) for a carer of an untreated ADHD patient (see below) and where the prevalence of ADHD can be estimated from the model prevalence parameters (see *Prevalence of attention deficit hyperactivity disorder in those referred for assessment*) as:

$$prev = \pi_{D6,C} p_{D6,C} + \pi_{No\ D6,\ C}(1 - p_{D6,C}) \quad (10)$$

Utilities for attention deficit hyperactivity disorder patients who do and do not respond to treatment

For the model, we need utilities for those who do and do not respond to treatment ($u_{responder}$ and $u_{non-responder}$ resp.). Our review of previous treatment models found that typically it was assumed that utilities for patients not on treatment were the same as non-responders to treatment, and we also make this assumption. We reviewed previous models for utility values and identified recent systematic reviews of quality of life in people with ADHD and searched the references for UK studies or studies using EuroQol-5 Dimensions (EQ-5D).[164–169] We considered the most appropriate source to be the van der Kolk *et al.* study,[143] used in the NICE Guideline NG87 models,[16] which was reasonably large and, although conducted in the Netherlands, used a UK value set. The estimated utilities were 0.83 and 0.74 for responders and non-responders, respectively, which we use in the model. In sensitivity analyses, we use the three sets of utilities that were used in sensitivity analyses by Zimovetz *et al.* (2016).[136]

Utility decrement for adverse events of treatment

We assume a utility decrement due to AEs of treatment based on Secnik (2005),[144] which was used in the NICE Guideline NG87 models,[16] and they report a reduction in utility (using adjusted standard gamble) for with versus without AEs of 0.01, which we use in the model.

Carer utilities

The quality of life for carers of patients with ADHD was based on Peasgood *et al.* (2021),[145] which was the most relevant study identified for a recent UK population and reports results for EQ-5D. Peasgood *et al.* compared the EQ-5D for carers of a child with ADHD with a matched control group, and we assume that this difference is a proxy for the difference in EQ-5D for carers of ADHD patients who are responding to treatment compared with non-responders. They report a difference in EQ-5D for carers of a child with ADHD versus matched controls of −0.071 when matching on standard covariates, −0.05 when also matching on results of ADHD screening for the carer and −0.018 when also matching on employment and relationship factors.[145] We use a value of 0.018 for carer disutility, $u_{carer-dis}$, in our base-case model, and vary it to 0.071 in a sensitivity analysis as well as running a sensitivity analysis with no carer disutility. A summary of all model inputs, assumed values, assumed distributions and evidence source is provided in *Table 26*.

Scenario and sensitivity analyses

A summary of the sensitivity and scenario analyses is given in *Table 27*, together with a rationale for each scenario.

Model results

Base-case results for strategy QbTestAll for diagnostic assessments

Under the base-case scenario, QbTestAll has higher costs and QALYs gained compared to standard assessment, with incremental costs of £238.35 and incremental QALYs of 0.0385 per person evaluated for ADHD (*Table 28*). The resulting ICER is £6183.71 per QALY gained, which is cost-effective at a WTP threshold of £20,000 per QALY. The mean INB is £532.55 and £918 at WTP of £20,000 and £30,000 per QALY, respectively. Exploring the impact of uncertainty in the input parameters, the QbTestAll intervention is cost-effective under 92% and 84% of model runs with a £20,000 and £30,000 WTP threshold, respectively (*Table 28*). It may appear counterintuitive that the probability that QbTestAll is cost-effective falls for higher WTP (*Figure 13*); however, the reason for this is due to uncertainty as to whether QbTestAll has higher or lower incremental costs, as can be seen from the cost-effectiveness plane (*Figure 14*). Some model runs fall within the bottom-left quadrant (lower costs and lower QALYs), and so a higher proportion of model runs lie under the WTP threshold line at £20,000 compared with £30,000 WTP per threshold. Most model runs (71%) fall in the top right quadrant (higher costs and higher QALYs), with 17% in the bottom left quadrant (lower costs and lower QALYs) and 12% in the bottom right quadrant (lower costs and higher QALYs, i.e. dominant).

Table 29 shows the breakdown of costs and QALYs accrued while on the waiting list, under assessment, and post assessment (for those that do or do not initiate treatment). In terms of costs, the QbTestAll strategy reduces the cost of assessment but increases the cost of treatment within the time horizon evaluated. This is due to diagnosis being reached sooner with QbTestAll, leading to patients being on treatment for a longer duration, and also due to a higher

TABLE 26 Summary of model inputs, values and distribution assumed in base-case analysis, and source of evidence

Model parameter	Value in base case	Distribution for PSA	Evidence source
Waiting time parameters			
Mean waiting time, standard	335 days (SE 25) 11.01 months (SE 0.8217)	Normal (mean = 11.01, SD = 0.8217)	Focus ADHD[31]
Rate waiting -> assessment, Standard	1/mean waiting time		Assumption
TR for clinical appointment time, QbTest vs. Standard, *TR*	0.85, 95% CI (0.77 to 0.93)	Log-normal (mean log = −0.163, SD log= 0.0482)	AQUA[18]
Rate waiting -> assessment, QbTest	$\lambda_{A,Q} = \frac{p_{D6,Q}\lambda_{A,C}}{TR} + (1 - p_{D6,Q})\lambda_{A,C}$		Assumption
Prevalence of ADHD parameters			
Prevalence of ADHD in those who have a diagnosis within 6 months, standard	65/76 = 85.5% (77.6% to 93.4%)	Beta (mean = 0.855, var = 0.0404[2])	AQUA[18]
Prevalence of ADHD in those who have a diagnosis within 6 months, QbTest	69/94 = 73.4% (64.5% to 82.3%)	Beta (mean = 0.734,var = 0.0456[2])	AQUA[18]
Prevalence of ADHD for those with no diagnosis within 6 months, standard	36.8%, 95% CI (15.2% to 58.5%)	Beta (mean = 0.368,var = 0.1107[2])	Vogt (2011)[86]
Prevalence of ADHD diagnosis given no diagnosis after 6 months, QbTest	$\pi_{No\ D6,\ Q} = \frac{\pi_{D6,\ C}p_{D6,\ C} + \pi_{No\ D6,\ C}(1-p_{D6,\ C}) - \pi_{D6,\ Q}p_{D6,\ Q}}{(1-p_{D6,\ Q})}$	N/A	Derived from the requirement that total prevalence of ADHD does not depend on assessment strategy
Subgroups			
Proportion with diagnosis within 6 months, standard	59.8%, 95% CI (51.3% to 68.4%)	Beta (mean = 0.598, var = 0.0435[2])	AQUA[18]
Proportion with diagnosis within 6 months, QbTest	76.4%, 95% CI (68.9% to 83.9%)	Beta (mean = 0.764, var = 0.0383[2])	AQUA[18]
Diagnosis rates			
Monthly diagnosis rate in those with diagnosis within 6 months, standard	0.76, 95% CI (0.706 to 0.831)	Log-normal (mean log = −0.269, SD log= 0.041)	Focus ADHD[31]
HR for diagnosis QbTest vs. standard (in those with diagnosis within 6 months)	1.44 (1.04 to 2.01)	Log-normal (mean log= 0.365, SD log= 0.168)	AQUA[18]
Monthly diagnosis rate in those with diagnosis within 6 months, QbTest	$\lambda_{D6,\ Q} = \lambda_{D6,\ C} \times HR$		Assumption
Diagnosis rate (in those with no diagnosis after 6 months)	0.12 (0.080 to 0.189)	Log-normal (mean log = −2.164, SD log= 0.222)	Focus ADHD[31]
Diagnostic accuracy			
Sensitivity of QbTest	1.0	N/A	Assumption
Specificity of QbTest	1.0	N/A	Assumption
Sensitivity standard clinical assessment	1.0	N/A	Assumed gold standard

continued

TABLE 26 Summary of model inputs, values and distribution assumed in base-case analysis, and source of evidence (*continued*)

Model parameter	Value in base case	Distribution for PSA	Evidence source
Specificity standard clinical assessment	1.0	N/A	Assumed gold standard
Proportion of those without diagnosis at 6 months who do not have further assessment	0.82	N/A	Assumption based on Focus ADHD[31] and AQUA[18]
Costs			
QbTest cost, including nurse time to administer the test	£50.86 per test	N/A	Manufacturer's submission, PSSRU (2023)[140]
Consultant paediatrician outpatient appointment	One appointment, £393.98	N/A	NHS reference costs 2021–2[141] Average of CAMHS and community services
Monthly average costs for responders	During titration month 1 £260.73, month 2 £380.57, post titration £38.06	N/A	NG87 appendix 2, PSSRU (2023)[115,140]
Monthly average costs for non-responders	During titration month 1 £293.12, post titration £76.11	N/A	NG87 appendix 2, PSSRU (2023)[115,140]
Drug costs MPH	During titration £21.25 per month After titration £33.31 per month	N/A	BNF
Drug costs LDX	During titration £67.31 per month After titration £73.50 per month	N/A	BNF
Drug costs ATX	During titration £52.96 per month After titration £54.42 per month	N/A	BNF
Treatment effects			
Proportion of responders on MPH	0.502, 95% CI (0.434 to 0.571)	Beta (alpha= 100.9, beta= 99.9)	Meta-analysis of studies from NICE NG87
Proportion of responders on LDX	0.814, 95% CI (0.751 to 0.877)	Beta (alpha= 118, beta= 27)	Meta-analysis of LDX studies from NG87
Proportion of responders on ATX	0.636, 95% CI (0.554 to 0.718)	Beta (alpha= 84, beta= 48)	Dittman (2013)[142]
Proportion with AEs on treatment	0.1435, 95% CI (0.0734 to 0.2186)	Beta (alpha= 13.2, beta= 78.7)	Meta-analysis of studies from NICE NG87
Proportion discontinuing due to adverse effects	MPH: 0.0244, 95% CrI (0.0127 to 0.0396) LDX: 0.0625, 95% CrI (0.0206 to 0.1044) ATX: 0.0746, 95% CrI (0.0301 to 0.1191)	Beta (alpha = 12.0, beta = 481.7) Beta (alpha = 8, beta = 120) Beta (alpha = 10, beta = 124)	Meta-analysis of studies from NICE NG87
Utilities			
Utility for patients waiting for assessment and diagnosis	0.76, 95% CrI (0.73 to 0.79), calculated using: $(u_{non-responder} - u_{carer-dis}) \times prev$ $+ u_{responder} \times (1 - prev)$, where $prev = \pi_{D6,C} p_{D6,C} + \pi_{No\ D6,C}(1 - p_{D6,C})$	Simulated based on distributions for utilities for responders and non-responders below	Van der Kolk (2014)[143]
Utility for ADHD patients responding to treatment, $u_{responder}$	0.83	Beta (alpha = 489.7, beta = 100.3)	Van der Kolk (2014)[143]

TABLE 26 Summary of model inputs, values and distribution assumed in base-case analysis, and source of evidence (*continued*)

Model parameter	Value in base case	Distribution for PSA	Evidence source
Utility for ADHD patients not on treatment or not responding to treatment, $u_{non-responder}$	0.74	Beta (alpha = 436.6, beta = 153.4)	
Disutility for AEs from treatment	0.01	N/A	Secnik (2005)[144]
Carer disutility for ADHD patients not on treatment, and for patients not responding to treatment, $u_{carer-dis}$	0.018	N/A	Peasgood (2021)[145]

N/A, not applicable.

TABLE 27 List of scenario analyses included

Scenario	Description	Base case	Sensitivity analysis	Rationale for analysis
1	Proportion of patients with less clear diagnoses	N/A	A threshold analysis for different values for the proportion of patients that the test is used for: varied from 0.5 up to 1	To explore objective 2 where sensor CPT is used in those with less clear diagnoses. We assume the only difference is in the cost of the test
2	HR for diagnosis rate, QbTest plus clinical assessment vs. clinical assessment alone	1.44 (1.04 to 2.01)	(a) 1.84 (1.23, 2.68) subgroup analysis from AQUA for children 6–12 years (b) 0.82 (0.37, 1.82) subgroup analysis from AQUA for adolescents 12+ years (c) 1 In all of above, the TR is assumed to vary linearly on a log-scale (passing through the base-case values and where TR= 1 when HR= 1), giving:	The HR for diagnosis rate from the AQUA trial differed in young children and adolescents
3	Mean waiting time under standard assessment	11.01 months (SE 0.8217)	3 months, 6 months and 18 months	There is wide variation in waiting times across regions
4	Sensor CPT cost (including nurse time)	£50.86	(a) £42.66 (QbTest lower range) (b) £115.66 (QbTest upper range) (c) £40.69 (Nesplora AULA single use) (d) £29.98 (Nesplora AULA quarterly plan for 22 uses) (e) £22.46 (Nesplora AULA annual professional plan, 40 assessments per month) (f) £13.14 (EF Sim assuming delivery model and costs proposed by company and 15 tests per monthly practice visit)	We do not have effectiveness data for sensor CPTs other than QbTest. To explore the impact of different test costs, we vary the price using the range of costs provided by the manufacturers of QbTest, Nesplora AULA and EF Sim, which vary according to volume of use
5	Higher response/non-response cost after dose titration period	Responders: £260.73 m1, £380.57 m2, £38.06 m3+ Non-responders £293.12 m1, £0 m2, £76.11 m3+	Zimovetz: Responders: £260.73 m1, £380.57 m2, £170.52 m3+ Non-responders £293.12 m1, £0 m2, £325.90 m3+ King: Responders: £260.73 m1, £380.57 m2, £191.45 m3+ Non-responders £293.12 m1, £0 m2, £285.71 m3+	We use the NG87 appendix 2 values in the base case, and for post dose titration (month 3+) responder and non-responder costs, we use the higher Zimovetz or King values as a scenario analysis

continued

TABLE 27 List of scenario analyses included (*continued*)

Scenario	Description	Base case	Sensitivity analysis	Rationale for analysis
6	Proportion with no further assessment after no diagnosis within 6 months,	0.82	0, 0.25, 0.5 A threshold analysis for different values for the proportion of patients that have no further assessments: varied from 0 up to 1	We have no evidence to inform the proportion without a diagnosis within 6 months who go on for further assessments, but model results are likely sensitive to assumptions on this parameter
7	Time horizon	10 years	15 and 20 years	The 10-year time horizon is in line with the longest time horizons of previous treatment models, but it is somewhat arbitrary. We therefore explore sensitivity of results to longer time horizons
8	Discount rate	3.5%	0%	Longer waiting times lead to lower test costs under discounting, which may benefit longer waiting times. We therefore run a scenario where discounting is not applied
9	TR	0.85	0.9, 0.95, 1	The TR is used to determine the impact of QbTest on waiting times, but this is based on assumption that the proportional effect in TR for number of appointments can be applied to waiting times. To explore the impact of this assumption, we vary the TR to reflect a smaller proportional effect on waiting times
10	Diagnostic accuracy	*sens*= 1 *spec*= 1	(a) $sens_Q = 0.9, spec_C = 1$ (b) $sens_Q = 1, spec_C = 0.9$ (c) $sens_Q = 0.9, spec_C = 0.9$ (d) $sens_Q = 0.9, spec_C = 0.9$ $\pi_{D6,test} = \frac{\pi_{AQUA,test} - (1 - spec_{test})}{sens_{test} - (1 - spec_{test})}$ We also run a threshold analysis varying $sens_Q$ from 0.6 to 1.	There was no evidence comparing test accuracy of QbTest plus clinical assessment vs. clinical assessment alone. We assumed that there is perfect diagnostic accuracy in our base case, but we relax this in this scenario
11	Prevalence of ADHD in those without a diagnosis within 6 months under clinical assessment, $\pi_{No\ D6,\ C}$	36.8%, 95% CI (15.2% to 58.5%)	20%, 50% Threshold analysis varying this from 0% to 100%	We did not find any studies reporting this directly, so made an assumption based on Vogt (2011)[86]
12	Carer disutility, $u_{carer-dis}$	0.018	0, 0.071	Peasgood (2021)[145] reports different estimates depending on what they match for. We run a sensitivity analysis using 0.071 rather than 0.018. We also run a sensitivity analysis that does not include a carer disutility

TABLE 27 List of scenario analyses included (*continued*)

Scenario	Description	Base case	Sensitivity analysis	Rationale for analysis
13	Waiting list costs	0	Waiting list costs for those with ADHD assumed equal to those for non-responding ADHD patients	Patients with ADHD may use additional NHS resource while waiting for assessment. For illustration, we set this to the resource costs for ADHD patients not responding to treatment, although we acknowledge this may be an upper bound as these patients will be monitored more closely
14	FP costs post dose titration	Non-responder health-state cost	0	Ideally, those without ADHD who initiate treatment (FPs) would stop treatment after a dose titration period due to lack of response. In practice, this does not happen and so we include a monitoring cost in our base case, but set this to 0 in a scenario
15	Proportion with diagnosis within 6 months, QbTest	0.764, 95% CI (0.689 to 0.839)	0.689, 0.598	To explore the impact of a lower proportion diagnoses within 6 months on QbTest, at the lower CI from AQUA and at the extreme with no difference between QbTestAll and standard
16	Test failure	Costs and QALYs unchanged	Test failure rate 5%, 11% who incur test administration cost, but all other costs and QALYs as for standard	Those who do not complete the QbTest are likely to have costs and QALYs similar to those under standard assessment
17	Utility for ADHD patients responding to treatment and not on treatment/ not responding to treatment	0.83 and 0.74	(a) 0.827 and 0.773 (b) 0.82 and 0.70 (c) 0.926 and 0.905	There are limited data on utilities. These scenarios cover those used in Zimovetz *et al.* (2016)[136] and cover assumptions made in previous models of treatment of ADHD

TABLE 28 Cost-effectiveness results comparing the QbTestAll strategy with standard for diagnostic assessment (probabilistic analysis)

						(£20,000 WTP)		(£30,000 WTP)	
Strategy	Total costs (discounted)	Total QALYs (discounted)	Incremental Costs	Incremental QALYs	ICER	Mean INB	Prob (CE)	Mean INB	Prob (CE)
Standard	£6004.78	6.9083	–	–	–	–	–	–	–
QbTestAll	£6243.14	6.9469	£238.35	0.0385	£6183.71	£532.55	0.922	£918.00	0.884

Prob (CE), probability of being most cost-effective.

TABLE 29 Costs and QALYs accrued while waiting for assessment, under assessment and post assessment for those who initiated treatment (TPs and FPs) and those who did not on initiate treatment (TNs and FNs); probabilistic analysis

	Total costs (discounted)				Total QALYs (discounted)			
Strategy	Waiting	Assessment	Post assessment: those who initiated treatment	Post assessment: those who did not initiate treatment	Waiting	Assessment	Post assessment: those who initiated treatment	Post assessment: those who did not initiate treatment
Standard	£0.00	£1462.02	£4542.76	£0.00	0.7150	0.4930	3.2551	2.4453
QbTestAll	£0.00	£1263.55	£4979.59	£0.00	0.6361	0.3457	3.5718	2.3933

TABLE 30 Proportion entering the postassessment states, base case

Strategy	QbTestAll	Standard
Proportion initiating treatment	0.579	0.538
Proportion not initiating treatment with ADHD	0.081	0.121
Proportion not initiating treatment with no ADHD	0.340	0.340

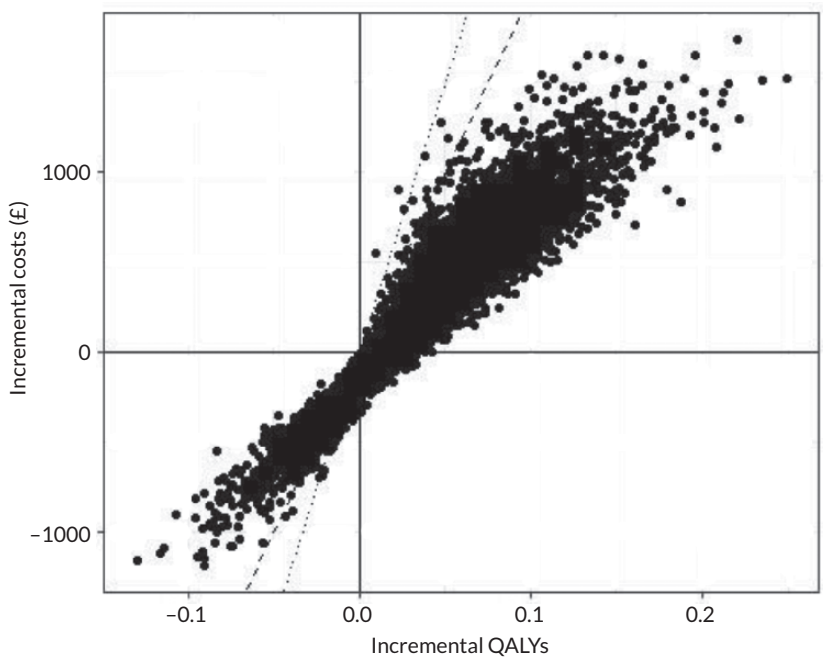

FIGURE 13 Cost-effectiveness acceptability curve for base-case model. Probability QbTestAll is cost-effective compared to standard.

proportion initiating treatment (*Table 30*) due to more patients receiving a diagnosis with QbTestAll. In terms of QALYs, fewer QALYs are accrued on the waiting list and under assessment with the QbTestAll strategy, again due to faster diagnosis. More total QALYs are accrued for those on treatment, while fewer are accrued for those who do not have ADHD and do not receive treatment. Note that in the base case, sensitivity and specificity of both tests are 1, so there are no FPs or FNs. However, the proportion who do not receive a diagnosis is lower with QbTestAll, which is why there are fewer QALYs accrued in the no-treatment group under QbTestAll. Overall, QALYs gained under QbTestAll are due to patients diagnosed with ADHD getting on treatment sooner and a higher proportion receiving a diagnosis.

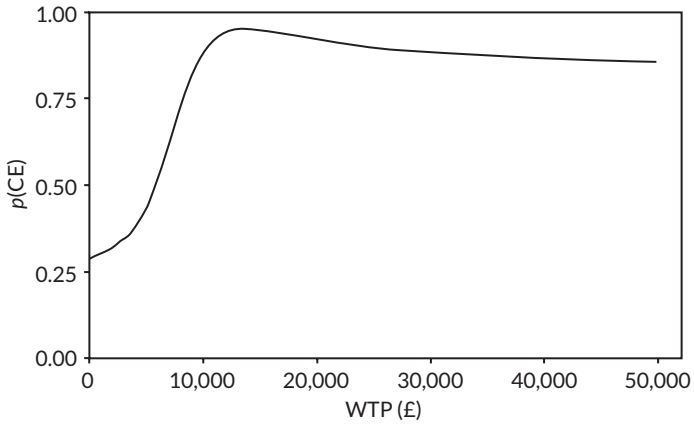

FIGURE 14 Cost-effectiveness plane for base-case model QbTestAll vs. standard (probabilistic analysis), with dashed line showing WTP threshold of £20,000/QALY and dotted line showing WTP threshold of £30,000/QALY.

Scenario and sensitivity analyses for diagnostic assessment

The main scenario analysis results are shown in *Table 31*, which we describe below. For the majority of scenarios examined, the QbTestAll scenario remains cost-effective.

Scenarios relating to parameters for time waiting for assessment

Varying the mean time on the waiting list under standard assessment (scenario 3) has little impact on the cost-effectiveness of QbTestAll, with the ICERs changing by < 8% under the changes in waiting list time examined. Increasing the TR parameter for the impact of increased rate of diagnosis due to QbTest on time spent on the waiting list (scenario 9) reduces both incremental costs and incremental QALYs, slightly reducing the ICER and the INB estimates, but QbTestAll remains cost-effective.

Scenarios relating to parameters for time from assessment to diagnosis

In scenario 2, we explore the impact of varying the HRs for rate of diagnosis for QbTestAll compared to standard assessment. For the HR of children aged 6–12 years (scenario 2a), the cost-effectiveness of QbTestAll is improved, with slightly higher incremental costs and higher incremental QALYs due to patients accessing treatment more quickly on average. When the HR is lower and highly uncertain as in children aged 12+ years (scenario 2b), costs are increased and QALYs are reduced compared to the base case. The mean INB is positive, but there is more uncertainty with only 65.9% or 68.9% of runs cost-effective at £20,000 or £30,000 per QALY WTP thresholds, respectively. We also included a scenario in which we assume the HR is 1 (scenario 2c) such that QbTestAll does not increase the rate of diagnosis compared to standard of care. In this case, incremental costs are higher and incremental QALYs are lower than the base case; however, the mean INB is still positive with probability of being cost-effective of 84.1% or 81.4% at £20,000 or £30,000 per QALY, respectively. This is because, it is still assumed that a higher proportion receives a diagnosis within 6 months for QbTestAll.

Decreasing the proportion with diagnosis within 6 months (scenario 15) under QbTestAll results in negative incremental costs (cost-saving) for both parameter values tested. When $p_{D6,Q} = 0.689$ (the lower CI from AQUA, scenario 15a), the incremental QALYs are positive and so QbTestAll dominates standard assessment, with positive INB, but only 71% or 62% of model runs cost-effective. When $p_{D6,Q} = 0.598$ (scenario 15b), there is no difference between the proportion with diagnosis within 6 months, and the mean incremental QALYs are negative and so the results represent the south–west quadrant, where we require the ICER to be less than −WTP thresholds, which is not the case indicating that QbTestAll is not cost-effective in this scenario. We also see that the INB is negative at both WTP thresholds, and there is only a 20% or 14% probability of being cost-effective at WTP £20,000 and £30,000, respectively.

Scenarios relating to diagnostic test accuracy

In scenario 10, we explored different sensitivity and specificity assumptions. Reduced sensitivity of QbTestAll (scenario 10a) to 0.9 slightly reduces the ICER and the INB values for QbTestAll but does not affect the overall results (90% or 85% of runs are cost-effective). Reducing the specificity of QbTest to 0.9 increases the mean ICER to £10,296/QALY,

TABLE 31 Incremental cost-effectiveness results for QbTestAll vs. standard clinical assessment for the sensitivity and scenario analyses (probabilistic analysis)

QbTestAll (or QbTestUnclear) vs. standard				£20,000 WTP		£30,000 WTP	
Scenario	Incremental costs	Incremental QALYs	ICER	Mean INB	Prob CE	Mean INB	Prob CE
Base case	£238.35	0.0385	£6183.71	£532.55	0.922	£918.00	0.884
1a. Proportion with less clear diagnoses: 0.5	£212.71	0.0385	£5531.25	£556.41	0.933	£940.97	0.895
1b. Proportion with less clear diagnoses: 0.6	£217.35	0.0384	£5655.30	£551.32	0.926	£935.66	0.889
1c. Proportion with less clear diagnoses: 0.7	£222.75	0.0384	£5807.51	£544.36	0.923	£927.92	0.885
1d. Proportion with less clear diagnoses: 0.8	£229.30	0.0388	£5912.07	£546.41	0.924	£934.27	0.887
1e. Proportion with less clear diagnoses: 0.9	£236.91	0.0387	£6114.70	£537.97	0.926	£925.42	0.890
2a. HR for diagnosis rate, QbTestAll vs. standard: 1.84 (1.23, 2.68)	£241.84	0.0432	£5593.45	£622.89	0.947	£1055.26	0.915
Subgroup analysis from AQUA for children aged 6–12 years							
2b. HR for diagnosis rate, QbTestAll vs. standard: 0.82 (0.37, 1.82)	£312.42	0.0248	£12604.07	£183.32	0.651	£431.19	0.689
Subgroup analysis from AQUA for adolescents 12+ years							
2c. HR for diagnosis rate, QbTestAll vs. standard: 1	£256.05	0.0306	£8356.34	£356.78	0.841	£663.20	0.814
3a. Mean waiting time under standard assessment: 3 months	£208.89	0.0367	£5692.11	£525.09	0.903	£892.08	0.860
3b. Mean waiting time under standard assessment: 6 months	£222.91	0.0375	£5947.33	£526.71	0.911	£901.52	0.874
3c. Mean waiting time under standard assessment: 18 months	£259.36	0.0398	£6511.20	£537.31	0.933	£935.64	0.903
4a. Sensor CPT cost (including nurse time): £42.66 (QbTest lower range)	£236.95	0.0390	£6083.24	£542.08	0.924	£931.59	0.887
4b. Sensor CPT cost (including nurse time): £115.66 (QbTest upper range)	£304.57	0.0387	£7862.25	£470.19	0.887	£857.58	0.863
4c. Sensor CPT cost (including nurse time): £40.69 (Nesplora AULA single use)	£217.52	0.0374	£5812.26	£530.97	0.921	£905.22	0.880
4d. Sensor CPT cost (including nurse time): £29.98 (Nesplora AULA quarterly plan for 22 uses)	£220.28	0.0386	£5705.75	£551.85	0.927	£937.91	0.889
4e. Sensor CPT cost (including nurse time): £22.46 (Nesplora AULA annual professional plan, 40 assessments per month)	£210.11	0.0387	£5433.40	£563.29	0.931	£950.00	0.895

TABLE 31 Incremental cost-effectiveness results for QbTestAll vs. standard clinical assessment for the sensitivity and scenario analyses (probabilistic analysis) (*continued*)

QbTestAll (or QbTestUnclear) vs. standard				£20,000 WTP		£30,000 WTP	
Scenario	Incremental costs	Incremental QALYs	ICER	Mean INB	Prob CE	Mean INB	Prob CE
4f. Sensor CPT cost (including nurse time): £13.14 (EF Sim assuming delivery model and costs proposed by company and 15 tests per monthly practice visit)	£204.74	0.0386	£5302.40	£567.52	0.936	£953.65	0.900
5a. Higher response/non-response cost after dose titration period from Zimovetz (2016): responder £170.52; non-responder £325.90	£845.24	0.0382	£22,109.05	−£80.63	0.481	£301.67	0.853
5b. Higher response/non-response cost after dose titration period from King (2006): responder £191.45; non-responder £285.71	£959.50	0.0392	£24,471.75	−£175.33	0.37	−£216.76	0.80
6a. Proportion with no further assessment after no diagnosis within 6 months, p_{missed} = 0	−£676.47	0.0132	−£51,211.14 (Dominates)	£940.66	0.998	£1072.76	0.998
6b. Proportion with no further assessment after no diagnosis within 6 months, p_{missed} 0.25	−£401.94	0.0209	−£19,193.86 (Dominates)	£820.76	0.991	£1030.16	0.980
6c. Proportion with no further assessment after no diagnosis within 6 months, p_{missed} 0.5	−£120.77	0.0289	−£4176.33 (Dominates)	£699.11	0.978	£988.28	0.948
7a. Time horizon 15 years	£385.65	0.0526	£7326.94	£667.04	0.891	£1193.39	0.856
7b. Time horizon 20 years	£483.85	0.0623	£7771.24	£761.39	0.872	£1384.01	0.839
8. Discount rate 0%	£290.25	0.0451	£6440.65	£611.05	0.910	£1061.69	0.875
9a. TR 0.9	£216.46	0.0365	£5929.82	£513.60	0.914	£878.63	0.873
9b. TR 0.95	£193.33	0.0346	£5591.57	£498.16	0.909	£843.91	0.866
9c. TR 1.0	£174.75	0.0331	£5282.59	£486.85	0.904	£817.65	0.860
10a. Diagnostic accuracy = 0.9, = 1	£157.18	0.0316	£4969.54	£475.41	0.898	£791.70	0.851
10b. Diagnostic accuracy = 1, = 0.9	£246.32	0.0239	£10,296.31	£232.15	0.713	£471.38	0.721
10c. Diagnostic accuracy = 0.9, = 0.9	£167.06	0.0220	£7583.93	£273.51	0.749	£493.79	0.726
10d. Diagnostic accuracy = 0.9, = 0.9	£157.54	0.0381	£4131.62	£605.08	0.943	£986.39	0.901

continued

TABLE 31 Incremental cost-effectiveness results for QbTestAll vs. standard clinical assessment for the sensitivity and scenario analyses (probabilistic analysis) (*continued*)

QbTestAll (or QbTestUnclear) vs. standard				£20,000 WTP		£30,000 WTP	
Scenario	**Incremental costs**	**Incremental QALYs**	**ICER**	**Mean INB**	**Prob CE**	**Mean INB**	**Prob CE**
11a. Prevalence of ADHD in those without a diagnosis within 6 months under clinical assessment, $\pi_{No\ D6,\ C} = 0.2$	£229.47	0.0375	£6115.25	£521.01	0.913	£896.24	0.877
11b. Prevalence of ADHD in those without a diagnosis within 6 months under clinical assessment, $\pi_{No\ D6,\ C} = 0.5$	£241.51	0.0389	£6215.75	£535.57	0.925	£924.11	0.887
12a. Carer disutility, $u_{carer-dis} = 0$	£241.48	0.0323	£7485.01	£403.76	0.934	£726.39	0.903
12b. Carer disutility, $u_{carer-dis} = 0.071$	£241.50	0.0580	£4165.87	£917.93	0.881	£1497.65	0.861
13. Waiting list costs for those with ADHD assumed equal to those for non-responding ADHD patients	£177.88	0.0390	£4565.72	£601.33	0.943	£990.94	0.898
14. FP costs post dose titration = £0	£244.20	0.0391	£6253.25	£536.83	0.921	£927.34	0.882
15a. Proportion with diagnosis within 6 months, QbTest, $p_{D6,Q} = 0.689$	−£86.52	0.0035	−£24,709.03 (Dominates)	£156.56	0.710	£191.57	0.624
15b. Proportion with diagnosis within 6 months, QbTest, $p_{D6,Q} = 0.598$	−£497.18	−0.0408	£12,198.13 (South–west Quadrant)	−£317.99	0.201	−£725.58	0.138
16a. Test failure rate 5% who incur test administration cost, but all other costs and QALYs as for standard	£237.49	0.0364	£6520.86	£490.91	0.905	£855.10	0.867
16b. Test failure rate 11% who incur test administration cost, but all other costs and QALYs as for standard	£235.09	0.0338	£6938.35	£442.57	0.878	£781.40	0.846
17a. Utilities for responders and non-responders/not on treatment: 0.827 and 0.773	£237.34	0.0254	£9336.56	£271.07	0.984	£525.28	0.916
17b. Utilities for responders and non-responders/not on treatment: 0.82 and 0.70	£235.61	0.0489	£4814.03	£743.25	0.887	£1232.68	0.862
17c. Utilities for responders and non-responders/not on treatment: 0.926 and 0.905	£239.20	0.0136	£17,570.07	£33.08	0.591	£169.22	0.997

while decreasing the INB such that 71% or 72% of runs are cost-effective. When both sensitivity and specificity of QbTestAll are 0.9, the ICER is £7584 and 75% or 73% of runs are cost-effective. When sensitivity of QbTestAll is 0.9 and specificity of standard assessment is 0.9, QbTestAll becomes more cost-effective than the base case with 94% or 90% of runs cost-effective and a mean ICER of £4131. We conducted a threshold analysis varying the sensitivity of QbTestAll, which shows that INB is positive and increases with sensitivity (*Figure 15*). A breakdown of the costs and QALYs accrued while on the waiting list, under assessment and post assessment (for those that do or do not initiate treatment) shows that, as sensitivity decreases, not only lower treatment costs are accrued but also lower QALYs on treatment, as the proportion of FNs increases (*Table 32*).

Reducing the proportion with no further assessment after no diagnosis within 6 months p_{missed} (scenario 6) drastically impacts the results for the values explored, with incremental costs becoming negative (cost-saving) so that QbTestAll dominates standard assessment. All three values in scenarios 6a–6c make the QbTestAll cost-saving, with 95–100% of model runs cost-effective. This occurs because there are more patients who do not have a diagnosis after 6 months under standard assessment, and if p_{missed} is small, most of these patients will incur the costs of further assessment, leading to higher further assessment costs under standard assessment compared with QbTestAll. The incremental QALYs decrease for QbTestAll compared with standard when p_{missed} is small due to a higher number of ADHD cases being diagnosed after 6 months for standard compared with QbTestAll. A threshold analysis varying this parameter shows that while INB decreases asp_{missed} increases, it remains positive, indicating that the conclusion that QbTestAll is cost-effective is robust to changes in this parameter (*Figure 16*).

The cost-effectiveness results are not sensitive to changes in the prevalence of ADHD in those without a diagnosis within 6 months (scenario 11), also shown by the threshold analysis (*Figure 17*). Increasing the proportion who fail to complete the test slightly increases the ICER and reduced INB, but overall conclusions do not change (scenarios 16a–b).

Scenarios relating to costs

Results were robust to varying the sensor CPT cost using values that represent the range of costs for QbTest or Nesplora (scenario 4).

TABLE 32 Costs and QALYs accrued while waiting for assessment, under assessment and post assessment for those who initiated treatment (TPs and FPs) and those who did not on initiate treatment (TNs and FNs); probabilistic analysis

Strategy	Total costs (discounted)				Total QALYs (discounted)			
QbTestAll sensitivity	Waiting	Assessment	Post assessment: those who initiated treatment	Post assessment: those who did not initiate treatment	Waiting	Assessment	Post assessment: those who initiated treatment	Post assessment: those who did not initiate treatment
0.6	£0	£1263.73	£4468.91	£0	0.6357	0.3874	3.2090	2.6716
0.65	£0	£1263.73	£4566.18	£0	0.6357	0.3794	3.2782	2.6188
0.7	£0	£1263.73	£4649.55	£0	0.6357	0.3726	3.3375	2.5736
0.75	£0	£1263.73	£4721.81	£0	0.6357	0.3666	3.3890	2.5343
0.8	£0	£1263.73	£4785.03	£0	0.6357	0.3614	3.4340	2.5000
0.85	£0	£1263.73	£4840.82	£0	0.6357	0.3568	3.4737	2.4697
0.9	£0	£1263.73	£4890.41	£0	0.6357	0.3528	3.5090	2.4428
0.95	£0	£1263.73	£4934.78	£0	0.6357	0.3491	3.5406	2.4187
1	£0	£1263.73	£4974.71	£0	0.6357	0.3459	3.5690	2.3971

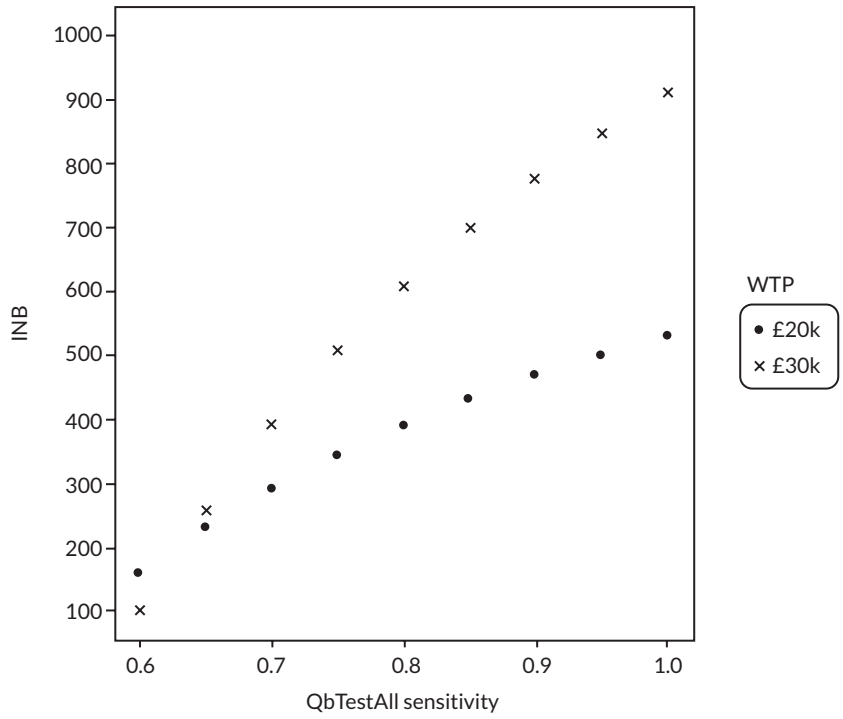

FIGURE 15 Threshold analysis for sensitivity of QbTest plus clinical assessment vs. clinical assessment alone.

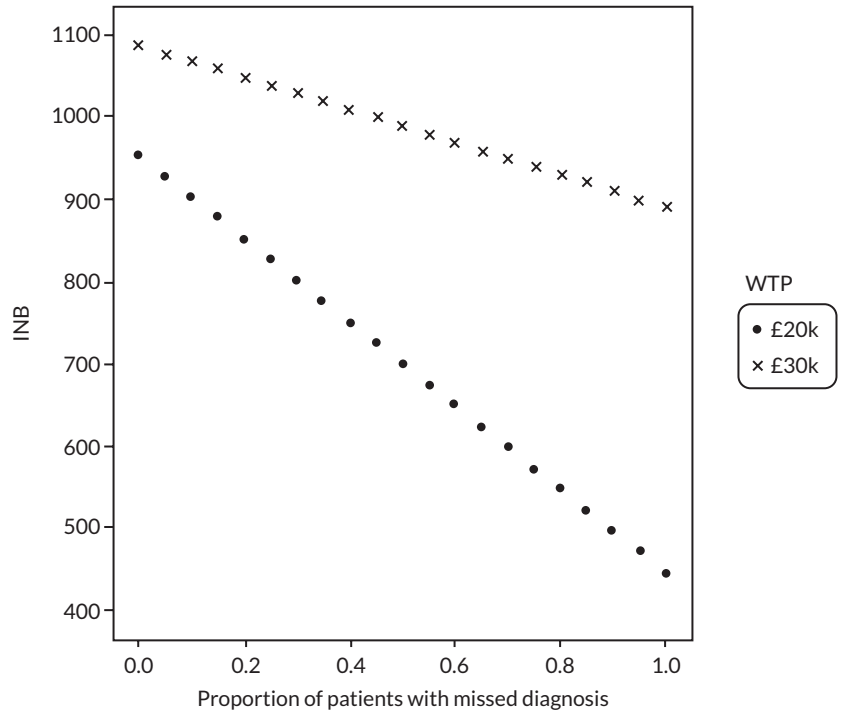

FIGURE 16 Threshold analysis for the proportion with no further assessment after no diagnosis within 6 months. QbTestAll vs. Standard for WTP thresholds of £20,000 and £30,000.

Using the higher response and non-response costs for patients on ADHD treatment (scenario 5) has a large impact on the results. For scenario 5a, using the costs based on resource use reported in Zimovetz (2016),[136] the ICER increased to £22,109/QALY with 48% of runs cost-effective at £20,000/QALY (85% at £30,000/QALY). For scenario 5b, using the King HTA resource use costs,[121] the ICER increased to £24,472/QALY with 37% of runs cost-effective at £20,000/QALY

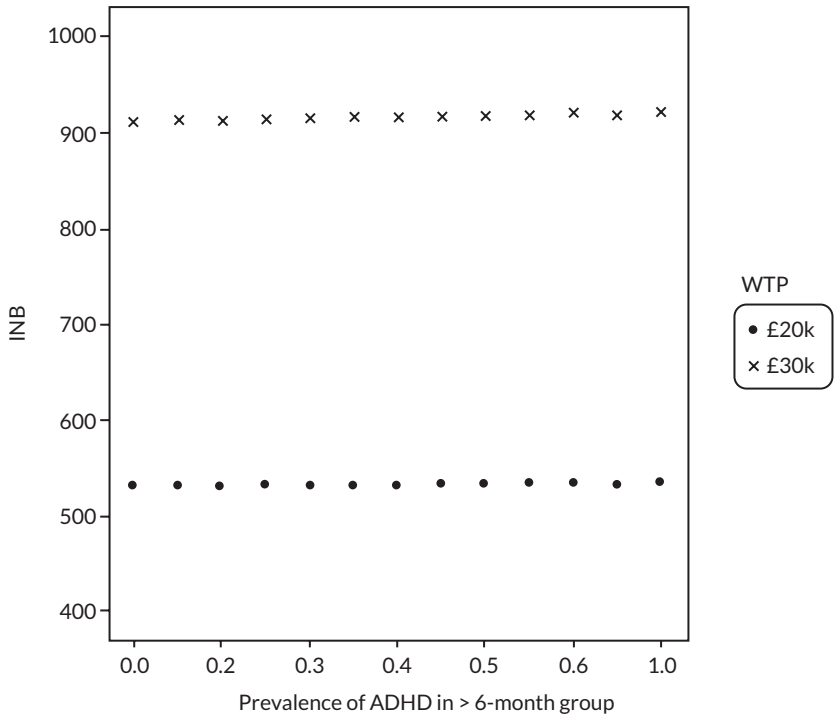

FIGURE 17 Threshold analysis for the prevalence of ADHD in those who do not have a diagnosis within 6 months. QbTestAll vs. Standard for WTP thresholds of £20,000 and £30,000.

(80% at £30,000/QALY). The reason results are so sensitive to these costs is because a higher proportion of patients initiate treatment and start treatment more quickly under QbTestAll and incur these costs.

If patients on the waiting list with ADHD are assumed to have resource use and costs equivalent to non-responding ADHD patients (scenario 13), then the cost-effectiveness is increased slightly with 94% or 90% of runs cost-effective. The results were also not sensitive to the removal of FP costs after dose titration (scenario 14), with this changing the ICER by no more than 1%.

Increasing the time horizon to 15 or 20 years (scenario 7) increases both incremental costs and incremental QALYs, but QbTestAll remains cost-effective. Using a discount rate of 0% slightly increases the ICER by 4% (scenario 8).

Scenarios relating to utilities

Results were robust to varying the utilities for responders and non-responders/not on treatment (scenario 17). When the difference in utility between responders and non-responders was small (scenario 17c), QbTestAll had a small increase in the mean INB of £33.08 and the probability of being cost-effective was only 0.59 at the £20,000 WTP threshold. Scenario 17c was a scenario used by Zimovetz (2016),[136] representing the values from a trial of LDX versus ATX in patients who had an inadequate response to MPH, and it was calculated using the Health Utilities Index Mark 2, whereas our base-case uses utilities were calculated using EQ-5D, did not restrict to those with an inadequate response to MPH and were based on a larger sample size with more precise estimates.

Removing the carer disutility (scenario 12a) reduces the QALYs gained and increases the mean ICER to £7,485, while increasing the carer disutility to 0.071 increases the QALYs gained and the mean ICER is reduced to £4166 (scenario 12b). In both cases, the overall result is similar to the base case.

Scenario for Objective 2: QbTestUnclear

Scenario 1 examines the QbTestUnclear scenario to address objective 2, in which only a proportion of patients with unclear diagnosis receive QbTest. Due to QbTest being used for only a subset of patients, the QbTestUnclear scenario is slightly more cost-effective (lower ICER, higher INB) than the base-case QbTestAll scenario. For example, if only

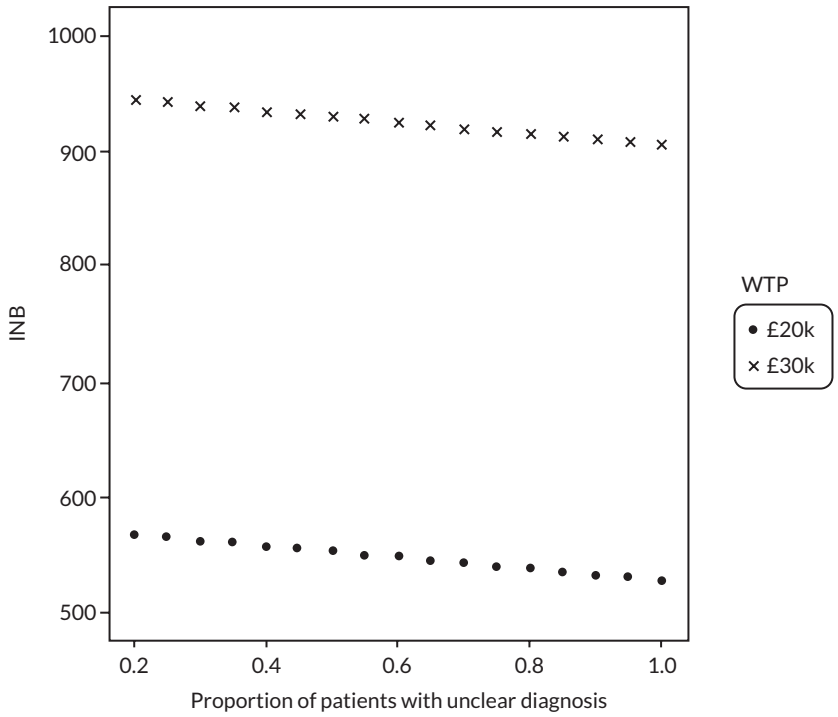

FIGURE 18 Threshold analysis for proportion with less clear diagnoses in whom QbTest is administered for objective 2. QbTestUnclear vs. Standard for WTP thresholds of £20,000 and £30,000.

50% of people receive QbTest, the mean INB at £20,000/QALY increases to 556.41 with 93.3% of model runs cost-effective. As the proportion of patients who receive the QbTest decreases, the INB increases (*Figure 18*). Note, however, that this scenario assumes no impact on diagnosis rates or other parameters than test cost and so needs to be interpreted accordingly.

Chapter 6 Assessment of factors relevant to the National Health Service and other parties

Due to the subjective nature of diagnosis for ADHD, there may be concerns regarding the reliability and consistency of the diagnosis,[17] which can lead to appeals which are time-consuming for all involved. A potential benefit of sensor CPTs is that they may lead to a lower proportion of cases being appealed. This potential benefit has not been directly captured in the economic modelling in this report.

The economic model estimates that there would be a higher proportion of those referred for assessment who initiate treatment when sensor CPTs are used due to lower numbers without any diagnosis. This may have implications for the availability of pharmacological medication.

To administer the sensor CPTs, a private and quiet room with a computer, desk and chair would be needed and staff would need to undergo training in order to be able to administer sensor CPTs.

If sensor CPTs were to be used for dose titration and long-term monitoring of treatment where appointments are held remotely, administration of the sensor CPT would need to be held in person in advance of the remote appointment so that the results are in place to inform the assessment. Similarly, for use in diagnosis for adults where assessments are typically conducted remotely.

Qualitative data suggested some concerns with the length and repetitive content of the QbTest, it may be that other tests are more interactive and engaging for patients. This should be explored further when making a decision regarding which sensor CPT to recommend.

Chapter 7 Discussion

Statement of principal findings

Therse were limited data on the clinical effectiveness of sensor CPTs for diagnosing ADHD. The majority of the evidence was focused on objective 1 (diagnostic accuracy and clinical effectiveness and cost-effectiveness of technologies that combine measures of cognition and motor activity for the diagnosis of ADHD in people referred with suspected ADHD), with some evidence for objective 3, but no data to address objectives 2 or 4. Most evidence was for the QbTest, mostly the in-person versions (6–12 and 12–60) with single studies on QbMini (the version for children aged 4–5 years) and the online QbCheck. There were two studies of EF Sim and two of Nesplora Kids – there were no studies of EF Sim web or of Nesplora Adults. Overall, the limited data suggest that diagnosis with QbTest is likely to have similar accuracy compared to diagnosis based on clinical information alone, with some evidence of improvements in the number of consultations required to make a diagnosis. These findings are based primarily on the AQUA trial, which had some methodological limitations. There are insufficient data on the EF Sim or Nesplora tests.

Only one small feasibility study provided information on the impact of the QbTest on clinical outcomes. However, due to the feasibility design, small size and very low FU, impacted by the COVID-19 pandemic, it was not possible to draw conclusions regarding clinical effectiveness from this study.

Data on the accuracy of the tests were also very limited, particularly in combination with clinical assessment, which is how the test is intended to be used in practice. Overall, the populations enrolled in DTA studies varied, with nine studies being conducted in adults (one of these focused on older adults) and eight in children (five studies in 6–12 years; one in 5 years; one in 5–15 years and one in children but age not specified). Two studies enrolled children and adolescents (12–18 years), one adolescents and one adolescents and adults. Most studies included more male than female participants, particularly in the ADHD groups, although five studies included slightly higher proportions of female participants. Studies were conducted almost exclusively in Europe. Most studies reported baseline data on at least one PROGRESS-Plus characteristic, most commonly sex (18 studies), neurodevelopmental/learning disorders (13 studies) and mental health disorders (11 studies). Some studies also reported on education (8 studies), occupation (4 studies), socioeconomic status (3 studies), features of relationships (3 studies), place of residence (2 studies) and ethnicity (1 study). However, we have highlighted that the reported data were not stratified by these characteristics, so we could not explore test accuracy within specific population subgroups.

Only five studies evaluated the accuracy of the QbTest in combination with clinical information, and only one of these (the AQUA trial) evaluated the accuracy in combination with clinical judgement. All others used prediction models to combine data from specific clinical measures with QbTest results – this is unlikely to reflect how the test would be used in practice. Data from the AQUA trial were limited as the diagnostic substudy was restricted to children in whom a diagnosis was made at 6 months, resulting in the exclusion of 80/250 children. It is likely that the restricted population may have represented a more 'easy–to-diagnose' population, as more complex cases may have been more likely to withdraw from the study or to have been discharged without a diagnosis. However, as there was no information available on these participants, this is difficult to judge. There were also limitations in the reference standard. This consisted of independent consensus criteria based on the DAWBA criteria, which is considered as an accepted reference standard. However, in 123/241 participants, DAWBAs were missing from one informant (i.e. either parent or teacher), meaning the independent assessors did not have access to this information when making a diagnosis. This is likely to have resulted in an underestimate of specificity and possible overestimate of sensitivity as the reference standard will have failed to diagnose some cases, and these may have been more likely to be complex cases. This is supported by the results, as the estimated specificity was very low (40%, 95% CI 25% to 56%) for this study. There is therefore no reliable data on the accuracy of any of the sensor CPTs when used in combination with clinical judgement.

Estimates of the accuracy of the sensor CPTs alone were heterogeneous, and so results should be interpreted with caution. Summary estimates of the accuracy of the QBTest suggested that sensitivity was highest when the

subcomponents were combined into an overall measure (summary sensitivity 79%, 95% CI 69% to 86%), but specificity was lower (summary specificity 59%, 95% CI 42% to 74%) than when subcategories were assessed individually. There was little evidence of a difference between the accuracy of the three subcategories of activity, impulsivity and inattention. The single study of the QbMini suggested that this test had very poor discriminatory ability, but this is based on a single study which was judged at high risk of bias. The single study that evaluated the QbCheck suggested that this was at least as accurate as the in-person version of the test, but this study was judged at high risk of bias and so results should be interpreted with caution. The single studies of Neslora Kids and EF Sim also suggested that accuracy was similar to that of the QbTest, but this was based on very limited information from studies at high risk of bias and no direct comparisons between tests were available.

Three studies provided a direct comparison between non-sensor CPT and sensor CPTs (two of QbTest and one of EF Sim), one study (the AQUA trial) provided a direct comparison between clinical diagnosis combined with QbTest with the accuracy of clinical diagnosis alone and one compared the accuracy of the QbTest alone to the accuracy of QbTest plus clinical information. There were no consistent results to suggest that the accuracy of QbTest or EF Sim differed from that of standard CPT. One study reported that an overall measure from EF Sim was more sensitive than the non-sensor CPT omission errors measure ($p = 0.03$), but it was less specific ($p = 0.07$). There was no difference between the overall EF Sim measure and the other two CPT measures. One study reported that Qb measures were more sensitive ($p \leq 0.01$) but less specific than the two Connors' CPT measures, while the other reported that the QbTest was less sensitive ($p < 0.01$) with no difference in specificity. The AQUA trial compared QbTest plus clinical judgement to a control group using the standard diagnostic process. The two groups had very similar specificity ($p = 0.80$), but sensitivity was slightly higher in the clinical diagnosis alone group compared to the group where diagnosis incorporated the QbTest (96% vs. 86%), but there was no statistical evidence of a difference between groups ($p = 0.14$). A study in older adults presented a comparison between models based on the QbTest alone and a model that incorporated a clinical measure of ADHD symptoms. The model that incorporated the clinical information was much more sensitive than the QbTest alone (91% vs. 56%; $p < 0.01$). There was no evidence for a difference in specificity ($p = 0.11$).

Five studies evaluated the impact of the QbTest on process measures. The AQUA trial randomised children to be assessed for ADHD with or without the QbTest as part of the diagnostic process. This study was judged at high risk of bias for time-to-event outcomes as a large proportion of participants (80/250) were uninformatively censored from the analysis, as they dropped out or were discharged from the clinic and so did not have a diagnosis at 6 months. It was at low risk of bias for other outcomes, except cost of clinic appointments, where risk of bias was judged to be unclear. It is likely that this is reflective of what would happen in practice, but no details were available of the proportion of those that were censored who dropped out and what proportion was discharged without a diagnosis. It is also unclear why participants were discharged without a diagnosis and what the next steps would be for these children. The other four studies were retrospective record reviews, where data for those evaluated for ADHD prior to implementation of the QbTest were compared to data for those evaluated after the implementation of the QbTest. The largest of these studies, Focus ADHD, was affected by the COVID-19 pandemic as the QbTest was implemented over the same period as the pandemic. All four studies were judged at serious risk of bias as none adjusted for potential confounding factors. These studies also had other methodological limitations, including lack of detail on how children were selected for inclusion in the assessments and very limited numerical and statistical data. Results from these studies should therefore be interpreted with extreme caution.

The AQUA trial reported a number of benefits associated with adding QbTest to the diagnostic process, including fewer appointments to reach a diagnosis, reduced consultation time, increased proportion of patients with a diagnosis, greater clinician confidence in the diagnostic decision and exclusion of the diagnosis in a greater proportion of children. They also reported that costs of clinic appointments were less in the QbTest arm compared to the control arm. The AQUA trial findings were supported by the limited data from the before–after studies, which found that following implementation of the QbTest, fewer consultations were required to reach a diagnosis. These studies also reported other benefits included reduced time to reach a diagnosis (two studies) and reduced costs of testing. Focus ADHD reported an increased time to make a diagnosis, fewer children having school observations as part of the diagnostic process and fewer patients with an ADHD diagnosis, but these data are likely to have heavily confounded by the COVID-19 pandemic and so are unlikely to be reliable.

Eight studies provided data on the clinician and/or patient and carer views of sensor CPTs for the diagnosis of ADHD. Most of the studies were judged to have some concerns of risk of bias due to a lack of detail reported about the methodology used. Five evaluated the QbTest through interviews, surveys or focus groups. Findings were in line with process measures data; clinicians felt it increased confidence in clinical decision-making, and both clinicians and families felt it may reduce the time to diagnostic decision. Clinicians and families also felt that the test helped to improve communication, although, some families felt that the test results were not properly explained to them and did not help them to understand symptoms or how diagnoses were made. Barriers to implementation included staffing, training and technology requirements. Patients and caregivers highlighted concerns with the length and repetitive content of the test, and staff in one study reported that patients struggled with sensory discomfort and stress during the test. One study of QbCheck reported that participants found it easy to use; however, this was from a brief three-question survey conducted as part of a DTA study. Additionally, two survey studies evaluated the EF Sim. Of these, one study, funded by the test manufacturer, reported positive findings concerning acceptability for teachers (confidential information has been removed) who had implemented the test. The other study also reported positive acceptability from a short survey to children who had used the test in a DTA study.

We did not identify any previous models evaluating the cost-effectiveness of diagnostic tests for ADHD, and so we developed a de novo model for sensor CPTs for the diagnosis of ADHD in people referred with suspected ADHD (objective 1). We only evaluated the QbTest in addition to clinical assessment versus clinical assessment alone due to lack of evidence on the inputs needed for our model for other sensor CPTs. We found that QbTest in addition to clinical assessment is likely to be cost-effective, with incremental costs of £238.35 and incremental QALYs of 0.0385 per person evaluated for ADHD. The resulting ICER is £6183 per QALY gained, which is cost-effective at a WTP threshold of £20,000 per QALY. The mean INB (probability of being cost-effective) is £532.55 (92%) and £918 (84%) at WTP of £20,000 and £30,000 per QALY, respectively. These findings were driven by reduced time waiting for assessment, reduced appointments until diagnosis and a higher proportion receiving a diagnosis so that more patients with ADHD receive treatment benefits.

Due to data limitations, we made several assumptions in our model, which we tested with a wide range of scenario analyses. We found that our overall conclusions were robust to most of our modelling assumptions. However, if the state costs for responders/non-responders on treatment were assumed to be higher, then QbTest in addition to clinical assessment would not be cost-effective due to the higher proportion who initiate treatment and incur the higher costs. Also, if the proportion of patients with a diagnosis within 6 months for QbTest in addition to clinical assessment is lower (closer to that for clinical assessment alone), then QbTest in addition to clinical assessment becomes cost-saving but also incurs lower or even less QALYs than clinical assessment alone. In this scenario, the cost savings do not justify the quality-of-life reductions.

As we did not identify any relevant studies for objective 2, we were unable to properly model the impact of sensor CPTs for the diagnosis of ADHD in people referred with suspected ADHD for whom current assessment cannot reach a diagnosis. We ran some exploratory analyses that demonstrated that if there are no consequences in terms of diagnostic accuracy, then using sensor CPTs on the subset of those where a diagnosis is not reached after one or two appointments would be more cost-effective than using sensor CPTs on all patients, because the test cost is incurred for only some patients.

Six studies provided data for objective 3; all evaluated the QbTest. One DTA study evaluated the accuracy of QbTest as part of dose titration to against the reference standard of ‘good outcome’ at 1-year FU. However, the QbTest formed part of the reference standard that is likely to overestimate the accuracy of the test and so it is not possible to draw strong conclusions from this study. One study (the QUOTA trial) provided data on process measures; however, it was a small feasibility trial that was not designed and powered to formally evaluate the impact on outcomes. Three RCTs (the AQUA trial and two feasibility RCTs: FACT and QUOTA) and two implementation studies provided interview or survey data on the patient/clinician views of the QbTest for medication management and dose titration. Most of the studies had concerns regarding quality due to lack of information reported on study design. Findings suggested that healthcare staff and families mostly valued the role of the test for dose titration, checking medication utility and improving medication adherence. However, some studies reported survey data from patients to suggest that the results of the

QbTest may not have helped them to understand medication decisions, and some clinicians highlighted that using the QbTest for medication management can present logistical challenges due to having to schedule more appointments.

Due to the limited data on clinical effectiveness for objective 3 and lack of data for objective 4, we did not have sufficient evidence to model the impact of sensor CPTs for dose titration and treatment decisions or long-term treatment monitoring for people with a diagnosis of ADHD.

Strengths and limitations of the assessment

Systematic review strengths and limitations

Our systematic review followed published guidance on the conduct of systematic reviews of DTA studies[35] and is reported according to PRISMA-2020 guidance[38] and PRISMA-DTA guidance,[170] making our review processes transparent and robust. The protocol was pre-registered on the PROSPERO database (CRD42023482963). The only changes that we made to the protocol were to clarify that, although we had specified that studies need to be conducted in secondary care or remote settings, we would also include studies in which some participants (e.g. control groups) were enrolled in other settings. The other change was to broaden our inclusion criteria for comparative studies to also include data from studies that compared the accuracy of sensor CPTs (alone or in combination with clinical diagnosis) with the accuracy of clinical diagnosis alone. We also included one study of the QbMini test despite this not being explicitly mentioned in the protocol, as this test is a version of the QbTest but for very young children (aged 5 years). We conducted extensive literature searches designed to maximise retrieval of relevant studies and did not apply any language, date or publication restrictions to these searches or to inclusion in the review. We identified one study reported only as an abstract in Spanish, and all other studies were reported in English. We used Google Translate to translate the Spanish abstract, and we asked a native Spanish language speaker to verify the accuracy of the translation. We pre-specified clearly defined, objective inclusion criteria. These specified that studies should be conducted in a population with suspected ADHD. We interpreted this broadly such that studies that used a multigate design incorporating patients with known ADHD and a group of controls without ADHD (either with another condition or healthy controls) were also included. We conducted a formal assessment of the risk of bias of included studies using the RoB 2 tool for RCTs,[44] the ROBINS–I study for non-randomised studies,[46] the QUADAS-2 tool for DTA studies,[45] its extension QUADAS-C[171] for comparative accuracy studies, the CASP checklist for qualitative studies and the Q-SSP tool for survey studies. Our synthesis included a meta-analysis where more than one study evaluated the same test. We stratified out analyses based on test, test component and age. There were insufficient data to formally investigate heterogeneity or to look at the impact of other study features such as quality on estimates of accuracy. We did not include a formal assessment of publication bias due to the small number of included studies and due to the difficulties in assessing publication bias for DTA studies where there is no clear threshold for 'significance'. Our synthesis also included formal synthesis of qualitative and survey data to supplement the more formal quantitative evaluations. We used the meta-aggregative approach to qualitative synthesis based on guidance from the JBI to synthesise data from qualitative studies. Using a mixed-methods approach in our review allowed us to add contextual insights from the qualitative data to help understand the findings from the quantitative studies on process measures.

Limitations of the evidence base

The evidence based for this assessment was limited. The most relevant study for our appraisal was the AQUA trial, both in terms of the accuracy data and the information on process measures. However, as highlighted above, this study had methodological limitations both for the main trial and for the diagnostic substudy. There were no good-quality data on the EF Sim or Nesplora tests.

There was very limited RCT data – we only identified three RCTs across the four objectives, and two of these were small feasibility studies that were not powered to assess clinical effectiveness. Although we identified a relatively large number of DTA studies, most were at high risk of bias, and only the AQUA trial evaluated the test in the context that it would be used in clinical practice. The majority of DTA studies used a multigate design that is likely to lead to overoptimistic estimates of accuracy. A challenge in this area is the identification of an appropriate reference standard, particularly for the evaluation of sensor tests in combination with clinical practice. We considered a diagnosis based

on DSM-IV or DSM-V or ICD-10 criteria to be an appropriate reference standard. However, most diagnoses made in clinical practice adhere to these criteria, making it difficult to assess the accuracy of sensor CPT in combination with clinical diagnosis. The AQUA trial used independent assessment by two experienced child psychiatrists based on the DAWBA to make a diagnosis. This combines a range of data, including parent interviews, interviews with the young person, a teacher questionnaire and a computer-assisted clinical diagnostic rating to generate an ICD-10 or DSM-V diagnosis. The use of two independent raters to confirm the DAWBA diagnosis is an attempt to separate the reference standard diagnosis from the routine clinical diagnosis using multiple experienced assessors to make this more robust. This is an appropriate reference standard; however, in the AQUA trial, the missing information from one informant for more than half of participants means that it cannot be considered as a gold standard diagnosis in this trial. A further limitation with the AQUA trial was that it was restricted to those in whom a diagnosis was made by 6-month FU. This lead to the exclusion of a large proportion of participants. Very few of the other DTA studies provided any information on whether any of the participants were missing a diagnosis. The multigate design will, by the nature of the design, have been restricted to those in whom a diagnosis was made. However, it is possible that other one-gate studies were also restricted to those with a diagnosis even though this was not explicitly reported.

The QbTest does not specify a threshold to define a positive test result or provide explicit guidance on how results from the different subcomponents should be combined to create an overall diagnosis of ADHD. This means that studies had to define their own threshold and define how to combine subcomponents to create an overall measure of ADHD. There was therefore some variation in thresholds reported across studies, and as many studies did not pre-specify the threshold used, it is possible that data-driven thresholds selected to optimise sensitivity and/or specificity may have been used. This has the potential to introduce bias. Studies also used different methods to derive an overall QbTest results – three studies, all by the same authors, defined a measure based on qualitative analyses of raw scores from the different QbTest subcategories, others used a mean of the three subcategory scores. Where studies combined the QbTest with clinical information, most did so based on prediction models that combined QbTest subcategory results with specific clinical scales. The AQUA trial allowed clinicians to make their own diagnosis based on the full results of the QbTest and their clinical assessment. This is reflective of how the test is likely to be used in practice, but it is difficult to standardise to allow comparison of accuracy across different studies.

The AQUA trial and the before–after implementation studies provided important information on process measures. However, all studies were restricted to children and so it is not clear whether similar results would be obtained in adults. They also included broad, general population and so it is not possible to determine whether similar results would be obtained in specific subpopulations such as those with comorbidities, including other neurodevelopmental conditions such as autism. The largest of the implementation studies, Focus ADHD, was severely impacted by the COVID-19 pandemic which coincided with the period in which the QbTest was implemented, making it very difficult to interpret results on measures such as number of appointments and waiting times. All implementation studies were judged at high risk of bias, mainly due to lack of adjustment for confounding. The numerical results data reported by the implementation studies were limited in most studies, and few provided formal statistical comparison of results or reporting data such as means and SDs, which would have allowed us to compare between groups.

Other studies were also limited by poor reporting. We contacted the authors of nine studies with requests for additional data where information was lacking or difficult to understand in the study reports, with five providing further information. However, four did not respond and so we are limited to the data reported in the study reports. Three of these were reported only as abstract and so very limited data were available for these studies,

There was no good-quality data on the clinical effectiveness of the use of the QbTest for dose titration; there were one data on the accuracy of QbTest, but results from these were difficult to interpret as the QbTest formed part of the reference standard. All other data were qualitative or survey data and that suggested some benefits and challenges of using the QbTest in this role, but high-quality quantitative studies are needed to assess the clinical effectiveness of using QbTest for dose titration and treatment decisions.

Economic model strengths and limitations

This is the first economic model developed to evaluate the cost-effectiveness of diagnostic tests in people referred with suspected ADHD. We capture the time waiting for assessment, initial period of assessment, further assessment for a

proportion of those without diagnosis following the initial period of assessment, diagnostic accuracy and initiation of pharmacological treatment in those diagnosed with ADHD. We populated the model using evidence identified in our clinical effectiveness review, our review of cost-effectiveness studies of diagnostic tests for the assessment of ADHD and previous cost-effectiveness models of treatment for ADHD and using targeted searches for specific inputs required in the economic model. Despite the comprehensive search for model inputs, there was a lack of evidence for some of the assumptions and inputs to our model, which we outline below.

The health economic model for the use of sensor CPT in diagnosis of ADHD was only able to include the QbTest CPT and largely relied on data from a single study (the AQUA trial) for the impact of the addition of QbTest to clinical assessment. The AQUA trial recruited children and adolescents from a mix of CAMHS (48%) and community paediatric clinics (52%) in England[18] and so our results are applicable for patients referred through these routes with a similar case-mix. In a scenario where we used the HR for diagnosis specific to adolescents, there was a reduction in INB and an increase in the ICER; however, the addition of QbTest to clinical assessment was still found to be cost-effective. This does rely on all other model inputs being unchanged for adolescents, in particular the proportion who receive a diagnosis within 6 months, which is a big driver of the cost-effectiveness results. We were unable to model the use of sensor CPT in the diagnosis of ADHD in adults due to a lack of evidence on use of sensor CPTs in this context. Due to the difference in diagnostic assessment between adult and paediatric services, with adult assessment taking place remotely in one extended session, we did not consider that the results from the AQUA trial could be applied in the adult setting. There was insufficient evidence to conduct subgroup analyses for sex, ethnicity, people with mental health, behavioural and neurodevelopmental conditions, people with developmental trauma, looked-after children or people in the Youth Justice System or Adult Criminal Justice System.

We assumed that the sensor CPT would be administered just once during the assessment process, but it is possible that it could be administered again in cases where diagnosis remains unclear after several appointments. In the East Midlands AHSN study,[113] one of the sites implemented QbTest in complex cases only, and we heard from our clinical advisors that this is how sensor CPTs may be used in practice. Due to limited data, we were only able to explore the cost-effectiveness of sensor CPTs used for complex cases only by making a strong assumption that sensor CPTs would be used for those where a diagnosis was not made in two appointments (including the initial appointment) and the benefits seen in the AQUA study were generated by those who had more than two appointments so that the findings would not change if QbTest were only administered after two appointments. We found that if this were the case, then QbTest in addition to clinical assessment became more cost-effective due to reducing the number of patients for whom it is administered. While this analysis was exploratory and makes assumptions, we hypothesise that using QbTest for those where diagnosis is unclear is likely to be cost-effective.

Waiting times for assessment can be long, and vary across regions, and we had to make assumptions about this. We found that QbTest was cost-effective across the range of mean waiting times we varied (even when we assumed no impact on waiting time), but it was more cost-effective when waiting times were longer due to the impact of QbTest on reducing waiting times.

As noted above, issues with the reference standard in the AQUA study meant that there was high uncertainty of the diagnostic accuracy of QbTest in addition to clinical assessment compared with clinical assessment alone. We assumed in the model that clinical assessment alone was a gold standard and explored different assumptions on the diagnostic accuracy of QbTest in addition to clinical assessment. We found that results were robust to assumptions on test accuracy, but they were driven by the proportion who received a diagnosis within 6 months, which was assumed higher for QbTest in addition to clinical assessment based on the findings of the AQUA trial. If there are no differences in the proportion who receive a diagnosis within 6 months, then QbTest in addition to clinical assessment is not cost-effective compared with clinical assessment alone. We also had to make assumptions about outcomes for those that did not receive a diagnosis within 6 months, including the prevalence of ADHD in this group and the proportion who undergo further assessment, and eventually a diagnosis is reached, which represents further uncertainties in the model results.

Patients for whom ADHD is excluded, or not diagnosed, may go onto have further assessments for other conditions, or they may appeal the diagnosis and undergo further assessment for ADHD. Our model does not capture this, although we base the number of appointments after 6 months on audit data from the Focus ADHD study.[31]

We identified three different sources for the post-titration costs incurred by responders and non-responders to treatment, and our results were sensitive to which we used. We preferred the figures used in the NICE guideline CG87, which give an ICER of £6184/QALY, but if the costs based on Zimovetz (2016)[136] are used, then the ICER increased to £22,109/QALY, and if the costs based on King HTA[121] were used, then the ICER increased to £49,079/QALY which is not cost-effective at normal WTP per QALY thresholds. This is a key uncertainty in the model.

To administer the QbTest, a private and quiet room with a computer, desk and chair is needed, but we did not include additional costs for space; but, note that appropriate space has to be available, which may be an issue for implementation. We also did not include costs of time spent completing training for QbTest, as it is a start-up cost that is not allocated per patient treated, but time will be required for staff to complete the training.

Our model took an NHS PSS perspective and so did not include the impact of sensor CPTs on education services or educational outcomes. However, reducing school visits to collect evidence was found to be a benefit in the East Midlands AHSN study,[113] which may reduce the burden on schools to provide reports. Appropriate diagnosis and treatment of ADHD are expected to have benefits on educational attainment,[172] forming and maintaining relationships and self-esteem[8,9] and wide-ranging long-term outcomes, including social function, education, criminality, alcohol use, substance use and occupational outcomes,[10,11] which were not captured in our model.

We used a 10-year time horizon, which only captures the period that children are managed within paediatric services. However, benefit of appropriate diagnosis continues into adulthood, and this benefit is not captured in our model. While we acknowledge that it is important to capture lifetime costs and benefits, this would have meant extrapolating very short-term data into the long term. Nearly all previous treatment models for ADHD used a 1-year time horizon, with just a few models using a 10-year time horizon, and so our model is in line with the longer of these.

Uncertainties

A key uncertainty, affecting both the clinical effectiveness and cost-effectiveness reviews, is the accuracy of the QbTest in combination with clinical judgement. Data from the AQUA trial suggest that these are equivalent, but this is based on a single study judged at high risk of bias. The accuracy of the EF Sim, Nesplora Attention and web versions of the sensor CPT in combination with clinical judgement has not been evaluated and so remains a key uncertainty. There are also insufficient data on the accuracy of any sensor CPT for medication management and so the clinical effectiveness of these tests in this role is unclear. There were also no data for any of the sensor CPT in subgroups of patient, such as sex, ethnicity, people with mental health, behavioural and neurodevelopmental conditions, people with developmental trauma, looked-after children or people in the Youth Justice System or Adult Criminal Justice System. Whether the tests perform differently in any of these subgroups remains a key uncertainty.

Another important area of uncertainty is the relative accuracy of sensor and non-sensor CPT for diagnosing ADHD. Limited data included in the review suggest that accuracy may be similar, and as non-sensor CPTs are likely to be less costly than sensor CPT, it is possible that it may be more cost-effective to use non-sensor CPT. However, this would depend on whether the non-sensor CPT also have the benefits associated with sensor CPT such as fewer appointments to reach a diagnosis, reduced consultation time, greater clinician confidence in the diagnostic decision, exclusion of the diagnosis in a greater proportion of children and improved communication. Evaluation of non-sensor CPT was beyond the scope of this appraisal, as we only evaluated data on non-sensor CPT when a direct comparison was made with a sensor CPT.

All data on the impact of sensor CPT on process measures were for the QbTest and were in children; it is unclear whether similar results would be seen in adults and for other sensor CPT. Given the differences between the diagnostic pathways between adults and children, it is possible that the QbTest would affect process measures in different ways for these different groups. Limited data from the AQUA trial suggested that effects on time to diagnosis may be greatest in younger children (ages 7–12 years) than in adolescents.

Key uncertainties driving the cost-effectiveness results were related to resource costs for patients who do or do not respond to treatment and the proportion of patients who do not receive a diagnosis following an initial period of assessment (6 months). The AQUA trial found a higher proportion of patients who received a diagnosis for QbTest in addition to clinical assessment compared with clinical assessment alone, but it is unclear what is driving these differences and if they would be seen in practice to the same degree. Those patients without a diagnosis were a mixture of those who were 'lost to clinic' and those who were discharged, but it was unknown what proportion of them would return for further assessment at a later date, and the prevalence of ADHD in those that do and do not undergo further assessment was also not known. The Focus ADHD study[31] showed that there are a proportion of patients who do undergo further assessment beyond six appointments, but unfortunately due to the COVID pandemic, this study does not provide reliable data on the impact of QbTest on this.

While we did not find evidence on the use of sensor CPTs in those patients where diagnosis is unclear, it is likely that use of QbTest in those where diagnosis is unclear is cost-effective compared with the standard clinical assessment. However, we are uncertain how the cost-effectiveness of using QbTest in addition to clinical assessment in all patients would compare with just using it in those where diagnosis is unclear.

Equality, diversity and inclusion

Our research was based on existing literature and so we had no control over the participants enrolled. We were broad in our inclusion criteria such that studies from any country and in any language of publication were eligible. We had intended to investigate how the accuracy of included tests varied across different populations, but there were insufficient data to allow us to do this.

Our team included researchers with a broad range of experience and expertise. The lead authors are junior researchers within Bristol Technology Appraisal Group, who were given the opportunity to lead on the writing of this report to help develop their research skills and portfolio. They were supported by the two senior authors, who provided advice and mentorship to the junior researchers leading on the reviews and health economic modelling. The team included those with expertise in systematic reviews, health economics and medical statistics.

Patient and public involvement

We involved two patient representatives with lived experience of ADHD in this project. One of the coauthors also has recent lived experience of the diagnostic process for ADHD and the QbTest as her son has been evaluated with the QbTest (6–12). They attended team meetings (one at the beginning of the project and one closer to the end of the project), gave feedback on the plain language summary report and wrote the section below about the difference sensor CPT may have for patients with ADHD. Involvement of patients had a positive impact on this project, as they also contributed to the section on research priorities.

Impact on patients

The process of gaining a diagnosis of ADHD, whether for your child or yourself, can be complex, lengthy and difficult to negotiate. Therefore, any improvements to the diagnostic pathway are very welcome. However, it is important to us that any changes to the current process are based on robust evidence of effectiveness as well as being acceptable to patients/carers and are valued by the clinical team. We appreciate the careful work the academic team have put into reviewing the evidence and are disappointed that there is not more robust evidence about their effectiveness and acceptability.

Speed of diagnosis is important to us, but accuracy is the most important factor so that people can be supported throughout their lives. Additionally, we feel that cost-effectiveness may reduce waiting times for diagnosis (which can

be considerable on the NHS – one patient representative waited for 4 years) and give more people access to diagnosis. The wait between referral and assessment can be a stressful, uncertain time. Likewise, we feel that support with dose titration could be valuable – we are not aware of any formal clinical process for measuring the effectiveness of medication, and people often do not know what to expect or how to really tell whether it is working. We are hopeful that the Qb/other systematic testing tools might contribute to the better detection and timely treatment of ADHD in the future and whole-heartedly support the recommendation of the review team that proper evaluation, including the cost-effectiveness to the NHS, is an important next step.

Chapter 8 Conclusions

Implications for practice

There was a lack of good-quality data on all tests, both for diagnosis and medication management, particularly when evaluated in combination with clinical information. Our results suggest that QbTesting as part of the diagnostic workup for ADHD in children (age < 18 years), when used in combination with clinical assessment, is cost-effective. We found this finding was robust to nearly all assumptions made in the model. It also appears likely that QbTest would be cost-effective if used for the subgroup of patients who are not diagnosed on initial clinical assessment. It is unclear whether it would be more cost-effective to perform the test only in this subgroup of patients when compared to using the test in all patients.

There are insufficient data to draw conclusions regarding the clinical or cost-effectiveness of any of the other sensor CPTs (QbCheck, EF Sim, EF Sim Web Version, Nesplora Kids and Nesplora adults), including web-based CPT. There are also insufficient data to draw conclusions regarding the use of CPT tests for dose titration, medication selection and long-term treatment management.

As highlighted in *Assessment of factors relevant to the National Health Service and other parties*, the following factors may need to be considered when implementing the test in practice in the NHS; these include:

- Potential benefits of sensor CPTs in reducing time-consuming appeals.
- Higher proportion of patients initiating treatment if sensor CPTs are used, which could have implications for availability of pharmacological medication.
- Need for private room and training for staff to be able to administer sensor CPTs.
- If sensor CPTs were to be used for medication management where appointments are held remotely, administration of the sensor CPT would need to be help in-person.
- Qualitative data suggested concerns with the length and repetitive content of the QbTest, it may be that other tests are more interactive and engaging for patients.

Suggested research priorities

The section on uncertainties (see *Uncertainties*) highlights a number of area where further research is needed. There is a clear need for a robust DTA study comparing sensor CPT plus clinical assessment and clinical assessment alone with an appropriate reference standard. Such a study could include a direct comparison of the different sensor CPTs (including web-based CPT) and could also include a comparison with a non-sensor CPT such as the Conners CPT II. It should be powered to compare the accuracy of the test across different subgroups of patients, including age, sex, ethnicity, people with mental health, behavioural and neurodevelopmental conditions, people with intellectual disability, people with developmental trauma, and, if data are available, could also consider whether accuracy varies in looked-after children, or people in the Youth Justice System or Adult Criminal Justice System.

There is also a need for further studies to look at the impact of CPT on process measures, patient outcomes and costs. Such studies should use a similar randomised design to the AQUA trial, include studies of adults and studies of children and evaluate other sensor CPTs and non-sensor CPTs, not just the QbTest. They should measure patient outcomes as well as process measures and costs and should also collect quantitative data on outcomes shown to be important to patients and clinicians in the qualitative evaluations, for example, confidence in diagnostic decision-making, communication between patients, clinicians and schools, patient understanding and acceptance of diagnostic decision and acceptability of the test to patients. Studies evaluating the effectiveness and cost-effectiveness of sensor CPTs in children should provide all results separately by setting (CAMHS or community paediatric services). It would also be valuable to consider subgroups of patients who are more difficult to diagnose separately from the whole population being evaluated for ADHD. For example, following up patients who do not receive a diagnosis after an initial period

of assessment (beyond 6 months) would be useful to estimate the proportion who subsequently receive further assessment for ADHD and the proportion of those with further assessment who are diagnosed with ADHD or have ADHD excluded and whether this differs if a sensor CPT is used as part of the diagnosis process.

There is currently no good-quality quantitative data on the use of sensor CPT for medication management, both for initial dose titration and medication selection and for longer-term medication FU. Studies are therefore also needed to address this question. A similar design to that test in the QUOTA feasibility study[110] could be employed with participants randomised to treatment arms with and without sensor CPT as part of initial dose titration and medication selection. FU should be sufficiently long to also consider longer-term medication management and provide information on longer-term costs to inform the economic model. Important outcomes to consider would be whether patients respond optimally or suboptimally to treatment, adherence to treatment, control of ADHD symptoms, quality of life, executive function and resource costs for patients depending on response to treatment as well as process measures, including number and length of appointments.

As the area of technology-assisted diagnosis and progress monitoring in ADHD encompasses more than the methods assessed in this review, future research should also explore whether integrating various technologies and leveraging artificial intelligence can enhance the diagnostic accuracy (and precision monitoring).

Additional information

CRediT contribution statement

Eve Tomlinson (https://orcid.org/0000-0002-0969-602X): Conceptualisation, Formal analysis, Methodology, Investigation, Project administration, Visualisation, Writing – original draft, Writing – reviewing and editing.

Mary Ward (https://orcid.org/0000-0002-2321-2546): Conceptualisation, Formal analysis, Methodology, Investigation, Project administration, Visualisation, Writing – original draft, Writing – reviewing and editing.

Josephine Walker (https://orcid.org/0000-0002-9732-5738): Conceptualisation, Formal analysis, Methodology, Investigation, Project administration, Visualisation, Writing – original draft, Writing – reviewing and editing.

Melissa Benevente (https://orcid.org/0000-0002-1908-2742): Investigation, Writing – reviewing and editing.

Hanyu Wang (https://orcid.org/0000-0001-7933-0667): Formal analysis, Writing – original draft.

Chris Cooper (https://orcid.org/0000-0003-0864-5607): Investigation, Writing – original draft.

Hayley E Jones (https://orcid.org/0000-0002-4265-2854): Formal analysis, Supervision, Writing – original draft.

Amanda Owen-Smith (https://orcid.org/0000-0003-1188-2371): Formal analysis, Methodology, Visualisation, Supervision, Writing – original draft, Writing – reviewing and editing, PPI contributions.

Catalina Lopez Manzano (https://orcid.org/0009-0009-8883-0167): Writing – reviewing and editing, PPI contributions.

Sara James: Writing – reviewing and editing, PPI contributions.

Dietmar Hank (https://orcid.org/0000-0001-8686-8078): Writing – reviewing and editing, Clinical advice.

Nicky J Welton (https://orcid.org/0000-0003-2198-3205): Conceptualisation, Formal analysis, Funding acquisition, Methodology, Investigation, Project administration, Supervision, Visualisation, Writing – original draft, Writing – reviewing and editing.

Penny Whiting (https://orcid.org/0000-0003-1138-5682): Conceptualisation, Formal analysis, Funding acquisition, Methodology, Investigation, Project administration, Supervision, Visualisation, Writing – original draft, Writing – reviewing and editing.

Richard Lee-Kelland: Writing – reviewing and editing, Clinical advice.

Acknowledgements

We would like to thank Laura Wade and Lisa Connell from Avon and Wiltshire Mental Health Partnership NHS Trust for clinical advice during the early stages of this project. We thank Jelena Savovic for her time and guidance on the application on the ROBINS–I tool. We would also like to thank the following study authors for providing us with additional information on their studies: Eva Billstedt, Oliver Grimm, Chris Hollis, Pia Tallberg, Bolian Guo and Nicole McGlennon.

Data-sharing statement

All data extracted for the systematic review and the results of the risk-of-bias assessments are provided in full in the appendices to this report. The economic model can be obtained from the corresponding author and will be shared upon reasonable request for academic collaboration.

Ethics statement

The research included in this report is secondary research and as such did not require ethical approval.

Information governance statement

There were no personal data involved in the production of this report.

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at https://doi.org/10.3310/DRDR7171.

Primary conflicts of interest: None of the authors have any competing interests.

Disclaimer

Every effort has been made to obtain the necessary permissions for reproduction, to credit original sources appropriately and to respect copyright requirements. However, despite our diligence, we acknowledge the possibility of unintentional omissions or errors and we welcome notifications of any concerns regarding copyright or permissions.

References

1. Bristol TAG. *[FINAL PROTOCOL] Technologies for the Assessment of Attention Deficit Hyperactivity Disorder*. 2023. URL: www.nice.org.uk/guidance/gid-dg10088/documents/final-protocol (accessed November 2023).
2. NHS. *Symptoms – Attention Deficit Hyperactivity Disorder (ADHD)*. 2021. URL: www.nhs.uk/conditions/attention-deficit-hyperactivity-disorder-adhd/symptoms/ (accessed October 2023).
3. Sayal K, Prasad V, Daley D, Ford T, Coghill D. ADHD in children and young people: prevalence, care pathways, and service provision. *Lancet Psychiatry* 2018;**5**:19–20.
4. Cybulski L, Ashcroft DM, Carr MJ, Garg S, Chew-Graham CA, Kapur N, Webb RT. Temporal trends in annual incidence rates for psychiatric disorders and self-harm among children and adolescents in the UK, 2003–2018. *BMC Psychiatry* 2021;**21**:229.
5. Leffa DT, Caye A, Rohde LA. ADHD in children and adults: diagnosis and prognosis. *Curr Top Behav Neurosci* 2022;**57**:1–18.
6. Baggio S, Fructuoso A, Guimaraes M, Fois E, Golay D, Heller P, *et al*. Prevalence of attention deficit hyperactivity disorder in detention settings: a systematic review and meta-analysis. *Front Psychiatry* 2018;**9**:331.
7. Fazel S, Favril L. *Prevalence of ADHD in Adult Prisoners: A Re-analysis*. 2023. URL: https://psyarxiv.com/ajsxm/download?format=pdf (accessed October 2023).
8. Peasgood T, Anupam B, Biggs K, Brazier JE, Coghill D, Cooper CL, *et al*. The impact of ADHD on the health and well-being of ADHD children and their siblings. *Eur Child Adolesc Psychiatry* 2016;**25**:1217–31.
9. Katragadda S, Schubiner H. ADHD in children, adolescents, and adults. *Primary Care* 2007;**34**:317–41; abstract viii.
10. Robin AL, Payson E. The impact of ADHD on marriage. *ADHD Report* 2002;**10**:9–14.
11. Franke B, Michelini G, Asherson P, Banaschewski T, Bilbow A, Buitelaar JK, *et al*. Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *Eur Neuropsychopharmacol* 2018;**28**:1059–88.
12. Watters C, Adamis D, McNicholas F, Gavin B. The impact of attention deficit hyperactivity disorder (ADHD) in adulthood: a qualitative study. *Ir J Psychol Med* 2018;**35**:173–9.
13. NHS. *Attention Deficit Hyperactivity Disorder (ADHD)*. 2021. URL: www.nhs.uk/conditions/attention-deficit-hyperactivity-disorder-adhd/ (accessed October 2023).
14. BMJ Best Practice. *Attention Deficit Hyperactivity Disorder in Children*. 2023. URL: https://bestpractice.bmj.com/topics/en-gb/142 (accessed October 2023).
15. Biederman J, Mick E, Faraone SV. Age-dependent decline of symptoms of attention deficit hyperactivity disorder: impact of remission definition and symptom type. *Am J Psychiatry* 2000;**157**:816–8.
16. National Institute for Health and Care Excellence. *NG87: Attention Deficit Hyperactivity Disorder: Diagnosis and Management*. 2019. URL: www.nice.org.uk/guidance/ng87 (accessed October 2023).
17. Ogundele MO, Ayyash HF, Banerjee S. Role of computerised continuous performance task tests in ADHD. *Prog Neurol Psychiatry* 2011;**15**:8–13.
18. Hollis C, Hall CL, Guo B, James M, Boadu J, Groom MJ, *et al*. 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. *J Child Psychol Psychiatry* 2018;**59**:1298–308.
19. Royal College of Psychiatrists in Scotland. *ADHD in Adults: Good Practice Guidelines*. 2017. URL: www.rcpsych.ac.uk/docs/default-source/members/divisions/scotland/adhd_in_adultsfinal_guidelines_june2017.pdf?sfvrsn=40650449_2 (accessed October 2023).

20. House of Commons. *Petitions Committee: ADHD and ASD (Autism) Assessment Waiting Times*. URL: https://committees.parliament.uk/writtenevidence/117330/default/ (accessed October 2023).
21. Smith MCF, Mukherjee RAS, Müller-Sedgwick U, Hank D, Carpenter P, Adamou M. UK adult ADHD services in crisis. *BJPsych Bulletin* 2024;**48**:1–5.
22. Health Innovation East Midlands. *Faster and More Cost-effective ADHD Diagnosis for Children*. URL: https://healthinnovation-em.org.uk/our-work/innovations/focus-adhd (accessed October 2023).
23. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision*. 5th edn. Washington, DC: American Psychiatric Association Publishing; 2022.
24. World Health Organization. *International Classification of Diseases (ICD-11) 11th Revision*. URL: www.who.int/standards/classifications/classification-of-diseases (accessed November 2023).
25. Mulraney M, Arrondo G, Musullulu H, Iturmendi-Sabater I, Cortese S, Westwood SJ, *et al*. Systematic review and meta-analysis: screening tools for attention-deficit/hyperactivity disorder in children and adolescents. *J Am Acad Child Adolesc Psychiatry* 2022;**61**:982–96.
26. Hall CL, Valentine AZ, Groom MJ, Walker GM, Sayal K, Daley D, Hollis C. The clinical utility of the continuous performance test and objective measures of activity for diagnosing and monitoring ADHD in children: a systematic review. *Eur Child Adolesc Psychiatry* 2016;**25**:677–99.
27. NHS. *Treatment – Attention Deficit Hyperactivity Disorder (ADHD)*. 2021. URL: www.nhs.uk/conditions/attention-deficit-hyperactivity-disorder-adhd/treatment/ (accessed October 2023).
28. Mechler K, Tobias B, Hohmann S, Häge A. Evidence-based pharmacological treatment options for ADHD in children and adolescents. *Pharmacol Ther* 2022;**230**:107940.
29. Hamadache S, Kathrin H, Labarga SZ, Gunther T. Is the QbMini a valid instrument for ADHD assessment? [References]. DP – Aug 2021. *J Atten Disord* 2021;**25**:1384–94.
30. Bellato AH, Charlotte L, Groom MJ, Emily S, Thapar A, Hollis C, Cortese S. Practitioner review: clinical utility of the QbTest for the assessment and diagnosis of attention-deficit/hyperactivity disorder – a systematic review and meta-analysis. *J Child Psychol Psychiatry* 2023;**65**:845–61. https://doi.org/10.1111/jcpp.13901
31. McKenzie C, Ingall BR, Hall C. *Focus ADHD National Programme Evaluation*. Nottingham: The Institute of Mental Health; 2022. URL: https://healthinnovation-em.org.uk/component/rsfiles/download-file/files?path=our-work%252Four-innovations%252FADHD%2BFOCUS%2Bevaluation%2Breport%2B-%2BFINAL%2Bv.1.0%2B18.10.22.pdf&Itemid=1457 (accessed October 2023).
32. National Institute for Health and Care Excellence. *[MIB318] QbTest for the Assessment of Attention Deficit Hyperactivity Disorder (ADHD)*. 2023. URL: www.nice.org.uk/advice/mib318 (accessed October 2023).
33. Centre for Reviews and Dissemination. *CRD's Guidance for Undertaking Reviews in Health Care*. York: University of York; 2009.
34. National Institute for Health Care Excellence. *NICE Health Technology Evaluations: The Manual (PMG36)*. 2022. URL: www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation (accessed November 2023).
35. Deeks J, Bossuyt P, Leeflang M, Takwoingi Y, Flemyng E. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 2022. URL: https://training.cochrane.org/handbook-diagnostic-test-accuracy (accessed November 2023).
36. Salameh JP, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, *et al*. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. *BMJ* 2020;**370**:m2632.
37. Welch V, Petticrew M, Tugwell P, Moher D, O'Neill J, Waters E, White H; PRISMA-Equity Bellagio group. PRISMA-equity 2012 extension: reporting guidelines for systematic reviews with a focus on health equity. *PLOS Med* 2012;**9**:e1001333.

38. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;**372**:n71.
39. Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, *et al*. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;**69**:225–34.
40. O'Neill JT, Welch H, Petticrew V, Pottie M, Clarke K, Evans M. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *J Clin Epidemiol* 2014;**67**:56–64.
41. Stevanovic D, Nasic S, Doric A, Wentz E, Knez R. The structure and diagnostic accuracy of the QbTest in pediatric ADHD: a retrospective clinical study. *J Atten Disord* 2023;**27**:1296–305.
42. Gardner J. *Statistical Testing for AUC Data [R Package auctestr Version 1.0.0]. R-Project.org*; 2017. URL: https://cran.rproject.org/package=auctestr (accessed December 2023).
43. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.
44. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, *et al*. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;**366**:l4898.
45. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al*.; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36.
46. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, *et al*. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;**355**:i4919.
47. Critical Appraisal Skills Programme (CASP). *CASP Checklists – Critical Appraisal Checklists*. 2024. URL: https://casp-uk.net/casp-tools-checklists/ (accessed May 2024).
48. Protogerou C, Hagger MS. A checklist to assess the quality of survey studies in psychology. *Methods Psychol* 2020;**3**:100031.
49. Nyaga VN, Arbyn M. Metadta: a Stata command for meta-analysis and meta-regression of diagnostic test accuracy data – a tutorial. *Arch Pub Health* 2022;**80**(1):216.
50. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;**58**:982–90.
51. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006;**59**:1331–2; author reply 1332.
52. RDocumentation [DataCamp]. *metagen: Generic Inverse Variance Meta-analysis*. n.r. URL: www.rdocumentation.org/packages/meta/versions/4.9-6/topics/metagen (accessed December 2023).
53. Fisher RA. Statistical Methods for Research Workers. In Kotz S, Johnson NL, editors. *Breakthroughs in Statistics: Methodology and Distribution*. New York, NY: Springer New York; 1992. pp. 66–70. http://doi.org/10.1007/978-1-4612-4380-9_6
54. Hannes K, Lockwood C. Pragmatism as the philosophical foundation for the Joanna Briggs meta-aggregative approach to qualitative evidence synthesis. *J Adv Nurs* 2011;**67**:1632–42.
55. Marshall P, Hoelzle J, Nikolas M. Diagnosing attention-deficit/hyperactivity disorder (ADHD) in young adults: a qualitative review of the utility of assessment measures and recommendations for improving the diagnostic process. *Clin Neuropsychol* 2021;**35**:165–98.
56. Peñuelas-Calvo I, Jiang-Lin LK, Girela-Serrano B, Delgado-Gomez D, Navarro-Jimenez R, Baca-Garcia E, Porras-Segovia A. Video games for the assessment and treatment of attention-deficit/hyperactivity disorder: a systematic review. *Eur Child Adolesc Psychiatry* 2022;**31**:5–20.

57. Valentine AZ, Brown BJ, Groom MJ, Young E, Hollis C, Hall CL. A systematic review evaluating the implementation of technologies to assess, monitor and treat neurodevelopmental disorders: a map of the current evidence. *Clin Psychol Rev* 2020;**80**:101870.

58. Cabas-Hoyos K, Figueroa PA, Gonzalez-Bracamonte Y. Programas de intervención basados en tecnologías para niños y adolescentes diagnosticados con TDAH: Una revisión sistemática. *Tesis Psicológica* 2022;**17**:1–32.

59. Gustafsson U, Hansen M. QbTest for monitoring medication treatment response in ADHD: a systematic review. *Clin Pract Epidemiol Ment Health* 2023;**19**:e17450179276630.

60. Wiebe A, Kyra K, Selaskowski B, Mehren A, Thöne AK, Pramme L, *et al*. Virtual reality in the diagnostic and therapy for mental disorders: a systematic review. *Clin Psychol Rev* 2022;**98**:102213.

61. Gustafsson U, Hansen M. QbTest in the clinical assessment of attention deficit hyperactivity disorder: a review of the evidence. *Mental Health Sci* 2023;**1**:270–81.

62. Sharma A, Singh B. Evaluation of the role of Qb testing in attention deficit hyperactivity disorder. *Arch Dis Child* 2009;**94**.

63. Rufo-Campos M, Cueto E, Iriarte Y, Rufo Muñoz M. Estudio de sensibilidad de un nuevo método diagnóstico para el TDAH: Aula Nesplora. *Rev Neurol (Paris)* 2012;**54**.

64. Sharma R, Waghorn A, Lacey S, Spiewakowski D. Implementing QB testing for ADHD: evaluating value in a DGH setting. *Arch Dis Child* 2022;**107**:A70.

65. Soff C, Sotnikova A, Siniatchkin M, Christiansen H. Additiver Nutzen des QbTests in der Diagnostik der Aufmerksamkeitsdefizit-/Hyperaktiv- itätsstörung im Kindesalter (Manuscript in preparation). 2017;**30**:656–72.

66. Tallberg P, Råstam M, Wenhov L, Eliasson G, Gustafsson P. Incremental clinical utility of continuous performance tests in childhood ADHD – an evidence-based assessment approach. *Scand J Psychol* 2019;**60**:26–35.

67. Hult N, Kadesjo J, Kadesjo B, Gillberg C, Billstedt E. ADHD and the QbTest: diagnostic validity of QbTest. *J Atten Disord* 2018;**22**:1074–80.

68. Brunkhorst-Kanaan N, Verdenhalven M, Kittel-Schneider S, Vainieri I, Reif A, Grimm O. The quantified behavioral test-a confirmatory test in the diagnostic process of adult ADHD? *Front Psychiatry* 2020;**11**:216.

69. Humphreys C, Sitton-Kent L. *Transforming ADHD Care across the East Midlands: An Evaluation Report*. *East Midlands Academic Health Network*. 2018. URL: https://healthinnovation-em.org.uk/component/rsfiles/download-file/files?path=our-work%252Four-innovations%252FTransforming-ADHD-Care%252FFinal_Overall_Evaluation_Report_31May18.pdf&Itemid=1457 (accessed March 2024).

70. Johansson V, Norén Selinus E, Kuja-Halkola R, Lundström S, Durbeej N, Anckarsäter H, *et al*. The quantified behavioral test failed to differentiate ADHD in adolescents with neurodevelopmental problems. *J Atten Disord* 2021;**25**:312–21.

71. NCT05846815. *Assessment of ARVO Web-based Software for Evaluating Attention-Deficit/Hyperactive Disorder (ADHD) Symptoms in Children*. 2023. URL: https://clinicaltrials.gov/show/NCT05846815 (accessed November 2023).

72. Chitsabesan P, Hall CL, Carter LA, Reeves M, Mohammed V, Beresford B, *et al*. Using an objective computer task (QbTest) to aid the identification of attention deficit hyperactivity disorder (ADHD) in the Children and Young People Secure Estate (CYPSE): a feasibility randomised controlled trial. *BMJ Open* 2022;**12**:e064951.

73. Zulueta A, Díaz-Orueta U, Crespo-Eguilaz N, Torrano F. Virtual reality-based assessment and rating scales in ADHD diagnosis. *Psicologia Educativa* 2019;**25**:13–22.

74. Söderström S, Pettersson R, Nilsson KW. Quantitative and subjective behavioural aspects in the assessment of attention-deficit hyperactivity disorder (ADHD) in adults. *Nord J Psychiatry* 2014;**68**:30–7.

75. Seesjärvi E, Puhakka J, Aronen ET, Lipsanen J, Mannerkoski M, Hering A, *et al*. Quantifying ADHD symptoms in open-ended everyday life contexts with a new virtual reality task. *J Atten Disord* 2022;**26**:1394–411.
76. Edebol H, Helldin L, Norlander T. Objective measures of behavior manifestations in adult ADHD and differentiation from participants with bipolar II disorder, borderline personality disorder, participants with disconfirmed ADHD as well as normative participants. *Clin Pract Epidemiol Ment Health* 2012;**8**:134–43.
77. Ulberstad F, Bostrom H, Chavanon ML, Knollmann M, Wiley J, Christiansen H, Thorell LB. Objective measurement of attention deficit hyperactivity disorder symptoms outside the clinic using the QbCheck: reliability and validity. *Int J Methods Psychiatr Res* 2020;**29**:e1822.
78. Bijlenga D, Ulberstad F, Thorell LB, Christiansen H, Hirsch O, Kooij JJS. Objective assessment of attention-deficit/hyperactivity disorder in older adults compared with controls using the QbTest. *Int J Geriatr Psychiatry* 2019;**34**:1526–33.
79. Edebol H, Helldin L, Norlander T. Measuring adult attention deficit hyperactivity disorder using the quantified behavior test plus. *Psychology J* 2013;**2**:48–62.
80. Groom MJ, Young Z, Hall CL, Gillott A, Hollis C. The incremental validity of a computerised assessment added to clinical rating scales to differentiate adult ADHD from autism spectrum disorder. *Psychiatry Res* 2016;**243**:168–73.
81. Adamou M, Jones SL, Marks L, Lowe D. Efficacy of continuous performance testing in adult ADHD in a clinical sample using QbTest. *J Atten Disord* 2022;**26**:1483–91.
82. Pettersson R, Söderström S, Nilsson KW. Diagnosing ADHD in adults: an examination of the discriminative validity of neuropsychological tests and diagnostic assessment instruments. *J Atten Disord* 2018;**22**:1019–31.
83. Emser TS, Johnston BA, Steele JD, Kooij S, Thorell L, Christiansen H. Assessing ADHD symptoms in children and adults: evaluating the role of objective measures. *Behav Brain Funct* 2018;**14**:11.
84. Edebol H, Helldin L, Holmberg E, Gustafsson SA, Norlander T. In search for objective measures of hyperactivity, impulsivity and inattention in adult attention deficit hyperactivity disorder using the quantified behavior test plus. *Eur J Psychol* 2011;**7**:443–457.
85. Hall CL, Selby K, Guo B, Valentine AZ, Walker GM, Hollis C. 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 Adolesc Ment Health* 2016;**21**:175–8.
86. Vogt C, Shameli A. Assessments for attention-deficit hyperactivity disorder: use of objective measurements. *Psychiatrist* 2011;**35**:380–3.
87. Pellegrini S, Murphy M, Lovett E. The QbTest for ADHD assessment: impact and implementation in child and adolescent mental health services. *Child Youth Serv Rev* 2020;**114**:105032.
88. Peli Vision Oy. *Research Behind EFSim and Feedback from Pilot Tests [Unpublished Report]*; (n.r.).
89. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 2000;**41**:645–55.
90. Bell CC. DSM-IV: diagnostic and statistical manual of mental disorders. *J Am Med Assoc* 1994;**272**:828.
91. American Psychiatric Association. *Highlights of Changes from DSM-IV-TR to DSM-5*. 2013. URL: www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/APA_DSM_Changes_from_DSM-IV-TR_-to_DSM-5.pdf (accessed April 2024).
92. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, DC: American Psychiatric Association; 2013.
93. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization; 1992.

94. Thomas R, Mitchell GK, Batstra L. Attention-deficit/hyperactivity disorder: are we helping or harming? *BMJ* 2013;**347**:f6172.

95. Kaufman J, Birmaher B, Brent D, Rao U, Ryan N. *Kiddie-Sads-Present and Lifetime Version (K-SADS-PL)*. Pittsburgh: University of Pittsburgh, School of Medicine; 1996.

96. Szomlaiski N, Dyrborg J, Rasmussen H, Schumann T, Koch SV, Bilenberg N. Validity and clinical feasibility of the ADHD rating scale (ADHD-RS): a Danish nationwide multicenter study. *Acta Paediatr* 2009;**98**:397–402.

97. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi EVA, *et al*. The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. *Psychol Med* 2005;**35**:245–56.

98. Christiansen H, Kis B, Hirsch O, Matthies S, Hebebrand J, Uekermann J, *et al*. German validation of the Conners Adult ADHD Rating Scales (CAARS) II: reliability, validity, diagnostic sensitivity and specificity. *Eur Psychiatric Assoc* 2012;**27**:321–8.

99. Zimmermann P, Fimm B. *[TAP] Testbatterie zur Aufmerksamkeitsprüfung.Herzogenrath*. Herzogenrath: Psytest; 2012.

100. Zimmermann P, Gondan M, Fimm B. *KiTAP: Testbatterie zur Aufmerksamkeitsprüfungfür Kinder KiTAP*. Herzogenrath: Psytest; 2003.

101. Conners CK, Erhadt D, Sparrow M. *Conners Adult Rating Scale and Autism Quotient-10*. New York, NY: Multihealth Systems, Inc.; 1999.

102. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The Autism-Spectrum Quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord* 2001;**31**:5–17.

103. Rosvold HE, Mirsky AF, Sarason I, Bransome Jr ED, Beck LH. A continuous performance test of brain damage. *J Consult Psychol* 1956;**20**:343–50.

104. Conners C, Staff M, Conners C, Conners K. *Conners' Continuous Performance Test (CPT-II) Computer Programs for Windows Technical Guide and Software Manual*. North Tonawanda, NY: Multi-Health Systems; 2000.

105. McKenzie C, Ingall BR, Hall C. *Focus ADHD National Programme Evaluation*. 2022. URL: https://healthinnovation-em.org.uk/component/rsfiles/download-file/files?path=our-work%252Four-innovations%-252FADHD%2BFOCUS%2Bevaluation%2Breport%2B-%2BFINAL%2Bv.1.0%2B18.10.22.pdf&Itemid=1457 (accessed March 2024).

106. Booth JN, Tomporowski PD, Boyle JM, Ness AR, Joinson C, Leary SD, Reilly JJ. Associations between executive attention and objectively measured physical activity in adolescence: findings from ALSPAC, a UK cohort. *Ment Health Phys Act* 2013;**6**:212–9.

107. Hall CL, Valentine AZ, Walker GM, Ball HM, Cogger H, Daley D, *et al*. Study of user experience of an objective test (QbTest) to aid ADHD assessment and medication management: a multi-methods approach. *BMC Psychiatry* 2017;**17**:66.

108. Lennox C, Hall CL, Carter LA, Beresford B, Young S, Kraam A, *et al*. FACT: a randomised controlled trial to assess the feasibility of QbTest in the assessment process of attention deficit hyperactivity disorder (ADHD) for young people in prison – a feasibility trial protocol. *BMJ Open* 2020;**10**:e035519.

109. ISRCTN17402196. *Feasibility Trial to Assess Attention Deficit Hyperactivity Disorder (ADHD) in the Criminal Justice System by Using QbTest (A Computer Task)*. 2019. URL: http://isrctn.com/ISRCTN17402196 (accessed November 2023).

110. Williams L, Hall CL, Brown S, Guo B, James M, Franceschini M, *et al*. Optimising medication management in children and young people with ADHD using a computerised test (QbTest): a feasibility randomised controlled trial. *Pilot Feasibility Stud* 2021;**7**:68.

111. Bussing R, Fernandez M, Harwood M, Wei H, Garvan CW, Eyberg SM, *et al*. Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: psychometric properties and normative ratings from a school district sample. *Assessment* 2008;**15**:317–28.

112. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press; 2015.

113. Kent Surrey Sussex Academic Health Science Network. *QbTest for ADHD Cost-Benefit Analysis*. n.r. URL: https://healthinnovation-em.org.uk/component/rsfiles/preview?path=our-work%252Four-innovations%252FKent%2BSurrey%2Band%2BSussex%2BAHSN%2BQbTest%2Bcost%2Bbenefit%2Banalysis.pdf&Itemid=1457 (accessed December 2023).

114. National Institute for Health and Care Excellence. *CG72: Attention Deficit Hyperactivity Disorder (update) Appendix 1: Cost-effectiveness Analysis: What Is the Cost Effectiveness of Parent Training Compared to No Treatment for Children with ADHD?* 2018. URL: www.nice.org.uk/guidance/ng87/evidence/appendix-1-parent-training-model-writeup-pdf-6606256285 (accessed April 2024).

115. National Institute for Health and Care Excellence. *CG72: Attention Deficit Hyperactivity Disorder (Update) Appendix 2: Cost-effectiveness Analysis: Combination Treatment in Children and Adolescents*. 2018. URL: www.nice.org.uk/guidance/ng87/evidence/appendix-2-children-combo-model-writeup-pdf-6606256286 (accessed April 2024).

116. Cottrell S, Tilden D, Robinson P, Bae J, Arellano J, Edgell E, *et al*. A modeled economic evaluation comparing atomoxetine with stimulant therapy in the treatment of children with attention-deficit/hyperactivity disorder in the United Kingdom. *Value Health* 2008;**11**:376–88.

117. Erder MH, Xie J, Signorovitch JE, Chen KS, Hodgkins P, Lu M, *et al*. Cost effectiveness of guanfacine extended-release versus atomoxetine for the treatment of attention-deficit/hyperactivity disorder: application of a matching-adjusted indirect comparison. *Appl Health Econ Health Policy* 2012;**10**:381–95.

118. Faber A, van Agthoven M, Kalverdijk LJ, Tobi H, de Jong-van den Berg LT, Annemans L, *et al*. Long-acting methylphenidate-OROS in youths with attention-deficit hyperactivity disorder suboptimally controlled with immediate-release methylphenidate: a study of cost effectiveness in the Netherlands. *CNS Drugs* 2008;**22**:157–70.

119. Freriks RD, Mierau JO, van der Schans J, Groenman AP, Hoekstra PJ, Postma MJ, *et al*. Cost-effectiveness of treatments in children with attention-deficit/hyperactivity disorder: a continuous-time markov modeling approach. *MDM Policy Pract* 2019;**4**:2381468319867629.

120. Hong J, Dilla T, Arellano J. A modelled economic evaluation comparing atomoxetine with methylphenidate in the treatment of children with attention-deficit/hyperactivity disorder in Spain. *BMC Psychiatry* 2009;**9**:15.

121. King S, Griffin S, Hodges Z, Weatherly H, Asseburg C, Richardson G, *et al*. A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents. *Health Technol Assess* 2006;**10**:**iii–iv**, xiii–146.

122. Klein G, Greenhouse J, Stein B, Seltman H. Characterizing patterns of care using administrative claims data: ADHD treatment in children. *Health Serv Outcomes Res Methodol* 2011;**11**:115–33.

123. Lachaine J, Sikirica V, Mathurin K. Is adjunctive pharmacotherapy in attention-deficit/hyperactivity disorder cost-effective in Canada: a cost-effectiveness assessment of guanfacine extended-release as an adjunctive therapy to a long-acting stimulant for the treatment of ADHD. *BMC Psychiatry* 2016;**16**:11.

124. Maia CR, Stella SF, Wagner F, Pianca TG, Krieger FV, Cruz LN, *et al*. Cost-utility analysis of methylphenidate treatment for children and adolescents with ADHD in Brazil. *Rev Bras Psiquiatr* 2016;**38**:30–8.

125. Marchetti A, Magar R, Lau H, Murphy EL, Jensen PS, Conners CK, *et al*. Pharmacotherapies for attention-deficit/hyperactivity disorder: expected-cost analysis. *Clin Ther* 2001;**23**:1904–21.

126. Nagy B, Setyawan J, Coghill D, Soroncz-Szabó T, Kaló Z, Doshi JA. A conceptual framework for a long-term economic model for the treatment of attention-deficit/hyperactivity disorder. *Expert Rev Pharmacoecon Outcomes Res* 2017;**17**:283–92.

127. Narayan S, Hay J. Cost effectiveness of methylphenidate versus AMP/DEX mixed salts for the first-line treatment of ADHD. *Expert Rev Pharmacoecon Outcomes Res* 2004;**4**:625–34.

128. Prasad S, Arellano J, Steer C, Libretto SE. Assessing the value of atomoxetine in treating children and adolescents with ADHD in the UK. *Int J Clin Pract* 2009;**63**:1031–40.

129. Schawo S, van der Kolk A, Bouwmans C, Annemans L, Postma M, Buitelaar J, *et al*. Probabilistic Markov model estimating cost effectiveness of methylphenidate osmotic-release oral system versus immediate-release methylphenidate in children and adolescents: which information is needed? *PharmacoEconomics* 2015;**33**:489–509.

130. Sikirica V, Haim Erder M, Xie J, Macaulay D, Diener M, Hodgkins P, Wu EQ. Cost effectiveness of guanfacine extended release as an adjunctive therapy to a stimulant compared with stimulant monotherapy for the treatment of attention-deficit hyperactivity disorder in children and adolescents. *PharmacoEconomics* 2012;**30**:e1–15.

131. Sohn M, Talbert J, Moga DC, Blumenschein K. A cost-effectiveness analysis of off-label atypical antipsychotic treatment in children and adolescents with ADHD who have failed stimulant therapy. *Atten Defic Hyperact Disord* 2016;**8**:149–58.

132. Tajik A, Nikfar S, Elyasi S, Rajabi O, Varmaghani M. Cost-effectiveness and budget impact analysis of lisdexamfetamine versus methylphenidate for patients under 18 with attention-deficit/hyperactivity disorder in Iran. *Child Adolesc Psychiatry Ment Health* 2023;**17**:115.

133. Tockhorn-Heidenreich A, Televantou F, Karlsdotter K, Dilla T. Atomoxetine for the treatment of adults with attention deficit hyperactivity disorder: a cost-utility analysis in Spain. *Pharmacoeconomics Open Access* 2015;**1**:104.

134. van der Schans J, Kotsopoulos N, Hoekstra PJ, Hak E, Postma MJ. Cost-effectiveness of extended-release methylphenidate in children and adolescents with attention-deficit/hyperactivity disorder sub-optimally treated with immediate release methylphenidate. *PLOS ONE* 2015;**10**:e0127237.

135. Vanoverbeke N, Annemans L, Ingham M, Adriaenssen I. A cost analysis of the management of attention-deficit/hyperactivity disorder (ADHD) in children in the UK. *J Med Econ* 2003;**6**:79–94.

136. Zimovetz EA, Beard SM, Hodgkins P, Bischof M, Mauskopf JA, Setyawan J. A cost-utility analysis of lisdexamfetamine versus atomoxetine in the treatment of children and adolescents with attention-deficit/hyperactivity disorder and inadequate response to methylphenidate. *CNS Drugs* 2016;**30**:985–96.

137. Zimovetz EA, Joseph A, Ayyagari R, Mauskopf JA. A cost-effectiveness analysis of lisdexamfetamine dimesylate in the treatment of adults with attention-deficit/hyperactivity disorder in the UK. *Eur J Health Econ* 2018;**19**:21–35.

138. Miller A, Lee SK, Raina P, Klassen A, Zupancic J, Olsen L. *A Review of Therapies for Attentiondeficit/Hyperactivity Disorder*. 1998. URL: www.cadth.ca/sites/default/files/pdf/adhd_tr_e.pdf (accessed April 2024).

139. Büyükkaramikli NC, Rutten-van Mölken MPMH, Severens JL, Al M. TECH-VER: a verification checklist to reduce errors in models and improve their credibility. *PharmacoEconomics* 2019;**37**:1391–408.

140. Personal Social Services Research Unit (PSSRU). *Unit Costs of Health and Social Care programme (2022 – 2027)*. n.r. URL: www.pssru.ac.uk/unitcostsreport/ (accessed April 2024).

141. NHS. *National Cost Collection for the NHS*. n.r. URL: www.england.nhs.uk/costing-in-the-nhs/national-cost-collection/ (accessed April 2024).

142. Dittmann RW, Cardo E, Nagy P, Anderson CS, Bloomfield R, Caballero B, *et al*. Efficacy and safety of lisdexamfetamine dimesylate and atomoxetine in the treatment of attention-deficit/hyperactivity disorder: a head-to-head, randomized, double-blind, phase IIIb study. *CNS Drugs* 2013;**27**:1081–92.

143. van der Kolk A, Bouwmans CA, Schawo SJ, Buitelaar JK, van Agthoven M, Hakkaart-van Roijen L. Association between quality of life and treatment response in children with attention deficit hyperactivity disorder and their parents. *J Ment Health Policy Econ* 2014;**17**:119–29.

144. Secnik K, Matza LS, Cottrell S, Edgell E, Tilden D, Mannix S. Health state utilities for childhood attention-deficit/hyperactivity disorder based on parent preferences in the United Kingdom. *Med Decis Making* 2005;**25**:56–70.

145. Peasgood T, Bhardwaj A, Brazier JE, Biggs K, Coghill D, Daley D, *et al*. What is the health and well-being burden for parents living with a child with ADHD in the United Kingdom? *J Atten Disord* 2021;**25**:1962–76.

146. NHS England. *2023-25 NHS Payment Scheme (Amended)*. 2024. URL: www.england.nhs.uk/publication/2023-25-nhs-payment-scheme/ (accessed May 2024).

147. Office for National Statistics (ONS). *CPI Index 06.2.1/3: Medical Services and Paramedical Services: 2015=100*. n.r. URL: www.ons.gov.uk/economy/inflationandpriceindices/timeseries/d7fa/mm23 (accessed May 2024).

148. National Institute for Health and Care Excellence. *British National Formulary (BNF)*. 2024. URL: https://bnf.nice.org.uk/ (accessed April 2024).

149. Specialist Pharmacy Service. *Considerations When Prescribing Modified-Release Methylphenidate*. 2023. URL: www.sps.nhs.uk/articles/considerations-when-prescribing-modified-release-methylphenidate/ (accessed May 2024).

150. Hervas A, Huss M, Johnson M, McNicholas F, van Stralen J, Sreckovic S, *et al*. Efficacy and safety of extended-release guanfacine hydrochloride in children and adolescents with attention-deficit/hyperactivity disorder: a randomized, controlled, phase III trial. *Eur Neuropsychopharmacol* 2014;**24**:1861–72.

151. Martenyi F, Zavadenko NN, Jarkova NB, Yarosh AA, Soldatenkova VO, Bardenstein LM, *et al*. Atomoxetine in children and adolescents with attention-deficit/hyperactivity disorder: a 6-week, randomized, placebo-controlled, double-blind trial in Russia. *Eur Child Adolesc Psychiatry* 2010;**19**:57–66.

152. Newcorn JH, Kratochvil CJ, Allen AJ, Casat CD, Ruff DD, Moore RJ, Michelson D; Atomoxetine/Methylphenidate Comparative Study Group. Atomoxetine and osmotically released methylphenidate for the treatment of attention deficit hyperactivity disorder: acute comparison and differential response. *Am J Psychiatry* 2008;**165**:721–30.

153. Takahashi M, Takita Y, Yamazaki K, Hayashi T, Ichikawa H, Kambayashi Y, *et al*. A randomized, double-blind, placebo-controlled study of atomoxetine in Japanese children and adolescents with attention-deficit/hyperactivity disorder. *J Child Adolesc Psychopharmacol* 2009;**19**:341–50.

154. Wehmeier PM, Schacht A, Ulberstad F, Lehmann M, Schneider-Fresenius C, Lehmkuhl G, *et al*. Does atomoxetine improve executive function, inhibitory control, and hyperactivity? Results from a placebo-controlled trial using quantitative measurement technology. *J Clin Psychopharmacol* 2012;**32**:653–60.

155. Montoya A, Hervas A, Cardo E, Artigas J, Mardomingo MJ, Alda JA, *et al*. Evaluation of atomoxetine for first-line treatment of newly diagnosed, treatment-naïve children and adolescents with attention deficit/hyperactivity disorder. *Curr Med Res Opin* 2009;**25**:2745–54.

156. Childress AC, Arnold V, Adeyi B, Dirks B, Babcock T, Scheckner B, *et al*. The effects of lisdexamfetamine dimesylate on emotional lability in children 6 to 12 years of age with ADHD in a double-blind placebo-controlled trial. *J Atten Disord* 2014;**18**:123–32.

157. Findling RL, Childress AC, Cutler AJ, Gasior M, Hamdani M, Ferreira-Cornwell MC, Squires L. Efficacy and safety of lisdexamfetamine dimesylate in adolescents with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2011;**50**:395–405.

158. Coghill D, Banaschewski T, Lecendreux M, Soutullo C, Johnson M, Zuddas A, *et al*. European, randomized, phase 3 study of lisdexamfetamine dimesylate in children and adolescents with attention-deficit/hyperactivity disorder. *Eur Neuropsychopharmacol* 2013;**23**:1208–18.

159. Findling RL, Childress AC, Krishnan S, McGough JJ. Long-term effectiveness and safety of lisdexamfetamine dimesylate in school-aged children with attention-deficit/hyperactivity disorder. *CNS Spectr* 2008;**13**:614–20.

160. Wolraich ML, Greenhill LL, Pelham W, Swanson J, Wilens T, Palumbo D, *et al*. Randomized, controlled trial of OROS methylphenidate once a day in children with attention-deficit/hyperactivity disorder. *Pediatrics* 2001;**108**:883–92.

161. Palumbo DR, Sallee FR, Pelham WE Jr, Bukstein OG, Daviss WB, McDermott MP; CAT STUDY TEAM. Clonidine for attention-deficit/hyperactivity disorder: I. Efficacy and tolerability outcomes. *J Am Acad Child Adolesc Psychiatry* 2008;**47**:180–8.

162. Wang Y, Zheng Y, Du Y, Song DH, Shin YJ, Cho SC, *et al*. Atomoxetine versus methylphenidate in paediatric outpatients with attention deficit hyperactivity disorder: a randomized, double-blind comparison trial. *Aust N Z J Psychiatry* 2007;**41**:222–30.

163. Jain R, Babcock T, Burtea T, Dirks B, Adeyi B, Scheckner B, Lasser R. Efficacy of lisdexamfetamine dimesylate in children with attention-deficit/hyperactivity disorder previously treated with methylphenidate: a post hoc analysis. *Child Adolesc Psychiatry Ment Health* 2011;**5**:35.

164. Mahjoob M, Paul T, Carbone J, Bokadia H, Cardy RE, Kassam S, *et al*. Predictors of health-related quality of life in neurodivergent children: a systematic review. *Clin Child Fam Psychol Rev* 2024;**27**:91–129.

165. Danckaerts M, Sonuga-Barke EJ, Banaschewski T, Buitelaar J, Döpfner M, Hollis C, *et al*. The quality of life of children with attention deficit/hyperactivity disorder: a systematic review. *Eur Child Adolesc Psychiatry* 2010;**19**:83–105.

166. Wanni Arachchige Dona S, Badloe N, Sciberras E, Gold L, Coghill D, Le HND. The impact of childhood attention-deficit/hyperactivity disorder (ADHD) on children's health-related quality of life: a systematic review and meta-analysis. *J Atten Disord* 2023;**27**:598–611.

167. Coghill DR, Banaschewski T, Soutullo C, Cottingham MG, Zuddas A. Systematic review of quality of life and functional outcomes in randomized placebo-controlled studies of medications for attention-deficit/hyperactivity disorder. *Eur Child Adolesc Psychiatry* 2017;**26**:1283–307.

168. Galloway H, Newman E. Is there a difference between child self-ratings and parent proxy-ratings of the quality of life of children with a diagnosis of attention-deficit hyperactivity disorder (ADHD)? A systematic review of the literature. *Atten Defic Hyperact Disord* 2017;**9**:11–29.

169. Dey M, Paz Castro R, Haug S, Schaub MP. Quality of life of parents of mentally-ill children: a systematic review and meta-analysis. *Epidemiol Psychiatr Sci* 2019;**28**:563–77.

170. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, *et al*.; and the PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *J Am Med Assoc* 2018;**319**:388–96.

171. Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, *et al*.; QUADAS-C Group†. QUADAS-C: a tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med* 2021;**174**:1592–9.

172. Jangmo A, Stålhandske A, Chang Z, Chen Q, Almqvist C, Feldman I, *et al*. Attention-deficit/hyperactivity disorder, school performance, and effect of medication. *J Am Acad Child Adolesc Psychiatry* 2019;**58**:423–32.

173. 2014-001488-11. *Effects of Expectations, Medication and Placebo during the Quantified Behavior Test in Patients with Untreated ADHD and Substance Use Disorder*. 2014. URL: www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2014-001488-11 (accessed October 2023).

174. Bussmann JB, Ebner-Priemer U, Fahrenberg J. Ambulatory activity monitoring: progress in measurement of activity, posture, and specific motion patterns in daily life. *Eur Psychol* 2009;**14**:142–52.

Appendix 1 Literature search strategies

Appendix 1.1: Clinical-effectiveness searches

Resource	*N*
MEDLINE	100
EMBASE	143
PsycInfo	362
CINAHL	15
ClinicalTrails.gov	13
ICTRP	30
Total	663
−Duplicates	−155
To screen	508

Database: MEDLINE (MEDALL)

Host: Ovid

Data parameters: 1946–16 November 2023

Date of search: 17 November 2023

#	Searches	Results
1	(QbTest* or "Qb Test*" or "(Qb) Test*" or "Qb Mini*" or "QbMini*" or (("Quantified Behavior*" or "Quantified Behaviour*") adj5 test*) or QbTech).af.	68
2	(QbCheck* or "Qb Check*" or "(Qb) Check*").af.	1
3	(Nesplora* or "Giunti psychometrics").af.	21
4	(ARVO* or EFSim* or "EF Sim*" or EPELI or "Peili Vision Company").af.	1507
5	Attention Deficit Disorder with Hyperactivity/ or ADHD.af.	44,434
6	4 and 5	5
7	((motion* adj5 senso*) and (hyperactivity or ADHD)).ti,ab,kf.	6
8	1 or 2 or 3 or 6 or 7	99
9	NCT03368573.af. or (QUOTA and adhd).ti,kf. [QB test]	3
10	NCT02209116.af. or ((AQUA and ADHD) or AQUA2).ti,kf. [QB test]	5
11	NCT02473185.af. [QB test]	1
12	NCT02477280.af. [QB test]	0
13	NCT05846815.af. [ARVO Test]	0
14	9 or 10 or 11 or 12 or 13	9
15	8 or 14	100

Database: EMBASE

Host: Ovid

Data parameters: 1974–16 November 2023

Date of search: 17 November 2023

#	Searches	Results
1	(QbTest* or "Qb Test*" or "(Qb) Test*" or "Qb Mini*" or "QbMini*" or (("Quantified Behavior*" or "Quantified Behaviour*") adj5 test*) or QbTech).af.	89
2	(QbCheck* or "Qb Check*" or "(Qb) Check*").af.	3
3	(Nesplora* or "Giunti psychometrics").af.	24
4	(ARVO* or EFSim* or "EF Sim*" or EPELI or "Peili Vision Company").af.	62,360
5	Attention Deficit Disorder with Hyperactivity/ or ADHD.af.	53,113
6	4 and 5	21
7	((motion* adj5 senso*) and (hyperactivity or ADHD)).ti,ab,kf.	10
8	1 or 2 or 3 or 6 or 7	143
9	NCT03368573.af. or (QUOTA and adhd).ti,kf. [QB test]	3
10	NCT02209116.af. or ((AQUA and ADHD) or AQUA2).ti,kf. [QB test]	6
11	NCT02473185.af. [QB test]	1
12	NCT02477280.af. [QB test]	0
13	NCT05846815.af. [ARVO Test]	0
14	9 or 10 or 11 or 12 or 13	10
15	8 or 14	143

Database: PsycInfo

Host: Ovid

Data parameters: 1806–Current

Date of search: 17 November 2023

#	Searches	Results
1	(QbTest* or "Qb Test*" or "(Qb) Test*" or "Qb Mini*" or "QbMini*" or (("Quantified Behavior*" or "Quantified Behaviour*") adj5 test*) or QbTech).af.	126
2	(QbCheck* or "Qb Check*" or "(Qb) Check*").af.	6
3	(Nesplora* or "Giunti psychometrics").af.	85
4	(ARVO* or EFSim* or "EF Sim*" or EPELI or "Peili Vision Company").af.	5417
5	Attention Deficit Disorder with Hyperactivity/ or ADHD.af.	92,604
6	4 and 5	50
7	((motion* adj5 senso*) and (hyperactivity or ADHD)).ti,ab,kf.	100
8	1 or 2 or 3 or 6 or 7	362

#	Searches	Results
9	NCT03368573.af. or (QUOTA and adhd).ti,kf. [QB test]	0
10	NCT02209116.af. or ((AQUA and ADHD) or AQUA2).ti,kf. [QB test]	0
11	NCT02473185.af. [QB test]	0
12	NCT02477280.af. [QB test]	0
13	NCT05846815.af. [ARVO Test]	0
14	9 or 10 or 11 or 12 or 13	0
15	8 or 14	362

Database: CINAHL

Host: EBSCO*host*

Data parameters: 1981–Current

Date of search: 17 November 2023

#	Searches	Results
1	TI ((QbTest* or "Qb Test*" or "(Qb) Test*" or "Qb Mini*" or "QbMini*" or (("Quantified Behavior*" or "Quantified Behaviour*") N4 test*) or QbTech)) OR AB ((QbTest* or "Qb Test*" or "(Qb) Test*" or "Qb Mini*" or "QbMini*" or (("Quantified Behavior*" or "Quantified Behaviour*") N4 test*) or QbTech))	21
2	TI ((QbCheck* or "Qb Check*" or "(Qb) Check*")) OR AB ((QbCheck* or "Qb Check*" or "(Qb) Check*"))	1
3	TI ((Nesplora* or "Giunti psychometrics")) OR AB ((Nesplora* or "Giunti psychometrics"))	2
4	TI ((ARVO* or EFSim* or "EF Sim*" or EPELI or "Peili Vision Company")) OR AB ((ARVO* or EFSim* or "EF Sim*" or EPELI or "Peili Vision Company"))	16,100
5	TI (("Attention Deficit Disorder" or ADHD)) OR AB (("Attention Deficit Disorder" or ADHD))	92,604
6	S4 and S5	1
7	TI (((motion* N4 senso*) and (hyperactivity or ADHD))) OR AB (((motion* N4 senso*) and (hyperactivity or ADHD)))	1
8	S1 OR S2 OR S3 OR S6 OR S7	26
9	S1 OR S2 OR S3 OR S6 OR S7 [remove MEDLINE studies]	15

Database: ClinicalTrials.gov

Host: www.clinicaltrials.gov/ct2/results/refine?show_xprt=Y

Date of search: 17 November 2023

13 Studies found for: (QBTest OR "QB Test" OR QBMini OR "QB Mini" OR QBCheck OR "Qb Check" OR Nesplora OR ARVO OR EFSim OR "EF Sim" OR EPELI)

Database: WHO ICTRP

Host: https://trialsearch.who.int/Default.aspx

Date of search: 17 November 2023

30 Studies found for: (QBTest OR "QB Test" OR QBMini OR "QB Mini" OR QBCheck OR "Qb Check" OR Nesplora OR ARVO OR EFSim OR "EF Sim" OR EPELI)

Appendix 1.2: Supplemental cost-effectiveness searches

Resource	*N*
MEDLINE	491
EMBASE	319
PsycInfo	284
EconLit	5
Total	1099
−Duplicates	−470
Total to screen	629

EconLit, American Economic Association electronic bibliography.

Database: MEDLINE (MEDALL)

Host: Ovid

Data parameters: 1946–Current

Date of search: 12 February 2024

#	Searches	Results
1	*Attention Deficit Disorder with Hyperactivity/	29,998
2	((attention and deficit and disorder and hyperact*) or adhd).ti,ab,kf.	41,992
3	1 or 2	46,558
4	*economics/ or exp *"costs and cost analysis"/	90,572
5	((economic$ or cost or costs or costly or costing or budget*) adj3 (effect* or utility or minimisation or consequence or analysis or evaluat* or model or impact*)).ti,ab,kf.	274,166
6	("decision tree" or Markov or "semi Markov" or "partitioned adj2 survival" or "discrete event" or "conceptual* adj2 model*" or (decision adj2 model*) or "outcome model*" or "causal model*" or (simulat* adj2 model*) or QALY*).ti,ab,kf.	113,692
7	4 or 5 or 6	430,699
8	3 and 7	491

Database: EMBASE

Host: Ovid

Data parameters: 1980–Current

Date of search: 12 February 2024

#	Searches	Results
1	*attention deficit hyperactivity disorder/	5557
2	((attention and deficit and disorder and hyperact*) or adhd).ti,ab,kf.	59,518
3	1 or 2	59,716
4	*economic evaluation/ or *health economics/	23,367
5	((economic$ or cost or costs or costly or costing or budget*) adj3 (effect* or utility or minimisation or consequence or analysis or evaluat* or model or impact*)).ti,ab,kf.	376,011
6	("decision tree" or Markov or "semi Markov" or "partitioned adj2 survival" or "discrete event" or "conceptual* adj2 model*" or (decision adj2 model*) or "outcome model*" or "causal model*" or (simulat* adj2 model*) or QALY*).ti,ab,kf.	147,640
7	4 or 5 or 6	497,249
8	3 and 7	581
9	Limit 8 to EMBASE	319

Database: PsycInfo

Host: Ovid

Data parameters: 1908–Current

Date of search: 12 February 2024

#	Searches	Results
1	*Attention Deficit Disorder with Hyperactivity/	28,552
2	((attention and deficit and disorder and hyperact*) or adhd).ti,ab,kf.	39,303
3	1 or 2	40,289
4	((economic$ or cost or costs or costly or costing or budget*) adj3 (effect* or utility or minimisation or consequence or analysis or evaluat* or model or impact*)).ti,ab,kf.	35,826
5	("decision tree" or Markov or "semi Markov" or "partitioned adj2 survival" or "discrete event" or "conceptual* adj2 model*" or (decision adj2 model*) or "outcome model*" or "causal model*" or (simulat* adj2 model*) or QALY*).ti,ab,kf.	21,271
6	4 or 5	55,090
7	3 and 6	284

Database: EconLit

Host: EBSCO*host*

Data parameters: 1981–Current

Date of search: 12 February 2023

#	Searches	Results
1	TI (("Attention Deficit Disorder with Hyperactivity" or ADHD)) OR AB (("Attention Deficit Disorder with Hyperactivity" or ADHD))	105
2	TI (((economic* or cost or costs or costly or costing or budget*) N2 (effect* or utility or minimisation or consequence or analysis or evaluat* or model or impact*))) OR AB (((economic* or cost or costs or costly or costing or budget*) N2 (effect* or utility or minimisation or consequence or analysis or evaluat* or model or impact*)))	78,438
3	TI (("decision tree" or Markov or "semi Markov" or "partitioned N1 survival" or "discrete event" or "conceptual* N1 model*" or (decision N1 model*) or "outcome model*" or "causal model*" or (simulat* N1 model*) or QALY*)) OR AB (("decision tree" or Markov or "semi Markov" or "partitioned N1 survival" or "discrete event" or "conceptual* N1 model*" or (decision N1 model*) or "outcome model*" or "causal model*" or (simulat* N1 model*) or QALY*))	19,172
S4	S2 or S3	95,519
S5	S1 AND S4	8
S6	S1 and S4 [remove MEDLINE studies]	5

Appendix 2 Tables of included, ongoing or excluded studies

TABLE 33 Studies included in the review showing primary and secondary reports

Study name	Primary report	Secondary reports	Identified from
NR	Sharma A. SB. Evaluation of the role of QbTest in attention deficit hyperactivity disorder. *Archives of Disease in Childhood* 2009;**94**:A72	None	Checking references of included studies
NR	Hamadache SH, Kathrin Labarga, Sara Zaplana Gunther, Thomas. Is the QbMini a valid instrument for ADHD assessment? [References]. DP – Aug 2021. *Journal of Attention Disorders* 2021;**25**(10):1384–94	Labarga SZH, Kathrin Hamadache, Salsabil Gunther, Thomas. Validation of the QbMini Test to diagnose Attention Deficit and Hyperactivity Disorder (ADHD) in 5-year-old children. *Zeitschrift fur Neuropsychologie* 2019;**30**(3):149–56 Gunther TL, S. V. N. Z. Hoberg, K. First validation of the QbMini to measure symptoms of ADHD in 5-year-old children. *ADHD Attention Deficit and Hyperactivity Disorders* 2017;**9**(1 Supplement):S15	Main searches
NR	Hult NK, Josefin Kadesjo, Bjorn Gillberg, Christopher Billstedt, Eva. ADHD and the QbTest: Diagnostic Validity of QbTest. *Journal of Attention Disorders* 2018;**22**(11):1074–80	None	Main searches
NR	Ulberstad FB, Hans Chavanon, Mira-Lynn Knollmann, Martin Wiley, James Christiansen, Hanna Thorell, Lisa B. Objective measurement of attention deficit hyperactivity disorder symptoms outside the clinic using the QbCheck: reliability and validity. *International Journal of Methods in Psychiatric Research* 2020;**29**(2):e1822	None	Main searches
NR	Adamou MJ, Sarah L. Marks, Laura Lowe, Deborah. Efficacy of continuous performance testing in adult ADHD in a clinical sample using QbTest. *Journal of Attention Disorders* 2022;**26**(11):1483–91	None	Main searches
NR	Bijlenga DU, Fredrik Thorell, Lisa B. Christiansen, Hanna Hirsch, Oliver Kooij, J. J. Sandra. Objective assessment of attention-deficit/hyperactivity disorder in older adults compared with controls using the QbTest. *International Journal of Geriatric Psychiatry* 2019;**34**(10):1526–33	None	Main searches
NR	Brunkhorst-Kanaan NV, Moritz Kittel-Schneider, Sarah Vainieri, Isabella Reif, Andreas Grimm, Oliver. The quantified behavioral test – a confirmatory test in the diagnostic process of adult ADHD? *Frontiers in Psychiatry* 2020;**11**:216	None	Main searches

continued

TABLE 33 Studies included in the review showing primary and secondary reports (*continued*)

Study name	Primary report	Secondary reports	Identified from
NR	Edebol HH, Lars Holmberg, Ebba Gustafsson, Stig-Arne Norlander, Torsten. In search for objective measures of hyperactivity, impulsivity and inattention in adult attention deficit hyperactivity disorder using the Quantified Behavior Test Plus. Europe's *Journal of Psychology* 2011;**7**(3):443–57	None	Checking included studies in systematic reviews
NR	Edebol HH, Lars Norlander, Torsten. Measuring adult attention deficit hyperactivity disorder using the Quantified Behavior Test Plus. *Psychology Journal* 2013;**2**(1):48–62	None	Main searches
NR	Edebol HH, Lars Norlander, Torsten. Objective measures of behavior manifestations in adult ADHD and differentiation from participants with bipolar II disorder, borderline personality disorder, participants with disconfirmed ADHD as well as normative participants. *Clinical Practice and Epidemiology in Mental Health* 2012;8134–43	None	Main searches
NR	Groom MJY, Zoe Hall, Charlotte L. Gillott, Alinda Hollis, Chris. The incremental validity of a computerised assessment added to clinical rating scales to differentiate adult ADHD from autism spectrum disorder. *Psychiatry Research* 2016;**243**:168–73	None	Main searches
NR	Johansson VNS, Eva Kuja-Halkola, Ralf Lundstrom, Sebastian Durbeej, Natalie Anckarsater, Henrik Lichtenstein, Paul Hellner, Clara. The quantified behavioral test failed to differentiate ADHD in Adolescents with neurodevelopmental problems. *Journal of Attention Disorders* 2021;**25**(3):312–21	None	Main searches
AQUA	Hollis CH, Hall Charlotte L., Guo Boliang, James Marilyn, Boadu Janet, Groom Madeleine J., Brown Nikki, Kaylor-Hughes Catherine, Moldavsky Maria, Valentine Althea Z, Walker Gemma M, Daley David, Sayal Kapil, Morriss Richard. 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* 2018;**59**(12):1298–308	Hall CLV, Althea Z. Walker, Gemma M. Ball, Harriet M. Cogger, Heather Daley, David Groom Madeleine J,. Sayal Kapil, Hollis Chris. Study of user experience of an objective test (QbTest) to aid ADHD assessment and medication management: a multi-methods approach. *BMC Psychiatry* 2017;**17**(1):66 Hall CLW, Walker Gemma M, Valentine Althea Z., Guo Boliang, Kaylor-Hughes Catherine, James Marilyn, Daley David, Sayal Kapil, Hollis Chris. Protocol investigating the clinical utility of an objective measure of activity and attention (QbTest) on diagnostic and treatment decision-making in children and young people with ADHD – 'Assessing QbTest Utility in ADHD' (AQUA): a randomised controlled trial. *BMJ Open* 2014;**4**(12):e006838 ISRCTN11727351. 2016. Comparing the effects of providing clinicians and patients with the results of an objective measure of activity and attention (QbTest) versus usual care on diagnostic and treatment decision making in children and young people with ADHD. www.isrctn.com/ISRCTN11727351 (accessed November 2023) NCT02209116. 2014. Assessing QbTest utility in ADHD: a randomised controlled trial. https://clinicaltrials.gov/show/NCT02209116 (accessed November 2023)	Main searches

TABLE 33 Studies included in the review showing primary and secondary reports (*continued*)

Study name	Primary report	Secondary reports	Identified from
NR	Pettersson RS, Staffan Nilsson, Kent W. Diagnosing ADHD in adults: an examination of the discriminative validity of neuropsychological tests and diagnostic assessment instruments. *Journal of Attention Disorders* 2018;**22**(11):1019–31	None	Main searches
NR	Söderström SP, Richard Nilsson, Kent W. Quantitative and subjective behavioural aspects in the assessment of attention-deficit hyperactivity disorder (ADHD) in adults. *Nordic Journal of Psychiatry*. 2014;**68**(1):30–7	None	Main searches
NR	Stevanovic DN, Salmir Doric, Ana Wentz, Elisabet Knez, Rajna. The structure and diagnostic accuracy of the QbTest in pediatric ADHD: a retrospective clinical study. *Journal of Attention Disorders* 2023;**27**(11):1296–305	None	Main searches
NR	Tallberg PR, Maria Wenhov, Lena Eliasson, Glen Gustafsson, Peik. Incremental clinical utility of continuous performance tests in childhood ADHD – an evidence-based assessment approach. *Scandinavian Journal of Psychology* 2019;**60**(1):26–35	Gustafsson PT, P. Towards evidence-based assessments: Clinical utility of rating scales and cognitive test methods in diagnostic assessment and treatment evaluations in children and adolescents with attention-deficit/hyperactivity disorder. *ADHD Attention Deficit and Hyperactivity Disorders* 2017;**9**(1 Supplement):S15	Main searches
NR	Seesjärvi EP, Jasmin Aronen, Eeva T. Lipsanen, Jari Mannerkoski, Minna Hering, Alexandra Zuber, Sascha Kliegel, Matthias Laine, Matti Salmi, Juha. Quantifying ADHD symptoms in open-ended everyday life contexts with a new virtual reality task. *Journal of Attention Disorders* 2022;**26**(11):1394–411	None	Main searches
NR	Zulueta ADO, Unai Crespo-Eguilaz, Nerea Torrano, Fermin. Virtual reality-based assessment and rating scales in ADHD diagnosis. *Psicologia Educativa* 2019;**25**(1):13–22	None	Main searches
NA	Rufo-Campos, M., Cueto, E., Iriarte, Y., Rufo-Muñoz, M. Sensitivity study of a new diagnostic method for ADHD: Aula Nesplora. *Revue Neurologique* 2012;**54**(Suppl3):S67–93	None	Nesplora manufacturer's submission
NR	Emser TSJ, Blair A. Steele, J. Douglas Kooij, Sandra Thorell, Lisa Christiansen, Hanna. Assessing ADHD symptoms in children and adults: evaluating the role of objective measures. *Behavioral and Brain Functions* 2018;**14**:11	None	Main searches
FACT	Chitsabesan PH, C. L. Carter, L. A. Reeves, M. Mohammed, V. Beresford, B. Young, S. Kraam, A. Trowse, S. Wilkinson-Cunningham, L. Lennox, C. Using an objective computer task (QbTest) to aid the identification of attention deficit hyperactivity disorder (ADHD) in the Children and Young People Secure Estate (CYPSE): a feasibility randomised controlled trial. *BMJ Open* 2022;**12**(12):e064951	ISRCTN17402196. 2019. Feasibility trial to assess attention deficit hyperactivity disorder (ADHD) in the criminal justice system by using QbTest (a computer task). http://isrctn.com/ISRCTN17402196 (accessed November 2023). Lennox CH, C. L. Carter, L. A. Beresford, B. Young, S. Kraam, A. Brown, N. Wilkinson-Cunningham, L. Reeves, M. Chitsabesan, P. FACT: a randomised controlled trial to assess the feasibility of QbTest in the assessment process of attention deficit hyperactivity disorder (ADHD) for young people in prison – a feasibility trial protocol. *BMJ Open* 2020;**10**(1):035519	Main searches

continued

TABLE 33 Studies included in the review showing primary and secondary reports (*continued*)

Study name	Primary report	Secondary reports	Identified from
QUOTA	Williams LH, Charlotte L. Brown, Susan Guo, Boliang James, Marilyn Franceschini, Matilde Clarke, Julie Selby, Kim Vijayan, Hena Kulkarni, Neeta Brown, Nikki Sayal, Kapil Hollis, Chris Groom, Madeleine J. Optimising medication management in children and young people with ADHD using a computerised test (QbTest): a feasibility randomised controlled trial. *Pilot and Feasibility Studies* 2021;**7**(1):68	Hall CLB, Susan James, Marilyn Martin, Jennifer L. Brown, Nikki Selby, Kim Clarke, Julie Williams, Laura Sayal, Kapil Hollis, Chris Groom, Madeleine J. Consensus workshops on the development of an ADHD medication management protocol using QbTest: developing a clinical trial protocol with multidisciplinary stakeholders. *BMC Medical Research Methodology* 2019;**19**(1):126 Hall CLJ, Marilyn Brown, Sue Martin, Jennifer L. Brown, Nikki Selby, Kim Clarke, Julie Vijayan, Hena Guo, Boliang Sayal, Kapil Hollis, Chris Groom, Madeleine J. Protocol investigating the clinical utility of an objective measure of attention, impulsivity and activity (QbTest) for optimising medication management in children and young people with ADHD 'QbTest Utility for Optimising Treatment in ADHD' (QUOTA): a feasibility randomised controlled trial. *BMJ Open* 2018;**8**(2):e021104 ISRCTN69461593. 2018. QbTest utility for optimising treatment in ADHD (QUOTA). www.isrctn.com/ISRCTN69461593 (accessed November 2023) NCT03368573. 2017. QbTest utility for optimising treatment in ADHD (QUOTA). https://clinicaltrials.gov/show/NCT03368573 (accessed November 2023)	Main searches
NR	Hall Charlotte L, Selby Kim, Guo Boliang, Valentine Althea Z, Walker Gemma M, Hollis Chris,. 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* 2016;**21**(3):175–8	None	Main searches
NR	Pellegrini SM, Mike Lovett, Ella. The QbTest for ADHD assessment: Impact and implementation in child and adolescent mental health services. *Children and Youth Services Review* 2020;**114**:n.r.	None	Main searches
NR	Sharma RW, A. Lacey, S. Spiewakowski, D. Implementing QbTesting for ADHD: evaluating value in a DGH setting. *Archives of Disease in Childhood* 2022;**107**(Supplement 2):A70	None	Main searches
NR	Vogt CS, A. Assessments for attention-deficit hyperactivity disorder: use of objective measurements. *Psychiatrist* 2011;**35**(10):380–3	None	Main searches
NR	Catriona Humphreys, Lucy Sitton-Kent. Transforming ADHD care across the East Midlands: an evaluation Report. East Midlands Academic Health Network. 2018. URL: https://healthinnovation-em.org.uk/component/rsfiles/download-file/files?path=our-work%252Four-innovations%252FTransforming-ADHD-Care%252FFinal_Overall_Evaluation_Report_31May18.pdf&Itemid=1457 (accessed March 2024)	None	QbTest manufacturer's submission

TABLE 33 Studies included in the review showing primary and secondary reports (*continued*)

Study name	Primary report	Secondary reports	Identified from
NR	Caitlin McKenzie, Benjamin-Rose Ingall, [Dr] Charlotte Hall. Focus ADHD National Programme Evaluation. 2022. URL: https://healthinnovation-em.org.uk/component/rsfiles/download-file/files?path=our-work%-252Four-innovations%252FADH-D%2BFOCUS%2Bevaluation%2Bre-port%2B-%2BFINAL%2Bv.1.0%2B18.10.22.pdf&Itemid=1457 (accessed March 2024)	None	QbTest manufacturer's submission
NR	Peli Vision Oy. Research behind EFSim and feedback from pilot tests [unpublished report]. n.r.	None	Peili Vision manufacturer's submission

Note
Primary reports are the primary publication for the study and are used to refer to that study throughout text and tables.

TABLE 34 Ongoing studies that appear to meet inclusion criteria for the review

Author	Identified from	Test	Study details	Estimated completion date
NCT05846815 (Sponsors: Peili Vision).[71]	Our searches	ARVO 2.0	Cross-over RCT in Finland, aiming to assess the performance and safety of web-based ARVO 2.0 for evaluating possible ADHD symptoms in children aged 8–13 years with ADHD and typically developing children of the same age. Comparison of results from ARVO to results from Conners CPT	June 2024
Peili Vision	Manufacturer's submission	EF Sim	Several pilots are being set up for spring 2024 in the UK using learnings from rolling out EF Sim in Finland to implement in the UK as part of an early triage tool	Not reported

TABLE 35 Studies excluded at full-text screening from the identification of studies via databases and registers

Report	Reason for exclusion
2014-001488-11. Effects of expectations, medication and placebo during the quantified behavior test in patients with untreated ADHD and substance use disorder. 2014. URL: www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2014-001488-11 (accessed October 2023)	Not an evaluation of the test
Areces DD, Julie Garcia, Trinidad Gonzalez-Castro, Paloma Rodriguez, Celestino. Analysis of cognitive and attentional profiles in children with and without ADHD using an innovative virtual reality tool. *PLOS ONE* 2018;**13**(8):e0201039	Not an evaluation of the test
Areces DG, Trinidad Cueli, Marisol Rodriguez, Celestino. Is a virtual reality test able to predict current and retrospective ADHD symptoms in adulthood and adolescence? *Brain Sciences* 2019;**9**(10):n.r.	Does not report on one of the outcomes of interest
Areces DR, Celestino Garcia, Trinidad Cueli, Marisol Gonzalez-Castro, Paloma. Efficacy of a continuous performance test based on virtual reality in the diagnosis of ADHD and its clinical presentations. *Journal of Attention Disorders* 2018;**22**(11):1081–91	Does not report on one of the outcomes of interest
Baader AK, B. Brunkhorst-Kanaan, N. Kittel-Schneider, S. Reif, A. Grimm, O. A within-sample comparison of two innovative neuropsychological tests for assessing adhd. *Brain Sciences* 2021;**11**(1):1–21	Does not report on one of the outcomes of interest

continued

TABLE 35 Studies excluded at full-text screening from the identification of studies via databases and registers (*continued*)

Report	Reason for exclusion
Baader AK, B. Brunkhorst-Kanaan, N. Kittel-Schneider, S. Reif, A. Grimm, O. A within-sample comparison of two innovative neuropsychological tests for assessing adhd. Brain Sciences 2021;**11**(1):1–21	Duplicate report
Baader AK, B. Brunkhorst-Kanaan, N. Kittel-Schneider, S. Reif, A. Grimm, O. P.632 A within-sample comparison of two innovative neuropsychological tests for diagnosing ADHD. European neuropsychopharmacology 2020;**40**(Supplement 1):S355–6	Does not report on one of the outcomes of interest
Bellato AH, Charlotte L. Groom, Madeleine J. Simonoff, Emily Thapar, Anita Hollis, Chris Cortese, Samuele. Practitioner review: clinical utility of the QbTest for the assessment and diagnosis of attention-deficit/hyperactivity disorder – a systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 2023;n.r.	SR
Berger I, Slobodin O, Cassuto H. Usefulness and validity of continuous performance tests in the diagnosis of attention-deficit hyperactivity disorder children. *Archives of Clinical Neuropsychology* 2017;**32**(1):81–93	Did not report on test of interest
Bhattacharyya NS, S. Banerjee, A. Ghosh, R. Sinha, O. Das, N. Gayen, *et al.* Integration of electroencephalogram (EEG) and motion tracking sensors for objective measure of attention-deficit hyperactivity disorder (MAHD) in pre-schoolers. *The Review of Scientific Instruments* 2022;**93**(5):054101	Did not report on test of interest
Bijlenga DJ, M. Gehlhaar, S. K. Sandra Kooij, J. J. Objective QbTest and subjective evaluation of stimulant treatment in adult attention deficit-hyperactivity disorder. *European Psychiatry: The Journal of the Association of European Psychiatrists* 2015;**30**(1):179–185	Does not report on one of the outcomes of interest
Brancaccio RK, J. Ayearst, L. E. Using wearables and artificial intelligence to improve diagnostic decisions and treatment in youth with attention-deficit hyperactivity disorder. *Innovations in Clinical Neuroscience* 2021;**18**(10–12 SUPPL):S2–3	Did not report on test of interest
Brocki KCT, Carin M. Bohlin, Gunilla. CPT performance, motor activity, and continuous relations to ADHD symptom domains: a developmental study. *European Journal of Developmental Psychology* 2010;**7**(2):178–197	Not an evaluation of the test
Camacho-Conde JAC, Gema. Attentional profile of adolescents with ADHD in virtual-reality dual execution tasks: a pilot study. *Applied Neuropsychology: Child* 2022;**11**(1):81–90	Not an evaluation of the test
Cedergren K, Östlund S, Åsberg Johnels J, Billstedt E, Johnson M. Monitoring medication response in ADHD: what can continuous performance tests tell us? *European Archives of Psychiatry and Clinical Neuroscience* 2022;**272**(2):291–99	Duplicate report
Cedergren K, Östlund S, Åsberg Johnels J, Billstedt E, Johnson M. Monitoring medication response in ADHD: What can continuous performance tests tell us? *European Archives of Psychiatry and Clinical Neuroscience* 2022;**272**(2): 291–99	Not an evaluation of the test
Climent GR, Celestino Garcia, Trinidad Areces, Debora Mejias, Miguel Aierbe, Amaia Moreno, Marta Cueto, Eduardo Castella, Judit Feli Gonzalez, Mari. New virtual reality tool (Nesplora Aquarium) for assessing attention and working memory in adults: a normative study. *Applied Neuropsychology Adult* 2021;**28**(4):403–15	Does not report on one of the outcomes of interest
Climent GR, Celestino Garcia, Trinidad Areces, Debora Mejias, Miguel Aierbe, Amaia Moreno, Marta Cueto, Eduardo Castella, Judit Feli Gonzalez, Mari. New virtual reality tool (Nesplora Aquarium) for assessing attention and working memory in adults: a normative study. *Applied Neuropsychology Adult* 2021;**28**(4): 403–15	Duplicate report
Cole E. Qb test improves diagnosis of attention deficit disorder. *Nursing Children and Young People* 2015;**27**(2):10–11	Not a primary study or SR
Diaz-Orueta U. Advances in neuropsychological assessment of attention: from initial computerized continuous performance tests to AULA. *The Role of Technology in Clinical Neuropsychology* 2017;**103**:n.r.	Not a primary study or SR
Diaz-Orueta UFF, M. A. Morillo-Rojas, M. D. Climent, G. [Efficacy of lisdexamphetamine to improve the behavioural and cognitive symptoms of attention deficit hyperactivity disorder: treatment monitored by means of the AULA Nesplora virtual reality test]. *Eficacia de la lisdexanfetamina en la mejora sintomatica conductual y cognitiva del trastorno por deficit de atencion/ hiperactividad: tratamiento monitorizado mediante el test AULA Nesplora de realidad virtual* 2016;**63**(1):19–27	Not an evaluation of the test
Diaz-Orueta UGL, Cristina Crespo-Eguilaz, Nerea Sanchez-Carpintero, Rocio Climent, Gema Narbona, Juan. AULA virtual reality test as an attention measure: convergent validity with Conners' Continuous Performance Test. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence* 2014;**20**(3): 328–42	Does not report on one of the outcomes of interest
DRKS00030766. Identification of objective markers for the evaluation and prediction of the treatment of children and adolescents with ADHD. 2022. URL: http://drks.de/search/en/trial/DRKS00030766 (accessed October 2024)	Not an evaluation of the test

TABLE 35 Studies excluded at full-text screening from the identification of studies via databases and registers (*continued*)

Report	Reason for exclusion
Faraone SV, Banaschewski T, Coghill D, Zheng Y, Biederman J, Bellgrove MA, *et al.* The World Federation of ADHD International Consensus Statement: 208 evidence-based conclusions about the disorder. *Neuroscience and Biobehavioral Reviews* 2021;128789–818	Background
Fernandez-Martin PL, J. J. Rodriguez-Herrera, R. Canovas, R. Martinez De Salazar, A. Cobos-Sanchez, L. Sanchez-Santed, F. Flores, P. Dimensional analysis of adolescent attention-deficit/hyperactivity disorder. *European Psychiatry* 2020;**63**(Supplement 1): S677	Not an evaluation of the test
Fernandez-Martin PRH, Rocio Canovas, Rosa Diaz-Orueta, Unai Martinez de Salazar, Alma Flores, Pilar. Data-driven profiles of attention-deficit/hyperactivity disorder using objective and ecological measures of attention, distractibility, and hyperactivity. *European Child and Adolescent Psychiatry* 2023 [Epub ahead of print]	Does not report on one of the outcomes of interest
Fischer SK, M. Lehfeld, H. Niklewski, G. Brandl, C. Influence of depressive symptoms on Qb test performance in adult ADHD patients. *ADHD Attention Deficit and Hyperactivity Disorders* 2015;**7**(Supplement. 1):S77	Does not report on one of the outcomes of interest
Garcia Murillo LC, S. Anderson, D. Di Martino, A. Castellanos, F. Meta-analysis of locomotor activity measures in attention-deficit/hyperactivity disorder. *European Child and Adolescent Psychiatry* 2015;**24**(1 SUPPL. 1):S154	Did not report on test of interest
Hager LAO, Geir Danielsen, Maria Billstedt, Eva Gillberg, Christopher Johnels, Jakob Asberg. Indexing executive functions with test scores, parent ratings and ERPs: how do the measures relate in children vs. adolescents with ADHD? [References]. DP – Feb 17, 2020. *Neuropsychiatric Disease and Treatment* 2020;16465–477	Not an evaluation of the test
Hall CLB, A. Kirk, J. D. Hollis, C. 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 (United Kingdom)* 2023;**33**(9):259–64	Background
Hall CLV, Althea Z. Groom, Madeleine J. Walker, Gemma M. Sayal, Kapil Daley, David Hollis, Chris. 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 and Adolescent Psychiatry* 2016;**25**(7): 677–99	SR
Hall CLW, G. M. Valentine, A. Z. Correction. Protocol investigating the clinical utility of an objective measure of activity and attention (QbTest) on diagnostic and treatment decision-making in children and young people with ADHD – 'Assessing QbTest Utility in ADHD' (AQUA): a randomised controlled trial. *BMJ Open* 2015;**5**(5): e006838corr006831	Erratum
Hall CLW, G. M. Valentine, A. Z. Erratum: Protocol investigating the clinical utility of an objective measure of activity and attention (QbTest) on diagnostic and treatment decision-making in children and young people with ADHD - 'Assessing QbTest Utility in ADHD' (AQUA): a randomised controlled trial [BMJ Open (2014) 4 (e006838))] *BMJ Open* 2015;**5**(5):006838corr006831	Erratum
Hall CLW, G. M. Valentine, A. Z. Erratum: Protocol investigating the clinical utility of an objective measure of activity and attention (QbTest) on diagnostic and treatment decision-making in children and young people with ADHD-'Assessing QbTest Utility in ADHD'(AQUA): a randomised controlled trial [BMJ Open (2014) 4 (e006838)]. *BMJ Open* 2016;**6**(1):e006838	Erratum
Hamadache SH, Kathrin Labarga, Sara Zaplana Gunther, Thomas. Is the QbMini a valid instrument for ADHD assessment? [References]. DP – Aug 2021. *Journal of Attention Disorders* 2021;**25**(10):1384–94	Duplicate report
Hirsch OC, Hanna. Factorial structure and validity of the quantified behavior test plus (Qb+©). *Assessment* 2017;**24**(8):1037–49	Does not report on one of the outcomes of interest
Iriarte YDO, Unai Cueto, Eduardo Irazustabarrena, Paula Banterla, Flavio Climent, Gema. AULA – advanced virtual reality tool for the assessment of attention: normative study in Spain. *Journal of Attention Disorders* 2016;**20**(6):542–68	Does not report on one of the outcomes of interest
Jansson LL, Monica Ostlund, Mona Domingo, Blanca. Effects of one single-dose methylphenidate compared to one single-dose placebo on QbTest performance in adults with untreated ADHD: a randomized controlled trial. *BMC Psychiatry* 2023;**23**(1):762	Not an evaluation of the test
Jylkka JR, Liisa Merzon, Liya Kangas, Suvi Kliegel, Matthias Zuber, Sascha Hering, Alexandra Laine, Matti Salmi, Juha. Assessment of goal-directed behavior and prospective memory in adult ADHD with an online 3D videogame simulating everyday tasks. *Scientific Reports* 2023;**13**(1):9299	Did not report on test of interest

continued

TABLE 35 Studies excluded at full-text screening from the identification of studies via databases and registers (*continued*)

Report	Reason for exclusion
Knez RS, Dejan Nasic, Salmir Doric, Ana Wentz, Elisabet. The Impact of methylphenidate on QbTest performance of children with ADHD: a Retrospective Clinical Study. *Neuropsychiatric Disease and Treatment* 2021;1719–32	Not an evaluation of the test
Kooij JJS, Bijlenga D, Salerno L, Jaeschke R, Bitter I, Balázs J, *et al.* Updated European Consensus Statement on diagnosis and treatment of adult ADHD. *European Psychiatry* 2019;5614–34	Background
Kuhle H. J., Lefering R. Video-assisted behavior observation as a tool for methylphenidate dose finding in ADHD: longer term outcome. *Neuropediatrics* 2013;**44**(2):PS20–1146	Did not report on test of interest
Kvitland LRJ, K. Achkhan, H. Berg, T. Dahlen, N. R. Kirkholt, G. M. Koren, K. N. Naess, M. F. The CPT-3 vs. the QB-test: A task-oriented computerized assessment of attention-related problems in out-patient children: will diagnosis predict the atypical attention scores? *ADHD Attention Deficit and Hyperactivity Disorders* 2019;**11**(1 Supplement):S18–9	Does not report on one of the outcomes of interest
Lindhiem OG, Mayank Shaaban, Sam Mak, Kristie J. Chikersal, Prerna Feldman, Jamie Harris, Jordan L. Objective measurement of hyperactivity using mobile sensing and machine learning: pilot study. *JMIR Formative Research* 2022;**6**(4):e35803	Did not report on test of interest
Lohman MD, Blanca Ostlund, Mona Jansson, Lennart. Contrasting expectancy effects with objective measures in adults with untreated ADHD during QbTest. *Scandinavian Journal of Psychology* 2023;**64**(4):461–69	Not an evaluation of the test
Luderer MS, Johanna Gerhardt, Sarah Hoffmann, Sabine Vollstadt-Klein, Sabine Reif, Andreas Sobanski, Esther. Drinking alcohol to cope with hyperactive ADHD? Self-reports vs. continuous performance test in patients with ADHD and/or alcohol use disorder. *Frontiers in Psychiatry* 2023:141112843	Does not report on one of the outcomes of interest
Manning D, Olety S. Qb technology – evaluating its use in ADHD diagnosis within a child and adolescent mental health service. *European Psychiatry* 2021;**64**(Supplement 1):S225	Does not report on one of the outcomes of interest
Marshall P, Hoelzle J, Nikolas M. Diagnosing attention-deficit/hyperactivity disorder (ADHD) in young adults: a qualitative review of the utility of assessment measures and recommendations for improving the diagnostic process. *The Clinical Neuropsychologist* 2021;**35**(1):165–98	SR
Martin-Key NA, Stevenson A, Roy P. Investigating the clinical utility of the combined use of objective and subjective measures of ADHD during treatment optimization. *Journal of Clinical Psychopharmacology* 2022;**42**(2):146–53	Does not report on one of the outcomes of interest
NCT02473185. Effects of expectation, medication and placebo on objective and self-rated performance during the QbTest. 2015. URL: https://clinicaltrials.gov/show/NCT02473185 (accessed October 2023)	Not an evaluation of the test
NCT02477280. Effects of expectation, medication and placebo on objective and self-rated performance. 2015. URL: https://clinicaltrials.gov/show/NCT02477280 (accessed October 2023)	Not an evaluation of the test
Nylander Elin, Sparding Timea, Floros Orestis, Ryden Eleonore, Landen Mikael, Hansen Stefan. The quantified behavioural test plus (qbtest+) in adult ADHD. *Nordic Psychology* 2022;**75**(1):20–34	Does not report on one of the outcomes of interest
Peñuelas-Calvo I, Jiang-Lin LK, Girela-Serrano B, Delgado-Gomez D, Navarro-Jimenez R, Baca-Garcia E, *et al.* Video games for the assessment and treatment of attention-deficit/hyperactivity disorder: a systematic review. *European Child and Adolescent Psychiatry* 2022;**31**(1):5–20	SR
Prasad V, Rezel-Potts E, White P, Downs J, Boddy *N*, Sayal K, *et al.* Use of healthcare services before diagnosis of attention-deficit/hyperactivity disorder: a population-based matched case-control study. *Archives of Disease in Childhood* 2023;**109**(1):46–51	Background
Puzzo IS, Ottilie Kelly, Rachel Greer, Ben Kumari, Veena Gujonsson, Gisli Young, Susan. Attention problems predict risk of violence and rehabilitative engagement in mentally disordered offenders. *Frontiers in Psychiatry* 2019;**10**:279	Not an evaluation of the test
Ramtvedt BE, Sundet K. Relationships between computer-based testing and behavioral ratings in the assessment of attention and activity in a pediatric ADHD stimulant crossover trial. *The Clinical Neuropsychologist* 2014;**28**(7):1146–61	Does not report on one of the outcomes of interest
Reh VS, Martin Lam, Le Schimmelmann, Benno G. Hebebrand, Johannes Rief, Winfried Christiansen, Hanna. Behavioral assessment of core ADHD symptoms using the QbTest. *Journal of Attention Disorders* 2015;**19**(12):1034–45	Does not report on one of the outcomes of interest

TABLE 35 Studies excluded at full-text screening from the identification of studies via databases and registers (*continued*)

Report	Reason for exclusion
Rodriguez CA, Debora Garcia, Trinidad Cueli, Marisol Gonzalez-Castro, Paloma. Comparison between two continuous performance tests for identifying ADHD: Traditional vs. virtual reality. *International Journal of Clinical and Health Psychology* 2018;**18**(3):254–63	Does not report on one of the outcomes of interest
Santosh P, Cortese S, Hollis C, Bölte S, Daley D, Coghill D, *et al.* Remote assessment of ADHD in children and adolescents: Recommendations from the European ADHD guidelines group following the clinical experience during the Covid-19 pandemic. *European Child and Adolescent Psychiatry* 2023;**32**(6):921–35	Background
Sanwo O, Huzair H. What's new in attention-deficit/hyperactivity disorder: updates on assessment and management. *Paediatrics and Child Health (United Kingdom)* 2022;**32**(8):282–89	Background
Schworer M, Jascenoka J, Nitkowski D, Petermann F, Vasileva M, Petermann U. Deficits in executive functions of children with ADHD: clinical validity of a diagnostic instrument for ADHD in children and adolescents (ADHS-KJ). *Kindheit und Entwicklung: Zeitschrift fur Klinische Kinderpsychologie* 2019;**28**(2):96–105	Did not report on test of interest
Selaskowski BA, Laura Marie Wiebe, Annika Kannen, Kyra Aslan, Behrem Gerding, Thiago Morano Sanchez, Dario Ettinger, Ulrich Kolle, Markus Lux, Silke Philipsen, Alexandra Braun, Niclas. Gaze-based attention refocusing training in virtual reality for adult attention-deficit/hyperactivity disorder. *BMC Psychiatry* 2023;**23**:74	Did not report on test of interest
Slobodin O, Davidovitch M. Gender differences in objective and subjective measures of ADHD among clinic-referred children. *Frontiers in Human Neuroscience* 2019;**13**:441	Did not report on test of interest
Stevanovic DW, Elisabet Nasic, Salmir Knez, Rajna. ASD with ADHD vs. ASD and ADHD alone: a study of the QbTest performance and single-dose methylphenidate responding in children and adolescents. *BMC Psychiatry* 2022;**22**(1):282	Does not report on one of the outcomes of interest
Stuart E, Torres S, Gutierrez B. B - 04 Evaluating the efficacy of a virtual reality neuropsychological assessment in detecting ADHD subtypes. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists* 2023;**38**(7):1368	Does not report on one of the outcomes of interest
Valentine AZ, Brown BJ, Groom MJ, Young E, Hollis C, Hall CL. A systematic review evaluating the implementation of technologies to assess, monitor and treat neurodevelopmental disorders: a map of the current evidence. *Clinical Psychology Review* 2020;**80**:101870	SR
Vogt C. Clinical conundrums when integrating the QbTest into a standard ADHD assessment of children and young people. *Neuropediatrics* 2021;**52**(3):155–62	Background
Wang XQ, Albitos PJ, Hao YF, Zhang H, Yuan LX, Zang YF. A review of objective assessments for hyperactivity in attention deficit hyperactivity disorder. *Journal of Neuroscience Methods* 2022;**370**:109479	Not a primary study or SR
Wehmeier P, Bender M. ADHD core symptom assessment in adults with ADHD, depression, addiction or borderline personality disorder using the Qb test. *ADHD Attention Deficit and Hyperactivity Disorders* 2017;**9**(1 Supplement):S13	Does not report on one of the outcomes of interest
Wehmeier P, Wolff J, Cabanas *N*, Bender M. ADHD core symptom assessment in adults with ADHD compared to adults with ADHD and comorbid borderline personality disorder using a computer-based continuous performance test (cb-CPT) combined with an infra-red motion-tracking device. *ADHD Attention Deficit and Hyperactivity Disorders* 2019;**11**(1 Supplement):S22	Does not report on one of the outcomes of interest
Wehmeier PM, Dittmann RW, Banaschewski T, Schacht A. Does stimulant pretreatment modify atomoxetine effects on core symptoms of ADHD in children assessed by quantitative measurement technology? *Journal of Attention Disorders* 2014;**18**(2):105–16	Not an evaluation of the test
Wehmeier PM, Schacht A, Ulberstad F, Lehmann M, Schneider-Fresenius C, Lehmkuhl G, *et al.* Does atomoxetine improve executive function, inhibitory control, and hyperactivity? Results from a placebo-controlled trial using quantitative measurement technology. *Journal of Clinical Psychopharmacology* 2012;**32**(5):653–60	Not an evaluation of the test
Wehmeier PMK, Laura Banaschewski, Tobias Dittmann, Ralf W. Schacht, Alexander. Does comorbid disruptive behavior modify the effects of atomoxetine on ADHD symptoms as measured by a continuous performance test and a motion tracking device? [References]. DP – Jul 2015. *Journal of Attention Disorders* 2015;**19**(7):591–602	Not an evaluation of the test
Wehrmann T, Jorg M. An objective measure of hyperactivity aspects with compressed webcam video. *Child and Adolescent Psychiatry and Mental Health* 2015;**945**	Did not report on test of interest

continued

TABLE 35 Studies excluded at full-text screening from the identification of studies via databases and registers (*continued*)

Report	Reason for exclusion
Williams LH, Charlotte L. Brown, Susan Guo, Boliang James, Marilyn Franceschini, Matilde Clarke, Julie Selby, Kim Vijayan, Hena Kulkarni, Neeta Brown, Nikki Sayal, Kapil Hollis, Chris Groom, Madeleine J. Correction to: optimising medication management in children and young people with ADHD using a computerised test (QbTest): a feasibility randomised controlled trial. *Pilot and Feasibility Studies* 2021;**7**(1):94	Erratum
Young SA, Nicoletta Asgeirsdottir, Bryndis Bjork Branney, Polly Beckett, Michelle Colley, William Cubbin, Sally Deeley, Quinton Farrag, Emad Gudjonsson, Gisli Hill, Peter Hollingdale, Jack Kilic, Ozge Lloyd, Tony Mason, Peter Paliokosta, Eleni Perecherla, Sri Sedgwick, Jane Skirrow, Caroline Tierney, Kevin van Rensburg, Kobus Woodhouse, Emma. Females with ADHD: an expert consensus statement taking a lifespan approach providing guidance for the identification and treatment of attention-deficit/ hyperactivity disorder in girls and women. *BMC Psychiatry* 2020;**20**:404	Background
Young SA, Philip Lloyd, Tony Absoud, Michael Arif, Muhammad Colley, William Andrew Cortese, Samuele Cubbin, Sally Doyle, Nancy Morua, Susan Dunn Ferreira-Lay, Philip Gudjonsson, Gisli Ivens, Valerie Jarvis, Christine Lewis, Alexandra Mason, Peter Newlove-Delgado, Tamsin Pitts, Mark Read, Helen van Rensburg, Kobus Zoritch, Bozhena Skirrow, Caroline. Failure of healthcare provision for attention-deficit/hyperactivity disorder in the United Kingdom: a consensus statement. *Frontiers in Psychiatry* 2021;**12**:649399	Background

TABLE 36 Studies excluded at full-text screening from checking manufacturer's websites

Study details	Manufacturer's website	Reason for exclusion
Lis S, Baer N, Stein-en-Nosse C, Gallhofer B, Sammer G, Kirsch P. Objective measurement of motor activity during cognitive performance in adults with attention-deficit/ hyperactivity disorder. *Acta Psychiatrica Scandinavica* 2010 Oct;**122**(4):285–94	QbTech	Does not report on one of the outcomes of interest
Merzon L. Real-world goal-directed behavior reveals aberrant functional connectivity in children with ADHD. 2023	Peili Vision	Does not report on one of the outcomes of interest
Salmi J, Merzon L, Eräste T, Seesjärvi E, Huhdanpää H, Aronen ET, *et al.* Fluctuations of attention during self-paced naturalistic goal-directed behavior in attention-deficit/ hyperactivity disorder. *JAACAP Open* 2023 Dec 21.	Peili Vision	Does not report on one of the outcomes of interest
Merzon L, Pettersson K, Aronen ET, Huhdanpää H, Seesjärvi E, Henriksson L, *et al.* Eye movement behavior in a real-world virtual reality task reveals ADHD in children. *Scientific Reports* 2022 Nov 24;**12**(1):20308	Peili Vision	Does not report on one of the outcomes of interest
Seesjärvi E, Puhakka J, Aronen ET, Hering A, Zuber S, Merzon L, *et al.* EPELI: a novel virtual reality task for the assessment of goal-directed behavior in real-life contexts. *Psychological Research* 2023 Sep;**87**(6):1899–916	Peili Vision	Did not include population with suspected or confirmed ADHD
Rebon F, Altuna I, Lobo A, Salillas E, Climent G. Validity performance in the AULA Nesplora test	Nesplora	Does not report on one of the outcomes of interest
Teruel MA, Sanchis J, Ruiz-Robledillo N, Albaladejo-Blázquez N, Ferrer-Cascales R, Trujillo J. Measuring attention of ADHD patients by means of a computer game featuring biometrical data gathering. *Heliyon* 2024 Feb 23	Nesplora	Did not report on test of interest
Zakani Z, Moradi H, Ghasemzadeh S, Riazi M, Mortazavi F. The validity of a machine learning-based video game in the objective screening of attention deficit hyperactivity disorder in children aged 5 to 12 years. arXiv preprint arXiv:2312.11832. 2023 Dec 19	Nesplora	Did not report on test of interest
https://nesplora.com/investigaci%C3%B3n/head-mounted-display-vs.-computer-monitor-for-visual-attention-screening-a-comparative-study/	Nesplora	Did not report on test of interest

TABLE 37 Studies excluded at full-text screening from checking the studies included in systematic reviews

Study details	Reason for exclusion
Delgado-Gomez D, Peñuelas-Calvo I, Masó-Besga AE, VallejoOñate S, Tello IB, Duarte EA, *et al.* Microsoft kinect-based continuous performance test: an objective attention deficit hyperactivity disorder assessment. *J Med Internet Res* 2017;**19**(3):e79	Did not report on test of interest
Faraone SV, Newcorn JH, Antshel KM, Adler L, Roots K, Heller M. The groundskeeper gaming platform as a diagnostic tool for attention-defcit/hyperactivity disorder: sensitivity, specificity, and relation to other measures. *J Child Adolesc Psychopharmacol* 2016;**26**(8):672–85	
Heller MD, Roots K, Srivastava S, Schumann J, Srivastava J, Hale TS. A machine learning-based analysis of game data for attention deficit hyperactivity disorder assessment. *Games Health J* 2013;**2**(5):291–98	
Pollak Y, Weiss PL, Rizzo AA, Weizer M, Shriki L, Shalev RS *et al.* The utility of a continuous performance test embedded in virtual reality in measuring ADHD-related deficits. *J Dev Behav Pediatr* 2009;**30**(1):2–6	
Shaw R, Grayson A, Lewis V. Inhibition, ADHD, and computer games: the inhibitory performance of children with ADHD on computerized tasks and games. *J Atten Disord* 2005;**8**(4):160–168	
Eom, H., Kim, K. K., Lee, S., Hong, Y. J., Heo, J., Kim, J. J., Kim, E. Development of virtual reality continuous performance test utilizing social cues for children and adolescents with attention-deficit/hyperactivity disorder. *Cyberpsychol Behav Soc Netw* 2019;**22**(3):198–204. https://doi.org/1089/cyber.2018.0377	
Shema-Shiratzky, S., Brozgol, M., Cornejo-Thumm, P., Geva-Dayan, K., Rotstein, M., Leitner, Y., Hausdorff, J. M., Mirelman, A. Virtual reality training to enhance behavior and cognitive function among children with attention-deficit/hyperactivity disorder: brief report. *Dev Neurorehabil* 2018;**22**(6):431–36. https://doi.org/10.1080/17518423.2018.1476602	
Wehmeier PM, Schacht A, Wolff C, Otto WR, Dittmann RW, Banaschewski T. Neuropsychological outcomes across the day in children with attention-deficit/hyperactivity disorder treated with atomoxetine: results from a placebo-controlled study using a computer-based continuous performance test combined with an infra-red motion-tracking device. *J Child Adolesc Psychopharmacol* 2011 Oct 1;**21**(5):433–44	Not an evaluation of the test
Reh V, Schmidt M, Lam L, Schimmelmann BG, Hebebrand J, Rief W, Christiansen H. Behavioral assessment of core ADHD symptoms using the QbTest. *J Atten Disord* 2013. https://doi.org/10.1177/1087054712472981	Does not report on one of the outcomes of interest

TABLE 38 Studies excluded at full-text screening from checking the QbTech manufacturer's submission

Study details	Reason for exclusion
Ulberstad *et al.*, the 6th World Congress on ADHD, 20–23 April 2017, Vancouver, Canada	Does not report on one of the outcomes of interest
Wehmeier PM, Schacht A, Wolff C, Otto WR, Dittmann RW, Banaschewski T. Neuropsychological outcomes across the day in children with attention-deficit/hyperactivity disorder treated with atomoxetine: results from a placebo-controlled study using a computer-based continuous performance test combined with an infra-red motion-tracking device. *J Child Adolesc Psychopharmacol* 2011;**21**:433–44. https://doi.org/10.1089/cap.2010.0142	Not an evaluation of the test
Roughan LA, Stafford J. Demand and capacity in an ADHD team: reducing the wait times for an ADHD assessment to 12 weeks. *BMJ Open Qual* 2019 Oct 30;**8**(4):e000653. https://doi.org/10.1136/bmjoq-2019-000653. PMID: 31750403; PMCID: PMC6830462	Did not report on test of interest
Gustafsson U, Hansen M. QbTest in the clinical assessment of attention deficit hyperactivity disorder: a review of the evidence. *Mental Health Sci* 2023.	Systematic review (we screened the studies)
Gustafsson U, Hansen M. QbTest for monitoring medication treatment response in ADHD: a systematic review. *Clin Pract Epidemiol Ment Health* 2023.	Systematic review (we screened the studies)

Note
This table reports studies included in manufacturer's submission. We report the citation, as provided by the manufacturer, and record how the study has been processed in this review.

TABLE 39 Studies excluded at full-text screening from checking the Peili Vision manufacturer's submission

Study details	Reason for exclusion
Seesjärvi, E., Puhakka, J., Aronen, E. T., Hering, A., Zuber, S., Kliegel, M., Laine., M, Salmi, J. (lähetty arvioita-vaksi). EPELI: a novel virtual reality task for the assessment of goal-directed behavior in real-life contexts. https://psyarxiv.com/aqbwt/	Does not report on one of the outcomes of interest
Toplak, M. E., West, R. F., Stanovich, K. E. Practitioner review: do performance-based measures and ratings of executive function assess the same construct? *Journal of Child* 2013	Not a primary study or SR
Seesjärvi E, Laine M, Kasteenpohja K, Salmi J. Assessing goal-directed behavior in virtual reality with the neuropsychological task EPELI: Children prefer head-mounted display but flat screen provides a viable performance measure for remote testing. *Frontiers in Virtual Reality* 2023 May 26;**4**:1138240	Does not report on one of the outcomes of interest

Note
This table reports studies included in manufacturer's submissions. We report the citation, as provided by the manufacturer, and record how the study has been processed in this review.

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission

Study details	Reason
Fernandez M, Morillo Rojas MD. *[Test-retest validation of AULANESPLORA. (Virtual reality continuous performance test) for ADHD]*. 2012. URL: https://giuntipsy-my.sharepoint.com/personal/crodriguez_nesplora_com/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fcrodriguez%5Fnesplora%5Fcom%2FDocuments%2FDatos%20adjuntos%2FTest%2Dretest%20validation%20of%20AULANESPLORA%20%28virtual%20reality%20continuous%20performance%20test%29%20for%20adhd%2Epdf&parent=%2Fpersonal%2Fcrodriguez%5Fnesplora%5Fcom%2FDocuments%2FDatos%20adjuntos&ga=1 (accessed March 2024)	Does not report on one of the outcomes of interest
Daniel Ursu, Z., Ahmed, R. Assessing the learning effect of the aquarium test on ADHD: a test-retest study with adults; in press n.d. https://doi.org/In press	Does not report on one of the outcomes of interest
Climent, G., Moreno Oyarzabal, M., González, M., Mejías, M., Redondo, M. Nesplora Aquarium: Utilidad de la herramienta para la identificación y evaluación del TDAH en adultos 2019	Not a primary study or SR
Voinescu, A., Petrini, K., Stanton Fraser, D. *et al.* The effectiveness of a virtual reality attention task to predict depression and anxiety in comparison with current clinical measures. *Virtual Reality* 2021. https://doi.org/10.1007/s10055-021-00520-7	Does not include population with suspected or confirmed ADHD
J.L. González. Aplicación de realidad virtual (Nesplora Aquarium) en la valoración cognitiva y control de incapacidad temporal por contingencia común en pacientes con trastorno psiquiátrico menor. *Rev Asoc Esp Espec Med Trab* 2020;**29**(3):223–35	Does not include population with suspected or confirmed ADHD
Díaz-Orueta, U., Climent-Martínez, G., otros autores (in press). Los Tests de Rendimiento Continuo en Neurofeedback. Utilidad y Aplicaciones. En: I. Moreno (Ed.). Neurofeedback aplicado al TDAH/Use of Neurofeedback at ADHD	Not a primary study or SR
Koch, M., Becker, N., Spinath, F., Greiff, S. Assessing intelligence without intelligence tests. Future perspectives. *Intelligence* 2021:101596. https://doi.org/10.1016/j.intell.2021.101596	Not a primary study or SR
Gettman, J. *Best Practices in School Neuropsychology: Guidelines for Effective Practice, Assessment, and Evidence-Based Intervention* (D. Miller, D. Maricle, C. Bedford, Eds.). 1st edn. Wiley; 2022.	Not a primary study or SR
Parsons, T., Duffield, T., Mcmahan, T., Diaz-Orueta, U. *Virtual School Environments for Neuropsychological Assessment and Training: Learning in the Age of Emerging Technologies* 2019. pp. 123–57.	Not a primary study or SR
Mejías, M., Redondo, M., Fernández, M., Díaz-Orueta, U. Eficacia del metilfenidato de liberación prolongada en la mejora sintomática cognitiva y conductual del TDAH monitorizado a través del Test AULA Nesplora. XXIV Congreso de la Academia Iberoamericana de Neurología Pediátrica (AINP). Madrid, España, 8-10 de septiembre 2016; 2016	Does not report on one of the outcomes of interest
Zulueta, A., Iriarte, Y., Díaz-Orueta, U., Climent, G. AULA NESPLORA: AVANCE EN LA EVALUACIÓN DE LOS PROCESOS ATENCIONALES. ESTUDIO DE LA VALIDEZ CONVERGENTE CON EL TEST DE PERCEPCIÓN DE DIFERENCIAS 'CARAS' (VERSIÓN AMPLIADA). 04, 8.; 2013	Does not report on one of the outcomes of interest

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Díaz-Orueta, U., Alonso-Sánchez, B., Climent-Martínez, G. AULA vs. d2 Test of Attention: Convergent validity and applicability of virtual reality in the study of reading disorders. 42nd Annual Meeting of the International Neuropsychological Society. Seattle, Washington, USA, 12th–15th February, 2014; 2014	Does not include population with suspected or confirmed ADHD
Díaz-Orueta, U., García-Cueto, E., Alonso-Sánchez, B., Crespo-Eguílaz, N., Fernández-Fernández, M.A., Otaduy, C., P érezLozano, C., Zulueta, A. AULA Virtual Reality based attention test: factorial validity and convergent validity with EDAH scale and DSM criteria. 9th Conference of the International Test Commission, San Sebastián, Spain, 2nd–5th July, 2014; 2014	Does not include population with suspected or confirmed ADHD
Moreno-García, I., Espinosa-Oneto, N., Camacho-Vara, C., Díaz-Orueta, U. Evaluación del trastorno por dé cit de atención e hiperactividad mediante realidad virtual. Comparación con escalas conductuales. *Comunicación y Pedagogía* 2015;**287-288**:33–37	Does not report on one of the outcomes of interest
Díaz-Orueta, U., Iriarte, Y., Climent-Martínez, G. Banterla, F. An ecological virtual reality test with distractors for attention in children and adolescents. *Journal of Virtual Reality* 2012;5:1–20	Does not report on one of the outcomes of interest
Redondo, M., González, *N*., Mejias, M., González, MF., Aierbe, A., Moreno, M., P érez, C. Validez convergente entre las herramientas Nesplora Aula y el CPT de Conners 3. [Convergent validity between the tools Nesplora Aula and the CPT of Conners 3.] Oral communication presented at the II Ibero-American Congress Of Neuropsychology, Almería, 3–5 May 2018; 2018.	Does not include population with suspected or confirmed ADHD
Rebon Ortiz, F., Altuna, I., Lobo, A., Climent, G. Validity Performance in the AULA Nesplora Test; 2022.	Does not include population with suspected or confirmed ADHD
Climent-Martínez, G., Banterla, F. AULA. Theoretical Manual. San Sebastian: Nesplora; 2011.	Not a primary study or SR
Mujika, J., Climent, G., Banterla F. Classroom a virtual reality task for attention assessment and ADD diagnosis support. *Revue Neurologique* 2011;**53**(10):619–35.	Not an evaluation of the test
Herman, H., Díaz-Orueta, U. Rehabilitation Gaming. In S. Arnab, I. Dunwell, K. Debattista, editors. *Serious Games for Healthcare: Applications and Implication. United States of America: Medical Information Science Reference*; 2013. pp. 50-–5.	Not a primary study or SR
Díaz-Orueta, U. *Processes and Programmes to Develop Attention and Improve Attention Deficit and Hyperactivity. Processes and Programmes in Educational Neuropsychology*. General Technical Secretariat. Publications Centre. Ministry of Education, Culture and Sport; 2015. pp. 154–68.	Not a primary study or SR
Moreno, I., Díaz-Orueta, U., Others (in press). Assessment of ADHD based on virtual reality. Monographic review on ADHD and virtual reality.	Not a primary study or SR
Iriarte, Y., Climent, G., Banterla, F. AULA, the latest innovation in the neuropsychological measurement of ADHD. Oral communication at the Colegio de Psicólogos de Madrid y de Asturias. November 2011;2011.	Not an evaluation of the test
Sánchez-Carpintero, R., Crespo-Eguílaz, N., Banterla, F., Climent-Martínez, G. Cognitive profiles of executive dysfunction in attention deficit disorder according to performance in the AULA virtual reality test. XV International Refresher Course in Neuropediatrics and Child Neuropsychology. Valencia, Spain, 28 February–1 March 2013; 2013.	Not an evaluation of the test
Zulueta, A., Díaz-Orueta, U., Crespo-Eguilaz, *N*. and Ruiz de Eguino, S. AULA virtual reality test and EDAH scale: complementary resources in the identification of ADHD. Communication presented at the VII National Congress of Neuropsychology: Neuropsychology 3.0. Bilbao, Spain, 15–17 October 2014; 2014	Not an evaluation of the test
Díaz-Orueta, U., Fernández-Fernández, M.A., Climent-Martínez, G. Objectivity in Clinical Diagnosis of ADHD by means of AULA virtual reality based neuropsychological test: Initial findings. 5th World Conference on ADHD. Glasgow, Scotland, UK. 28–31 May 2015; 2015. https://doi.org/0.1007/s12402-015-0169-y/89	Does not report on one of the outcomes of interest
U. Diaz-Orueta, A. Zulueta, N. Crespo-Eguılaz. AULA virtual reality test and EDAH observation scale: Complementary resources in the identification of ADHD. 5th World Conference on ADHD. Glasgow, Scotland, United Kingdom. 28–31 May 2015; 2015. https://doi.org/0.1007/s12402-015-0169-y/89	Does not report on one of the outcomes of interest
Zulueta, A., Redondo, M., Mejías, M., González, E. Reaction time in GO/NO GO task of AULA in children aged 6 to 16 years with and without ADHD. 60th Congress of Child and Adolescent Psychiatry (AEPNYA). San Sebastian, Spain, 1–4 June 2016; 2016.	Not an evaluation of the test

continued

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
González, M.F., Zulueta, A., Redondo, M., Mejías, M., Otaduy, C. and González-Fraile, E. Differential pattern of responses of children with ADHD to visual and auditory stimuli. IX International and XIV National Congress of Clinical Psychology. Santander, Spain, 17–20 November 2016; 2016.	Not an evaluation of the test
Redondo, M., Mejías, M., González, M.F., Zulueta, A, Lizarazu, B. Effects of impulsivity (commissions) on reaction times in children with ADHD. II International Congress of Clinical and Health Psychology on Children and Adolescents. Barcelona, Spain, 17–19 November 2016; 2016.	Not an evaluation of the test
Redondo, M., González, M.F., Mejías, M., Lizarazu, B., Rebón, F. Ceiling and floor effect in a test (NESPLORA Attention AULA) for the assessment of attentional processes. II International Congress of Clinical and Health Psychology on Children and Adolescents. Barcelona, Spain, 17–19 November 2016; 2016.	Does not include population with suspected or confirmed ADHD
González, M.F., Mejías, M., Redondo, M., Otaduy, C., Crespo, N. and Pérez, C. . Per les of impulsivity and inattention in children with ADHD according to age. XIX International Conference on Neurodevelopmental Disorders. Valencia, Spain, 3-4 March 2017; 2017	Not an evaluation of the test
Mejias, M., Delgado-Mejía, I.D., González, M.F., Redondo, C., Abadi, A. and Lalor, S. Comparison between processing speed of WISC-IV and response time of the CPT NESPLORA AULA in children with ADHD. Poster presented at 6th World Conference on ADHD, Vancouver, Canada, 20–23 April 2017; 2017	Not an evaluation of the test
Fernández, Fernández, M., Redondo, Zaballos, M., Mejías, M., González, Pérez, M.F. and Díaz-Orueta, U. Differential effect of methylphenidate and lisdexamfetamine on the performance of the AULA Nesplora neuropsychological test in children under treatment for ADHD. Poster presented at the XI SENEP Annual Meeting, Madrid, Spain, 25–27 May 2017; 2017	Not an evaluation of the test
Moreno, M., Aierbe, A., González, M.F. and Mejías, M. Convergent validity between computerized and virtual reality continuous performance tasks. Poster presented at the XX Congreso Internacional de Actualización en Trastornos del Neurodesarrollo, Valencia, Spain, 9-10 March 2018; 2018	Not an evaluation of the test
Mejías, M., Redondo, M., Moreno, M., Aierbe, A. and González, M. Nesplora Aula School: development of a neuropsychological tool for the educational eld. Oral presentation held at the I Congreso de Psicología, Innovación Tecnológica y Emprendimiento, Almería, Spain, 19–21 April 2018; 2018	Not an evaluation of the test
Mejías, M., Climent, G., González, M., Moreno, M., Aierbe, A. VRMIND: Development of neuropsychological assessment tools in Virtual Reality. Oral presentation held at the I Congreso de Psicología, Innovación Tecnológica y Emprendimiento, Almería, Spain, 19-21 April 2018; 2018	Not an evaluation of the test
Moreno, M., Rebón, F., Aierbe, A., Mejías, M., González, M., Climent, G. Attentional development in childhood: Stability in the results of cognitive exploration. XIX International Congress of Psychology and Education, 20–23 June, Logroño, Spain; 2018	Did not report on test of interest
Ruiz-Ruano García, A.M., López Puga, J., Lizarazu Rodrigo, B., Moreno Oyarzabal, M., Aierbe Pombo, A., Mejías Pérez, M. and Climent Martínez, G. Comparing attentional performance with Calibrated Bayes Factors in virtual reality-based continuous performance test. Oral Communication presented at ICERI2018 - International Conference Of Education, Research and Innovation. Seville, Spain, 12–14 November 2018; 2018	Not an evaluation of the test
Moreno, M. Nesplora Aula School: a neuropsychological test for the educational environment. Oral communication at the IX Encuentro Nacional de Orientadores, Zaragoza, 18–20 May 2018; 2018	Not an evaluation of the test
Ruiz-Ruano García, A., López Puga, J., Rodrigo, B., Moreno Oyarzabal, M., Pombo, A., Mejías, M., Climent, G. Comparing attentional performance with calibrated bayes factors in a virtual reality-based continuous performance test 2018. https://doi.org/10.21125/iceri.2018.0557	Not an evaluation of the test
García, A. I. EVALUACIÓN PSICOMÉTRICA DEL TEST DE AULA NESPLORA: APLICACIÓN [PhD Thesis] 2021	Does not report on one of the outcomes of interest
Areces, D., Rodríguez, C., García, T., Cueli, M. Is an ADHD Observation-scale based on DSM criteria able to predict performance in a virtual reality continuous performance test? *Applied Sciences* 2020;**10**(7):2409. https://doi.org/10.3390/app10072409	Not an evaluation of the test
Areces, D., Rodríguez, C., Garcia, T., Cueli, M., Gonzalez-Castro, P. The influence of state and trait anxiety on the achievement of a virtual reality continuous performance test in children and adolescents with ADHD symptoms. *Journal of Clinical Medicine* 2021;**10**. https://doi.org/10.3390/jcm10122534	Not an evaluation of the test
Corrigan, N., Păsărelu, C.-R., Voinescu, A. (2023). Immersive virtual reality for improving cognitive deficits in children with ADHD: a systematic review and meta-analysis. *Virtual Reality* 1–20. https://doi.org/10.1007/s10055-023-00768-1	Did not report on test of interest

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Doulou, A., Skianis, C. VR and electronic games based assessment for ADHD. *Dialogues in Clinical Neuroscience and Mental Health* 2023;**6**(4):Article 4. https://doi.org/10.26386/obrela.v6i4.276	Not a primary study or SR
Parsons, T. D., Kane, R., Duffield, T. C. Virtual-reality-based neuropsychological assessments of everyday functioning 2022;**37**.	Not an evaluation of the test
Wiguna, T., Bahana, R., Dirgantoro, B., Minayati, K., Teh, S. D., Ismail, R., Kaligis, F., Wigantara, N. Developing attention deficits/hyperactivity disorder-virtual reality diagnostic tool with machine learning for children and adolescents. *Frontiers in Psychiatry* 2022;**13**:984481	Did not report on test of interest
K., Pramme, L., Blumenthal, N., Li, M., Asché, L., Jonas, S., Bey, K., *et al.* Virtual reality in the diagnostic and therapy for mental disorders: a systematic review. *Clinical Psychology Review* 2022;**98**:102213. https://doi.org/10.1016/j.cpr.2022.102213	Systematic review (we screened the studies)
Yongmei. Application of virtual reality technology in pediatric clinical practice. 2022.	Not a primary study or SR
A review on machine learning approaches in diagnosis of ADHD based on big data. (*n*.d.). Nesplora. Retrieved December 12, 2023, from	Not a primary study or SR
Adabla, S., Nabors, L., Hamblin, K. A scoping review of virtual reality interventions for youth with attention-deficit/hyperactivity disorder. *Advances in Neurodevelopmental Disorders* 2021;**5**. https://doi.org/10.1007/s41252-021-00207-9	Not a primary study or SR
Alam, S., Raja, P., Gulzar, Y. Investigation of machine learning methods for early prediction of neurodevelopmental disorders in children. *Wireless Communications and Mobile Computing* 2022;**2022**:1–12. https://doi.org/10.1155/2022/5766386	Not a primary study or SR
Alava Sordo, S. Relación entre diagnóstico de TDAH y los procesos intelectuales y atencionales en muestra clínica comparación entre TDAH y Trastorno de Aprendizaje.pdf [PhD Thesis]; 2018.	Did not report on test of interest
Alberca. TDAH, Diagnóstico, prácticas y estrategias de tratamiento_2014.pdf [PhD Thesis]; 2014.	Did not report on test of interest
Alcañiz, M., Parra, E., Giglioli, I. A. C. Virtual reality as an emerging methodology for leadership assessment and training. *Frontiers in Psychology* 2018. https://doi.org/10.3389/fpsyg.2018.01658	Not a primary study or SR
Alqithami, S. A serious-gamification blueprint towards a normalized attention. *Brain Informatics* 2021;**8**(1):6. https://doi.org/10.1186/s40708-021-00127-3	Not a primary study or SR
Alqithami, S., Alzahrani, M., Alzahrani, A., Mustafa, A. AR-therapist: design and simulation of an AR-game environment as a CBT for patients with ADHD. *Healthcare* 2019;**7**:146. https://doi.org/10.3390/healthcare7040146	Not a primary study or SR
Araiza-Alba, P., Keane, T., Beaudry, J., Kaufman, J. Immersive virtual reality implementations in developmental psychology. *International Journal of Virtual Reality* 2020;**20**. https://doi.org/10.20870/IJVR.2020.20.2.3094	Not a primary study or SR
Arboleda Gil, S. V. Desempeño ejecutivo y procesos de monitoreo y control metacognitivo en niños. *Tempus Psicológico* 2020;**3**(2). https://doi.org/10.30554/tempuspsi.3.2.3405.2020	Not an evaluation of the test
Baggio, S., Hasler, R., Giacomini, V., El-Masri, H., Weibel, S., Perroud, N., Deiber, M.-P. Does the continuous performance test predict ADHD symptoms severity and ADHD presentation in adults? *Journal of Attention Disorders* 2020;**24**(6):840–848. https://doi.org/10.1177/1087054718822060	Did not report on test of interest
Bahana, R., Abdurachman, E., Lumban Gaol, F., Wiguna, T., Hutagalung, F., Dirgantoro, B., Nugroho, E. A therapy game for elementary students with ADHD. *AIP Conference Proceedings* 2023;**2508**. https://doi.org/10.1063/5.0114939	Did not report on test of interest
Bassano, C., Chessa, M., Solari, F. Visualization and interaction technologies in serious and exergames for cognitive assessment and training: a survey on available solutions and their validation. *IEEE Access*, 2022;**10**:104295–312. https://doi.org/10.1109/ACCESS.2022.3210562	Not a primary study or SR
Batlle. Test_d_Atencio_Selectiva_i_Sostinguda_TA.pdf [PhD Thesis]; 2020	Did not report on test of interest
Bejarano, A., Correa, J., Figueroa, P. Escape room virtual reality: a tool for diagnosis and treatment of attention deficit disorder. *IEEE* 2020:338. https://doi.org/10.1109/SVR51698.2020.00056	Did not report on test of interest

continued

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Bernardelli, G., Flori, V., Greci, L., Scaglione, A., Zangiacomi, A. A Virtual Reality Based Application for Children with ADHD: Design and Usability Evaluation. In L. T. De Paolis, P. Arpaia, P. Bourdot, editors. *Augmented Reality, Virtual Reality, and Computer Graphics*. Springer International Publishing; 2021. Vol. 12980, pp. 363–375. https://doi.org/10.1007/978-3-030-87595-4_27	Did not report on test of interest
Binz, T., Williner, E., Strajhar, P., Dolder, P., Liechti, M., Baumgartner, M., Kraemer, T., Steuer, A. Chiral analysis of amphetamines in hair by liquid chromatography-tandem mass spectrometry: compliance-monitoring of attention deficit hyperactivity disorder (ADHD) patients under Elvanse® therapy and identification after controlled low dose application: compliance-monitoring of amphetamine in hair. *Drug Testing and Analysis* 2017;**10**. https://doi.org/10.1002/dta.2208	Did not report on test of interest
Biomedical Engineering Research Group, Stellenbosch University, Swarts, R., Fourie, P. R., Biomedical Engineering Research Group, Stellenbosch University, van den Heever, D.; Biomedical Engineering Research Group, Stellenbosch University. ADHD screening tool: investigating the effectiveness of a tablet-based game with machine learning. *Global Health Innovation* 2019;**2**(2). https://doi.org/10.15641/ghi.v2i2.809	Did not report on test of interest
Björling, E. A., Sonney, J., Rodriguez, S., Carr, N., Zade, H., Moon, S. H. Exploring the effect of a nature-based virtual reality environment on stress in adolescents. *Frontiers in Virtual Reality* 2022;**3**. www.frontiersin.org/article/10.3389/frvir.2022.831026	Did not report on test of interest
Boechi, L. C., Encina Benítez, F. L., Rodas Jara, R. L., Rodas Jara, L. R., Villagra, M. D. R., Báez, D., *et al*. Tecnologías para la Evaluación, Diagnóstico y Tratamiento del Trastorno por Déficit de Atención e Hiperactividad: Una Revisión Preliminar e Integradora. *Revista Científica Ciencias de La Salud* 2023;**5**:01–07. https://doi.org/10.53732/rccsalud/2023.e5301	Not a primary study or SR
Borgnis, F., Baglio, F., Pedroli, E., Rossetto, F., Meloni, M., Riva, G., Cipresso, P. EXIT 360°—EXecutive-Functions Innovative Tool 360°—a simple and effective way to study executive functions in Parkinson's disease by using 360° videos. *Applied Sciences* 2021;**11**(15):6791. https://doi.org/10.3390/app11156791	Did not report on test of interest
Bozkir, E. Towards everyday virtual reality through eye tracking. 2021:192	Did not report on test of interest
Cabas-Hoyos, K., Figueroa, P., Bracamonte, Y. Programas de intervención basados en tecnologías para niños y adolescentes diagnosticados con TDAH: Una revisión sistemática; 2022	Systematic review (we screened the studies)
calzón, lopez. ANÁLISIS Y VALORACIÓN DE ALGUNOS PATRONES DIAGNÓSTICOS DIFERENCIALES EN LOS SUBTIPOS DEL TDAH [PhD Thesis]; 2012	Did not report on test of interest
Candela, G. Candela_Attentional variables and BCI_InPACT_2018.pdf. 2018	Not an evaluation of the test
Castellaano. INTERVENCIÓN EN EL AULA PARA LA MEJORA DE LA ATENCIÓN Y EL RENDIMIENTO EN EL ALUMNADO DE SEGUNDO NIVEL DE EDUCACIÓN PRIMARIA: EFICACIA DE LAS AUTOINSTRUCCIONES Y DE LA AUTOOBSERVACIÓN.pdf [PhD Thesis]; 2015	Not an evaluation of the test
Castilla, N., Higuera-Trujillo, J. L., Llinares, C. The effects of illuminance on students' memory. a neuroarchitecture study. *Building and Environment* 2023;**228**:109833. https://doi.org/10.1016/j.buildenv.2022.109833	Did not report on test of interest
Chen, I.-C., Chen, C.-L., Chang, C.-H., Fan, Z.-C., Chang, Y., Lin, C.-H., Ko, L.-W. Task-rate-related neural dynamics using wireless EEG to assist diagnosis and intervention planning for preschoolers with ADHD exhibiting heterogeneous cognitive proficiency. *Journal of Personalized Medicine* 2022;**12**:731. https://doi.org/10.3390/jpm12050731	Did not report on test of interest
Cho, Y., Yum, J., Kim, K., Shin, B., Eom, H., Hong, Y., *et al*. Evaluating attention deficit hyperactivity disorder symptoms in children and adolescents through tracked head movements in a virtual reality classroom: the effect of social cues with different sensory modalities. *Frontiers in Human Neuroscience* 2022;**16**. https://doi.org/10.3389/fnhum.2022.943478	Did not report on test of interest
Cibrian, F., Hayes, G., Lakes, K. Research advances in ADHD and technology. *Synthesis Lectures on Assistive, Rehabilitative, and Health-Preserving Technologies* 2020;**9**:i–156.https://doi.org/10.2200/S01061ED1V01Y202011ARH015	Not a primary study or SR
Company, R. Estatus socioeconnómico y desarrollo cognitivo en la infancia y adolesecnecia [PhD Thesis]; 2022	Not an evaluation of the test

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Crepaldi, M., Colombo, V., Mottura, S., Baldassini, D., Sacco, M., Cancer, A., Antonietti, A. The use of a serious game to assess inhibition mechanisms in children. *Frontiers in Computer Science* 2020;**2**. https://doi.org/10.3389/fcomp.2020.00034	Did not report on test of interest
De La Fuente, J. LIBRO RESUMENES CIPI 2018.pdf, 2018	Did not report on test of interest
de la Fuente, J., González-Torres, M. C., Aznárez-Sanado, M., Martínez-Vicente, J. M., Peralta-Sánchez, F. J., Vera, M. M. Implications of unconnected micro, molecular, and molar level research in psychology: the case of executive functions, self-regulation, and external regulation. *Frontiers in Psychology* 2019;**10**:1919. https://doi.org/10.3389/fpsyg.2019.01919	Not a primary study or SR
delgado, G. Anuario de Psicología Clínica y de la Salud Annuary of Clinical and Health Psychology_ Monografico sobre Realidad Virtual.pdf, 2012	Not a primary study or SR
Delgado Reyes, A., Lopez, J. Escenarios Virtuales para la evaluación Neuropsicológica: Una Revisión de Tema 2021:2–196. https://doi.org/10.7714/CNPS/15.2.216	Did not report on test of interest
Delgado-Reyes, A. C. REALIDAD VIRTUAL: EVALUACIÓN E INTERVENCIÓN EN EL TRASTORNO POR DÉFICIT DE ATENCIÓN/HIPERACTIVIDAD (TDAH) 2021;**28**.	Did not report on test of interest
Diaz-Orueta, U., Facal, D., Nap, H. H., Ranga, M.-M. What Is the key for older people to show interest in playing digital learning games? initial qualitative findings from the LEAGE project on a multicultural European sample. *Games for Health Journal* 2012;**1**(2):115–123. https://doi.org/10.1089/g4h.2011.0024	Does not include population with suspected or confirmed ADHD
Dittmann, R., Cardo, E., Nagy, P., Anderson, C., Adeyi, B., Caballero, B., *et al.* Treatment response and remission in a double-blind, randomized, head-to-head study of lisdexamfetamine dimesylate and atomoxetine in children and adolescents with attention-deficit hyperactivity disorder. *CNS Drugs* 2014;**28**. https://doi.org/10.1007/s40263-014-0188-9	Did not report on test of interest
Dönmez, A., Türk, A. Çocukluk Dönemi Korkuları ve Bir Müdahale Aracı Olarak Sanal Gerçeklik Uygulamasının Kullanımı. *Hemşirelik Bilimi Dergisi*; 2023. https://doi.org/10.54189/hbd.1088650	Not an evaluation of the test
Doulou, A., Drigas, A. Virtual reality and electronic games for assessment in ADHD. *International Journal of Recent Contributions from Engineering, Science and IT (iJES)* 2022;**10**(02):4–15. https://doi.org/10.3991/ijes.v10i02.29735	Did not report on test of interest
Drane, D., Pedersen, N., Sabsevitz, D., Block, C., Dickey, A., Alwaki, A., Kheder, A. Cognitive and emotional mapping with SEEG. *Frontiers in Neurology* 2021;**12**:627981. https://doi.org/10.3389/fneur.2021.627981	Did not report on test of interest
Drigas, A., Mitsea, E., Skianis, C. Virtual reality and metacognition training techniques for learning disabilities. *Sustainability* 2022;**14**:1–19. https://doi.org/10.3390/su141610170	Did not report on test of interest
Duffield, T. C., Parsons, T. D., Landry, A., Karam, S., Otero, T., Mastel, S., Hall, T. A. Virtual environments as an assessment modality with pediatric ASD populations: a brief report. *Child Neuropsychology* 2018;**24**(8):1129–1136. https://doi.org/10.1080/09297049.2017.1375473	Did not report on test of interest
Edwards, J., Parsons, T. D. Virtual Reality Applications for Neuropsychological Assessment in the Military. In *The Role of Technology in Clinical Neuropsychology*. Oxford University Press; 2017. https://doi.org/10.1093/oso/9780190234737.003.0014	Did not report on test of interest
Emmelkamp, P., Meyerbröker, K. Virtual reality therapy in mental health. *Annual Review of Clinical Psychology* 2021;**17**. https://doi.org/10.1146/annurev-clinpsy-081219-115923	Not a primary study or SR
Fang, Y., Han, D., Luo, H. A virtual reality application for assessment for attention deficit hyperactivity disorder in school-aged children. *Neuropsychiatric Disease and Treatment* 2019;**15**:1517–1523. https://doi.org/10.2147/NDT.S206742	Did not report on test of interest
Faria, A. L., Pinho, M. S. DO PAPEL-E-LÁPIS À REALIDADE VIRTUAL: UMA NOVA ABORDAGEM PARA REABILITAÇÃO COGNITIVA PERSONALIZADA. 2016;**9**.	Did not report on test of interest
Fei, C., Sun, B., Li, Y., Zhang, Q. A Study of Virtual Reality Systems for Attention Stabilization. In H. Liu, Z. Yin, L. Liu, L. Jiang, G. Gu, X. Wu, W. Ren, editors. *Intelligent Robotics and Applications*. Springer International Publishing; 2022. pp. 105–113. https://doi.org/10.1007/978-3-031-13844-7_11	Did not report on test of interest

continued

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Fernandez. Fenández_Valoración_Aula_TDAH_Rev_Neurol_2012.pdf. *Revue Neurologique* 2012;54 (Supl 3):S67–S93. https://doi.org/10.33588/rn.54S03.2012203	Does not report on one of the outcomes of interest
Ferreira-Brito, F., Fialho, M., Virgolino, A., Neves, I., Miranda, A., Sousa-Santos, N., *et al.* Game-based interventions for neuropsychological assessment, training and rehabilitation: Which game-elements to use? A systematic review. *Journal of Biomedical Informatics* 2019;**98**:103287. https://doi.org/10.1016/j.jbi.2019.103287	Does not include population with suspected or confirmed ADHD
Feu, A. M. REALIDAD VIRTUAL APLICADA A LA EVALUACIÓN DEL TDAH EN EL DEPARTAMENTO DE ORIENTACIÓN. AULA NESPLORA [PhD Thesis]; 2017.	Does not report on one of the outcomes of interest
Flores, P. Neuropsychological Profiles of Attention and Inhibitory Control in Neurodevelopmental Disorders Through a Virtual Reality Test; n.d. www.unioviedo.es/psicobiologia/wp-content/uploads/2017/12/Abstract-book-%C3%81vila-2017.pdf (accessed 2 May 2023)	Did not report on test of interest
fuentes. TDaHpp: App para Android para detección temprana en TDAH.pdf [PhD Thesis]; 2019.	Not an evaluation of the test
Gabay, M., Schonberg, T. Passive identification of subjective preferences towards individual items using eye-tracking in a virtual reality environment (p. 2022.12.18.520570). *bioRxiv* 2022. https://doi.org/10.1101/2022.12.18.520570	Did not report on test of interest
Gao, H. Assessment of human behavior in virtual reality by eye tracking; n.d. (accessed 2 May 2023)	Did not report on test of interest
Gao, H., Bozkir, E., Hasenbein, L., Hahn, J.-U., Göllner, R., Kasneci, E. Digital transformations of classrooms in virtual reality. 2021.	Did not report on test of interest
García. García_Executive functions in kids_Int Journal Psychology, 2013.pdf. 2013	Did not report on test of interest
García. García_López_Validación convergente con el CPT_Curso Actualización Neuropediatría_2012.pdf. 2012	Did not report on test of interest
García, A. I. EVALUACIÓN PSICOMÉTRICA DEL TEST DE AULA NESPLORA: APLICACIÓN [PhD Thesis]; 2021.	Did not report on one of the outcomes of interest
García Fernández, T., Rodríguez Pérez, C., González Castro, M. P., González-Pienda García, J. A. The assessment of executive functioning in childhood and adolescence: current situation and future lines of research. *Executive Functioning: Role in Early Learning Processes, Impairments in Neurological Disorders and Impact of Cognitive Behavior Therapy* 2014	Not a primary study or SR
García Matilla, E. Caso Clínico: Trastorno por déficit de atención e hiperactividad con síntomas de ansiedad infantil [PhD Thesis]; 2022	Not an evaluation of the test
George, A. The connections between attention-deficit/hyperactivity disorder and levels of criminal behavior among adults. *Open Journal of Social Sciences* 2022;10:1–45. https://doi.org/10.4236/jss.2022.102001	Did not report on test of interest
Gettman, J. *Best Practices in School Neuropsychology: Guidelines for Effective Practice, Assessment, and Evidence-Based Intervention* (D. Miller, D. Maricle, C. Bedford, Eds.). 1st edn. Wiley; 2022. https://doi.org/10.1002/9781119790563	Not a primary study or SR
Gizatdinova, Y., Remizova, V., Sand, A., Sharma, S., Rantanen, K., Helminen, T., Kylliäinen, A. PigScape: An embodied video game for cognitive peer-training of impulse and behavior control in children with ADHD. *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*; 2022. pp. 1–4. https://doi.org/10.1145/3517428.3550401	Did not report on test of interest
Goharinejad, S., Goharinejad, S., Hajesmaeel Gohari, S., Bahaadinbeigy, K. 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* 2022;**22**:1–13. https://doi.org/10.1186/s12888-021-03632-1	Did not report on test of interest
gonzalez lajas, J. Trastorno por Déficit de Atención con Hiperactividad (TDAH)_algoritmos y GPCI_aepap. Aepap; 2016	Not an evaluation of the test
Gualtieri, L. (2021). METHODOLOGIES AND GUIDELINES FOR THE DESIGN OF SAFE AND ERGONOMIC COLLABORATIVE ROBOTIC ASSEMBLY SYSTEMS IN INDUSTRIAL SETTINGS [PhD Thesis]	Did not report on test of interest

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Guerrero, R. (2016). TRASTORNO POR DÉFICIT DE ATENCIÓN CON HIPERACTIVIDAD: ENTRE LA PATOLOGÍA Y LA NORMALIDAD | RAFAEL GUERRERO. www.casadellibro.com/libro-trastorno-por-deficit-de-atencion-con-hiperactividad-entre-la-patologia-y-la-normalidad/9788448022198/2939390	Not a primary study or SR
Gutiérrez-Maldonado, J. The use of virtual reality technology in the treatment of psychopathological disorders. *Journal of Clinical Medicine* 2022;**11**:5358. https://doi.org/10.3390/jcm11185358	Did not report on test of interest
Halder, S., Halder, S. Application of virtual reality in cognitive rehabilitation: a road ahead. 2022. https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-7998-8371-5.ch013. www.igi-global.com/gateway/chapter/www.igi-global.com/gateway/chapter/294210	Not a primary study or SR
Harstad, E., Weaver, A., Katusic, S., Colligan, R., Kumar, S., Chan, E., Voigt, R., Barbaresi, W. ADHD, Stimulant treatment, and growth: a longitudinal study. *Pediatrics* 2014;**134**. https://doi.org/10.1542/peds.2014-0428	Did not report on test of interest
Hayden, A., Hooley, J. M., Dougherty, D. D., Camprodon, J. A., Chou, T. Neuroticism modulates the qualitative effects of inferior parietal tDCS on negatively-valenced memories. *Journal of Psychiatric Research* 2023;**161**:467–75. https://doi.org/10.1016/j.jpsychires.2023.04.005	Did not report on test of interest
Herrán Paz, M. E., Ortiz Monasterio, R., Herrán Ramírez, M. A., Rodríguez-Díaz, A., García Villalpando, A. K. Narrative review of scales assessing attention-deficit/hyperactivity disorder in children and adolescents. *Medwave* 2014;**14**(01):e5887–e5887.	Did not report on test of interest
Higuera-Trujillo, J. L., Millán, C. L., Aviñó, A. M. i, Cueco, J. T., Omarrementeria, C. S. The cognitive effect of university classroom geometry. A virtual reality study focused on memory and attention. *INNODOCT 2020*; 2021, January 21.	Did not report on test of interest
Hosfelt, D. Making ethical decisions for the immersive web. 2019	Did not report on test of interest
Jung, E., Eun, S., Cho, S., Kim, H., Park, D. Virtual reality in psychiatry. *SPG BioMed*; 2019. https://doi.org/10.32392/biomed.55.1	Not a primary study or SR
Junior, F. Transtorno de déficit de atenção e hiperatividade (TDAH): Informações gerais e os jogos como uma das principais técnicas para o ensino de crianças com esse transtorno. 2019. https://doi.org/10.29327/710987	Did not report on test of interest
Kakoulidou, M. Understanding the role of motivation in the reading of children with ADHD-related characteristics [PhD Thesis]. 2022.	Does not report on one of the outcomes of interest
Kakoulidou, M., Knight, F., Filippi, R., Hurry, J. The effects of choice on the reading comprehension and enjoyment of children with severe inattention and no attentional difficulties. *Journal of Abnormal Child Psychology* 2021;**49**. https://doi.org/10.1007/s10802-021-00835-8	Not an evaluation of the test
Kállai J. A komputer által létrehozott virtuális valóság pszichológiai mechanizmusai: Téri reprezentációs sajátosságok. *Magyar Pszichológiai Szemle* 2019;**74**(2):181–200. https://doi.org/10.1556/0016.2019.74.2.4	Did not report on test of interest
Kim, E., Han, J., Choi, H., Prie, Y., Vigier, T., Bluteau, S., and Kwon, G. H. Examining the academic trends of neuropsychological tests for executive functions using virtual reality: systematic literature review (Preprint). *JMIR Serious Games* 2021;**9**. https://doi.org/10.2196/30249	Did not report on test of interest
Knight, F. L. C., and Dimitriou, D. Poor sleep has negative implications for children with and without ADHD, but in different ways. *Behavioral Sleep Medicine* 2019;**17**(4):423–436. https://doi.org/10.1080/15402002.2017.1395335	Did not report on test of interest
Krieger, V., and Amador-Campos, J. Clinical presentations of attention-deficit/hyperactivity disorder (ADHD) in children and adolescents: comparison of neurocognitive performance. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence* 2021;**27**:1–30. https://doi.org/10.1080/09297049.2021.1917530	Did not report on test of interest
Kwan, H., Lin, L., Fahy, C., Shell, J., Pang, S., and Xing, Y. Designing VR training systems for children with attention deficit hyperactivity disorder (ADHD). 2022:89. https://doi.org/10.1109/VRW55335.2022.00030	Did not report on test of interest
León, J. M. R. S. D. Manual de neuropsicología pediátrica. *José María Ruiz Sánchez de León* 2016. https://doi.org/10.13140/RG.2.1.3492.6968	Not a primary study or SR

continued

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Liberatore, M., and Wagner, W. Virtual, mixed, and augmented reality: a systematic review for immersive systems research. *Virtual Reality* 2021;**25**:1–27. https://doi.org/10.1007/s10055-020-00492-0	Does not include population with suspected or confirmed ADHD
limachi. APLICACIÓN DE LA PRUEBA DE REALIDAD VIRTUAL 'AULA' EN NIÑOS CON TRASTORNO DE DÉFICIT DE ATENCIÓN CON HIPERACTIVIDAD DEL CENTRO DE DESARROLLO INTEGRAL NEUROGYM[PhD Thesis]; 2019.	Does not report on one of the outcomes of interest
Lin, H.-Y., Chang, W.-D., Hsieh, H.-C., Yu, W.-H., and Lee, P. Relationship between intraindividual auditory and visual attention in children with ADHD. *Research in Developmental Disabilities* 2021;**108**:103808. https://doi.org/10.1016/j.ridd.2020.103808	Did not report on test of interest
Liu, T.-C., Lin, Y.-C., Wang, T.-*N*., Yeh, S.-C., and Kalyuga, S. Studying the effect of redundancy in a virtual reality classroom. *Educational Technology Research and Development* 2021;**69**. https://doi.org/10.1007/s11423-021-09991-6	Did not report on test of interest
Lopez, J. V. S. Andrés Camilo Delgado-Reyesa 2021;**15**:21.	Not a primary study or SR
Loyer Carbonneau, M., Demers, M., Bigras, M., and Guay, M.-C. Meta-Analysis of sex differences in ADHD symptoms and associated cognitive deficits. *Journal of Attention Disorders* 2020:1087054720923736. https://doi.org/10.1177/1087054720923736	Did not report on test of interest
Lozano-Álvarez, M., Rodríguez-Cano, S., Delgado-Benito, V., and Mercado Val, E. A systematic review of literature on emerging technologies and specific learning difficulties. *Education Sciences* 2023;**13**:298. https://doi.org/10.3390/educsci13030298	Did not report on test of interest
Lv, Z., Wang, J.-Y., Kumar, *N*., and Lloret, J. Special issue on 'augmented reality, virtual reality and semantic 3D reconstruction'. *Applied Sciences* 2021;**11**:8590. https://doi.org/10.3390/app11188590	Not an evaluation of the test
Maciá, D. TDAH_en_la_infancia_y_la_adolescencia_co.pdf. 2012	Not a primary study or SR
Madalena, I., Paskakulis, M., Torres, C., Queiroz, A., Paula-Silva, F. Use of midazolam for behavioral management in dental care of a child with attention deficit hyperactivity disorder: a case report. *RSBO* 2021;**18**:368–74. https://doi.org/10.21726/rsbo.v18i2.1617	Did not report on test of interest
Maddalon, L. A voice recognition application for the semantic and prosodic analysis of ASD caregivers. *Annual Review of Cybertherapy and Telemedicine 2021*; 2022.	Did not report on test of interest
Mader. Game_design_methods_for_therapeutic_games.pdf. 2015	Did not report on test of interest
Management Association, I. R., editor. Virtual and augmented reality: concepts, methodologies, tools, and applications. *IGI Global* 2018. https://doi.org/10.4018/978-1-5225-5469-1	Not a primary study or SR
Martínez-Álvarez, I. Neuropsychology applied to education: theoretical framework and intervention areas for the reading competence and attention difficulties. 2018;**14**.[AU: Please provide journal details for reference 'Martínez-Álvarez (2018)' in Table 40, if available.	Did not report on test of interest
Mash, L. E., Klein, R. M., Townsend, J. Brief report: a gaming approach to the assessment of attention networks in autism spectrum disorder and typical development. *Journal of Autism and Developmental Disorders* 2018. https://doi.org/10.1007/s10803-018-3635-5	Did not report on test of interest
McKay, E., Kirk, H., Coxon, J., Courtney, D., Bellgrove, M., Arnatkeviciute, A., Cornish, K. Training inhibitory control in adolescents with elevated attention deficit hyperactivity disorder traits: a randomised controlled trial of the Alfi Virtual Reality programme. *BMJ Open* 2022;**12**:e061626. https://doi.org/10.1136/bmjopen-2022-061626	Did not report on test of interest
Mühlberger, A., Jekel, K., Probst, T., Schecklmann, M., Conzelmann, A., Andreatta, M., *et al.* The influence of methylphenidate on hyperactivity and attention deficits in children with ADHD: a virtual classroom test. *Journal of Attention Disorders* 2020;**24**(2):277–289. https://doi.org/10.1177/1087054716647480	Did not report on test of interest
Muñoz, A. REVISION CPTS PARA EVALAUCION DE LA ATENCION_TFM_MUÑOZ CASTILLEJO, ANGELA.pdf [PhD Thesis]. 2018	Does not report on one of the outcomes of interest
Musalek, M., Kovar, I., Sysala, T. Use of virtual reality for the therapy of children with attention deficit hyperactivity disorder. *MATEC Web of Conferences* 2019;**292**:01042. https://doi.org/10.1051/matecconf/201929201042	Did not report on test of interest

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Nasiri, E., Khalilzad, M., Hakimzadeh, Z., Isari, A., Faryabi-Yousefabad, S., Sadigh-Eteghad, S., Naseri, A. A comprehensive review of attention tests: can we assess what we exactly do not understand? *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery* 2023;**59**. https://doi.org/10.1186/s41983-023-00628-4	Did not report on test of interest
Neguț, A., Jurma, A. M., David, D. Virtual-reality-based attention assessment of ADHD: ClinicaVR: classroom-CPT vs. a traditional continuous performance test. *Child Neuropsychology* 2017;**23**:692–712. https://doi.org/10.1080/09297049.2016.1186617	Did not report on test of interest
Neguț, A., Matu, S.-A., Sava, F. A., David, D. Virtual reality measures in neuropsychological assessment: a meta-analytic review. *The Clinical Neuropsychologist* 2016;**30**(2):165–184. https://doi.org/10.1080/13854046.2016.1144793	Did not report on test of interest
Nolé Fajardo, M. L., Higuera-Trujillo, J. L., Llinares, C. Lighting, colour and geometry: Which has the greatest influence on students' cognitive processes? *Frontiers of Architectural Research* 2023;**12**(4):575–586. https://doi.org/10.1016/j.foar.2023.02.003	Did not report on test of interest
Nolé, M. L., Soler, D., Higuera-Trujillo, J. L., Llinares, C. Optimization of the cognitive processes in a virtual classroom: a multi-objective integer linear programming approach. *Mathematics* 2022;**10**(7):1184. https://doi.org/10.3390/math10071184	Did not report on test of interest
Obrist, V. U., Martínez, E. A. Application of virtual reality in a learning experience. 2016;**6**(2):5	Did not report on test of interest
Ortiz de Gortari, A., Panagiotidi, M. The interplay between executive function deficits, psychopathological traits and dysfunctional gaming habits in the context of Game Transfer Phenomena. *Computers in Human Behavior* 2022;**138**:107469. https://doi.org/10.1016/j.chb.2022.107469	Did not report on test of interest
Ortiz Pérez, A. Evaluación de la sintomatología, comorbilidad e impacto del trastorno por déficit de atención con hiperactividad a partir de evaluación electroencefalográfica, tests de rendimiento continuo y escalas de valoración; 2017. https://idus.us.es/handle/11441/69053	Did not report on test of interest
Pardos, A. P. Análisis descriptivo de la batería Test of everyday attention for children (TEA-Ch) en niños españoles de educación primaria [PhD Thesis]. 2014	Not an evaluation of the test
Parsons, T. D., Carlew, A. R. Bimodal virtual reality stroop for assessing distractor inhibition in autism spectrum disorders. *Journal of Autism and Developmental Disorders* 2016;**46**(4):1255–1267. https://doi.org/10.1007/s10803-015-2663-7	Did not report on test of interest
Parsons, T. D., Carlew, A. R., Magtoto, J., Stonecipher, K. The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. *Neuropsychological Rehabilitation* 2017;**27**(5):777–807. https://doi.org/10.1080/09602011.2015.1109524	Did not report on test of interest
Parsons, T. D., Duffield, T. National Institutes of Health initiatives for advancing scientific developments in clinical neuropsychology. *The Clinical Neuropsychologist* 2019;**33**(2):246–270. https://doi.org/10.1080/13854046.2018.1523465	Not a primary study or SR
Parsons, T. D., Phillips, A. S. Virtual reality for psychological assessment in clinical practice. *Practice Innovations* 2016;**1**(3):197–217. https://doi.org/10.1037/pri0000028	Did not report on test of interest
Parsons, T. D., Riva, G., Parsons, S., Mantovani, F., Newbutt, N., Lin, L., Venturini, E., Hall, T. Virtual reality in pediatric psychology. *Pediatrics* 2017;**140**(Supplement_2):S86–91. https://doi.org/10.1542/peds.2016-1758I	Not a primary study or SR
Parsons, T., Duffield, T. Paradigm shift toward digital neuropsychology and high-dimensional neuropsychological assessments: review. *Journal of Medical Internet Research* 2020;**22**(12):e23777. https://doi.org/10.2196/23777	Not a primary study or SR
Parsons, T., Kane, R. The role of technology in clinical neuropsychology. 2017. https://academic.oup.com/book/40883?login=true#login-purchase#login-purchase	Not a primary study or SR
Perra, A. Virtual reality frontiers in bipolar disorders: a recovery oriented cognitive rehabilitation tool; n.d.	Did not report on test of interest
Perra, A., Riccardo, C. L., De Lorenzo, V., De Marco, E., Di Natale, L., Kurotschka, P. K., Preti, A., Carta, M. G. Fully immersive virtual reality-based cognitive remediation for adults with psychosocial disabilities: a systematic scoping review of methods intervention gaps and meta-analysis of published effectiveness studies. *International Journal of Environmental Research and Public Health* 2023;**20**(2):1527. https://doi.org/10.3390/ijerph20021527	Did not report on test of interest

continued

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Pflueger, M., Mager, R., Graf, M., Stieglitz, R.-D. Encoding of everyday objects in older adults: episodic memory assessment in virtual reality. *Frontiers in Aging Neuroscience* 2023;**15**:1100057. https://doi.org/10.3389/fnagi.2023.1100057	Did not report on test of interest
Pinnow, D., Hubbard, H., Meulenbroek, P. Assessment of attention and memory utilizing ecologically valid distractions: a scoping review. *Frontiers in Virtual Reality* 2021;**2**:685921. https://doi.org/10.3389/frvir.2021.685921	Not a primary study or SR
Polanczyk, G., Salum, G., Sugaya, L., Caye, A., Rohde, L. Annual research review: a meta-analysis of the world-wide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 2015;**56**. https://doi.org/10.1111/jcpp.12381	Did not report on test of interest
Rasoulian Kasrineh, M., Tabatabaei, S. M. Virtual reality among children with mental disorders: a mini-review. *Advances in Health and Behavior* 2021;**4**:177–181. https://doi.org/10.25082/AHB.2021.01.004	Not a primary study or SR
Rodríguez, C., Areces, D., Garcia, T., Cueli, M., Gonzalez-Castro, P. Neurodevelopmental disorders: an innovative perspective via the response to intervention model. *World Journal of Psychiatry* 2021;**11**:1017–1026. https://doi.org/10.5498/wjp.v11.i11.1017	Not a primary study or SR
Rodríguez, C., Garcia, T., Areces, D. New and future challenges concerning the use of virtual reality tools for assessing ADHD. *Current Developmental Disorders Reports* 2017;**4**. https://doi.org/10.1007/s40474-017-0103-4	Not a primary study or SR
Rodrıguez-Barranco, M., Gil, F., Herna, A. F., Alguacil, J., Lorca, A., Molina-Villalba, I., *et al.* Postnatal arsenic exposure and attention impairment in school children 2016;**13**	Not an evaluation of the test
Romero-Ayuso, D. Assessment of cognitive instrumental activities of daily living: a systematic review. 2021. www.tandfonline.com/doi/full/10.1080/09638288.2019.1665720	Not an evaluation of the test
Romero-Ayuso, D., Alcantara-Vázquez, P., Almenara, A., Núñez-Camarero, I., Triviño, J., Ariza-Vega, P., Molina Masso, J. P., González, P. Self-regulation in children with neurodevelopmental disorders 'SR-MRehab: Un Colegio Emocionante': a protocol study. *International Journal of Environmental Research and Public Health* 2020;**17**. https://doi.org/10.3390/ijerph17124198	Did not report on test of interest
Romero-Ayuso, D., Toledano-González, A., Rodríguez-Martínez, M., Arroyo Castillo, P., Triviño, J., González, P., *et al.* Effectiveness of virtual reality-based interventions for children and adolescents with ADHD: a systematic review and meta-analysis. *Children* 2021;**18**:70. https://doi.org/10.3390/children8020070	Not an evaluation of the test
Ruiz-Ruano-García, A. M., Sánchez-Kuhn, A., Flores, P., López-Puga, J. Social Expectancy Increases Skin Conductance Response in Mobile Instant Messaging Users. *Psicothema* 2023;**35**(4):414–422. https://doi.org/10.7334/psicothema2022.362	Did not report on test of interest
Salas-Bravo, S., Gonzalez-Arias, M., Araya-Piñones, A., Valencia-Jimenez, M., Oyarce-Cortes C., S. Uso del Testde Rendimiento Continuo de Conners para diferenciar niños normales y con TDAH en Chile. *Terapia psicológica* 2017;**35**(3):283–291. https://doi.org/10.4067/S0718-48082017000300283	Did not report on test of interest
Sánchez-Kuhn, A., León, J. J., Gôngora, K., Pérez-Fernández, C., Sánchez-Santed, F., Moreno, M., *et al.* Does the go/no-go task measure impulsivity or compulsivity? 2015;**82**.	Did not report on test of interest
Satu, P., Minna, L., Satu, S. Immersive VR assessment and intervention research of individuals with neurodevelopmental disorders is dominated by ASD and ADHD: a scoping review. *Review Journal of Autism and Developmental Disorders* 2023:1–19. https://doi.org/10.1007/s40489-023-00377-3	Not a primary study or SR
Schöne, B., Kisker, J., Sylvester, R. S., Radtke, E. L., Gruber, T. Library for universal virtual reality experiments (luVRe): a standardized immersive 3D/360° picture and video database for VR based research. *Current Psychology* 2021. https://doi.org/10.1007/s12144-021-01841-1	Did not report on test of interest
Schweitzer and Rizzo. Virtual reality and ADHD: clinical assessment and treatment in the metaverse. *The ADHD Report* 2022. https://guilfordjournals.com/doi/abs/10.1521/adhd.2022.30.3.1	Not a primary study or SR
Seesjärvi, E., Laine, M., Kasteenpohja, K., Salmi, J. Assessing goal-directed behavior in virtual reality with the neuropsychological task EPELI: children prefer head-mounted display but flat screen provides a viable performance measure for remote testing. *Frontiers in Virtual Reality* 2023;**4**.	Did not report on test of interest
Seivane, M. S., Brenlla, M. E. Aplicaciones de la realidad virtual en el campo de la evaluación psicológica: Una revisión sistemática. *Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport* 2022;**40**(2), Article 2. https://doi.org/10.51698/aloma.2022.40.2.21-31	Not an evaluation of the test

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Sempere-Tortosa, M., Fernández-Carrasco, F., Mora-Lizán, F., Rizo-Maestre, C. Objective analysis of movement in subjects with ADHD. Multidisciplinary control tool for students in the classroom. International *Journal of Environmental Research and Public Health* 2020;**17**(15):5620. https://doi.org/10.3390/ijerph17155620	Did not report on test of interest
Seo, S., Kim, E., Mundy, P., Heo, J., Kim, K. K. Joint attention virtual classroom: a preliminary study. *Psychiatry Investigation* 2019;**16**(4):292–299. https://doi.org/10.30773/pi.2019.02.08	Did not report on test of interest
Serrano-Barroso, A., Siugzdaite, R., Guerrero-Cubero, J., Molina-Cantero, A. J., Gomez-Gonzalez, I. M., Lopez, J. C., Vargas, J. P. Detecting attention levels in ADHD children with a video game and the measurement of brain activity with a single-channel BCI headset. *Sensors* 2021;**21**(9):3221. https://doi.org/10.3390/s21093221	Did not report on test of interest
Simões, E. *N*., Carvalho, A. L. *N*., Schmidt, S. L. The role of visual and auditory stimuli in continuous performance tests: differential effects on children with ADHD. *Journal of Attention Disorders* 2021;**25**(1):53–62. https://doi.org/10.1177/1087054718769149	Did not report on test of interest
Sinha, S. A step towards the design of a collaborative virtual reality-based story-telling environment [PhD Thesis]. 2018. https://doi.org/10.13140/RG.2.2.35115.92965	Did not report on test of interest
Skalski, S. Impact of placebo-related instruction on HEG biofeedback outcomes in children with ADHD. *Applied Neuropsychology Child* 2020. https://doi.org/10.1080/21622965.2020.1861546	Did not report on test of interest
Skalski, S., Konaszewski, K., Pochwatko, G., Balas, R., Surzykiewicz, J. Effects of hemoencephalographic biofeedback with virtual reality on selected aspects of attention in children with ADHD. *International Journal of Psychophysiology* 2021;**170**:59–66. https://doi.org/10.1016/j.ijpsycho.2021.10.001	Did not report on test of interest
Son, H., Lee, D., Joung, Y.-S., Lee, J., Seok, E., Chung, T.-M., Oh, S. A novel approach to diagnose ADHD using virtual reality. *International Journal of Web Information Systems*, ahead-of-print 2021. https://doi.org/10.1108/IJWIS-03-2021-0021	Did not report on test of interest
Sordo, S., Garrido-Hernansaiz, H., Cantero-García, M., Sánchez-Iglesias, I., González-Moreno, J., Santacreu, J. Validez de las pruebas de atención para el diagnóstico diferencial de TDAH infantil y Trastornos del Aprendizaje///Validity of attention tests for differential diagnosis of childhood ADHD and learning disabilities. *Electronic Journal of Research in Educational Psychology* 2021;**19**:437–64.	Not an evaluation of the test
Soroa Martínez, G., Revert Sánchez, L., Aritzeta Galán, A. Familia eta hiperaktibitatea elkarrekin biziz: Atxikimendua eta heziketa-estiloak. *Uztaro: giza eta gizarte-zientzien aldizkaria* 2020;**114**:81–106.	Did not report on test of interest
Sowell, M. M. Diagnosis and assessment of adult attention deficit hyperactivity disorder: symptom severity and performance on cognitive and achievement testing [Thesis]; 2014. https://oaktrust.library.tamu.edu/handle/1969.1/153987	Did not report on test of interest
Spesialpedagogikk, M. I., Healy, C. There is something 'extra' with my child; n.d.	Not an evaluation of the test
Stokes, J. D., Rizzo, A., Geng, J. J., Schweitzer, J. B. measuring attentional distraction in children with ADHD using virtual reality technology with eye-tracking. *Frontiers in Virtual Reality* 2022;**3**. www.frontiersin.org/articles/10.3389/frvir.2022.855895	Did not report on test of interest
Sujar, A., Bayona, S., Delgado-Gómez, D., Miguélez-Fernández, C., Ardoy-Cuadros, J., Peñuelas-Calvo, I., Baca-García, E., Blasco-Fontecilla, H. Attention deficit hyperactivity disorder assessment based on patient behavior exhibited in a car video game: a pilot study. *Brain Sciences* 2022;**12**(7), Article 7. https://doi.org/10.3390/brainsci12070877	Did not report on test of interest
Sung, D., Park, B., Kim, B., Kim, H., Jung, K.-I., Lee, S.-Y., Kim, B., Park, S., Park, M.-H. Gray matter volume in the developing frontal lobe and its relationship with executive function in late childhood and adolescence: a community-based study. *Frontiers in Psychiatry* 2021;**12**:686174. https://doi.org/10.3389/fpsyt.2021.686174	Did not report on test of interest
Tärning, B., Ternblad, E.-M., Haake, M., Gulz, A., Nirme, J. Lessons learned from a study on distractions in virtual learning environments: reliability, ecological validity and an elusive social component. *Presence: Virtual and Augmented Reality* 2021;**28**:1–51. https://doi.org/10.1162/pres_a_00342	Did not report on test of interest
Ticknor, B. Virtual reality and correctional rehabilitation: a game changer. *Criminal Justice and Behavior* 2019;**46**:009385481984258. https://doi.org/10.1177/0093854819842588	Did not report on test of interest

continued

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Trigueiro, M. The effect of a virtual reality based intervention on processing speed and working memory in individuals with ADHD – a pilot-study. *Frontiers in Virtual Reality* 2023;**4**. https://doi.org/10.3389/frvir.2023.1108060	Did not report on test of interest
Valladares-Rodriguez, S., Fernández-Iglesias, M. J., Anido-Rifón, L., Facal, D., Pérez-Rodríguez, R. Episodix: a serious game to detect cognitive impairment in senior adults. A psychometric study. *PeerJ* 2018;**6**:e5478. https://doi.org/10.7717/peerj.5478	Not an evaluation of the test
Vaz de Carvalho, C., González González, C., Popescu, E., Rugelj, J. Serious games. 2021. https://doi.org/10.3389/978-2-88966-944-8	Not a primary study or SR
Vicente, raquel. Trabajo teórico de revisión, actualización y análisis de un tema Raquel Vicente García; n.d.	Not an evaluation of the test
Villani, D., editor. *Integrating Technology in Positive Psychology Practice*. Information Science Reference, an imprint of IGI Global; 2016.	Not a primary study or SR
Voinescu, A., David, D. The effect of learning in a virtual environment on explicit and implicit memory by applying a process dissociation procedure. *International Journal of Human–Computer Interaction* 2019;**35**(1):27–37. https://doi.org/10.1080/10447318.2018.1424102	Did not report on test of interest
Volkov, A., Obukhov, A., Nazarova, A., Patutin, K. Structural model of the microservice architecture of the control system for training complexes. *AIP Conference Proceedings* 2023;**2910**(1):020164. https://doi.org/10.1063/5.0166558	Did not report on test of interest
Wallisch, A., Little, L. M., Dean, E., Dunn, W. Executive function measures for children: a scoping review of ecological validity. *OTJR: Occupation, Participation and Health* 2018;**38**(1):6–14. https://doi.org/10.1177/1539449217727118	Did not report on test of interest
Wang, Z., Xu, H., Yuan, H. Research on design and experience of immersive virtual reality psychological relaxation game based on image. *IOP Conference Series: Materials Science and Engineering* 2020;**740**(1):012118. https://doi.org/10.1088/1757-899X/740/1/012118	Did not report on test of interest
Wiederhold, B. K., Riva, G., G. The virtual reality working-memory training program (VR WORK M): description of an individualized, integrated program. 2018. www.arctt.info/volume-16-summer-2018	Did not report on test of interest
Wiguna, T., Wigantara, *N*., Ismail, R., Kaligis, F., Minayati, K., Bahana, R., Dirgantoro, B. A four-step method for the development of an ADHD-VR digital game diagnostic tool prototype for children using a DL model. *Frontiers in Psychiatry* 2020;**11**:829. https://doi.org/10.3389/fpsyt.2020.00829	Not an evaluation of the test
Yeh, S.-C., Lin, S.-Y., Wu, E., Zhang, K.-F., Xiu, X., Rizzo, A., Chung, C.-R. A virtual-reality system integrated with neuro-behavior sensing for attention-deficit/hyperactivity disorder intelligent assessment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 2020;**28**: 1899–907. https://doi.org/10.1109/TNSRE.2020.3004545	Not an evaluation of the test
Yez Tellez, Ma. G. Neuropsicologia de los trastornos del neurodesarrollo: Diagnostico, evaluacion e intervencion; 2016	Did not report on test of interest
YILMAZ, *N*., Duran, F., Fidan, U. Psikiyatrik Rahatsızlıklarda Sanal Gerçeklik ve Artırılmış Gerçeklik. *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji* 2021;**9**. https://doi.org/10.29109/gujsc.961331	Not an evaluation of the test
Żyła, K. Attention deficit hyperactivity disorder detection – from psychological checklists to mobile solutions. *Studies in Logic, Grammar and Rhetoric* 2019;**60**(1):85–100. https://doi.org/10.2478/slgr-2019-0047	Not an evaluation of the test
Aierbe, A., Climent, G. Factorial structure of Nesplora Aquarium_INS 2018.pdf. International Neuropsychological Society 2018 Mid-Year Meeting, Praga, República Checa; 2018	Does not report on one of the outcomes of interest
Aierbe Pombo, A., Moreno Oyarzabal, M., Redondo, M., Mejías, M., González, M. Comparison of the execution in the Nesplora Aquarium test between monolingual and bilingual people. X Congreso Nacional de Neuropsicología FANPSE; 2018	Does not report on one of the outcomes of interest
Climent, G. (2018). Manual Nesplora Aquarium	Not a primary study or SR
González, M., Redondo, M., Mejías, M., Aierbe Pombo, A., Moreno Oyarzabal, M. Evolución de los procesos atencionales en función de la edad, medidos a través de una herramienta en realidad virtual. Congreso Nacional de Psicología, Oviedo, 3-7 de julio de 2017; 2017	Did not report on test of interest

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Mejias, M., Aierbe Pombo, A., Gonzalez, M. a F., Moreno Oyarzabal, M. Development of a virtual reality-based continuous performance test for the assessment of attention in adults. Nesplora Aquarium. I Congreso de Psicología, Innovación Tecnológica y Emprendimiento, Almería, España; 2018	Did not report on test of interest
Mejías, M., González, M., Redondo, M., Aierbe Pombo, A., Moreno Oyarzabal, M., Guinea, J. Attention assessment in adults through virtual reality. *6th Scientific Meeting of the Federation of the European Societies of Neuropsychology*, Maastricht, 13–15 de septiembre de 2017; 2017	Did not report on test of interest
Alshehri, A., Shehata, S., Almosa, K., Awadalla, N. Schoolteachers' knowledge of attention-deficit/hyperactivity disorder—current status and effectiveness of knowledge improvement program: a randomized controlled trial. *International Journal of Environmental Research and Public Health* 2020;**17**:5605. https://doi.org/10.3390/ijerph17155605	Did not report on test of interest
Duan, D., Wu, Z., Zhou, Y., Wan, X., Wen, D. Working memory training and evaluation based on brain-computer interface and virtual reality: our opinion. *Frontiers in Human Neuroscience* 2023;**17**. www.frontiersin.org/articles/10.3389/fnhum.2023.1291983	Not a primary study or SR
Fernández, M., Morillo, M., Gilibert, *N*., Carvalho, C., Bello, S. The technological tools of the diagnosis and treatment of attention deficit disorder and hyperactivity. *Medicina* 2020;**80**(Suppl 2):67–71.	Not primary study or SR
Geraets, C., Wallinius, M., Sygel, K. Use of virtual reality in psychiatric diagnostic assessments: a systematic review. *Frontiers in Psychiatry* 2022;**13**. https://doi.org/10.3389/fpsyt.2022.828410	Does not include population with suspected or confirmed ADHD
González Torrecillas, J. L., Marín, B., Alonso, B. Aplicación de realidad virtual (Nesplora Aquarium) en la valoración cognitiva y control de incapacidad temporal por contingencia común en pacientes con trastorno psiquiátrico menor. *Revista de La Asociación Española de Especialistas En Medicina Del Trabajo* 2020;**29**(3):223–235	Does not include population with suspected or confirmed ADHD
Neguț, A., Matu, S.-A., Sava, F. A., David, D. Virtual reality measures in neuropsychological assessment: a meta-analytic review. *The Clinical Neuropsychologist* 2016;**30**(2):165–184. https://doi.org/10.1080/13854046.2016.1144793	Not an evaluation of the test
Voinescu, A., Fodor, L. A., Fraser, D. S., David, D. Exploring attention in vr: Effects of visual and auditory modalities. *International Conference on Applied Human Factors and Ergonomics* 2020:677–683	Does not report on one of the outcomes of interest
Voinescu, A., Fodor, L.-A., Fraser, D. S., Mejías, M., David, D. Exploring the usability of Nesplora Aquarium, a virtual reality system for neuropsychological assessment of attention and executive functioning. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* 2019:1207–08. https://doi.org/10.1109/VR.2019.8798191	Does not include population with suspected or confirmed ADHD
Akram, U., Barclay, N., Milkins, B., Stevenson, J., Gardani, M. Sleep-related attentional and interpretive-bias in insomnia: a systematic review and meta-analysis. *Sleep Medicine Reviews* 2023;**67**. https://doi.org/10.1016/j.smrv.2022.101713	Does not include population with suspected or confirmed ADHD
Alam, F., Matava, C. A new virtual world? The future of immersive environments in anesthesiology. *Anesthesia and Analgesia* 2022;**135**(2):230–238. https://doi.org/10.1213/ANE.0000000000006118	Not a primary study or SR
Areces, D. Velocidad nombramiento dificultades lectoras y atencionales_2014.pdf [PhD Thesis]. 2014	Did not report on test of interest
Baertsch, T., Huang, Y.-Y., Menozzi, M. (2023). Head-mounted display vs. computer monitor for visual attention screening: a comparative study. *Heliyon* **0**. https://doi.org/10.1016/j.heliyon.2023.e16610	Did not report on test of interest
beristain, garcia. 7th World Congress on ADHD: From Child to Adult Disorder: 25th–28th April, Lisbon Portugal. *ADHD Attention Deficit and Hyperactivity Disorders* 2019;**11**(S1):1–89. https://doi.org/10.1007/s12402-019-00295-7	Does not report on one of the outcomes of interest
Borgnis, F., Baglio, F., Pedroli, E., Rossetto, F., Meloni, M., Riva, G., Cipresso, P. A psychometric tool for evaluating executive functions in Parkinson's disease. *Journal of Clinical Medicine* 2022;**11**. https://doi.org/10.3390/jcm11051153	Did not report on test of interest
Borgnis, F., Baglio, F., Pedroli, E., Rossetto, F., Uccellatore, L., Oliveira, J., Riva, G., Cipresso, P. Available virtual reality-based tools for executive functions: a systematic review. *Frontiers in Psychology* 2022;**13**. https://doi.org/10.3389/fpsyg.2022.833136	Does not include population with suspected or confirmed ADHD

continued

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
CIMA aeroespacial. Centro de Instrucción de Medicina Aeroespacial—Investigación—Investigaciones anteriores; n.d. (accessed 15 August 2022) https://ejercitodelaire.defensa.gob.es/EA/cima/investigacion/invAnteriores/#	Did not report on test of interest
Contreras-González, N., Téllez-Alanís, B., Haro, R., Jiménez-Correa, U., Poblano, A. Executive dysfunction in patients with chronic primary insomnia treated with clonazepam. *Neurological Research* 2015;**37**(12):1047–1053. https://doi.org/10.1080/01616412.2015.1114740	Did not report on test of interest
di, T. di D., Borgnis, F., Matricola, N. EXecutive-functions Innovative Tool—EXIT 360: Development and validation of a new 360-video instrument for executive functions; n.d.	Did not report on test of interest
Expósito, M. Á. F. TDaHpp: App para Android para detección temprana en TDAH [PhD Thesis]. 2019	Not an evaluation of the test
Fernández, M. A., Morillo, M. D., Gilibert, *N*., Carvalho, C., Bello, S. Herramientas tecnológicas del diagnóstico y tratamiento del trastorno por déficit de atención e hiperactividad. *Medicina (Buenos Aires)* 2020;**80**:67–71	Not a primary study or SR
Floris, M. La realtà virtuale nei disturbi affettivi: Uno studio pilota sulla prestazione cognitiva e i correlati elettroencefalografici della depressione; n.d.	Not an evaluation of the test
Friedenberg, J. *The Future of the Self: An Interdisciplinary Approach to Personhood and Identity in the Digital Age*. 1st edn. University of California Press; 2020.	Not an evaluation of the test
Hurtado-Pomares, M., Carmen Terol-Cantero, M., Sánchez-Pérez, A., Peral-Gómez, P., Valera-Gran, D., Navarrete-Muñoz, E. M. The frontal assessment battery in clinical practice: a systematic review. *International Journal of Geriatric Psychiatry* 2018;**33**(2):237–251. https://doi.org/10.1002/gps.4751	Did not report on test of interest
Jensen, T. D., Korbutt, W. K., Nedelev, G. P., Bemman, B. Towards diagnostic support of hyperactivity in adults with ADHD using a virtual reality based continuous performance test and motion sensor data. *International Conference on Pervasive Computing Technologies for Healthcare* 2022:505–521	Did not report on test of interest
kolk. Power of combined modern technology: multitouch-multiuser tabletops and virtual reality platforms (PowerVR) in social communication skills training for children with neurological disorders: a pilot study. *Applied Neuropsychology: Child* 2022;**0**. www.tandfonline.com/doi/abs/10.1080/21622965.2022.2066532	Not an evaluation of the test
Mazancová, F. Cognitive screening tests and their potential to detect cognitive impairment in neurodegenerative diseases.pdf [PhD Thesis]. n.d. https://dspace.cuni.cz/bitstream/handle/20.500.11956/152560/140095799.pdf?sequence=1 (accessed 23 May 2022)	Not an evaluation of the test
Montoya-Arenas, D. A., Arbeláez-Vargas, J. F., Díaz-Soto, C. M. Rendimiento frontal y ejecutivo en niños en proceso de restablecimiento de derechos en Antioquia, Colombia. *Cuadernos Hispanoamericanos de Psicología* 2018;**18**(2):1–16. https://doi.org/10.18270/chpsv18i2.3051	Did not report on test of interest
Oliveira, J., Gamito, P., Alghazzawi, D. M., Fardoun, H. M., Rosa, P. J., Sousa, T., *et al.* Performance on naturalistic virtual reality tasks depends on global cognitive functioning as assessed via traditional neurocognitive tests. *Applied Neuropsychology: Adult* 2018;**25**:555–561. https://doi.org/10.1080/23279095.2017.1349661	Did not report on test of interest
Panerai, S., Catania, V., Rundo, F., Ferri, R. Remote home-based virtual training of functional living skills for adolescents and young adults with intellectual disability: feasibility and preliminary results. *Frontiers in Psychology* 2018;**9**:1730. https://doi.org/10.3389/fpsyg.2018.01730	Did not report on test of interest
Parsons, T. D. Technologically enhanced neuropsychological assessments. 2019;**35**.	Not a primary study or SR
Parsons, T. D., Lin, L., Cockerham, D. Mind, *Brain and Technology: Learning in the Age of Emerging Technologies*. Springer; 2018. https://link.springer.com/book/10.1007/978-3-030-02631-8	Not a primary study or SR
Rodríguez, C., Garcia, T., Areces, D., Rodríguez-Díaz, F., Arteaga, G., Ramos-Quiroga, A. Retrospective symptoms and learning difficulties predicting ADHD in adults: differences between prison inmates and the clinical population. *Scandinavian Journal of Psychology* 2021. https://doi.org/10.1007/s10.1111/sjop.12716	Did not report on test of interest
Sahu, A., Bajaj, J. Evidence-Based Immersive Technology Use in Cognitive Assessments and Cognition-Based Interventions. In *Emerging Advancements for Virtual and Augmented Reality in Healthcare*. IGI Global; 2022. pp. 193–215	Not primary study or SR
Shams, S., Farhadi, H. Effectiveness of the virtual reality package on social panic and social lectures. 2021	Did not report on test of interest

TABLE 40 Studies excluded at full-text screening from checking the Nesplora manufacturer's submission (*continued*)

Study details	Reason
Simons, A., Wohlgenannt, I., Zelt, S., Weinmann, M., Schneider, J., Brocke, J. vom. Intelligence at play: game-based assessment using a virtual-reality application. *Virtual Reality* 2023:1–17. https://doi.org/10.1007/s10055-023-00752-9	Did not report on test of interest
Voinescu, A., Petrini, K., Stanton Fraser, D. Presence and simulator sickness predict the usability of a virtual reality attention task. *Virtual Reality* 2023. https://doi.org/10.1007/s10055-023-00782-3	Does not include population with suspected or confirmed ADHD
Parsons, T.D. Bowerly, T. Buckwalter, J.G. Rizzo, A.A. A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence* 2007;**13**(4):363–81	Did not report on test of interest
Adams, R. Finn, P. Moes, E. Flannery, K. Rizzo, A.A. Distractibility in attention/deicit/hyperactivity disorder (ADHD): the virtual reality classroom. *Child Neuropsychology* 2009;**15**:120–35. 120. hurstone, L.L. Yela, M. (2001) CARAS: Test de percepción de diferencias (9a Edición). Madrid: TEA Ediciones	Did not report on test of interest

Note
This table reports studies included in manufacturer's submissions. We report the citation, as provided by the manufacturer, and record how the study has been processed in this review.

Appendix 3 Data extraction tables and risk-of-bias tables

TABLE 41 Baseline details for DTA studies included for objective 1

Study details	Setting and population	Index test	Reference standard
Adamou (2022)[81] **Design** One-gate **Country** United Kingdom **Funding** Unfunded	**Setting:** secondary care **Population:** adults **Inclusion criteria:** adults (18+ years) referred to Specialist Adult ADHD and Autism service; good comprehension of the English language; IQ > 70 **Exclusion criteria:** age < 18 years; intellectual disability **Number enrolled (number analysed):** 71 (69)	**Sensor CPT** QbTest (12–60)	**Reference standard** DSM-V **Details** All patients underwent routine clinical evaluation which involved a 'thorough psychiatric assessment by a doctor with expertise in ADHD and General Psychiatry. Including full psychiatric history, mental state examination, observations during assessments, and informant history'. This included the Diagnostic Interview for ADHD in Adults 2.0. Assessment led to 38 ADHD diagnoses and 31 non-ADHD
Bijlenga (2019)[78] **Design** Two-gate **Country** The Netherlands; Germany; Sweden **Funding** Not reported – second author employed by QbTech	**Setting:** secondary care **Population:** older adults **Inclusion criteria:** ADHD group (n = 97): 55+ years and meet DSM-IV ADHD diagnostic criteria. Control group: healthy controls (n = 112): 55+ years and no ADHD diagnosis **Exclusion criteria:** both groups: concurrent diagnosis that may affect test performance; mini-mental state examination score ≤ 23; other conditions that could affect test performance (e.g. migraine/physical disability); concurrent medications that could affect test performance significantly. Control group: past or current ADHD diagnosis; scored below cut-off on self-report measure for ADHD symptom severity **Number enrolled (number analysed):** 234 (209)	**Sensor CPT** QbTest (12–60) **Sensor CPT** QbTest (12–60) + clinical judgement **Clinical component** Symptom severity self-report scales	**Reference standard** ADHD group (n = 97): DSM-V ADHD diagnosis Controls (n = 112): healthy controls, with score below cut-off on symptom severity measures **Details** No further details **Reference standard** ADHD group (n = 97): DSM-IV-TR ADHD diagnosis, based on Diagnostic Interview for ADHD in Adults (DIVA 2.0) and rating scales Controls (n = 112): healthy controls, with score below cut-off on symptom severity measures **Details** No further details

TABLE 41 Baseline details for DTA studies included for objective 1 (*continued*)

Study details	Setting and population	Index test	Reference standard
Brunkhorst-Kanaan (2020)[68] **Design** One-gate **Country** Germany **Funding** Non-industry + industry	**Setting:** secondary care **Population:** adults **Inclusion criteria:** patients referred for diagnostic assessment for adult ADHD between July 2018 and July 2018 at the Department of Psychiatry, Psychosomatic Medicine and Psychotherapy. Following ADHD assessment, patients were separated into ADHD group (*n* = 94): confirmed ADHD diagnosis, and control group (*n* = 20): ADHD disconfirmed during diagnostic process **Exclusion criteria:** none reported **Number enrolled (number analysed):** 114 (114)	**Sensor CPT** QbTest (12–60)	**Reference standard** DSM-V [Diagnostic Interview for ADHD in Adults (DIVA) interview] **Details** Clinical ADHD diagnosis: DIVA interview undertaken, in which if certain criteria are met then a diagnosis of ADHD is plausible using DSM-V criteria. Assessment led to 94 ADHD diagnoses and 20 non-ADHD
Edebol (2011)[84] **Design** One-gate **Country** Sweden **Funding** Not reported	**Setting:** secondary care **Population:** adults **Inclusion criteria:** clinic-referred adult patients awaiting clinical assessment of ADHD at the 'NU-health care' hospital group **Exclusion criteria:** none reported **Number enrolled (number analysed):** 19 (19)	**Sensor CPT** QbTest (12–60)	**Reference standard** DSM-IV **Details** Clinical assessment was conducted by 'trained clinicians in the NU-health care'. This usually involved observation, self-report symptom scales, childhood anamnesis and information from relatives. It also often involved psychological or occupational tests and occasionally additional batteries of tests conducted by neuropsychiatry specialists. DSM-IV was used for diagnostic considerations This led to 12 ADHD diagnoses and 7 non-ADHD
Edebol (2012)[76] **Design** Four-gate **Country** Sweden **Funding** Industry and non-industry	**Setting:** secondary care **Population:** adults **Inclusion criteria:** 306 participants were included belonging to four groups: ADHD (*n* = 53): confirmed ADHD, as per DSM criteria, following assessment at outpatient clinic. Borderline/Bipolar (*n* = 45): confirmed borderline personality disorder or bipolar disorder. Disconfirmed (*n* = 29): assessed for ADHD but disconfirmed diagnosis. Healthy controls (*n* = 179): people aged 18–65 years who had no known psychiatric diagnoses and were willing to sign consent and complete study **Exclusion criteria:** none reported **Number enrolled (number analysed):** 306 (306)	**Sensor CPT** QbTest (12–60)	**Reference standard** ADHD group: DSM diagnosis (version not specified) (*n* = 53) B/B group: diagnosed with borderline/bipolar (*n* = 45) Disconfirmed ADHD (*n* = 29) (retained for analysis) Healthy controls (*n* = 179) **Details** No further details

continued

TABLE 41 Baseline details for DTA studies included for objective 1 (*continued*)

Study details	Setting and population	Index test	Reference standard
Edebol (2013)[79] **Design** Two-gate **Country** Sweden; Germany **Funding** Industry and non-industry	**Setting:** secondary care **Population:** adults **Inclusion criteria:** ADHD (*n* = 55): aged 18–65 years; DSM-IV ADHD diagnosis; chronic ADHD symptamology from childhood to adulthood with some symptoms present before 7 years old; accepted withdrawal from central stimulant treatment 24 hours to QbTest. Non-ADHD controls (*n* = 202): 18–65 years; sign informed consent and complete procedures; no known psychiatric diagnoses **Exclusion criteria:** ADHD: clinically unstable psychiatric condition including acute mood disorder, acute bipolar disorder, acute OCD or not meeting DSM-IV ADHD diagnosis. Non-ADHD controls: known psychiatric diagnosis **Number enrolled (number analysed):** 261 (257)	**Sensor CPT** QbTest (12–60)	**Reference standard** ADHD group (*n* = 55): diagnosed with ADHD following clinical assessment adhering to DSM-IV Non-ADHD control group (*n* = 202): healthy controls with no known psychiatric diagnoses **Details** No further details
Emser (2018)[83] **Design** Two-gate **Country** Germany **Funding** 'N/A'	**Setting:** secondary care **Population:** adults and children **Inclusion criteria:** children ADHD: Meet DSM-IV criteria for ADHD; IQ ≥ 80 on short version of Wechsler Intelligence Scale for Children IV; stop taking medication 2 days before sensor CPTs. Adult ADHD: Same as for children, except IQ not assessed (but all estimated to have ≥ 80IQ due to completing middle school) **Exclusion criteria:** ADHD: symptoms of inattention, hyperactivity or impulsivity due to other medical conditions; any genetic/medical disorder associated with externalising behaviour. Controls: established or suspected ADHD diagnosis or family history of ADHD **Number enrolled (number analysed):** 136 (NR)	**Sensor CPT** QbTest (12–60) + clinical judgement **Clinical component** TAP **Sensor CPT** QbTest (6–12) + clinical judgement **Clinical component** KiTAP	**Reference standard** ADHD (*n* = 68): DSM-IV-oriented clinical interview by experienced clinician, including KSADS and rating scales Controls (*n* = 68): no established or suspected ADHD diagnosis or family history of ADHD, unclear how assessed. Age/gender matched at group level **Details** No further details

TABLE 41 Baseline details for DTA studies included for objective 1 (*continued*)

Study details	Setting and population	Index test	Reference standard
Groom (2016)[80] **Design** Two-gate **Country** UK **Funding** Industry and non-industry	**Setting:** Secondary care **Population:** Adults **Inclusion criteria:** ADHD group (*n* = 32): DSM-V diagnosis of ADHD. Autism (ASD) group (*n* = 25): ICD-10 diagnosis of Asperger syndrome **Exclusion criteria:** ADHD group: disconfirmed ADHD diagnosis; non-completion of the test; continuation of ADHD medication during trial; dual diagnosis of ADHD and ASD; unavailable AQ10 scores. Autism group: ADHD group: disconfirmed autism diagnosis; non-completion of the test; continuation of psychostimulant medication during trial; dual diagnosis of ADHD and ASD **Number enrolled (number analysed):** 84 (57)	**Sensor CPT** QbTest (12–60) + clinical judgement **Clinical component** Conners Adult Rating Scale and Autism Quotient-10	**Reference standard** ADHD group (*n* = 32): DSM-V diagnosis using DIVA interview, in addition to clinical rating scales CAARS and AQ10 Autism (ASD) group (*n* = 25): ICD-10 diagnosis of Asperger syndrome **Details** No further details
Hamadache (2021)[29] **Design** Three-gate **Country** Germany **Funding** Unfunded	**Setting:** secondary care **Population:** children (age 5 years) **Inclusion criteria:** healthy controls: tested at pre-schools within early research efforts and found to be normally developing. Cases and controls with SLI: 63 children recruited from hospital social-paediatric centre **Exclusion criteria:** none reported **Number enrolled (number analysed):** NR (119)	**Sensor CPT** QbMini	**Reference standard** ADHD based on DSM-IV (*n* = 37) SLI (*n* = 27) Healthy controls: tested at pre-schools and found to be normally developing (*n* = 55) **Details** ADHD assessment was done using Fremdbeurteilungsbogen für Vorschüler mit Aufmerksamkeits- und Hyperaktivitätsstörungen (FBB-ADHS-V). A questionnaire which consists of four parts, of which the second part checks diagnostic criteria per DSM-IV
Hollis (2018)[18] **Design** One-gate **Country** England **Funding** Non-industry	**Setting:** secondary care; community **Population:** children and adolescents (age 6–16 years) **Inclusion criteria:** children aged 6–17 years referred for their first ADHD assessment **Exclusion criteria:** previous or current ADHD diagnosis; non-fluent in English; suspected moderate/severe intellectual disability **Number enrolled (number analysed):** 267 (250)	**Sensor CPT** QbTest (6–12) or QbTest (12–60) + clinical judgement **Clinical component** Clinical judgement	**Reference standard** Consensus diagnosis using DAWBA[89] DSM-V and ICD-10 **Details** Independent consensus research diagnosis made blind to group allocation using the DAWBA. Two experienced child psychiatrists reached clinical consensus diagnoses using DSM-V and ICD-10. They had access, where available, to the Children's Global Assessment Scale and SNAP-IV, but not clinic records or structured proformas. Assessment led to 69 ADHD diagnoses and 25 non-ADHD

continued

TABLE 41 Baseline details for DTA studies included for objective 1 (*continued*)

Study details	Setting and population	Index test	Reference standard
Hult (2018)[67] **Design** One-gate **Country** Sweden **Funding** Unfunded	**Setting:** secondary care **Population:** children (age 6–12 years) **Inclusion criteria:** children (age 6–12 years) with suspected ADHD, autism or another neurodevelopmental disorder. Diagnosis based on DSM-IV; assessed by multiprofessional team. Following ADHD assessment, patients separated into ADHD group (*n* = 124; ADHD diagnosis confirmed) and non-ADHD group (*n* = 58; ADHD diagnosis disconfirmed) **Exclusion criteria:** Medication with central stimulants at time of assessment; not valid QbTest; Wechsler scale assessment for IQ < 70; syndromal medical disorder diagnosis **Number enrolled (number analysed):** 182 (182)	**Sensor CPT** QbTest (6–12)	**Reference standard** DSM-IV **Details** All participants were assessed by multi-professional team using LEAD procedure, with clinical diagnosis of ADHD based on behavioural criteria according to DSM-IV. This led to 124 ADHD diagnoses and 58 non-ADHD
Johansson (2018)[70] **Design** One-gate **Country** Sweden **Funding** Non-industry	**Setting:** community **Population:** adolescents (age 15 years) **Inclusion criteria:** individual twins recruited from the DOGSS study if they had suspected neurodevelopmental disorder(s) and had been clinically assessed, including completion of the QbTest. Following ADHD assessment, participants were grouped into ADHD confirmed and ADHD disconfirmed **Exclusion criteria:** incomplete diagnostic information; taken ADHD medication prior to testing procedure **Number enrolled (number analysed):** 356 (340)	**Sensor CPT** QbTest (12–60)	**Reference standard** K-SADS-PL interview **Details** Psychologists used the diagnostic interview Schedule for Affective Disorders and Schizophrenia in School-Age Children (K-SADS-PL). This led to 89 ADHD diagnoses and 248 non-ADHD
Pettersson (2018)[82] **Design** One-gate **Country** Sweden **Funding** Non-industry	**Setting:** secondary care **Population:** adults **Inclusion criteria:** referral for ADHD assessment; age 18+ years; informant who knew patient as child willing to participate in clinical interview. Following ADHD assessment, patients separated into ADHD group (*n* = 60; ADHD diagnosis confirmed) and non-ADHD group (*n* = 48; ADHD diagnosis disconfirmed) **Exclusion criteria:** treatment with medications targeting ADHD; IQ ≤ 70 on WAIS-IV; substance-related disorder **Number enrolled (number analysed):** 108 (108)	**Sensor CPT** QbTestPlus **Comparator CPT** CPT-II	**Reference standard** DSM-IV **Details** The reference standard was expert clinical consensus. Clinical assessment was undertaken by team of psychologists/occupational therapist/MD specialising in neuropsychology [including interview using DIVA 2.0 (based on DSM-IV criteria), SCID-I, SCID-II]. This led to 60 ADHD diagnoses and 48 non-ADHD

TABLE 41 Baseline details for DTA studies included for objective 1 (*continued*)

Study details	Setting and population	Index test	Reference standard
Rufo-Campos (2012)[63] **Design** Two-gate **Country** Not reported **Funding** Not reported	**Setting:** not reported **Population:** children (age not reported) **Inclusion criteria:** ADHD group (*n* = 62): children diagnosed with ADHD. Non-ADHD group (*n* = 62): children without diagnosis **Exclusion criteria:** not reported **Number enrolled (number analysed):** 124 (124)	**Sensor CPT** Nesplora AULA	**Reference standard** Not reported
Seesjärvi (2022)[75] **Design** Two-gate **Country** Finland **Funding** Non-industry (but authors developed test)	**Setting:** secondary care **Population:** children (age 9–12 years) **Inclusion criteria:** ADHD group (*n* = 38): ADHD diagnosis by licensed physician using ICD-10 (with mainly hyperactive/impulsive subtype or combined inattention and hyperactive/impulsive subtype); age 9–12 years when recruited; native language Finnish. Non-ADHD group (*n* = 38): no mental or behavioural disorder **Exclusion criteria:** ADHD group: Any nervous system disease (ICD-10, G00–G99); any mental/behavioral disorders (F00–F99) except a secondary diagnosis of emotional disorder with childhood onset and unspecified behavioral and emotional disorder. Non-ADHD group: same as ADHD group except any mental or behavioural disorder excluded **Number enrolled (number analysed):** 115 (76)	**Sensor CPT** EPELI **Comparator CPT** continuous performance task	**Reference standard** ADHD group (*n* = 38): ADHD diagnosis by licensed physician using ICD-10 Non-ADHD group (*n* = 38): no mental or behavioural disorder; matched to cases; identified from questionnaires to the parents of the child where they were asked to list any diagnoses the child had **Details** No further details
Sharma (2009)[62] **Design** Unclear **Country** UK **Funding** Not reported	**Setting:** secondary care **Population:** children and adolescents (age 5–15 years) **Inclusion criteria:** children and adolescents (age 5–15 years) selected from QbTest database, which were evaluated for ADHD as per local protocol or as diagnosed by child/family guidance **Exclusion criteria:** age < 5 years or > 15 years **Number enrolled (number analysed):** 50 (50)	**Sensor CPT** QbTest (6–12) or QbTest (12–60)	**Reference standard** Assessment of disruptive behaviour pathway used locally as standard **Details** No further information; number with/without ADHD not reported

continued

TABLE 41 Baseline details for DTA studies included for objective 1 (*continued*)

Study details	Setting and population	Index test	Reference standard
Söderström (2014)[74] **Design** One-gate **Country** Sweden **Funding** Non-industry	**Setting:** secondary care **Population:** adults **Inclusion criteria:** referred to Neuropsychological Clinic in Vasteras Sweden for ADHD assessment between 1 September 2009 and 1 March 2011. Following ADHD assessment, patients separated into ADHD group (ADHD confirmed; *n* = 41) and non-ADHD group (ADHD disconfirmed, *n* = 20) **Exclusion criteria –** none reported **Number enrolled (number analysed):** 61 (61)	**Sensor CPT** QbTest (12–60)	**Reference standard** DSM-IV **Details** Clinical assessment for ADHD, including self-rating scales, clinical interview, intelligence testing and general psychiatric assessment. These relate to DSM-IV criteria and led to 41 ADHD diagnoses and 20 non-ADHD
Stevanovic (2023)[41] **Design** One-gate **Country** Sweden **Funding** Non-industry	**Setting:** secondary care **Population:** children and adolescents (mean age 13.5 years) **Inclusion criteria:** age 6–18 years; undergone QbTest or QbTest Plus at department of CAP in one of a few general hospitals in Sweden; availability of reliable QbTest scores. Following ADHD assessment, participants separated into ADHD group (*n* = 708) **Exclusion criteria:** severe mental and/or neurodevelopmental disorders meaning could not understand or perform test accurately; inability to understand/perform test accurately **Number enrolled (number analysed):** 1274 (928)	**Sensor CPT** QbTest (12–60) QbTest (6–12)	**Reference standard** Diagnostic process according to clinic's standard diagnostic procedure – no further information. Process led to ADHD confirmed (*n* = 708); no ADHD (*n* = 220)
Tallberg (2019)[66] **Design** One-gate **Country** Sweden **Funding** Non-industry	**Setting:** secondary care **Population:** children (age 9–14 years) **Inclusion criteria:** diagnostic study: children who screened positive for ADHD and were referred for further assessments in CAP clinic in southern Sweden between 1 November 2009 and 31 December 2010 (*n* = 118, of which, following assessment, 80 were diagnosed with ADHD and 38 had disconfirmed diagnosis) **Exclusion criteria:** none reported **Number enrolled (number analysed):** 118 (118)	**Sensor CPT** QbTest (6–12) **Comparator CPT** Conners CPT II confidence index	**Reference standard** DSM-IV. Process led to ADHD confirmed (*n* = 80); no ADHD (*n* = 38)

TABLE 41 Baseline details for DTA studies included for objective 1 (*continued*)

Study details	Setting and population	Index test	Reference standard
Ulberstad (2020)[77] **Design** Two-gate **Country** Germany, Sweden, USA **Funding** Industry – authors employed by QbTech	**Setting:** secondary care **Population:** adolescents and adults (12–59 years) **Inclusion criteria** Cases (*n* = 69): meet ADHD diagnostic criteria according to DSM-V. Controls (*n* = 73): healthy controls (convenience sample) **Exclusion criteria** Control group: high levels of inattention or hyperactivity/impulsivity according to DSM-V **Numbers** 142 (142)	**Sensor CPT** QbCheck	**Reference standard** ADHD (*n* = 69): DSM-V diagnostic criteria Controls (*n* = 73): healthy controls; those with high levels of inattention/hyperactivity/impulsivity according to DSM-V excluded **Details** No further details
Zulueta (2019)[73] **Design** Two-gate **Country** Spain **Funding** Not reported	**Setting:** secondary care **Population:** children (age 6–16 years) **Inclusion criteria:** ADHD group recruited from outpatient services (*n* = 213): age 6–16 years; ADHD positive from ADHD diagnostic assessment at outpatient service (neuropsychology clinic or paediatric neurology clinic); IQ within the normal limits (IQ > 80); consent to participate; off stimulants medication for 48 hours prior to testing Typically developing controls recruited from schools (*n* = 194): age 6–16 years; IQ within the normal limits (IQ > 80); consent to participate; ADHD negative (/ minimal ADHD symptoms) from ADHD diagnostic assessment at outpatient service and no other behavioural disorder **Exclusion criteria** – none reported **Number enrolled (number analysed):** 407 (407)	**Sensor CPT** Nesplora Kids (AULA)	**Reference standard** ADHD (*n* = 213): DSM-V criteria, measured using ADHD Rating Scale-IV Healthy control group (*n* = 194): from schools and neurology clinics minimal ADHD symptoms and no other behavioural disorder **Details** No further details

B/B, borderline/bipolar; CAP, Child and Adolescent Psychiatry; CPT-II: Conners' Continuous Performance Test II; OCD, obsessive–compulsive disorder; SCID-I, Structured Clinical Interview for DSM-IV v1; SCID-II, Structured Clinical Interview for DSM-IV v2; TAP, Test of Attentional Performance for Children.

TABLE 42 PROGRESS-Plus information reported in DTA studies and RCTs included for objective 1

Study details	Progress plus item	Details
Adamou (2022)[81]	Age – mean (SD; range)	33 (9.9; not reported)
	Sex (% male)	65.2%
	Neurodevelopmental/ learning disorders	No intellectual disability
Bijlenga (2019)[78]	Age – mean (SD; range)	**ADHD:** 63.2 (4.8); **control:** 64.4 (5.4). **Total sample range:** 55–79
	Sex (% male)	**ADHD:** 46.4%; **control:** 45.5%
	Education (% highest education)	**ADHD:** primary school/none 10.6%; lower-level professional education 25.5%; higher level professional education 17%; college/university 46.8%. **Control**: primary school/ none 20.2%; lower level 13.1%; higher level 9.5%; college/university 57.1%
	Mental health disorders	**ADHD:** depression 26.8%; anxiety 12.4%; bipolar depression 4.1%; substance use or addiction 4.1%; other 13.4%; use of psychiatric medication 30.9%. **Control:** not reported
	Neurodevelopmental/ learning disorders	**ADHD:** ADHD combined subtype 79.4%; inattentive subtype 20.6%; symptom severity mean 56.7 (SD 16) **Control:** ADHD subtypes not reported; symptom severity mean 22.1 (SD 10.8)
Brunkhorst-Kanaan (2020)[68]	Age – mean (SD; range)	**ADHD:** 34.7 (11.05; not reported); **Control:** 35.8 (10.6; not reported)
	Sex (% male)	**ADHD:** 57.4%; **control:** 40%
	Mental health disorders	**ADHD:** depression 27.7%; substance use disorder 18.1%; bipolar 2.1%; other (e.g. PTSD, OCD and somatisation disorders) 8.5%. Overall, 47.9% had a comorbidity with ≥ 1 other psychiatric disorder, 4.4% had > 2 other psychiatric disorders. Patients with affective comorbidities all suffered moderate-severe depressive episodes at time of examination. **Control:** depression 45%; substance use disorder 10%; bipolar 5%; other 15%. 12/20 patients had a psychiatric disorder
Chitsabesan (2022)[72] (RCT)	Age – percentages in each age category	**QbTest (*n* = 30)**: age 16 years: 20%; a17: 26.7%; age 18 years: 50%; missing: 3.3% **Usual care (*n* = 30)**: age 16 years: 10%; age 17 years: 36.7%; age 18 years: 53.3%; missing: 0%
	Sex (% male)	100%
	Ethnicity	**QbTest (*n* = 30)**: white 76.7%; other 20%; missing 3.3%. **Usual care (*n* = 30)**: white 80%; other 20%; missing 0%.
	Education	**QbTest (*n* = 30):** mainstream 20%; pupil referral unit 10%; none 66.7%; other 0%; missing 3.3%. **Usual care (*n* = 30)**: mainstream 20%; pupil referral unit 16.7%; none 56.7%; other 6.7%; missing 0%
	Time-dependent relationships	All participants in youth justice system
Edebol (2011)[84]	Age – mean (SD; range)	31.7 (9.3; 20–54)
	Sex (% male)	47%
	Occupation	Majority were unemployed, on sick leave or carried sickness pension (*n* = 14), and the remaining had full-time or part-time work (*n* = 1), arranged daytime activities (*n* = 1), studied (*n* = 1), were on parents leave (*n* = 1) or retired (*n* = 1)
	Education	1 person had not begun high school, but a majority had completed it (*n* = 6) or made part of it (*n* = 9) and some (*n* = 3) had studied at postgraduate levels

TABLE 42 PROGRESS-Plus information reported in DTA studies and RCTs included for objective 1 (*continued*)

Study details	Progress plus item	Details
	Mental health disorders	Mean age for initial psychiatric contact was 20.2 (SD 10.9, *n* = 12), and 10 people had undertaken psychiatric hospitalisation one or more times starting at the mean age of 27.1 (SD 9.9, range 14–48). The sample indicated both serious symptoms and dysfunctions with Global Assessment of Functioning symptom severity at mean 49.9 (SD 6.9, range 40 = 60) and level of adaptive functioning at M48.2 (SD 8.8, range 35–60). Prior to ADHD assessment, all but two participants had at least one psychiatric diagnosis and some (*n* = 8) had two. In total, relapsing episodes of depression or dysthymia (seven); anxiety disorders or mixed anxiety/depression (five); bipolar disorder (three); substance use disorder (three); personality disorder (two); adaptive disorder (one); acute stress reaction (one)
	Neurodevelopmental/ learning disorders	Majority (*n* = 14) had no family or relative with ADHD, and none had undergone ADHD assessment before
	Relationship features	Nine were single, six either married or sharing household with a partner, three had a relationship and one person was divorced
Edebol (2012)[76]	Age – mean (SD; range)	**ADHD (*n* = 53)**: 35.89 (12.25; 18–64). **Bipolar/borderline personality (B/B; *n* = 45)**: 42.33 (11.63, 22–60). **Disconfirmed group (*n* = 29):** 35.21 (10.31, 20–54). **Normative group (*n* = 179):** 31.45 (10.33, 18–53)
	Sex (% male)	**ADHD:** 45%; **B/B:** 29%; **disconfirmed:** 45%; **normative**: 55%
	Occupation (% employed)	**ADHD:** 58%; **B/B**: 27%; **disconfirmed:** 38%; **normative:** not reported
	Education (% highest education)	**ADHD:** high school 23%; senior high school 57%; graduate school 19%; **B/B:** high school 27%; senior high school 62%; graduate school 9%; **disconfirmed:** high school 21%; senior high school 69%; graduate school 10%; **normative:** not reported
	Mental health disorders	**ADHD:** 9 participants had 1 (*n* = 7) or 2 (*n* = 2) psychiatric disorders, including dyslexia (3); social phobia (3); generalised anxiety disorder (1); depression (2); stress reaction (1), emotionally instable personality disorder (1). **B/B:** Bipolar disorder (27); borderline personality disorder (18). 13 participants had 1 or several additional diagnoses, including psychological and behavioural disturbances, because of substance use (4); generalised anxiety disorder (3); social phobia (2); panic disorders (1); anxiety and depression (2); adaption disorder (1); relapsing depression (1); 2 with borderline personality disorder also had bipolar disorders. **Disconfirmed:** no psychiatric diagnoses (8); 2 psychiatric diagnoses (12). Of the people with diagnoses, these included Asperger syndrome (6); dyslexia (4); personality disorders (4); borderline personality disorder (1); bipolar unspecified (2); OCD (1): PTSD (1); memory disorder unspecified (1); as well as secondary diagnoses of depression (3); dyscalulia (2); attention disorders unspecified (2); DCD (1); tics (1); social phobia (1); dysmorphobia (1); mixed substance use disorder (1). **Normative:** exclusion criteria were any known psychiatric diagnoses
	Relationship features (marital status)	**ADHD:** 51% single; 38% married/spouse; 9% divorced/separated. **Disconfirmed:** 62% single; 34% married/spouse; 3% divorced/separated. **B/B:** 38% single; 51% married/ spouse; 9% divorced/separated. **Normative:** not reported
	Place of residence	**ADHD:** live alone 60%; live with spouse 38%; group home 0%. **B/B:** live alone 33%; live with spouse 60%; group home 4%. **Disconfirmed:** live alone 55%; live with spouse 45%; group home 0%. **Normative:** Not reported
Edebol (2013)[79]	Age – mean (SD; range)	**ADHD:** 33.35 (8.84; not reported); **non-ADHD:** 31.06 (10.27; 18–53)
	Sex (% male)	**ADHD:** 45.5%. **Non-ADHD**: 56%
	Occupation (% employment type)	**ADHD:** sick leave 32.7%; full-time/part-time employment 23.6%; rehabilitation/practice 12.7%; unemployed 10.9%; studying 9.1%; retired 7.3%; parental leave 3.6%. **Non-ADHD**: not reported
	Education (% highest education)	**ADHD:** junior high school 21.8%; partial high school 27.3%; complete high school 34.6%; partial graduate school 9.1%; complete graduate school 7.3%. **Non-ADHD**: not reported
	Socioeconomic status (% income type)	**ADHD:** income by public maintenance 68.7%; employment 20.4%; student loans 7.4%; other income 3.7%. **Non-ADHD**: not reported

continued

TABLE 42 PROGRESS-Plus information reported in DTA studies and RCTs included for objective 1 (*continued*)

Study details	Progress plus item	Details
	Mental health disorders	**ADHD:** substance abuse 18.4%; relapsing/moderate depression 18.4%; anxiety disorders 21.1%; mixed anxiety/depression 5.3%; bipolar disorders 15.8%; personality disorders 10.5%; adjustment disorders 5.3%; 43.6% had no current psychiatric comorbidity, 43.6% had one and 12.7% had two. **Non-ADHD:** not reported, but an exclusion criterion is presence of 'unstable psychiatric condition'
	Neurodevelopmental/ learning disorders	**ADHD:** autism 2.6%; dyslexia 2.6%; adjustment disorders (5.3%); personality disorders (10.5%); bipolar disorders (15.8%); mixed anxiety/depression (5.3%); anxiety disorders (21.1%); relapsing/moderate depression (18.4%); substance abuse (18.4%). **Non-ADHD:** not reported
	Relationship features (marital status; household set-up)	**Total sample:** 38.2% married/common law; 41.9% single; 20% partner **Total sample:** 43.6% single household; 56.4% shared household
Emser (2018)[83]	Age – mean (SD; range)	**ADHD (adults):** 35.1 (11.7; 19–63); **control (adults):** 32.2 (9.6; 21–56). **ADHD (children):** 8.9 (1.4; 7–11). **Control (children):** 8.7 (1.2; 6.9–10.8)
	Sex (% male)	**ADHD (adults):** 65.8%; **control (adults):** 65.8%. **ADHD (children):** 70%; **control (children):** 63.3%
	Education	All adults included had completed middle school
	Neurodevelopmental/ learning disorders	**ADHD (adults):** not reported; **control (adults):** not reported; **ADHD (children):** mean IQ 113.1 (SD 11.6). **Control (children):** mean IQ 125.8 (SD 10.8)
	Place of residence	Small university town: high mean IQ said to be due to large proportion of children from academic families in small university town
Groom (2016)[80]	Age – mean (SD; range)	**ADHD:** 31.64 (10.17; not reported); **ASD:** 33.22 (11.74; not reported)
	Sex (% male)	**ADHD:** 63%; **ASD:** 76%
	Socioeconomic status (index of multiple deprivation categories; decile ranks; low ranks indicate high level of deprivation, high ranks indicate low deprivation)	**ADHD:** low 50%, middle 18%, high 32%. **ASD:** low 64%, middle 12%, high 24%.
	Mental health disorders	**ADHD:** depression (2); anxiety disorder (2); emotionally unstable personality disorder (i.e. borderline personality) (2). **ASD:** anxiety (4); depression (2); anxiety and depression (1); bipolar (1); substance misuse (1)
Hamadache (2021)[29]	Age – mean (SD; range)	**ADHD:** 5.53 (not reported); **controls:** 5.45 (not reported). All aged 5 years
	Sex (% male)	**ADHD:** 81% boys; **control** 56% boys; **SLI**: 67%
	Neurodevelopmental/ learning disorders	**ADHD:** motor disorder 8.3%; epilepsy 2.8%; language disorder 27.8%; tic disorder 5.5%; IQ 100.69. **SLI:** motor disorder 4%; epilepsy 0%; Language disorder 100%; tic disorder 0%; IQ 97.27
	Developmental trauma (% premature birth)	**ADHD:** 11%; **SLI**: 4%
Hollis (2018)[18]	Age – mean (SD; range)	**QbOpen:** 9.5 (2.8; 6.0–17.4); **QbBlind:** 9.4 (2.8; 5.9–16.2)
(RCT with DTA substudy)	Sex (% male)	**QbOpen:** 77%; **QbBlind:** 80%.
	Ethnicity (% white, mixed, other)	**QbOpen** (data from 83/123 participants): white 88%; mixed and other 12%. **QbBlind** (89/127 participants): white 90%; mixed and other 10%
	Neurodevelopmental/ learning disorders	Diagnoses (n = 241; allows more than one diagnosis per patient): 71% ADHD; 35% oppositional defiant disorder/conduct disorder; 20% any anxiety disorder; 17% chronic tic disorder/Tourette syndrome; 9% ASD; 3% depressive disorder; 11% learning difficulties; 0.4% attachment disorder; 19% no psychiatric diagnoses

TABLE 42 PROGRESS-Plus information reported in DTA studies and RCTs included for objective 1 (*continued*)

Study details	Progress plus item	Details
Hult (2018)[67]	Age – mean (SD; range)	**ADHD:** 10.3 (1.7; not reported); **non-ADHD:** 10.8 (1.8; not reported)
	Sex (% male)	**ADHD:** 97%; **non-ADHD** 53%
	Mental health disorders	**ADHD:** depression/anxiety 5%; **non-ADHD:** depression/anxiety 7%.
	Neurodevelopmental/ learning disorders	**ADHD:** ASDs 28%; tic disorders 4%; DCD 32%; borderline intellectual functioning 10%; dyslexia 31%; language disorder 9%; mean full-scale IQ (SD): 89.5 (13.2). **non-ADHD** ASDs 81%; tic disorders 12%; DCD 7%; borderline intellectual functioning 16%; dyslexia 10%; language disorder 10%; mean full scale IQ (SD) 92.2 (14.6)
Johansson (2018)[70]	Age – mean (SD; range)	15 (not reported; 14–16)
	Sex (% male)	**ADHD:** 70.79%; **non-ADHD:** 49.8%
	Education (parental education of mother and father)	**Mother – ADHD:** elementary school 8.99%; secondary school 59.55%; high school 29.21%; unknown 2.25%. **Mother – non-ADHD:** elementary school 8.76%; secondary school 49.41%; high school 37.06%; unknown 4.71%. **Father – ADHD**: elementary school 11.24%; secondary school 42.7%; high school 23.6%; unknown 22.47%. **Father – non-ADHD:** elementary school 15.14%; secondary school 37.45%; high school 28.69%; unknown 18.73%
	Features of relationships	**ADHD:** major school problems (failing to receive grades or repeating school year): 40.45%; antisocial behaviour (criminal or violent behaviour): 14.61%
		Non-ADHD: Major school problems: 11.95%; Antisocial behaviour 17.13%
	Mental health disorders	**ADHD:** psychiatric condition other than ADHD 77.5%; anxiety 23.6%; stress-related disorder 8.99%; depression life time 8.99%; OCD 6.74%; substance/alcohol misuse 6.74%; eating disorder 2.25%; bipolar disorder 0%; psychosis 0%. **Non-ADHD:** psychiatric condition other than ADHD 59%; Anxiety 20.72%; stress-related disorder 10.76%; depression life time 7.97%; OCD 3.19%; substance/alcohol misuse 2.39%; eating disorder 3.19%; bipolar disorder 0.8%; psychosis 0%
	Neurodevelopmental/ learning disorders	**ADHD:** language disorder 37.08%; tic disorder 22.47%; oppositional defiant disorder 12.36; conduct disorder 7.87%; autism 1.12%, total IQ < 70: mean 4 (SD 4.49). **Non-ADHD:** language disorder 23.9%; tic disorder 11.6%; oppositional defiant disorder 1.2%; conduct disorder 0.8%; autism 1.99%; total IQ < 70 mean 12 (SD 4.78)
Pettersson (2018)[82]	Age – mean (SD; range)	**ADHD:** 28.18 (9.09; not reported); **Non-ADHD:** 32.75 (10.61; not reported)
	Sex (% male)	**ADHD:** 53.3%; **non-ADHD:** 52.1%
	Occupation (employment type %)	**ADHD:** full-time work/studying 56.7%; part-time work/studying 15%; unemployment/ vocational training 21.7%; long-term sick leave/disability pension 6.7%. **Non-ADHD**: full-time work/studying 41.7%; part-time work/studying 22.9%; unemployment/vocational training 16.7%; long-term sick leave/disability pension 18.8%
	Education – mean years (SD)	**ADHD:** 11.72 (1.85); **non-ADHD:** 12.32 (1.60)
	Mental health disorders	**ADHD:** Beck Depression Inventory: mean 17.25 (SD 12.70); Beck Anxiety inventory mean 11.70 (SD 10.29); Mental health diagnoses – Axis 1 diagnosis (one or more) 50%; Axis II diagnosis (one or more) 16.7%. Distribution of Axis I and II diagnoses: mood disorder 25%; anxiety disorder 43.3%; other Axis I disorder 16.7%; Axis II Cluster A disorder 5%; Axis II Cluster B disorder 8.3%; Axis II Cluster C disorder 10%. Estimated IQ: mean 91.52 (SD 12.31)
		Non-ADHD: Beck Depression Inventory: mean 23.83 (SD 12.87); Beck Anxiety inventory mean 17.96 (SD 11.98); mental health diagnoses: Axis 1 diagnosis (one or more) 83.3%; Axis II diagnosis (one or more) 45.8%. Distribution of Axis I and II diagnoses: mood disorder 43.8%; anxiety disorder 68.8%; other Axis I disorder 47.9%; Axis II Cluster A disorder 12.5%; Axis II Cluster B disorder 8.3%; Axis II Cluster C disorder 31.2%; estimated IQ: 98.96 (SD 13.74)

continued

TABLE 42 PROGRESS-Plus information reported in DTA studies and RCTs included for objective 1 (*continued*)

Study details	Progress plus item	Details
Rufo-Campos (2012)[63]	No PROGRESS-Plus information reported (conference abstract)	
Seesjärvi (2022)[75]	Age – mean (SD; range)	**ADHD:** 10 years 4 months (1 year 1 month; not reported); **non-ADHD:** 10 years 9 months (1 year 1 month; not reported)
	Sex (% male)	Unclear
	Education	**ADHD:** 2.4 (0.6); **non-ADHD:** 2.7 (0.5)
	[Mean (SD) parental education: 1 comprehensive school, 2 high school/vocational school, 3 university degree or equivalent]	
	Socioeconomic status	**ADHD:** 3.7 (1); **non-ADHD:** 4 (1)
	[Mean (SD) parental income before tax per adult: 1: < 1500 €/month, 2: 1500–2200 €/month, 3: 2200–3000 €/month, 4: 3000–4000 €/month, 5: > 4000 €/month]	
	Neurodevelopmental/ learning disorders	**ADHD:** conduct disorder *n* = 3; oppositional defiant disorder *n* = 4; OCD *n* = 1; Tourette's *n* = 1; provisional tic disorder *n* = 1. **Non-ADHD:** exclusion criteria was any mental or behavioural disorder
Sharma (2009)[62]	Age – mean (SD; range)	Only range reported: 5–15 years
Söderström (2014)[74]	Age – mean (SD; range)	**ADHD:** 32.46 (8.99; not reported); **non-ADHD:** 30 (9.76; not reported)
	Sex (% male)	**ADHD (*n* = 41):** 43.9%: **non-ADHD (*n* = 20):** 40%
	Mental health disorders	**Total sample (*n* = 61):** 63.9% had previously had contact with psychiatric services and had one or more psychiatric diagnoses
		ADHD (*n* = 41): Axis I or Axis II (Cluster B diagnoses) 56.1%, of which: mood disorders 39%; anxiety disorders 31.7%; Axis II Cluster B disorders 4.9%; substance dependence disorders 7.3%.
		Non-ADHD (*n* = 20): Axis I or Axis II (Cluster B diagnoses) 80%, of which: mood disorders 45%; anxiety disorders 60%; Axis II Cluster B disorders 5%; substance dependence disorders 5%
Stevanovic (2023)[41]	Age – mean (SD; range)	**Total sample (*n* = 1274) –** 13.5 (3.2; not reported)
	Sex (% male)	**Total sample (*n* = 1274) –** 59.9%
	Neurodevelopmental/ learning disorders	**Total sample (*n* = 1274). ADHD:** ASD 31.9%; another mental behavioural or neurodevelopmental disorder other than ASD 31.6%. **Non-ADHD:** any mental behavioural or neurodevelopmental disorder other than ADHD 81.8%; no diagnosis assigned/clinical controls 18.2%. Intellectual difficulties: 32 people (excluded from analysis)
Tallberg (2019)[66]	Age – median (median first–third quartiles)	**ADHD**: 12.5 (9.6–14.4); **non-ADHD:** 11.2 (9.6–13.0)
	Sex (% male)	**ADHD**: 71%; **non-ADHD**: 63%
	Mental health disorders	**ADHD**: not reported. **Non-ADHD**: internalised problems such as mood disorder or anxiety disorder *n* = 12

TABLE 42 PROGRESS-Plus information reported in DTA studies and RCTs included for objective 1 (*continued*)

Study details	Progress plus item	Details
	Neurodevelopmental/ learning disorders	**ADHD**: % comorbid disorders not reported; Wechsler Intelligence Scale for Children IQ mean 87.15 CI 74.58 to 99.72. **Non-ADHD:** ASDs *n* = 5; tic disorders *n* = 3; language impairments or learning disorders *n* = 12; internalised problems such as mood disorder or anxiety disorder; no diagnostic criteria *n* = 14; Wechsler Intelligence Scale for Children IQ: mean 91.86 CI 78.59 to 105.13. 'Two cases had full scale IQ just below 70, but with uneven cognitive profiles'
Ulberstad (2020)[77]	Age – mean (SD; range)	**ADHD:** 27.58 (12.12); **control**: 26.16 (9.55). **Total sample:** range 12–60
	Sex (% male)	**ADHD:** 52.2%; **control**: 43.8%
Zulueta (2019)[73]	Age – mean (SD; range)	**ADHD-combined:** 9.78 (2.65; not reported). **ADHD-inattentive:** 10.62 (2.79; not reported). **Control:** 9.08 (2.66; not reported)
	Sex (% male)	**ADHD-combined:** 76.9%; **ADHD-inattentive:** 69.5%; **control:** 59.8%
	Neurodevelopmental/ learning disorders	**IQ mean (SD) – ADHD-combined:** 101.46 (SD 10.77); **ADHD-inattentive:** 98.78 (10.16); **control:** 101.44 (10.55). Controls had no other behavioural disorder and minimal symptoms of ADHD reported on parent and teacher rating scales

DCD, developmental co-ordination disorder; PTSD, post-traumatic stress disorder.
Note
Exchange rate of €1 = £0.84, as of 17 March 2025.

TABLE 43 Results for DTA studies included for objective 1

Study details	Index test	Measure and Subgroup	Threshold	Reference stand	TP	FP	FN	TN	Sensitivity	Specificity	AUC (95% CI)
Adamou (2022)[81]	QbTest (12–60)	Overall	1.5	DSM-V	27	18	11	13	0.71	0.42	NR
Bijlenga (2019)[78]	QbTest (12–60) + clinical judgement	QBHyperactivity + inattention	1.5	DSM-IV	88	10	9	102	0.91	0.91	NR
	QbTest (12–60)	QBHyperactivity + inattention		DSM-V	54	19	43	93	0.56	0.83	NR
Brunkhorst-Kanaan (2020)[68]	QbTest (12–60)	QBImpulsivity	1.5	DSM-V	NR						0.54 (0.52 to 0.56)
		QBInattention	1.5		NR						0.56 (0.54 to 0.57)
		QBActivity	2.35		45	5	49	15	0.48	0.75	0.65 (0.63 to 0.67)
		QBActivity	1.5		64	10	30	10	0.68	0.5	
		QBActivity	2.95		26	2	68	18	0.28	0.90	
Edebol (2013)[79]	QbTest (12–60)	Overall	NR	DSM-IV	47	35	8	167	0.85	0.83	NR
Edebol (2012)[76]	QbTest (12–60)	Overall; *all controls combined*	NR	DSM (version NR)	46	73		180	0.87	0.71	NR
		Overall; *disconfirmed ADHD Only*[a]				17		12	0.87	0.41	NR
		Overall; *bipolar group*				29		16	0.87	0.36	NR
		Overall; *healthy controls*				27		152	0.87	0.85	NR
Edebol (2011)[84]	QbTest (12–60)	Overall	> 1.3	DSM-IV	10	3	2	4	0.83	0.57	NR
Emser (2018)[83]	QbTest (12–60) + clinical judgement	Overall	NR	DSM-IV	31	9	7	29	0.82	0.76	NR
	QbTest (6–12) + clinical judgement	Overall			24	7	6	23	0.80	0.77	NR
Groom (2016)[80]	QbTest (12–60) + clinical judgement	Overall	NR	DSM-V	30	4	2	21	0.94	0.84	0.87

TABLE 43 Results for DTA studies included for objective 1 (*continued*)

Study details	Index test	Measure and Subgroup	Threshold	Reference stand	TP	FP	FN	TN	Sensitivity	Specificity	AUC (95% CI)
Hamadache (2021)[29]	QbMini	QBActivity; *healthy controls*	NR	DSM-IV	NR						0.800
		QBActivity; *SLI group control*[a]									0.506
		QBInattention; *healthy controls*									0.670
		QBInattention; *SLI group*[a]									0.524
		QBImpulsivity; *SLI group*[a]									0.594
		QBImpulsivity; *healthy controls*									0.589
Hollis (2018)[18]	QbTest (6–12) or QbTest (12–60) + clinical judgement	Overall	NR	DAWBA	37	26	6	17	0.86	0.40	NR
	Clinical judgement alone	Overall	NR	DAWBA	49	16	2	9	0.96	0.36	NR
Hult (2018)[67]	QbTest (6–12)	QBActivity: *total sample*	1.25	DSM-IV	78	15	46	43	0.63	0.74	0.74 (0.66 to 0.82)
		QBActivity: *ADHD combined subgroup*			59	15	29	43	0.67	0.74	0.74 (0.66 to 0.83)
		QBActivity: *ADHD inattentive subgroup*			18	15	12	43	0.60	0.74	0.73 (0.63 to 0.84)
		QBImpulsivity: total sample			52	16	72	42	0.42	0.72	0.62 (0.53 to 0.70)
		QBImpulsivity: *ADHD combined subgroup*			39	16	49	42	0.44	0.72	0.62 (0.53 to 0.71)
		QBImpulsivity: *ADHD inattentive subgroup*			11	16	19	42	0.37	0.72	0.62 (0.50 to 0.74)
		QBInattention: *total sample*			60	10	64	48	0.48	0.83	0.76 (0.69–0.84)
		QBInattention: *ADHD combined subgroup*			45	10	43	48	0.51	0.83	0.77 (0.69 to 0.85)
		QBInattention: *ADHD inattentive subgroup*			14	10	16	48	0.47	0.83	0.76 (0.66 to 0.86)

continued

TABLE 43 Results for DTA studies included for objective 1 (*continued*)

Study details	Index test	Measure and Subgroup	Threshold	Reference stand	TP	FP	FN	TN	Sensitivity	Specificity	AUC (95% CI)
Johansson (2018)[70]	QbTest (12–60)	QBInattention	NR	K-SADS-PL interview					0	0	0.59
		QBImpulsivity							0	0	0.58
		Overall			60	103	29	145	0.67	0.58	0.58
		QBActivity							0	0	0.49
Pettersson (2018)[82]	QbTestPlus	QBActivity	> 1.5	DSM-IV	46	27	14	21	0.77	0.44	0.664
		QBInattention	> 1.5	DSM-IV	35	16	25	32	0.58	0.67	0.673
		QbReactionTimeVariance	> 1.5	DSM-IV	26	12	34	36	0.43	0.75	0.674
		QbOmissionerrors	> 1.5	DSM-IV	44	21	16	27	0.73	0.56	0.725
	CPT-II	CPTIICom	> 1.5	DSM-IV	20	4	40	44	0.33	0.92	0.741
		CPTIIVar	> 1.5	DSM-IV	16	7	44	41	0.27	0.85	0.706
Rufo-Campos (2012)[63]	Nesplora Kids (AULA)	Overall	NR	NR	Overall accuracy 93.5%						
Seesjärvi (2022)[75]	EPELI	Overall	46.5	ICD-10	29	17	9	21	0.76	0.55	0.70 (0.59 to 0.82)
		EPELITaskEfficacy	0.29		25	4	13	34	0.66	0.89	0.83 (0.74 to 0.92)
		EPELINavigationEfficacy	0.06		29	13	9	25	0.76	0.66	0.75 (0.64 to 0.86)
		EPELIControllerMotion	68,588.85		27	13	11	25	0.71	0.66	0.73 (0.62 to 0.85)
		EPELIActions	463		23	4	15	34	0.61	0.89	0.78 (0.68 to 0.89)
	CPT	CPT omission errors	3.5		19	8	19	30	0.5	0.79	0.70 (0.57 to 0.82)
		CPT reaction time variability	150.3		33	9	5	29	0.87	0.76	0.85 (0.76 to 0.94)
		CPT commission errors	13.5		30	19	8	19	0.79	0.5	0.70 (0.58 to 0.82)

TABLE 43 Results for DTA studies included for objective 1 (*continued*)

Study details	Index test	Measure and Subgroup	Threshold	Reference stand	TP	FP	FN	TN	Sensitivity	Specificity	AUC (95% CI)
Sharma (2009)[62]	QbTest (6–12) or QbTest (12–60)	Overall	NR	NR	27	4	1	17	0.96	0.81	NR
Söderström (2014)[74]	QbTest (12–60)	QBActivity	1.5	DSM-IV	28	7	13	13	0.68	0.65	0.666
		QBImpulsivity			24	4	17	16	0.59	0.80	0.683
		QBInattention			15	0	26	20	0.37	1	0.693
Stevanovic (2023)[41]	QbTest (6–12)	QBActivity	1.5	Clinic's standard diagnostic procedure	73	4	264	85	0.22	0.96	0.59 (0.54 to 0.64)
		QBInattention			168	19	171	68	0.5	0.78	0.64 (0.59 to 0.69)
		QBImpulsivity			90	6	252	78	0.26	0.93	0.59 (0.54 to 0.64)
	QbTest (12–60)	QBActivity			143	23	218	118	0.40	0.84	0.62 (0.57 to 0.66)
		QBInattention			124	24	239	115	0.34	0.83	0.58 (0.54 to 0.63)
		QBImpulsivity			117	19	249	117	0.32	0.86	0.59 (0.55 to 0.63)
Tallberg (2019)[66]	QbTest (6–12)	QbActivity	NR	DSM-IV	45	20	35	18	0.56	0.47	0.48 (0.36 to 0.61)
		QbInattention			43	17	37	21	0.54	0.55	0.59 (0.46 to 0.72)
		QbImpulsivity			38	11	42	27	0.48	0.71	0.60 (0.49 to 0.72)
	Conners CPT II confidence index	NR			70	18	10	20	0.88	0.53	0.73 (0.62 to 0.84)

continued

TABLE 43 Results for DTA studies included for objective 1 (*continued*)

Study details	Index test	Measure and Subgroup	Threshold	Reference stand	TP	FP	FN	TN	Sensitivity	Specificity	AUC (95% CI)
Ulberstad (2020)[77]	QbCheck	Overall	NR	DSM-V	57	15	12	58	0.83	0.79	NR
		QbCheck reaction time			NR						0.73
		QbCheck commission errors									0.74
		QbCheck omission errors									0.75
		QbCheck microevents									0.80
		QbCheck reaction time variability	NR	DSM-V							0.81
Zulueta (2019)[73]	Nesplora Kids (AULA)	Overall	NR	DSM-V	145	48	68	146	0.68	0.75	NR

a Data selected for synthesis where multiple control groups were available for a single study.

TABLE 44 Detailed QUADAS-2 assessment showing judgements and rational for risk of bias and concerns regarding applicability for DTA studies included for objective 1

Study details	Risk of bias																	Concerns regarding applicability				
	Consecutive/random	Case-control avoided	No inappropriate excl	Patient selection bias	Index test blinded	Pre-specified threshold	Index test bias	Reference stand appropriate	Blinded reference stand	Reference stand bias	Time interval	All received reference stand	Same reference stand	All included in analysis	Patient flow bias	Overall bias	Rationale Bias	Patient applicability	Reference stand applicability	Index applicability	Overall applicability	Rationale
Adamou (2022)[81]	?	✓	✓	☺	✓	✓	☺	✓	?	?	✓	✓	✓	✗	☺	?	Unclear whether reference standard interpreted blind to QbTest results. Two patients excluded from analysis but considered unlikely to have impacted results	☺	☺	☺	☺	No concerns
Bijlenga (2019)[78] *QbTest alone*	✗	✗	✓	☹	?	✓	?	✓	✓	☺	?	✓	✗	✗	☹	☹	Two-gate design – matched on age and gender. Control group received different reference standards. High proportion of dropouts (25/234)	☹	☺	☺	☹	Two-gate design
QbTest + clinical					?	?	?										No information on threshold					
Brunkhorst-Kanaan (2020)[68]	?	✓	✓	☺	?	✓	☺	✓	?	☺	?	✓	✓	✓	☺	☺	No concerns – no explicit information on blinding, but QbTest was conducted in separate appointment, so it appears unlikely that this would have influenced reference standard	☺	☺	?	?	Limited details on test conduct and interpretation
Edebol (2013)[79]	✗	✗	✓	☹	✓	✓	☺	✓	?	☺	?	✓	✗	✗	☹	☹	Two-gate design. 4/55 ADHD group excluded from analysis	☹	☺	☺	☹	Two-gate design
Edebol (2011)[84]	?	✓	✓	☺	✓	✓	☺	✓	✓	☺	✓	✓	✓	✓	☺	☺	No concerns	☺	☺	☺	☺	No concerns
Edebol (2012)[76]	✗	✗	✓	☹	?	✓	☺	?	?	?	?	?	✗	✓	☺	☹	Four-gate design. Limited details on reference standard	☹	?	☺	☹	Four-gate design Limited details on reference standard
Emser (2018)[83]	✗	✗	✓	☹	?	?	?	✓	?	☺	?	✓	✗	✓	☺	☹	Case-control design. No information on threshold for QbTest + clinical assessment. No information on blinding of reference standard. Control group received different reference standards	☹	☺	?	☹	Case-control design. Limited details on test conduct and interpretation

continued

TABLE 44 Detailed QUADAS-2 assessment showing judgements and rational for risk of bias and concerns regarding applicability for DTA studies included for objective 1 (*continued*)

Study details	Risk of bias																Concerns regarding applicability				
	Consecutive/random	Case-control avoided	No inappropriate excl	Patient selection bias	Index test blinded	Pre-specified threshold	Index test bias	Reference stand appropriate	Blinded reference stand	Reference stand bias	Time interval	All received reference stand	Same reference stand	All included in analysis	Patient flow bias	Overall bias	Rationale Bias	Patient applicability	Reference stand applicability	Index applicability	Overall applicability	Rationale
Groom (2016)[80]	✗	✗	✓	☹	?	?	?	✓	?	☺	✓	✓	✓	✗	☹	☹	Case-control design. No information on blinding of QbTest to case/control status. No detail on threshold. High proportion of dropouts (5/37 in ADHD group)	☹	☺	?	☹	Case-control design. Limited details on test conduct and interpretation
Hamadache (2021)[29]	✗	✗	✓	☹	?	✓	☹	✓	✓	☺	?	✓	✗	?	☺	☹	Three-gate design. Limited details on QbMini. ROC analysis only so no thresholds	☹	☺	?	☹	Three-gate design. Limited details on test conduct and interpretation
Hollis (2018)[18]	✓	✓	✗	☹	✓	?	?	?	✓	☹	✓	?	✓	✓	☺	☹	Participants eligible for DTA substudy if diagnostic decision had been made at 6 months (QbOpen eligible sample *n* = 94/123; QbBlind *n* = 76/127). Reference standard diagnosis made using limited data for around 50% participants	☺	☺	☺	☺	No concerns
Hult (2015)[67]	✓	✓	✓	☺	?	✓	☺	✓	✓	☺	✓	✓	✓	✓	☺	☺	No concerns	☺	☺	☺	☺	No concerns
Johansson (2018)[70]	✓	✓	✓	☺	?	?	?	?	?	?	?	✓	✓	✗	☹	☹	Reference standard K-SADS-PL – not ADHD-specific and so may not correctly diagnose ADHD. High proportion of participants excluded from 2 × 2 table	☹	?	?	☹	Participants enrolled if at least one of twin pairs had pre-specified neurodevelopmental disorders. Unlikely to be reflective of population with symptoms of ADHD
Pettersson (2015)[82]	✓	✓	✓	☺	?	✓	☺	✓	?	?	?	✓	✓	✓	☺	?	Unclear if reference standard was blind to QbTest result	☺	☺	☺	☺	No concerns
Rufo-Campos (2012)[63]	✗	✗	?	☹	?	?	?	?	?	?	?	?	?	?	?	☹	Two-gate design; no details about conduct/interpretation of index test, reference standard or flow and timing	☹	?	?	☹	Two-gate design. Limited details on index test conduct and interpretation; no details about reference standard

TABLE 44 Detailed QUADAS-2 assessment showing judgements and rational for risk of bias and concerns regarding applicability for DTA studies included for objective 1 (*continued*)

Study details	Risk of bias																Concerns regarding applicability				
	Consecutive/random	Case-control avoided	No inappropriate excl	Patient selection bias	Index test blinded	Pre-specified threshold	Index test bias	Reference stand appropriate	Blinded reference stand	Reference stand bias	Time interval	All received reference stand	Same reference stand	All included in analysis	Patient flow bias	Overall bias	Rationale Bias	Patient applicability	Reference stand applicability	Index applicability	Overall applicability	Rationale
Seesjärvi (2022)[75]	✗	✗	✗	☹	?	?	?	✓	✓	☺	?	✗	✗	✗	☹	☹	Two-gate design; patients with other listed comorbidities excluded from cases and controls; controls matched to cases. No information on whether Epeli test interpreters were blinded to diagnosis; high proportion excluded from 2 × 2 table	☹	☺	?	☹	Limited details on test conduct and interpretation
Sharma (2009)[62]	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	Limited information on patient selection (selected 'semi-randomly' from database). Appropriateness of reference standard unclear; not clear if reference standard interpreters blinded to index test; not clear if all received same ref standard	?	?	?	?	Very limited details on patient population, reference standard and patient flow
Söderström (2014)[74]	?	✓	✓	☺	✓	✓	☺	✓	✗	☹	?	✓	✓	✓	☺	☹	Clinicians aware of QbTest results when interpreting reference standard	☺	☺	☺	☺	No concerns
Stevanovic (2023)[41]	✗	✓	✓	☺	?	✓	☺	?	?	☹	?	✓	✗	✗	☹	☹	Unlikely that reference standard interpreted blind to index test; insufficient details on reference standard but was based on clinic records not DSM criteria. High proportion of dropouts	☹	☹	☹	☹	Children referred for evaluation of various neuropsychological conditions (not just ADHD). Test conduct did not follow manufacturer's instructions (used only second part of QbTest to calculate scores)
Tallberg (2019)[66] - accuracy	✗	✓	✓	☺	?	✓	?	✓	?	?	?	✓	✓	✗	☹	☹	High proportion of missing data. Unclear if reference standard was blinded to QbTest; was not blinded to other tests evaluated (including CPT)	?	☺	?	?	Children had screened positive for ADHD and so were referred for further evaluation – unclear what screening was involved and if it was representative of our study population

continued

TABLE 44 Detailed QUADAS-2 assessment showing judgements and rational for risk of bias and concerns regarding applicability for DTA studies included for objective 1 (*continued*)

Study details	Risk of bias																Concerns regarding applicability				
	Consecutive/random	Case-control avoided	No inappropriate excl	Patient selection bias	Index test blinded	Pre-specified threshold	Index test bias	Reference stand appropriate	Blinded reference stand	Reference stand bias	Time interval	All received reference stand	Same reference stand	All included in analysis	Patient flow bias	Overall bias	Rationale Bias	Patient applicability	Reference stand applicability	Index applicability	Overall applicability	Rationale
Ulberstad (2020)[77]	✗	✗	?	☹	?	?	?	✓	✓	☺	?	✓	✗	✗	☹	☹	Two-gate design. Participants performed test at home; unclear who interpreted the test. 7/149 patients were not included in 2 × 2 table	☹	☺	?	☹	Two-gate design. Limited details on test conduct and interpretation
Zulueta (2019)[73]	✗	✗	✓	☹	?	?	?	✓	✓	☺	?	✓	✓	✓	☺	☹	Two-gate design. No information on test interpretation or threshold	☹	☺	?	☹	Two-gate design. Limited details on test conduct and interpretation

TABLE 45 QUADAS-C assessment showing judgements and rationale for risk of bias for comparative DTA studies included for objective 1

Study	Test A	Test B	Design	Patient	Index test	Reference standard	Flow and timing	Overall	Rationale
Bijlenga (2019)[78]	QbTest alone	QbTest + clinical	Fully paired	☹	?	☺	☹	☹	Two-gate design. No information on threshold. No information on blinding between tests. High proportion of missing data for both tests
Hollis (2018)[18]	QbTest + clinical	Clinical alone	Randomised	☺	☺	☹	☹	☹	Reference standard diagnosis made using limited data for around 50% participants as either parent or teacher assessment was missing. High proportion of missing data for both tests as those without diagnosis at 6 months were excluded
Pettersson (2018)[82]	QbTest	CPTII	Fully paired	☺	?	?	☺	?	No information on blinding between tests or if reference standard was blinded to test results
Seesjärvi (2022)[75]	EF Sim	CPT?	Fully paired	☹	?	☺	☹	☹	Two-gate design; patients with other listed comorbidities were excluded from cases and controls; cases matched to controls. No information on blinding to reference standard or between tests. High proportion was excluded from 2 × 2 table
Tallberg (2019)[66] – accuracy	QbTest	CPTII	Fully paired	☺	?	?	☹	☹	No information on blinding to reference standard or between tests

TABLE 46 Baseline details for RCTs included for objective 1

Study details	Participants	Group 1	Control
Author (year) Chitsabesan (2022)[72] **Study name** FACT **Country** England **Language** English **Setting** YOI **Study design** Single-centre feasibility RCT **Funding** Non-industry	**Population:** ADHD diagnosis in boys aged 15–18 years **Inclusion criteria:** boys aged 15–18 years from a YOI who had any ADHD symptom from the Comprehensive Health Assessment Tool **Exclusion criteria:** being on remand; not speaking English; previous/current ADHD diagnosis; risk to researcher/staff; unable to give informed consent (16 years+) or no guardian consent (<16 years) **Number participants included (analysed):** 60 (47 at 3 months, 19 at 6 months) **Age** QbTest – age 16: 20%; 17: 26.7%; 18: 50%; missing: 3.3% Control – age 16: 10%; 17: 36.7%; 18: 53.3%; missing: 0% **Sex (% male)** 100	**QbTest and usual care (*n* = 30 randomised; 20 completed test):** QbTest completed prior to first assessment by neurodevelopmental lead. Information from QbTest, plus clinical information, used to inform diagnostic decision	**Usual care (*n* = 30):** Assessed by neurodevelopmental lead. If potential ADHD symptoms are present, then these are also assessed by assistant mental health practitioner (questionnaires, developmental history and observation). Third assessment by neurodevelopmental lead for diagnostic decision
Author (year) Hollis (2018)[18] **Study name** AQUA trial **Country** England **Language** English **Setting** Secondary care/community: 10 CAMHS or community paediatric clinics **Study design** RCT with embedded qualitative evaluation and accuracy data **Funding** Non-industry	**Population** ADHD diagnosis in children aged 6–17 years **Inclusion criteria** Children aged 6–17 years referred for their first ADHD assessment **Exclusion criteria** Previous or current ADHD diagnosis; non-fluent in English; suspected moderate/severe intellectual disability **Number participants included (analysed)** 267 (250) **Age** QbOpen(*n* = 123): mean 9.5; range 6.0–17.4; SD 2.8 QbBlind (*n* = 127): mean 9.4; range 5.9–16.2; SD 2.8 **Sex (% male)** **QbOpen:** 77%; **QbBlind:** 80%	**QbOpen (*n* = 123):** Usual care, in addition to QbTest (7–12 years) or QbTestPlus (12 + years), with Qb results shared with clinician to inform diagnostic decision, alongside clinical assessment Usual care varied between sites but typically included interview with child and their family and one standardised informant-based behavioural assessment measure	**QbBlind (*n* = 127):** Same as group 1, but QbTest/QbTestPlus results were withheld from clinician

TABLE 47 Results from RCTs included for objective 1

Study	Outcome	Details	Group 1: QbTest + usual care		Group 2 vs. usual care		Effect measure – estimate (95% CI), *p*-value
			n	No. events	*n*	Number of events	
Chitsabesan (2022)[72]	Time to assessment (*n = 20 who completed the QbTest*)	Median number of days between randomisation and QbTest	20	Median (interquartile range) = 42 (26–93). Min = 1; max = 195	NR	NR	NR
	Impact on clinical decision-making	Diagnostic decision made (all decisions were exclusion of ADHD diagnosis)	30	8	30	6	NR
	Morbidity	SDQ baseline: close to average	30	7	30	5	NR
		SDQ baseline: slightly raised	30	4	30	8	NR
		SDQ baseline: high	30	2	30	5	NR
		SDQ baseline: very high	30	16	30	12	NR
		SDQ baseline: missing	30	1	30	0	NR
		SDQ 3 months: close to average	23	2	24	4	NR
		SDQ 3 months: slightly raised	23	0	24	5	NR
		SDQ 3 months: high	23	4	24	1	NR
		SDQ 3 months: very high	23	7	24	7	NR
		SDQ 3 months: missing	23	17	24	13	NR
		SDQ 6 months: close to average	9	2	10	3	NR
		SDQ 6 months: slightly raised	9	0	10	4	NR
		SDQ 6 months: high	9	0	10	1	NR
		SDQ 6 months: very high	9	7	10	1	NR
		SDQ 6 months: missing	9	21	10	21	NR
Hollis (2018)[18]	Impact on clinical decision-making	Diagnostic decision (confirming or excluding ADHD diagnosis) made	123	94	127	76	OR 2.43 (1.34 to 4.39)	*p* = 0.003

continued

Study	Outcome		Details		Group 1: QbTest + usual care		Group 2 vs. usual care		Effect measure – estimate (95% CI), *p*-value
					n	No. events	*n*	Number of events	
	Diagnostic status	ADHD confirmed	123	69	127	65	RRR 2.14 (1.00 to 4.59)		*p* = 0.049
		ADHD excluded	123	25	127	11			
		No decision made (dropped out or discharged from clinic)	123	29	127	51			
	Diagnostic confidence	Possible/uncertain	122	16	121	29	OR 1.77 (1.09 to 2.89)		*p* = 0.022
		Probable	122	32	121	34			
		Definitely	122	74	121	58			
	Stability	Stability in diagnosis (any change in diagnosis from first confirmed diagnosis throughout study)	123	Kappa (95% CI) = 1 (1 to 1)	127	Kappa (95% CI) = 0.90 (0.7 to 1)	(χ2(1)= 0.01)		*p* = 0.32
	Time to diagnostic decision	Number of minutes spent at clinic appointments until diagnosis	123	Mean (SD) = 141.97 (53.84) Observed median survival time (95% CI) = 150 (140 to 155)	127	Mean (SD) = 152.83 (75.88) Observed median survival time (95% CI) = 165 (150 to 180)	TR: 0.85 (0.77–0.93)		*p* = 0.001
		Number of days to diagnostic decision	123	Mean (SD): 82.54 (49.53) Observed median survival time (95% CI) = 96 (85 to 99)	127	Mean (SD): 83.94 (58.14) Observed median survival time (95% CI)= 108 (91–140)	TR: 0.90 (0.73–1.10)		*p* = 0.285
		Number of clinic appointments until diagnosis	123	Mean (SD): All participants: 2.69 (0.85) Those with diagnostic decision: 2.82 Those who dropped out or were discharged without a diagnosis: 2.28	127	Mean (SD): All participants: 2.72 (0.91) Those with diagnostic decision: 2.76 Those who dropped out or were discharged without a diagnosis: 2.67	NR		

TABLE 47 Results from RCTs included for objective 1 (*continued*)

Study	Outcome	Details		Group 1: QbTest + usual care		Group 2 vs. usual care		Effect measure – estimate (95% CI), *p*-value
				n	No. events	*n*	Number of events	
		Number of consultations to diagnostic decision (confirming or excluding ADHD diagnosis) by group over 6 months, *n* = 250	123	–	127	–	HR 1.44 (1.04 to 2.01)	*p* = 0.029
		Number of consultations to diagnostic decision (confirming or excluding ADHD diagnosis) by group over 6 months, in *n* = 198 aged 6–12 years (**using QbTest in intervention group)**	NR	–	NR	–	HR 1.84 (1.23 to 2.68)	*p* = 0.001
		Number of consultations to diagnostic decision (confirming or excluding ADHD diagnosis) by group over 6 months, in *n* = 52 aged > 12 years (**using QbTest (12–60) in intervention group)**	NR	–	NR	–	HR 0.82 (0.37 to 1.80)	*p* = 0.618
	Costs	Cost of clinic appointments	123	Mean (SD): £87.62 (£40.45)	127	Mean (SD): £90.06 (£41.19)	–	

TABLE 48 Risk-of-bias assessment for RCTs included for objective 1

Study details	Outcome	Domain[a] 1	2	3	4	5	Overall	Rationale
Hollis (2018)[18]	Diagnostic decision (confirming or excluding ADHD diagnosis) made OR 2.43 (1.34 to 4.39)	☺	☺	☺	☺	☺	☺	Appropriate randomisation and allocation concealment; participants blinded to allocation, clinicians not blinded, but it seems unlikely deviations took place due to trial context; appropriate measurement of the outcomes; pre-registered protocol, however, potential for selective reporting due to HRQoL pre-specified, but data not reported Outcome not impacted by censoring/withdrawals
	Diagnostic status	☺	☺	☺	☺	☺	☺	Outcome not impacted by censoring/withdrawals
	Diagnostic confidence	☺	☺	☺	☺	☺	☺	Outcome not impacted by censoring/withdrawals
	Stability of diagnosis	☺	☺	☺	☺	☺	☺	Outcome not impacted by censoring/withdrawals
	Number of consultations to diagnostic decision	☺	☺	☹	☺	☺	☹	Large proportions of participants (80/250) were censored from the analysis as they dropped out or were discharged from the clinic and so did not have a diagnosis at 6 months. This was a particular problem for time-to-event outcome data where the analysis assumed that participants were uninformatively censored and so had equivalent outcomes to those for whom full FU data were available
	Number of minutes spent at clinic appointments until diagnosis	☺	☺	☹	☺	☺	☹	
	Number of clinic appointments until diagnosis	☺	☺	☹	☺	☺	☹	
	Number of days to diagnostic decision	☺	☺	☹	☺	☺	☹	
	Cost of clinic appointments	☺	☺	?	☺	☺	?	Unclear how costs are calculated and so not clear how censored individuals contributed to this outcome

a 1: Randomisation process; 2: Deviation from intended intervention; 3: Missing outcome data; 4: Measurement of the outcome; 5: Selective outcome reporting.

Note
RoB 2 assessment for the AQUA trial – the only RCT included for objective 1 that was not a feasibility trial.

TABLE 49 Baseline details for implementation studies that contribute data on process measures for objective 1

Study details	Participants	Interventions and confounders
Author (year) Hall (2016)[85] **Study name** Not reported **Study location** Kent, UK **Language** English **Setting** Community paediatric ADHD clinic **Study design** Uncontrolled before–after implementation study **Funding** Non-industry	**Population** Children and adolescents diagnosed with ADHD in community paediatric clinic **Sample selection and inclusion criteria** Patient files selected using random number generator. Case notes included if case had received primary diagnosis of ADHD; for the post-test implementation evaluation, cases were only included if they had received a QbTest as part of diagnostic assessment. If a file was excluded, next available file was selected **Exclusion criteria** Not reported **Number of participants included (analysed)** 80 (80) **Age** Pre-QbTest group: mean 8.1; SD 2.4; range 4.5–14.6 QbTest group: mean 9.2; SD 2.3; range 6.2–13.10 **PROGRESS-Plus criteria reported by study** • **Sex (% male):** pre-QbTest group: 80%; QbTest group: 70% • **Neurodevelopmental:** number of participants with secondary diagnosis – pre-QbTest group: ASD 6; ASD and tic disorder 2; ASD and dyspraxia 2; ASD and OCD 1; oppositional defiance disorder 1; sensorineural deafness 1; mild epilepsy 1; Tourette syndrome 1. QbTest group: ASD 7, Tourette syndrome 1, sensory processing disorder 1, mild speech and language disorder 1, emotional difficulties1, dyslexia 1, learning difficulties 1 **Note: study results were not stratified by PROGRESS-Plus criteria.**	**Group 1 (pre-test implementation):** standard ADHD assessment (*n* = 40) **Group 2 (post-test implementation):** QbTest (6–12) or QbTest (12–60) + standard ADHD assessment (*n* = 40) **Confounders**: authors state that, 'During this time period, there was no change to the assessment process, except the QbTest. Methods of acquiring parent and teacher information, and the quantity and quality of information, remained unchanged, as did members of the clinical and administration team'
Author (year) Vogt (2011)[86] **Study location** Berkshire, UK **Language** English **Setting** CAMHS **Study design** Uncontrolled before–after implementation study **Funding** Not reported	**Population** Children and adolescents referred for ADHD assessment in CAMHS **Sample selection and inclusion criteria** Notes of 108 patients referred for ADHD to CAMHS clinic over 2-year period – 1 year before (2006–7) and 1 year after implementation of QbTest (2007–8). Unclear whether all children were assessed during eligible time periods enrolled or selected subsample **Exclusion criteria** Not reported **Number participants included (analysed)** 108 (108) **Age** Pre-QbTest group: mean 9; mode 10; median 9 QbTest group: mean 10.5; mode 8; median 10 **PROGRESS-Plus criteria reported by study** None reported	**Group 1 (pre-test implementation):** standard ADHD assessment (*n* = 46) **Group 2 (post-test implementation):** QbTest (6–12) or QbTest (12–60) + standard ADHD assessment (*n* = 62) **Confounders**: same child and adolescent psychiatrists conducted the assessments for both groups using the same protocol

continued

TABLE 49 Baseline details for implementation studies that contribute data on process measures for objective 1 (*continued*)

Study details	Participants	Interventions and confounders
Author (year) Sharma (2022)[64] **Study location** Swindon, UK **Language** English **Setting** Hospital paediatric clinic **Study design** Uncontrolled before–after implementation study **Funding** Not reported	**Population** Children and adolescents referred for ADHD assessment in hospital paediatric clinic **Sample selection and inclusion criteria** Patients assessed for ADHD between July 2020 and January 2022 in hospital paediatric clinic, who had been referred for ADHD/non-specific behavioural problems/ASD. Unclear how patients were selected **Exclusion criteria** Any patient who had not completed an ADHD assessment in the time frame or whose assessment resulted in inconclusive determination **Number participants included (analysed)** 40 (40) **Age** All participants: mean 11.7 (SD 2.4) **PROGRESS-Plus criteria reported by study** None reported	**Group 1 (pre-test implementation):** standard ADHD assessment (*n* = 20) **Group 2 (post-test implementation):** QbTest (6–12) or QbTest (12–60) + standard ADHD assessment (*n* = 20) **Subgroups:** ADHD cases – those referred for ADHD Complex cases – those originally referred for non-specific behavioural difficulties or ASD **Confounding factors**: none reported
Author (year) Humphreys (2018)[69] **Study location** East Midlands, UK **Language** English **Setting** Community paediatric mental health settings in three NHS trusts **Study design** Audit **Funding** Industry and non-industry: QbTech and East Midlands AHSN	**Population** Children and adolescents referred for ADHD assessment in community paediatric mental health settings **Sample selection inclusion criteria** Selection of children (method of selection not reported) referred for ADHD assessment in community paediatric mental health settings, before and after implementation of QbTest **Exclusion criteria** Not reported **Number participants included (analysed)** Unclear – 20–30 cases before QbTest implementation and 20–30 cases after test implementation, from each of the three trusts **Age** 5–16 years **PROGRESS-Plus criteria reported by study** None reported	**Group 1 (pre-test implementation):** standard ADHD assessment (60–90) **Group 2 (post-test implementation):** QbTest (6–12) or QbTest (12–60) + standard ADHD assessment (*n* = 60–90) **Confounding factors**: none reported; authors note that the postimplementation group is after introduction of the QbTest and pathway redesign in two sites

TABLE 49 Baseline details for implementation studies that contribute data on process measures for objective 1 (*continued*)

Study details	Participants	Interventions and confounders
Author (year) McKenzie (2022)[31] **Study name** Focus ADHD **Study location** England (sites throughout the country) **Language** English **Setting** CAMHS and paediatric sites (total of 20 sites) **Study design** Audit **Funding** Industry and non-industry: QbTech and AHSNs in England	**Population** Children and adolescents referred for ADHD in CAMHS and paediatric sites **Inclusion criteria** Selection of children referred for ADHD assessment in in CAMHS and paediatric sites across England, before and after implementation of QbTest; 61 potential sites were identified; usable data were obtained from 21 sites. Unclear how each site selected cases to report on. One site used test only for complex cases **Exclusion criteria** Not reported **Number participants included (analysed)** 1098 cases – these consist of 549 (10–30 cases per site) before QbTest implementation and 549 (10–30 cases) after test implementation, from each of the 21 included sites **Age** 6–18 years **PROGRESS-Plus criteria reported by study** None reported	**Group 1 (pre-test implementation):** Standard ADHD assessment (*n* = 549) **Group 2 (post-test implementation):** QbTest (6–12) or QbTest (12–60) + standard ADHD assessment (*n* = 549) **Subgroups**: CAMHS vs. paediatric sites Also report stratified data based on number of cases referred per site, and large vs. small test volume – stratified data not extracted for these **Confounding factors**: QbTest implementation occurred from April 2019 to March 2022 and so overlaps with COVID-19 pandemic (from March 2020)

TABLE 50 Results data on process measures from implementation studies that contributed to objective 1

Study	Outcome	Details	Group 1: standard ADHD assessment (pre implementation)		Group 2: QbTest (6–12) or QbTest (12–60) + standard ADHD assessment (post implementation)		Effect measure – estimate (95% CI)	*p*-Value	Other reported details
			n	Number of events	*n*	Number of events			
Hall (2016)[85]	Number of consultations to ADHD diagnosis	Number of consultations until ADHD diagnosis (mean, min, max)	40	Mean 3.05 (min 1, max 7)	40	Mean 2.18 (min 1, max 4)	Poisson regression IRR (95% CI) 0.71 (0.54 to 0.94)	$p = 0.02$	
	Reasons for delay in diagnosis	Clinician-reported reasons for delay in diagnosis, in those where => 5 consultations were needed to make a diagnosis (all in pre-QbTest group)	For 4/6 (66.6%) of cases, inconclusive or discrepancy outcomes from clinical rating scales were cited as the primary reason for delay; 1 case (17.0%) cited complex comorbidities and 1 (17.0%) clinician reluctance to make a diagnosis.		–	–	–	–	
	Consultation cost	Total cost spent on ADHD assessment for all 40 cases combined	40	£13,176	40	£10,636	Saving = £2540	–	Cost of a consultation within the trust at the time of audit = £108.00. A single QbTest cost the trust £31.00 (cost of the test as a proportion of the lease fee and a 30-minute nurse-led appointment to conduct the test)
Vogt (2011)[86]	Diagnoses revised to ADHD + in those with a diagnosis rejected at initial assessment at 1-year FU	–	19	7	19	0	–	$p = 0.0035$	

TABLE 50 Results data on process measures from implementation studies that contributed to objective 1 (*continued*)

Study	Outcome	Details	Group 1: standard ADHD assessment (pre implementation)		Group 2: QbTest (6–12) or QbTest (12–60) + standard ADHD assessment (post implementation)		Effect measure – estimate (95% CI)	*p*-Value	Other reported details
			n	Number of events	*n*	Number of events			
	Outcomes of those with ADHD at - year FU	ADHD diagnosis changed	27	1	43	1	–	$p = 0.24$	
		Continuing on medication		13		28			
		Discontinued medication		9		9			
		Medication trial		22		38			
		Lost to FU		3		4			
Sharma (2022)[64]	Number of contacts to diagnosis	All participants	20	Mean 2.7 (SD 0.7)	20	Mean 2.4 (SD 0.8)	–	$p > 0.05$	-
	Number of months to diagnosis	All participants	20	Mean 6.5 (SD 3)	20	Mean 5.5 (SD 1.8)	–	$p > 0.05$	–
	ADHD-confirmed diagnosis rate	All participants	20	90.6%	20	87.5%	–	$p > 0.05$	–
	Number of months to diagnosis	ADHD cases (those referred for ADHD)	NR	NR	NR	Mean 5.6 (SD 1.7)	–	–	–
	Number of months to diagnosis	Complex cases (those originally referred for non-specific behavioural difficulties or ASD)	NR	NR	NR	Mean 5.5 (SD 2.7)	–	–	–

continued

TABLE 50 Results data on process measures from implementation studies that contributed to objective 1 (*continued*)

Study	Outcome	Details	Group 1: standard ADHD assessment (pre implementation)		Group 2: QbTest (6–12) or QbTest (12–60) + standard ADHD assessment (post implementation)		Effect measure – estimate (95% CI)	*p*-Value	Other reported details
			n	Number of events	*n*	Number of events			
Humphreys (2018)[69]	Number of appointments to diagnostic decision	–	60–90	Range of 3–8 appointments	60–90	Reduction compared to control of between (on average) 0.24 and 1.04 appts per child. In two trusts, a diagnosis was often reached at the first contact with paediatrician	–	–	–
	Number of days to diagnostic decision	–	60–90	Average ranged from 161 to 453 (approx. 5–15 months)	60–90	Average ranged from 15 to 252 (approximately, 2 weeks–8.5 months)	–	–	The authors note for this outcome that the post-implementation group is after introduction of the QbTest AND pathway redesign in two sites
	Number of days from assessment to commencing medications	–	60–90	Range 42–179 days	60–90	Range 15–96 days			
	Release of clinical time required to reach a diagnostic decision	–				Range 20–33% reduction			
McKenzie (2022)[31]	Number of clinical appointments	All sites	549	Mean 3.22 (range 1–50)	549	Mean 2.85 (range 1–32)	Per cent change: 11.5% decrease	NR	Data in this study were likely affected by COVID-19 for all group 2 data and comparison between groups 2 and 1

TABLE 50 Results data on process measures from implementation studies that contributed to objective 1 (*continued*)

Study	Outcome	Details	Group 1: standard ADHD assessment (pre implementation)		Group 2: QbTest (6–12) or QbTest (12–60) + standard ADHD assessment (post implementation)		Effect measure – estimate (95% CI)	*p*-Value	Other reported details
			n	Number of events	*n*	Number of events			
	Number of days from initial referral to diagnosis	All sites	549	Mean 452 (range 15–3276)	549	Mean 507 (range 43–1281)	Per cent change: 12.2% increase	$p < 0.01$	
	Number of days to reach diagnostic decision	All sites	549	Mean 117 (range 0–1570)	549	Mean 129 (range 0–1378)	Per cent change: 10.3% increase	NR	
	Number of school observations utilised		549	120	549	49	Per cent change: 17% decrease		
	Number of ADHD diagnoses		549	445	549	418	Per cent change: 5% decrease		
	Number of clinical appointments	CAMHS services	326	Mean 4.13 (range 1–50)	326	Mean 3.75 (range 1–32)	Per cent change: 9.2% decrease		
	Number of days from initial referral to diagnosis	CAMHS services	326	Mean 442 (range 18–1161)	326	Mean 566 (range 43–1821)	Per cent change: 28.1% increase		
	Number of days to reach diagnostic decision	CAMHS services	326	Mean 119 (range 0–888)	326	Mean 135 (range 0–1378)	Per cent change: 13.4% increase		
	Number of clinical appointments	Paediatric clinics	194	Mean 2.01 (range 1–15)	194	Mean 1.63 (range 1–4)	Per cent change: 18.9% decrease		
	Number of days from initial referral to diagnosis	Paediatric clinics	194	Mean 444 (range 15–3276)	194	Mean 367 (range 1494)	Per cent change: 17.3% decrease		
	Number of days to reach diagnostic decision	Paediatric clinics	194	Mean 130 (range 0–1570)	194	Mean 138 (range 0–1036)	Per cent change: 6.2% decrease		

TABLE 51 The ROBINS–I risk of bias of implementation studies that contribute process measure data for objective 1

Study details	Domain[a] 1	2	3	4	5	6	7	Overall	Rationale
Hall (2016)[85]	☹	☺	☺	☺	😐	😐	😐	☹	Confounders not controlled for and potential for confounding of the effect of the intervention; only people who had final diagnosis within time frame were selected; outcome measure could have been influenced by knowledge of intervention received; no protocol. Note: selection of participants was random, hence exclusion of participants was covered under the missing data domain
Sharma (2022)[64]	?	☹	☺	?	☺	😐	😐	☹	Conference abstract with no information about whether confounders were controlled for, or about bias due to deviations from intended interventions; participants were excluded if assessment was inconclusive, or did not receive diagnosis in time frame; outcome measure could have been influenced by knowledge of intervention received; no protocol
Humphreys (2018)[69]	☹	?	☺	😐	?	😐	😐	☹	Confounders not controlled for and potential for confounding of the effect of the intervention; no information about participant selection; potential for bias due to deviations from intended interventions due to two sites having a pathway redesign after introduction of QbTest; no information about missing data (authors confirmed not only ADHD + cases selected); outcome measure could have been influenced by knowledge of intervention received; no protocol
McKenzie (2022)[31]	☹	?	☺	☺	☺	😐	😐	☹	Confounders not controlled for and potential for confounding of the effect of the intervention (COVID-19 only confounder, authors say would have impacted on the analysis); little information on participant selection; outcome measure could have been influenced by knowledge of intervention received; no protocol
Vogt (2011)[86]	☹	?	☺	☺	☺	😐	😐	☹	Confounders not controlled for and potential for confounding of the effect of the intervention; no information about participant selection; outcome measure could have been influenced by knowledge of intervention received; no protocol

a 1: Confounding [potential confounders for all studies: age at the point of seeking referral for ADHD; sex; comorbidities – e.g. autism, anxiety; nature and severity of symptoms at presentation – e.g. predominantly inattentive or hyperactive; Socioeconomic status; ethnicity and for McKenzie (2022) also COVID-19 pandemic]; 2: Selection of participants; 3: Classification of interventions; 4: Deviations from intended interventions; 5: Missing data; 6: Measurement of the outcome; 7: Selection of the reported result.

Note:
Sad face = serious risk of bias; smiling face = low risk of bias, question mark = no information.

TABLE 52 Results data on process measures for DTA studies that contribute data for objective 1

Study details	Number of patients with unavailable test result (%)	Details of missing results	Action taken post-test failure
Ulberstad (2020)[77] Test: QbCheck	7/149 (5%)	Seven participants failed to complete the test. Two had technical problems with camera; four ended the test in the middle of the session for unknown reasons; one intentionally did not follow through the test. Six of the non-completers belonged to the ADHD group; one belonged to the healthy controls group	Not reported – the participants were excluded from analyses
Bijlenga (2019)[78] QbTest (12–60)	25/234 (11%)	Two female ADHD patients (aged 63 and 73 years) did not perform QbTest because they did not understand the task. Twenty-three participants (7 ADHD; 16 healthy controls) were invalid due to being extreme outliers, not following instructions, technical errors, aborted test (data not stratified by reasons given)	
Groom (2016)[80] QbTest (12–60)	4/84 (5%)	Non-completion of the test by three people in the ADHD group. Failure to complete QbTest by one person in ASD group (no further information)	
Seesjärvi (2022)[75] Group 1: EPELI	22/115 (19%)	Five children with ADHD and 17 controls had technical failures or human error (scenarios accidentally presented in different order). (Data not stratified by reason given)	

TABLE 53 Baseline data for studies that reported on patient/clinician carer views of sensor CPTs for ADHD diagnosis

Study details	Study component	Participants and methodology
Author (year) Chitsabesan (2022)[72,108] **Study name** FACT **Country** England **Language** English **Setting** YOI **Study design** Interview and survey components of FACT feasibility RCT **Funding** Non-industry **Sensor CPT** QbTest + standard assessment	Interviews with young people	**Participants:** six adolescent boys in a YOI who participated in the FACT trial in the QbTest group **Sampling strategy:** purposive sampling used to select people considering age, completion of QbTest and scores on the 'Qb Opinion Questionnaire'. Unclear how many people were invited to participate in the interviews **Data collection:** Semistructured interviews completed 3 months into the FACT trial about acceptability of QbTest. At the time of interview, not all people had received the result of the test/ADHD assessment **Analysis:** thematic analysis, using inductive approach
	Interviews with staff from the YOI and the research assistant	**Participants:** one research assistant and five staff members from the YOI who used QbTest in the FACT trial **Sampling strategy:** all staff and the one researcher who used the QbTest in the trial were invited to participate **Data collection:** semistructured interviews completed at the end of the FACT trial about the acceptability and feasibility of administering and implementing QbTest within usual practice, barriers and facilitators to use and reasons for non-completion **Analysis:** thematic analysis, using inductive approach
	Survey to young people	**Participants:** 10 adolescent boys in a YOI who participated in the FACT trial in the QbTest group **Sampling strategy:** All 20 young people who completed QbTest in FACT trial invited to complete survey; 10 responded **Data collection:** 'Qb Opinion Questionnaire' completed at 3 months. The survey contains 12 items, e.g. 'the QbTest results were difficult to understand' and the young person rates each item on a 5-point scale **Analysis:** descriptive analysis

continued

TABLE 53 Baseline data for studies that reported on patient/clinician carer views of sensor CPTs for ADHD diagnosis (*continued*)

Study details	Study component	Participants and methodology
Author (year) Hall (2017)[107] **Study name** AQUA trial **Country** England **Language** English **Setting** Secondary care/community: 10 CAMHS or community paediatric clinics **Study design** Qualitative substudy of AQUA trial **Funding** Non-industry **Sensor CPT** Usual care + QbTest (6–12 and 12–60), with test results available to clinician ('QbOpen')	Interviews with clinicians	**Participants:** 10 clinician leads (20% male) from each of the 10 sites involved in the AQUA trial **Sampling strategy:** the clinical lead for the AQUA trial at each of the 10 sites was invited to interview (all accepted) **Data collection:** semistructured interviews conducted by a trained researcher regarding opinions of QbTest **Analysis:** thematic analysis, using an inductive, reflexive approach
	Interviews with families	**Participants:** 20 families from the AQUA trial (the main care giver was the primary interviewee, but where possible, the young person was encouraged to participate with their parents – all young people had been in the 'QbOpen' group). Sample characteristics: • Child mean age 10.7 years (SD 2.9; range 9–18) • 75% male • Confirmed primary diagnosis – ADHD 55%; not ADHD 25%, unconfirmed 25%. Comorbidities – ASD 5%; conduct disorder and oppositional defiance disorder 0%; Tourette's/tic 5%; attachment disorder 0%; learning difficulties 0%; anxiety and depression 0% **Sampling strategy:** two families per site who had participated in the AQUA trial 'QbOpen' group were invited to interview. Thirty-eight families were invited to interview and 18 declined to participate. Refusing families were replaced with the next family until 2 families from each site were enrolled **Data collection:** semistructured interviews conducted by a trained researcher regarding opinions of QbTest **Analysis:** thematic analysis using an inductive, reflexive approach
	Survey to clinicians and families	**Participants:** 10 clinician leads (20% male) from each site in the AQUA trial, and 76 families from the AQUA trial. The following details were reported for the families only: • Child mean age 10.2 years (SD 2.9; range 7–18) • 79% male • Confirmed primary diagnosis – ADHD 46%; not ADHD 14%, unconfirmed 39%. Comorbidities – ASD 5%; conduct disorder and oppositional defiance disorder 4%; Tourette's/tic 1%; attachment disorder 1%; learning difficulties 3%; anxiety and depression 1% **Sampling strategy:** all participants and the 10 lead clinicians from the trial were invited to participate; 10 clinicians and 76 families responded **Data collection:** quantitative online survey. Clinician questions centred on how best to administer QbTest, understanding results and communicating with families. Family questions focused on utility of QbTest in understanding symptoms and decisions and experience of completing test **Analysis:** descriptive analysis

TABLE 53 Baseline data for studies that reported on patient/clinician carer views of sensor CPTs for ADHD diagnosis (*continued*)

Study details	Study component	Participants and methodology
Author (year) McKenzie (2022)[31] **Study name** Focus ADHD **Study location** England (sites throughout the country) **Language** English **Setting** CAMHS and paediatric sites **Study design** Qualitative interview and survey components of an uncontrolled before–after implementation study **Funding** Industry and non-industry: QbTech and AHSNs in England **Sensor CPT** QbTest (6–12) or QbTest (12–60) + standard ADHD assessment	Interviews with healthcare staff	**Participants:** 21 healthcare staff involved in implementation of QbTest at their site, or conducting the test/interpreting test results, in the Focus ADHD study **Sampling strategy:** all sites were invited to participate – they aimed to include participants with different roles in the test implementation process, including those who delivered the test, interpreted the test and managers who were responsible for implementing the test at their site **Data collection:** semistructured interviews conducted to explore experience of using test, adoption of test at their site and sustainability of its use **Analysis:** thematic analysis, analysed using the non-adoption, abandonment, scale-up, spread, sustainability framework
	Survey for HCPs	**Participants:** 65 HCPs who attended audit training in the Focus ADHD study **Sampling strategy:** all HCPs who attended audit training in the Focus ADHD study invited (*n* = unknown), 65 responded **Data collection:** online survey about how best to administer the QbTest, understanding the results and communicating with families **Analysis:** descriptive analysis
	Survey for patients and their families	**Participants:** 22 patients who had been assessed with QbTest (and their parents) in the Focus ADHD study **Sampling strategy:** survey distributed to all patient families via text/e-mail and clinicians/key stakeholders asked to pass it on (*n* = unknown); 22 patients/families responded **Data collection:** online survey about the utility of the QbTest in understanding symptoms and diagnostic decisions and the experience of completing the test **Analysis:** descriptive analysis
Author (year) Pellegrini (2020)[87] **Study name** Not reported **Study location** Ireland **Language** English **Setting** Irish CAMHS – three CAMHS teams **Study design** Mixed-methods study of real-world impact of test implementation **Funding** Not funded **Sensor CPT** QbTest + standard ADHD assessment	Focus groups with clinicians	**Participants**: 19 clinicians who were working in one of the three CAMHS teams selected for this research in Ireland, and who were involved in using the QbTest as part of an ADHD assessment process. Professional disciplines included: administration, occupational therapy, nurses, psychology, psychiatry, social work and speech and language therapy **Sampling strategy:** all clinicians in the study who were using QbTest were invited (*n* = unknown) **Data collection:** three semistructured focus groups were conducted (*n* = 6; *n* = 6; *n* = 7), gathering information on their experiences with the QbTest **Analysis:** thematic analysis, using a six-step, reflexive process
	Survey to clinicians, service users and their families	**Participants:** 50 participants: 17 clinicians who had used QbTest in one of the three CAMHS involved in the study, 15 young people who had completed QbTest as part of ADHD assessment in one of the three CAMHS teams involved in the study and their parent/guardians (*n* = 18) **Sampling strategy:** young people and their parents/guardians were recruited during ADHD assessment – the clinician made the family aware of the survey study. Clinicians were sent the survey via e-mail by research staff. Number of people invited to participate not reported **Data collection:** quantitative survey on experience of using QbTest. The survey was based on a template provided by QbTech that had been used in the AQUA qualitative substudy[107] **Analysis:** descriptive analysis

continued

TABLE 53 Baseline data for studies that reported on patient/clinician carer views of sensor CPTs for ADHD diagnosis (*continued*)

Study details	Study component	Participants and methodology
Author (year) Humphreys (2018)[69] **Study name** Not reported **Study location** East Midlands, UK **Language** English **Setting** Community paediatric mental health settings in three NHS Trusts **Study design** Survey component of an uncontrolled before–after implementation study **Funding** Industry and non-industry: QbTech and East Midlands AHSN **Sensor CPT** QbTest (6–12) or QbTest (12–60) + standard ADHD assessment	Survey to patients and families	**Participants:** 48 patients (children who had ADHD assessment using QbTest in CAMHS in the before–after study) and their families **Sampling strategy:** surveys were distributed by clinic staff as paper version – 90 questionnaires distributed, 43% response rate (48 respondents) **Data collection:** survey on their experience of using QbTest (the same survey used in the AQUA trial[107]) **Analysis:** descriptive analysis
	Survey to staff	**Participants:** staff who had used QbTest (*n* = unknown) **Sampling strategy:** sampling strategy not reported. Number distributed was not reported, 76% response rate **Data collection:** survey on their experience of using QbTest (the same survey used in the AQUA trial[107]) **Analysis:** descriptive analysis
Author (year) Peili Vision (n.r.)[88] **Study name** Health Service Pilot **Study location** Finland **Language** English **Setting** 18 schools in Finland **Study design** Pilot cohort study **Funding** Industry – test manufacturer (Peili Vision) **Sensor CPT** EF Sim test + psychologist evaluation	Survey to teachers	**Participants:** 21 teachers of participating schools that used EF Sim for students in the Health Service Pilot **Sampling strategy:** not reported. Number of teachers invited to participate is unknown **Data collection:** feedback questionnaire for evaluating the main aspects of how they felt the EF Sim check went, containing 10 questions. Scores given on a scale of 1–5 **Analysis:** descriptive analysis
	(Confidential information has been removed.)	(Confidential information has been removed.) (Confidential information has been removed.) (Confidential information has been removed.) (Confidential information has been removed.)

TABLE 53 Baseline data for studies that reported on patient/clinician carer views of sensor CPTs for ADHD diagnosis (*continued*)

Study details	Study component	Participants and methodology
Author (year) Ulberstad (2020)[77] **Study name** Not reported **Study location** Germany, Sweden, USA **Language** English **Setting** Secondary care **Study design** Survey data from two-gate DTA study **Funding** Industry – authors employed by QbTech **Sensor CPT** QbCheck	Survey to patients	**Participants:** patients who used QbCheck in the DTA study and who completed the survey (n = 125; 59 ADHD and 69 healthy controls) **Sampling strategy:** all patients (142) from DTA study given survey, 125 completed it **Data collection:** survey about experience of using QbCheck – three questions assessed on a scale from 0 to 10 that assessed the usability of the test; one yes/no question about problems with using the test **Analysis:** t-tests (the dimensional variables) or χ^2 test (the categorical variable) to compare the group with ADHD to the controls
Author (year) Seesjärvi (2022)[75] **Study name** Not reported **Study location** Finland **Language** English **Setting** Secondary care **Study design** Survey data from two-gate DTA study **Funding** Non-industry (but, authors developed test) **Sensor CPT** EPELI	Survey to patients	**Participants:** children (some with ADHD; some healthy controls – n = not reported) who took part in the DTA study using EPELI test and completed the survey component **Sampling strategy:** not reported **Data collection:** survey about use of EPELI test – shortened version of the Presence Questionnaire 3.0 **Analysis:** descriptive analysis

TABLE 54 The CASP checklist quality assessment of studies included for objective 1 that reported qualitative data on patient/clinician carer views of sensor CPTs for ADHD diagnosis

CASP checklist questions	Quality assessment answers per study (answer options: yes, no, cannot tell)			
	Chitsabesan (2022)[72]	Hall (2017)[107]	McKenzie (2022)[31]	Pellegrini (2020)[87]
Was there a clear statement of the aims of the research?	Yes	Yes	Yes	Yes
Is a qualitative methodology appropriate?	Yes	Yes	Yes	Yes
Was the research design appropriate to address the aims of the research?	Yes	Yes	Yes	Yes
Was the recruitment strategy appropriate to the aims of the research?	Yes	Yes	Yes	Yes
Were the data collected in a way that addressed the research issue?	Yes	Yes	Yes	Yes
Has the relationship between researcher and participants been adequately considered?	Cannot tell	Yes	Cannot tell	Yes
Have ethical issues been taken into consideration?	Yes	Yes	Yes	Yes
Was the data analysis sufficiently rigorous?	Cannot tell	Yes	Cannot tell	Yes
Is there a clear statement of findings?	Yes	Yes	Yes	Yes

TABLE 55 Quality assessment of studies included for objective 1 that reported survey data on patient/clinician carer views of sensor CPTs for ADHD diagnosis

Questions (n = 20)	Quality assessment answers per study (answer options: yes, no, not stated clearly)							
	Chitsabesan (2022)[72]	Hollis (2018)[18]	McKenzie (2022)[31]	Pellegrini (2020)[87]	Humphreys (2018)[69]	Peili Vision (n.r.)[88]	Ulberstad (2020)[77]	Seesjärvi (2022)[109]
Was the problem or phenomenon under investigation defined, described and justified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was the population under investigation defined, described and justified?	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Were specific research questions and/or hypotheses stated?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were operational definitions of all study variables provided?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were participant inclusion criteria stated?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was the participant recruitment strategy described?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Was a justification/rationale for the sample size provided?	No	No	No	No	No	No	No	No
Was the attrition rate provided? (applies to cross-sectional and prospective studies)	Yes	Yes	Yes	Yes	No	No	Yes	No
Was a method of treating attrition provided? (applies to cross-sectional and prospective studies)	No	Yes	No	No	No	No	No	No
Were the data analysis techniques justified? (i.e. was the link between hypotheses/aims/research questions and data analyses explained?)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were the measures provided in the report (or in a supplement) in full?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was evidence provided for the validity of all the measures (or instrument) used?	No	No	No	No	No	No	No	No
Was information provided about the person(s) who collected the data (e.g. training, expertise and other demographic characteristics)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was information provided about the context (e.g. place) of data collection?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was information provided about the duration (or start and end date) of data collection?	No	Yes	No	No	No	No	No	No
Was the study sample described in terms of key demographic characteristics?	No	Yes	No	No	No	No	No	No
Was discussion of findings confined to the population from which the sample was drawn?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were participants asked to provide (informed) consent or assent?	Yes	Yes	Yes	Yes	Not stated clearly	Not stated clearly	Not stated clearly	Not stated clearly
Were participants debriefed at the end of data collection?	Not stated clearly	Not stated clearly	Not stated clearly	Yes	Not stated clearly	Not stated clearly	Not stated clearly	Not stated clearly
Were funding sources or conflicts of interest disclosed?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

TABLE 56 Baseline details for RCT included for objective 3

Study details	Participants	Group 1	Control
Author (year) Williams (2021)[110]	**Population** ADHD medication management for people aged 6–15 years	**QbTest + treatment as usual (*n* = 21):**	**Treatment as usual (*n* = 23):**
Study name QUOTA	**Inclusion criteria** 6–17 years; referred to CAMHS/community paediatric; clinical ADHD diagnosis; about to commence ADHD medication (MPH/LDX)	Treatment as usual, in addition to QbTest completed at baseline, 2–4 weeks later (FU 1) and 8–12 weeks later (FU 2). At each time point, the clinician reviewed QbTest results with other clinical tools to monitor medication	Treatment as usual was as listed for group 1 (usual care, without QbTest)
Country England	**Exclusion criteria** Non-fluent English; unable to provide written consent; suspected severe learning disability.	Treatment as usual varied between sites. Participants received their site's standard usual care, but all sites were asked to contact participant twice by the end of the 12 weeks (to ensure level of contact consistent between groups)	
Language English	**Number participants included (analysed)** 44 (44)		
Setting Secondary care/community: five CAMHS or community paediatric clinics	**Age** Mean (SD): QbTest: 9.29 (2.81); control: 9.22 (2.19). Full sample range: 6–15 years		
Study design Parallel group, single-blind, feasibility multisite RCT with embedded qualitative evaluation	**PROGRESS-Plus criteria reported by study**		
Funding Non-industry	• Sex (% male): QbTest – 95.24%. Control – 82.61% • **Ethnicity: QbTest –** white 76.19%; Bangladeshi 4.76%; dual heritage 4.76%; not given 4.76%; other 4.76%; Pakistani 4.76%. **Control –** white 91.30%; Bangladeshi 0%; dual heritage 0%; not given 0%; other 4.35%; Pakistani 4.35% • **Neurodevelopmental: QbTest** – ASD/social communication/speech/speech difficulties 14.28%; attachment disorder 0%; conduct disorder 0%; tic and neurological disorders 9.52%; mood disorders 4.76%. **Control –** ASD/social communication/speech/speech difficulties 21.75%; attachment disorder; 4.35% conduct disorder 8.70%; tic and neurological disorders 0%; mood disorders 0%		

TABLE 57 Results from RCT included for objective 3

Study	Comparison	Outcome	Details	Group 1		Group 2	
				n	Number of events	*n*	Number of events
Williams (2021)[110]	QbTest + treatment as usual vs. treatment as usual	Use of interventions, e.g. ADHD medication	Change to type or dose of ADHD medication at FU 1 (2–4 weeks)	18	10	21	7
			Change to type or dose of ADHD medication at FU 2 (8–12 weeks)	17	7	19	9
			Medication adherence at FU 1: taken medication most/every day	8	8	9	8
			Medication adherence at FU 2: taken medication most/every day	8	7	9	9

Note
Table reports number of participants and number of events in each group, unless stated otherwise.

TABLE 58 Results from qualitative substudy of RCT included for objective 3

Study details	Outcome	Details	Results
Williams (2021)[110]	Impact on clinical decision-making	Clinician-completed proforma (*n* = 33) in the intervention arm (QbTest + treatment as usual)	• Across both FUs, 73% (24/33 responses) of clinicians reported that the QbTest was useful in determining treatment; 18% (6) were neutral, and 9% (3) stated it was not helpful
			• More clinicians found the QbTest helpful at FU 1 (76.5%; 13/17) than FU 2 (68.8%; 11/16)
	Ease of use/ acceptability – patients/ carers	Interviews with the parents of eight children who took part in the trial (six interventions; two controls), about using the QbTest to monitor medication	• Needing to have multiple appointments for the QbTest means time out of school. Appointments before/after school or in the school holidays would be preferable, but ultimately attending the appointments was considered to be beneficial
			• QbTest was described by parents as increasing their confidence in the child's treatment
			• Parents considered repeated QbTests to be useful in increasing confidence in ongoing medication decisions as well as a tool the clinicians used to communicate changes in ADHD symptoms
			• Parents said the QbTest was not considered to be burdensome to children and young people, but some found it 'boring'
			• QbTest has potential to aid communication – parents described how a visual representation of the child's symptoms helped them to better understand treatment impact
	Ease of use/ acceptability – clinicians	Interviews with five clinicians (from four of five clinic sites) from the trial, about using the QbTest to monitor medication. Four community paediatricians and one psychiatrist (all female)	• Objectivity of the QbTest appreciated by clinicians in comparison to informant measures traditionally used to monitor medication

continued

TABLE 58 Results from qualitative substudy of RCT included for objective 3 (*continued*)

Study details	Outcome	Details	Results
			• Clinic appointments often occur during working hours, which has implications for children and their families. Running multiple QbTest appointments could increase these problems and it may be burdensome for children and young people • Preference to only run multiple QbTests when it was perceived to add value. It was described as one of a suite of tools to monitor ADHD symptoms and clinicians felt the additional resources needed to carry out QbTests (staffing, clinic time and test interpretation) are not necessary in routine cases, but may be of use in trickier/complex cases • QbTest has potential to aid communication, helps parents to better understand treatment impact and give extra weight to clinician advice during consultations. Clinicians note this appears to help parents to be more accepting of treatment recommendations
	Confidence of HCP in assessment		• QbTest was described by clinicians (and parents) as increasing their confidence in the child's treatment

TABLE 59 Baseline details for DTA study included for objective 3

Study details	Setting and population	Index test	Reference standard
Tallberg (2019)[66]	**Setting:** secondary care	**Sensor CPT**	**Reference standard**
Design	**Population:** children (age 9–14 years)	QbTest (6–12)	QbTest (6–12) + SNAP-IV
One-gate	**Inclusion criteria:** children with ADHD from a Child and Adolescent Psychiatry clinic in southern Sweden (*n* = 186)		
Country			
Sweden	**Exclusion criteria:** none reported		
Funding	**Numbers:** 186 (56)		
Non-industry			

TABLE 60 Results for DTA study included for objective 3

Study details	Index test	Measure and subgroup	Threshold	Reference stand	TP	FP	FN	TN	Sensitivity	Specificity	AUC (95% CI)
Tallberg (2019)[66]	QbTest (6–12)	QBInattention	Unsure	Unsure	41	4	9	6	0.82	0.60	NR
		QBActivity	Unsure	Unsure	38	6	12	4	0.76	0.40	NR

TABLE 61 Detailed QUADAS-2 assessment showing judgements and rational for risk of bias and concerns regarding applicability for DTA study included for objective 3

	Risk of bias																Concerns regarding applicability				
Study details	Consecutive/random	Case-control avoided	No inappropriate excl	Patient selection bias	Index test blinded	Pre-specified threshold	Index test bias	Reference stand appropriate	Blinded reference stand	Reference stand bias	Time interval	All received reference standard	Same ref stand	All included in analysis	Patient flow bias	Overall bias	Rationale bias	Patient applicability	Reference stand applicability	Index applicability	Overall applicability	Rationale
Talberg – dose titration	✗	✓	✓	☺	☺	✓	?	✗	✗	☹	✓	✓	✓	✗	☹	☹	Index test formed part of reference standard (improvement on SNAP-IV or decrease on QbTest). High proportion of dropouts (130/186)	☺	☺	?	☹	Limited details on test conduct and interpretation

TABLE 62 The CASP checklist quality assessment of study included for objective 3 and not objective 1 that reported qualitative data on patient/clinician carer views of sensor CPTs for ADHD medication management/titration

Study details: **Williams (2021)**[110]		
Question	**Answer (yes/cannot tell/no)**	**Comments**
Was there a clear statement of the aims of the research?	Yes	Aims of the interviews are listed in the measures section as not the main aim of the feasibility trial
Is a qualitative methodology appropriate?	Yes	
Was the research design appropriate to address the aims of the research?	Yes	
Was the recruitment strategy appropriate to the aims of the research?	Yes	All intervention and control participants and clinicians were invited to take part (random subsample – see protocol)
Were the data collected in a way that addressed the research issue?	Yes	
Has the relationship between researcher and participants been adequately considered?	Cannot tell	Not stated whether researcher examined their own potential bias
Have ethical issues been taken into consideration?	Yes	
Was the data analysis sufficiently rigorous?	Yes	
Is there a clear statement of findings?	Yes	

TABLE 63 Quality assessment of study included for objective 3 and not objective 1 that reported survey data on patient/clinician carer views of sensor CPTs for ADHD medication management/titration

Questions	**Williams (2021)**[110]
Was the problem or phenomenon under investigation defined, described and justified?	Yes
Was the population under investigation defined, described and justified?	Yes
Were specific research questions and/or hypotheses stated?	Yes
Were operational definitions of all study variables provided?	Yes
Were participant inclusion criteria stated?	Yes
Was the participant recruitment strategy described?	Yes
Was a justification/rationale for the sample size provided?	No
Was the attrition rate provided? (applies to cross-sectional and prospective studies)	Yes
Was a method of treating attrition provided? (applies to cross-sectional and prospective studies)	Yes
Were the data analysis techniques justified (i.e. was the link between hypotheses/aims/research questions and data analyses explained)?	Yes
Were the measures provided in the report (or in a supplement) in full?	No
Was evidence provided for the validity of all the measures (or instrument) used?	No
Was information provided about the person(s) who collected the data (e.g. training, expertise and other demographic characteristics)?	Yes
Was information provided about the context (e.g. place) of data collection?	Yes
Was information provided about the duration (or start and end date) of data collection?	No
Was the study sample described in terms of key demographic characteristics?	Yes

TABLE 63 Quality assessment of study included for objective 3 and not objective 1 that reported survey data on patient/clinician carer views of sensor CPTs for ADHD medication management/titration (*continued*)

Questions	Williams (2021)[110]
Was discussion of findings confined to the population from which the sample was drawn?	Yes
Were participants asked to provide (informed) consent or assent?	Yes
Were participants debriefed at the end of data collection?	Cannot tell
Were funding sources or conflicts of interest disclosed?	Yes

Appendix 4 Forest plots showing estimates of area under the receiver operating characteristic

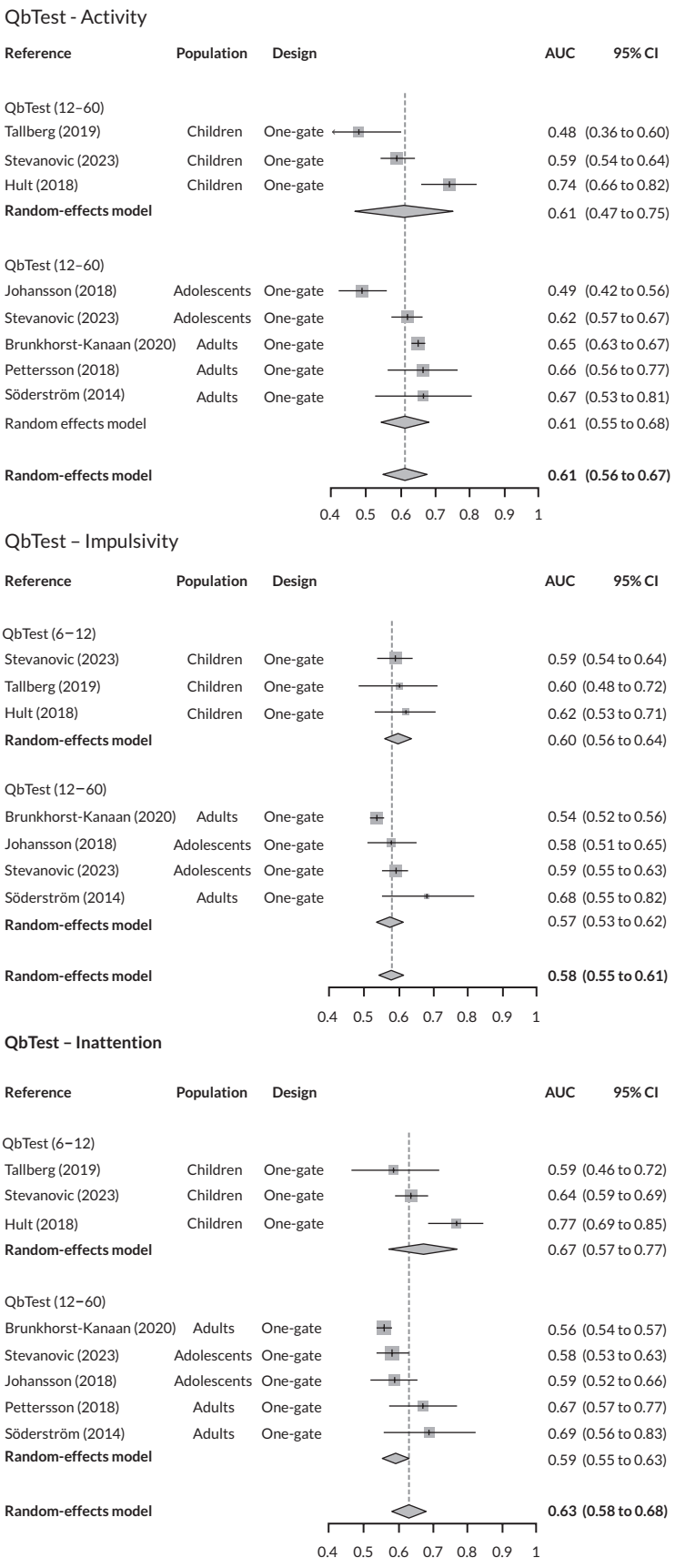

FIGURE 19 Forest plot showing estimates of AUC ROC with 95% CIs for studies that evaluated the QbTest stratified according to QbTest domain.

Reference	Population	Design	Subgroup	AUC	95% CI
EF Sim					
Seesjärvi (2022)	Children	Two-gate	Overall	0.70	(0.58 to 0.82)
Seesjärvi (2022)	Children	Two-gate	Controller motion	0.73	(0.61 to 0.85)
Seesjärvi (2022)	Children	Two-gate	Navigation efficacy	0.75	(0.64 to 0.86)
Seesjärvi (2022)	Children	Two-gate	Actions	0.78	(0.67 to 0.89)
Seesjärvi (2022)	Children	Two-gate	Task efficacy	0.83	(0.74 to 0.92)
QbCheck					
Ulberstad (2020)	Adolescents and adults	Two-gate	Reaction time	0.73	(0.65 to 0.81)
Ulberstad (2020)	Adolescents and adults	Two-gate	Commission errors	0.74	(0.66 to 0.82)
Ulberstad (2020)	Adolescents and adults	Two-gate	Omission errors	0.75	(0.67 to 0.83)
Ulberstad (2020)	Adolescents and adults	Two-gate	Microevents	0.80	(0.73 to 0.87)
Ulberstad (2020)	Adolescents and adults	Two-gate	Reaction time variability	0.81	(0.74 to 0.88)
QbMini					
Hamadache (2021)	Children	Two-gate	QBActivity	0.51	(0.38 to 0.63)
Hamadache (2021)	Children	Two-gate	Qblnattention	0.52	(0.40 to 0.65)
Hamadache (2021)	Children	Two-gate	QblmpulsivityQblmpulsivity	0.59	(0.47 to 0.71)

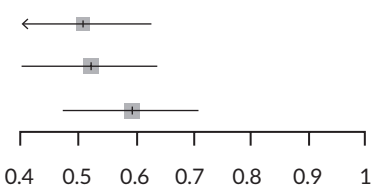

FIGURE 20 Forest plot showing estimates of AUC ROC with 95% CIs for sensor CPTs that were evaluated in single studies.

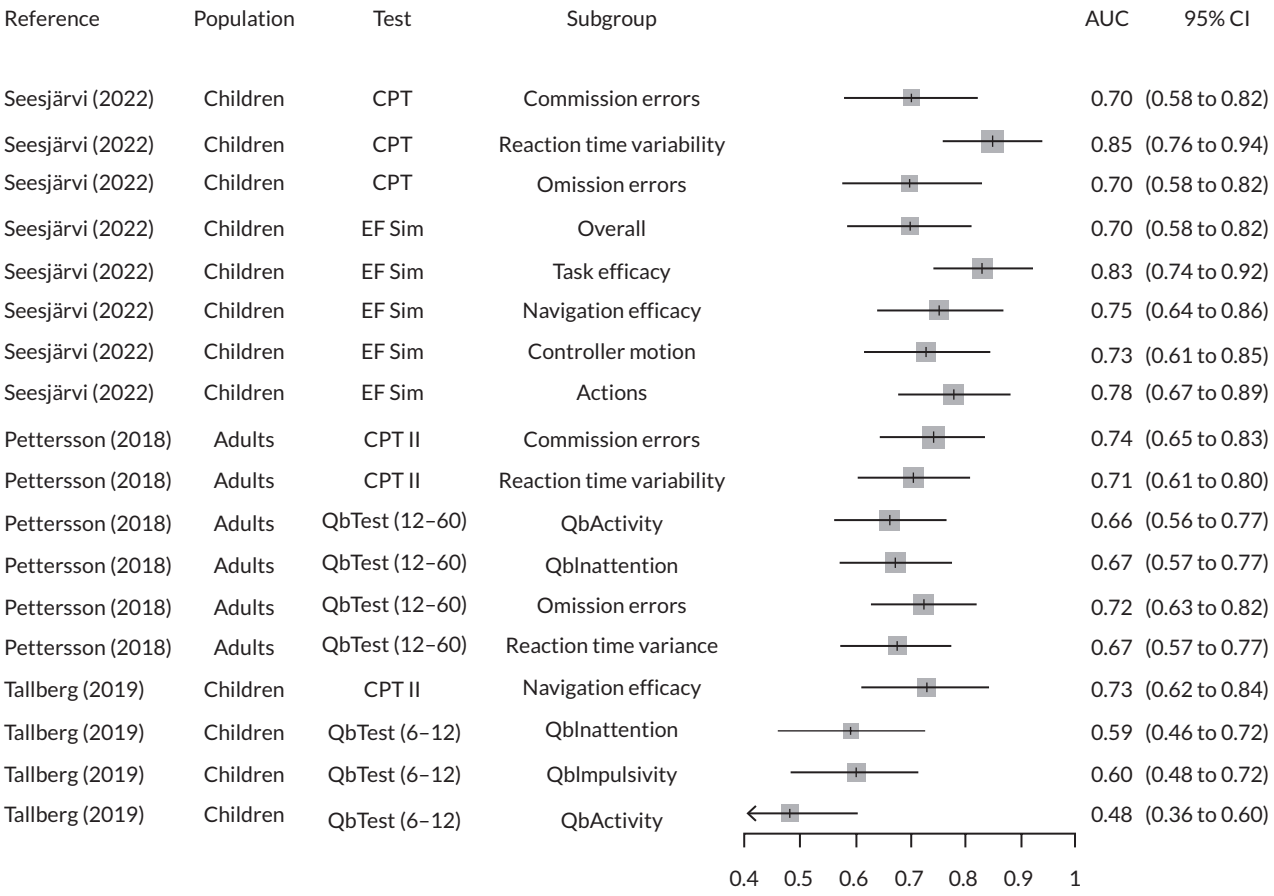

FIGURE 21 Forest plot showing estimates of AUC ROC with 95% CIs studies that estimated the accuracy of a non-sensor CPT and sensor CPT on the same participants.

Appendix 5 Synthesis of studies that reported on patient/clinician carer views of sensor continuous performance tests for attention deficit hyperactivity disorder diagnosis

Views around the helpfulness of the QbTest

Conceptual categories we identified regarding views around the helpfulness of the QbTest included contribution to ADHD diagnosis, communication with caregivers and understanding of subjective experience.

Contribution to attention deficit hyperactivity disorder diagnosis

Findings from qualitative data

Clinicians interviewed in the qualitative substudy of the AQUA trial reported that use of the QbTest led them to feel more confident in their diagnostic decision-making.[107]

> *I would move to the diagnosis more confidently and more quickly having evidence that something was wrong, you know objective evidence . . . reduced the amount of the anxiety of uncertainty.*
>
> *HCP on the use of QbTest*[107]

Increased confidence in the diagnostic decision was also reported in interviews with healthcare staff in the Focus ADHD study, who commented that the increased confidence was derived from the fact that the data provided by the test are objective, rather than scales and surveys that give subjective data.[31] Focus groups with clinicians in CAMHS also revealed that the QbTest gave them increased confidence in their decisions.[87]

> *I think it gives all clinicians a bit more confidence around making diagnosis, and I think for nurses, that's where its particularly helpful. Especially if they're nurse prescribers, because they have that responsibility of making the diagnosis and providing medication. So, they want it to be . . . they want to feel very, very sure that this is ADHD, that nothing is being missed.*
>
> *HCP on using QbTest*[31]

Despite the suggestion from studies that the QbTest could contribute positively to the ADHD diagnostic process, clinicians reported in focus groups that there is a need to establish where the QbTest falls on the ADHD assessment pathway.[87] Staff interviewed in the Focus ADHD study felt that the QbTest should be implemented early in the assessment pathway, and when this was done, the clinicians felt they had a clearer understanding of whether the young person had a profile indicative of ADHD.[31] In line with this, most clinicians and families interviewed in the AQUA substudy felt that the QbTest should be conducted before the initial appointment with clinician. One family suggested doing QbTest in GP surgery as initial screen and clinicians were supportive of this, whereas, some clinicians suggested using it only for complex cases.[107]

> *I would then also even put a QbTest in as a precursor to the initial consultation so that at the time you see the child, they've had all the relevant questionnaires completed from home and school and a QbTest and you could probably make a diagnosis on the first appointment.*
>
> *HCP on the use of QbTest*[107]

Some clinicians and families interviewed in the AQUA substudy questioned the validity of the clinical setting of the test and wondered if it is not representative of what happens, for example, in school.

> *He behaves differently at home and school to what he would do in a clinical office sort of thing . . . And of course for that twenty minutes that he was seen he was on his best behaviour.*
>
> *HCP on the use of QbTest*[107]

Findings from quantitative data

Some respondents to the patient/carer survey in the Focus ADHD study said that they think the QbTest should have been offered sooner.[31]

Diagnosis in complex cases

Findings from qualitative data

Interview findings from two studies suggested that the QbTest can be helpful in the diagnosis of individuals with comorbidities.[72,107] Clinicians interviewed in the AQUA trial substudy reported that the tests helped to discriminate ADHD from ASD, anxiety, depression and learning difficulties. Clinicians with more prior experience of using the QbTest were more positive about its abilities to help in the diagnosis of cases with comorbidities than those with less experience.[173] In the FACT RCT, one staff member interviewed also said that the QbTest was helpful in the assessment of young people where there might be concerns about comorbid diagnosis.[72]

> *I very often use it for children that I suspect have got ASD comorbidity. I think it's very clear that there's a group of children with just pure ADHD who do a QbTest in one way, and then the group that's got some degree of autism or autistic traits do it very differently, and I think that's really helpful.*
>
> *HCP on the use of QbTest*[107]

In the Focus ADHD study, interviews with staff also found that the QbTest can be helpful in cases where there is contradictory information between home and school settings, or cases where the young person has limited corroborating information due to being home schooled or a 'looked after child'.[31]

> *I think it works well with subtle presentations. Presentations maybe where there's a disagreement between school and home. Cases where there are parental disagreements. Cases where young people themselves are unsure.*
>
> *HCP on the use of QbTest*[31]

Clinicians interviewed in the AQUA trial substudy reported that the QbTest was useful in differentiating ADHD subtypes, but there was no consensus as to which symptom domain was particularly valuable. Some clinicians specifically commented on the utility of the attention measure for girls with the inattentive subtype who can be hard to diagnose.[107] This was also highlighted in the Focus ADHD study – healthcare staff commented that the addition of the QbTest into the assessment process helped to identify individuals with subtle presentation of ADHD (e.g. girls or older adolescents) and those who mask their difficulties.[31]

> *I think it can be helpful for picking out cases where there might be more subtle presentations, for example in girls or older adolescents.*
>
> *HCP on the use of QbTest*[31]

Findings from quantitative data

In two studies that surveyed HCPs, there was no consensus as to whether the QbTest should be reserved for use in cases where there is a diagnostic uncertainty.[31,69] However, in one of the studies,[31] some HCPs did report that the test was most useful in certain patient groups, including female patients, older children, cases where the parent or school does not agree with the clinician's decision and in identifying patients for ASD assessment by being able to rule out ADHD. Survey data from healthcare staff in the Focus ADHD study concurred with the interview findings from this study; healthcare staff reported that the QbTest is useful in those with subtle presentations who mask their symptoms.[174]

Time to diagnostic decision

Findings from qualitative data

Qualitative data (mostly from HCPs) from four studies suggested that the QbTest could be helpful in improving the time to diagnostic decision. Clinicians interviewed in the AQUA substudy reported that the QbTest may help to reduce delays in diagnosis and treatment onset. They also noted that time and cost savings may be made by replacing the lengthy and difficult-to-access school observations with the QbTest.[107]

> *What we did was because of QbTest results, I then stopped the school observations, so then we could confirm the diagnosis and go ahead with the medication.*
>
> *HCP on the use of QbTest*[107]

Families interviewed in the AQUA trial commented that there is a need for a quick decision to facilitate treatment initiation, particularly for children who were struggling in education, and to not prolong the emotionally overwhelming process. However, they also emphasised that they did not want the process to be rushed and their child should not be 'labelled' quickly.[107]

> *I just wished it were more like I say I was in and out, just wished it were more appointments and a bit more time.*
>
> *Parent of child who had used QbTest*[107]

Staff interviewed in the FACT RCT also felt that the QbTest could help to improve waiting times.[72] Focus groups with 19 clinicians who had used the QbTest in CAMHS highlighted that the QbTest was perceived to have resulted in time savings and felt that it has the potential to streamline and improve the service.[87]

> *[S]o on the ground level it's helping us with our picture of the child, but in the bigger picture of things, if we are dealing more efficiently and more correctly with each child, that's going to make the service more efficient and better for the next child coming in the door, so there's a bigger picture knock on effect happening with a tool like this.*
>
> *HCP on the use of QbTest*[87]

These views were shared by some healthcare staff interviewed in the Focus ADHD study, who felt that the addition of the QbTest into the assessment process led to a faster and more efficient process, which in turn reduces cost.[31] Most sites in the Focus ADHD study found that QbTest implementation had resulted in fewer appointments by replacing the school observation and that the quicker assessment pathway supported the young person in getting educational support quickly. Some sites also reported a reduction in rereferrals from caregivers who disagreed with a non-diagnosis decision.

> *I see it as a way of reducing the amount of time children are waiting to be seen. And thus, reducing the number of follow-ups, thus reducing the number of times they have to come back to the hospital so it's an opportunity to save the patients and parents time.*
>
> *HCP on the use of QbTest*[31]

Findings from quantitative data

Some patients/carers (*n* = not reported) surveyed in the Focus ADHD study reported that the QbTest helped to speed up the assessment process and to get a diagnosis.[31]

Communication with caregivers

Findings from qualitative data

Interview findings suggested that the QbTest helped to improve communication between clinicians and patients/families. Clinicians and families interviewed in the AQUA trial substudy felt that the output of the QbTest helped them to communicate to families information around diagnosis and medication effect. Specifically, clinicians reported that

being able to show a comparison of the child's performance to a normative sample helped them to communicate the diagnostic decision to families, and they thought that this helped families to accept the decision.

> *A lot of parents who previously would have probably shouted and screamed at you for not saying their child had ADHD will accept it if the computer is not showing the evidence.*
>
> *HCP on using QbTest*[107]

Mostly, families in the AQUA substudy felt that clinicians explained the QbTest reports well and they were easy to understand; however, some families felt that it was unclear how the report was being used to inform decision-making.[107]

> *I don't know if she explained, it felt like the QbTest had said it so that's what we're going with.*
>
> *Parent of child who had used QbTest*[107]

In two other studies, clinicians also felt that the QbTest helped to improve communication with young people and their families through improving clarity[87] and through providing an objective and visual aspect to use to evidence and justify diagnostic decisions.[31] However, some clinicians in the latter study (Focus ADHD) commented that families could still struggle to accept a diagnostic decision.[31] This study did not interview parents/carers.

> *I think they offer a very visual result for the parents, [. . .] especially the little chart that shows hyperactivity and stillness and the wild swinging round. So, I think that sort of aspect to it is really good to be able to communicate the diagnosis.*
>
> *HCP on use of QbTest*[31]

Findings also suggested that the implementation of the QbTest can help to improve communication between the clinician and school,[107] between clinical colleagues[87] and between the person with ADHD and their family.[107]

Findings from quantitative data

Survey data suggested that clinicians valued the QbTest for improving communication with the patient/family. In line with the results from the AQUA substudy interviews, all 10 clinicians surveyed reported that the QbTest helped to improve communication with patients and they all valued the QbTest in helping to explain why they had ruled out a diagnosis.[107] Likewise, the majority of clinicians surveyed in three other studies felt that the QbTest results improved the communication of diagnostic decision with the patient.[31,69,87]

However, the views of parents/carers were more mixed as to whether the QbTest improved communication. Only 31/68 families in the AQUA substudy said that the QbTest helped them to understand how the diagnosis was made, and answers were split regarding whether they thought the results of the test were difficult to understand. Families who received a diagnosis of ADHD were more likely to view the QbTest as useful for understanding how the diagnosis was made than those who were not diagnosed.[107] Similarly, in the Focus ADHD study, only 10/22 patients/carers surveyed felt that when the clinician talked through the QbTest results with them, it helped them to understand how they reached the diagnosis. The respondents did not have a strong opinion about whether the results were difficult to understand (votes were split and many voted 'neither agree/disagree'), but some respondents noted in free-text responses that they did not find the test to be helpful because the results were not properly explained to them.[31]

In two studies, parents/carers provided a more positive view on the QbTest for aiding communication, with the majority of survey respondents reporting that the clinician talking through the results helped them to understand how their diagnosis had been made.[69,87]

Understanding of subjective experience

Findings from qualitative data

Clinicians reported in focus groups that the test helped them to better understand the young person's subjective experience.[87] Additionally, one staff member interviewed as part of the FACT RCT reported that the QbTest helped the young person and the staff to better understand the young person's behaviours.[72]

> *It feels as if it brings another layer into knowing some of the children.*
>
> *CAMHS professional on the use of the QbTest*[87]

Clinicians and families interviewed in the qualitative substudy of the AQUA trial appreciated that the QbTest provided what they regarded as an objective and observable measure of symptoms.[107] This finding was echoed in focus groups with clinicians in CAMHS,[87] interviews with healthcare staff in the Focus ADHD study[31] and by one staff member interviewed in the FACT RCT.[72]

> *I think to be able to see something, it's that black and whiteness of it, to look at it and go yeah I can see that.*
>
> *Parent on the use of the QbTest*[107]

Findings from quantitative data

Findings from surveys with HCPs were in line with the interview data in suggesting that the QbTest can help staff to better understand the patient's symptoms. In the AQUA trial substudy, all 10 clinicians surveyed felt that the QbTest had helped them to better understand the patient's ADHD symptoms.[107] Likewise, most HCPs surveyed in the Focus ADHD study agreed that the QbTest results were helpful in understanding their client's symptoms[31] as did clinicians surveyed who had used the QbTest in CAMHS.[87]

Findings were more mixed from surveys with patients and carers. In the AQUA trial qualitative substudy, only 35/73 families surveyed felt that it helped them to understand their child's symptoms better.[107] Likewise, only 11 out of 22 patients/carers surveyed in the Focus ADHD study felt that the QbTest helped them to understand their symptoms.[31] In a survey of 10 adolescent boys in a YOI who used the QbTest in the FACT RCT, the majority of respondents reported that they neither agreed nor disagreed that the QbTest helped them to understand their ADHD symptoms or changes in their symptoms.[72] Two studies reported more beneficial effects of the QbTest on level of understanding. In one study, 13/15 children/adolescents reported that the QbTest helped them to understand their symptoms,[87] and in the other study, 41/48 children (and their families) felt that it helped them to understand their symptoms.[69]

Barriers to implementation of the QbTest

Conceptual categories we identified regarding views around barriers to the implementation of the QbTest included practical barriers, other barriers and acceptability to patients/carers.

Practical barriers

Space

Findings from qualitative data

Interview data from three studies highlighted that a room is required to be able to administer the QbTest, and sometimes, this is hard to arrange, which means the equipment may need to be moved between rooms.[31,72,107] Focus groups with clinicians in CAMHS highlighted concerns about managing environmental factors influencing the QbTest.[87]

> *The main [challenges] were just the practical side, like the room space and things. It's really competitive to get rooms here so making sure it was booked well in advance.*
>
> *HCP on the use of QbTest*[31]

Findings from quantitative data
None reported.

Staffing

Findings from qualitative data
Clinicians in the AQUA substudy said that use of the QbTest requires someone trained to administer the task and they thought it is best delivered by healthcare assistant, then interpreted by clinician. However, some HCPs noted that it was important to observe the test to assess the validity of the results.[107] Similarly, in focus groups conducted with clinicians in another study, while some clinicians felt that hiring an administrator to administer the test would be helpful, others felt that observing a young person complete the QbTest provided extremely valuable information and this superseded the value that a team would receive from a QbTest administrator.[87] Staff interviewed in the Focus ADHD study highlighted issues with training needs and staff capacity,[31] and interviews with clinicians in one other study flagged the need for continued supervision and learning about the test.[87]

> *If you're not aware of what's actually happening at that time, then I think it might be difficult . . . the actual observation, what's happening during that time, is very important.*
>
> *HCP on the use of QbTest*[107]

Findings from quantitative data
None reported.

Technology

Findings from qualitative data
Some clinicians in the AQUA trial had issues with technology (internet connection and access to printer) and lack of resources.[107] Likewise, focus groups with clinicians reported being intimidated by the technology and noted instances of QbTest reports disappearing, connectivity issues and components of the test breaking.[87] Staff in the FACT RCT also reported concerns because of equipment and information technology system needed,[72] and staff interviewed in the Focus ADHD similarly flagged issues with equipment and Wi-Fi, including challenges with finding a room with a Wi-Fi connection, accessing laptops and sharing passwords.[31]

> *There was a lot of IT [Information Technology] governance issues to get it set up.*
>
> *HCP on the use of QbTest*[107]

Findings from quantitative data
None reported.

Other barriers

Findings from qualitative data
Funding was mentioned as a resource need in the Focus ADHD study.[31] Additionally, a lack of FU was highlighted in the AQUA substudy. Some families interviewed felt abandoned by the service after diagnosis and those who received medication reported they should have been more closely monitored. Additionally, those who did not receive medication were unclear of what options were available.[107] However, it is not clear how this relates to the QbTest as opposed to the general diagnostic process.

> *Like I just feel like maybe my child by the doctors and stuff has been let down a bit by not being seen and just like he said he should have been seen really after the medication and he hasn't.*
>
> *Parent of child who used QbTest*[107]

Findings from quantitative data

None reported.

Acceptability to patients and caregivers

Findings from qualitative data

Two studies reported qualitative data concerning the acceptability of the QbTest. In the FACT RCT, some of the adolescent boys interviewed reported that they found the QbTest boring or felt exhausted by it and one person felt cross that they had to repeat the test. However, one person did report that they would recommend the test to others (no quotes provided).[72]

In the Focus ADHD study, interviews with healthcare staff highlighted that particular groups struggled to use the test. Some young people experienced sensory discomfort during the QbTest and some individuals with autism also struggled with having the tight headband around their head. In some instances, the individual could adapt the test (e.g. to wear a hoodie underneath the headband); however, these issues did prevent some individuals from completing the test. Staff also reported that some young people (particularly 6-year-olds) struggled with anxiety during the test due to the test itself and/or being without their caregivers. Additionally, some of the younger children struggled to follow the instructions and some older teenagers disengaged from the test and became disruptive. Further issues were raised about the language used in the assessment (e.g. use of the word 'test' made people stressed), the length and repetitive nature of the test, the lack of representation of different ethnicities in the explanation video and the requirement to choose biological sex before conducting the test.[31]

> *A lot of our young people that come in for both an autism and an ADHD assessment can experience difficulty with the plastic covering of the headband, because it's quite a sensory thing on the head and that can be quite uncomfortable. It's quite tight on the forehead and around the head.*
>
> *HCP on the use of QbTest*[31]

Findings from quantitative data

Four studies provided information about the acceptability of the QbTest from surveys to patients/carers.[31,69,72,87] Findings were mixed between studies, with some participants finding the QbTest difficult to complete and others not having issues with the test.

In a survey of 10 adolescent boys assessed for ADHD in the FACT RCT (based in a YOI), the majority (9/10) of respondents said that they found the QbTest assessment to be very stressful and that the task took too long to complete. Additionally, 8 out of 10 respondents agreed that the task was difficult to complete.[72]

By contrast, in a survey of 48 children (and their families) who had used QbTest in CAMHS, the majority of respondents reported that the results were not difficult to understand and they did not find the task difficult to complete.[69] Additionally, in a survey of 15 children/adolescents who had used the QbTest in a study conducted in CAMHS, 67% did not find the task difficult to complete and most (93%) agreed that, overall, the experience of using the test was helpful. There was no clear consensus in this study on whether respondents found the stool/chair very uncomfortable or whether the QbTest results were difficult to understand.[87]

In the Focus ADHD study, there was no clear consensus on whether the QbTest was difficult to complete (3/22 said it was, 9 neither agree/disagree and 10 strongly disagree/disagree).[31] Although, some of the participants surveyed reported issues with the test, including that their child could not sit through the full test, the QbTest machine did not work in their appointment and that they felt the staff member delivering the test did not know what they were doing.

Two studies provided information from surveys about the acceptability of the QbTest for clinicians.[69,87] In a survey of 17 clinicians who had used the QbTest in CAMHS, 13/17 clinicians agreed that the QbTest was easy to use. Additionally, all clinicians agreed that the test helps them to visualise and quantify symptoms, it is a great addition to other investigative techniques and it is helpful to monitor the effects of treatment and to standardise assessment and treatment.[87] Whereas, in another study that involved a survey to clinicians in CAMHS (*n* = not reported), 30% of respondents found the results were difficult to understand.[69]

QbCheck

One study provided survey data on the acceptability of the QbCheck from a short survey given to 125 patients (56 with ADHD; 69 healthy controls) in a DTA study.[77] The participants reported that they found the test easy to use, including performing the preparations before starting the test and understanding and following the test rules during the test. The questions were scored on a scale of 0–10, with higher scores indicating higher ease of use, and mean values were all ≥ 8.06. The most common reason for a score < 8 was that the test took a long time, so it was hard to stay focused.

Patient/carer/clinician views of the EF Sim Test

Two studies reported survey data for the EF Sim test, mainly focusing on the acceptability of the test. As there are only two studies, which reported fairly limited data, we summarise them in turn, below.[75,88]

One study, run by the test manufacturer, surveyed 21 teachers of participating schools that had implemented the EF Sim test for students in a pilot study. On average, the majority of the teachers found the test results usable and reported that they can support communication with guardians and that they are helpful to identify executive functioning challenges in students that may otherwise go unnoticed.[88] (Confidential information has been removed.)88

The other study was a DTA study of the EF Sim test (previous version named EPELI) in children (some with ADHD, some healthy controls, *n* = not reported). The short survey was answered on a scale from 1 'no' to 7 'completely/very much'. On average, children appeared to feel enthusiastic about the tasks (ADHD mean score 5 SD, 1.95; healthy control 5.45, SD 1.59), found them to be interesting (ADHD mean score 4.82, SD 2.17; healthy control 5.32, SD 1.54) and they put effort into their performance (ADHD mean score 5.87, SD 1.23; healthy control 6.21, SD 0.96).

Appendix 6 Review of economic models: Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagrams

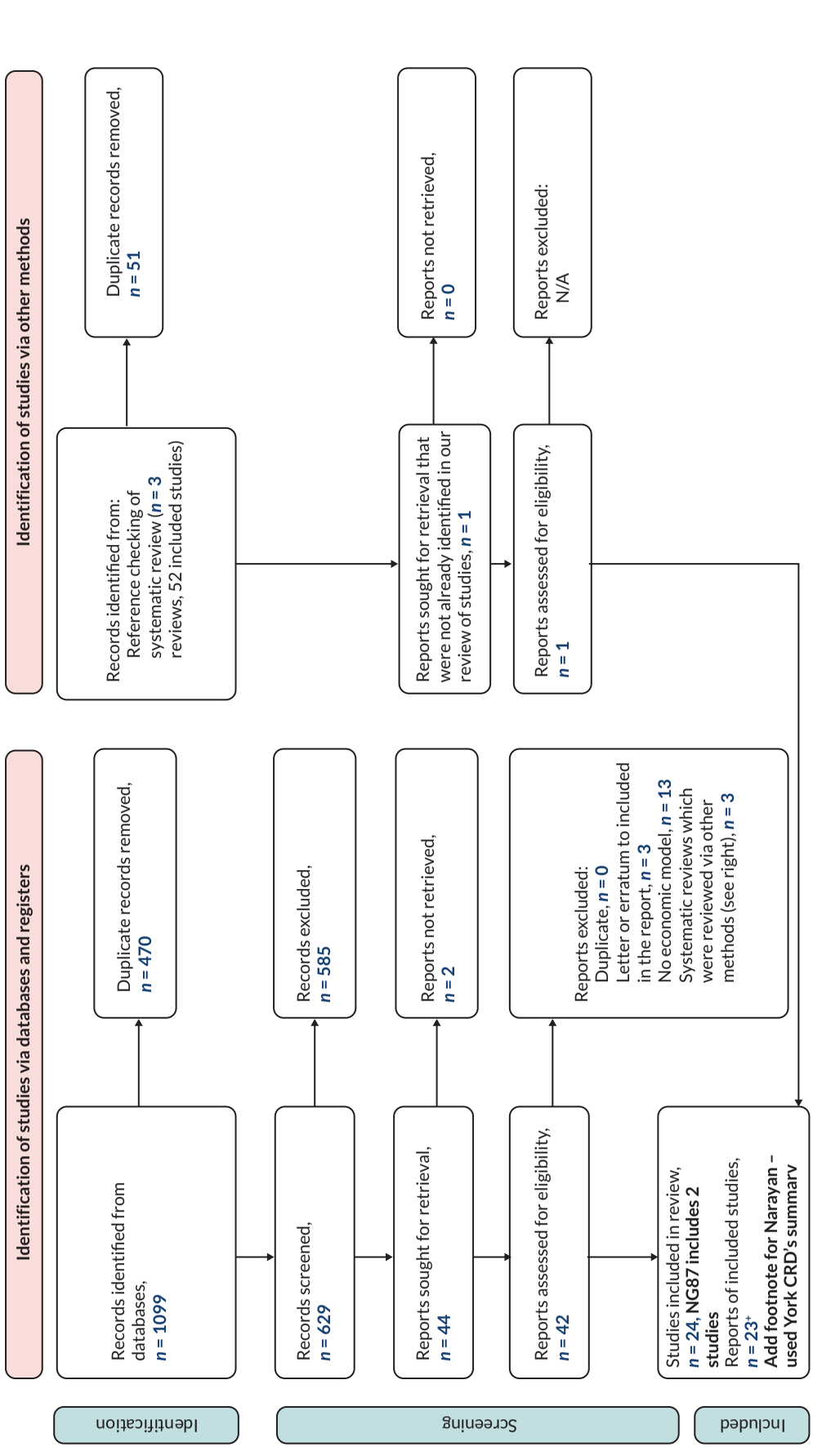

FIGURE 22 The PRISMA diagram for the review of economic models for diagnosis and treatment of ADHD.

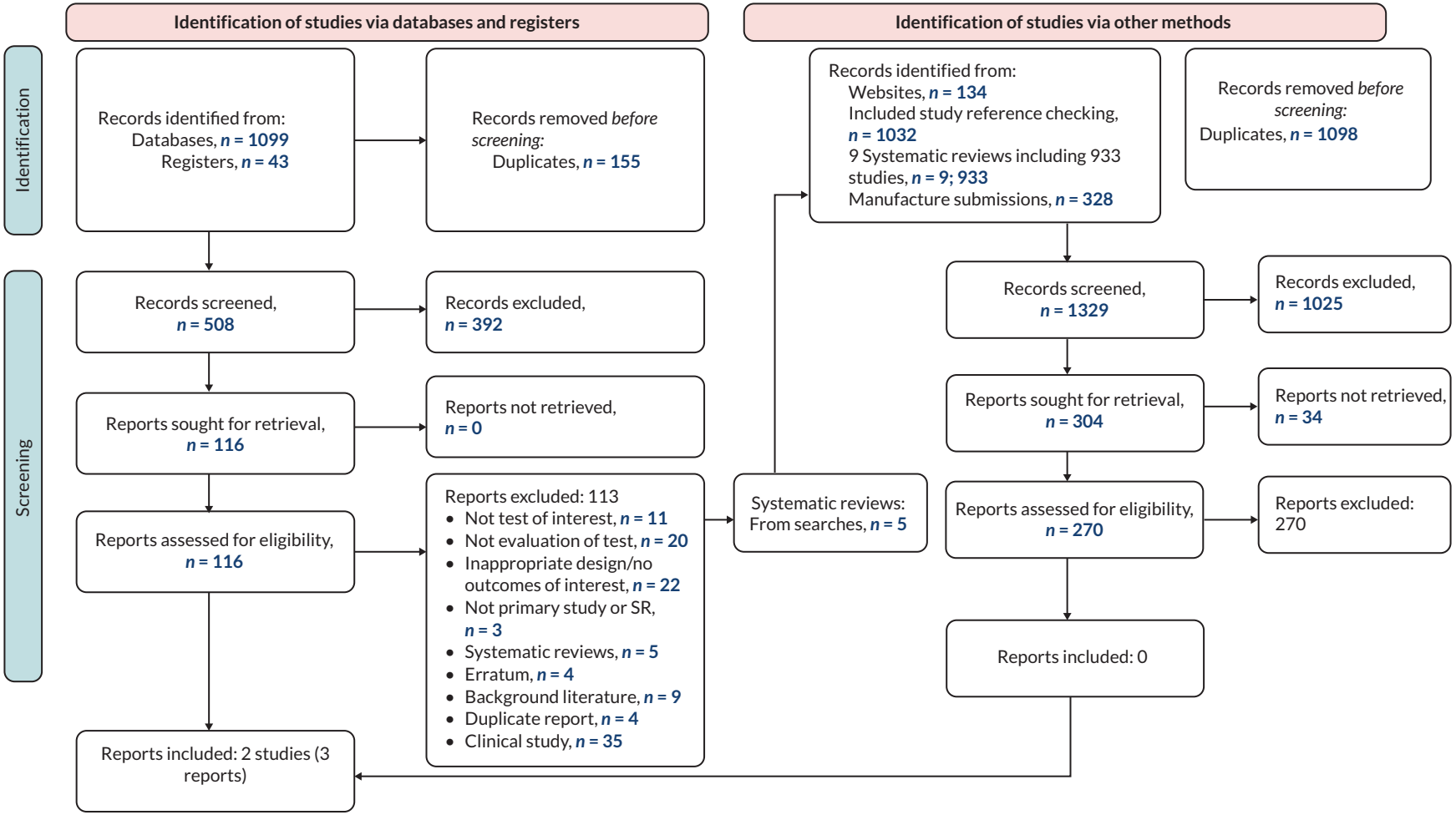

FIGURE 23 The PRISMA diagram for the identification of economic evaluations of sensor CPTs for diagnosis of ADHD.

Appendix 7 Quality assessment economic evaluations of sensor continuous performance tests for the diagnosis of attention deficit hyperactivity disorder

TABLE 64 Quality assessment using the Drummond checklist[112] for the two economic evaluations of sensor CPTs for ADHD

Drummond criteria	AQUA trial[18]	AHSN study[69]
Study design		
1. The research question is stated	Yes	Yes
2. The economic importance of the research question is stated	Yes	Yes
3. The viewpoints of the analysis are clearly stated and justified	Yes	Yes
4. The rationale for choosing the alternative programmes or interventions compared is stated	Yes	Yes
5. The alternatives being compared are clearly described	Yes	Yes
6. The form of economic evaluation used is stated	Yes	Yes
7. The choice of form of economic evaluation is justified in relation to the questions addressed	Yes	Yes
Data collection		
8. The sources of effectiveness estimates used are stated	Yes	Yes
9. Details of the design and results of effectiveness study are given (if based on a single study)	Yes	Yes, but more detail would be useful
10. Details of the method of synthesis or meta-analysis of estimates are given (if based on an overview of a number of effectiveness studies)	N/A	N/A
11. The primary outcome measure(s) for the economic evaluation are clearly stated	Yes	Yes
12. Methods to value health states and other benefits are stated	No, but EQ5DY data were collected	Yes
13. Details of the subjects from whom evaluations were obtained are given	Yes	No or N/A
14. Productivity changes (if included) are reported separately	N/A	N/A
15. The relevance of productivity changes to the study question is discussed	N/A	Yes
16. Quantities of resources are reported separately from their unit costs	Yes	Yes, but not very clearly
17. Methods for the estimation of quantities and unit costs are described	Yes	Yes
18. Currency and price data are recorded	Yes	No
19. Details of currency or price adjustments for inflation or currency conversion are given	Yes	Yes
20. Details of any model used are given	N/A	Yes
21. The choice of model used and the key parameters on which it is based are justified	N/A	N/A
Analysis and interpretation of results		
22. Time horizon of costs and benefits is stated	Yes	No (only in results and it is inconsistent)
23. The discount rate(s) is stated	Yes	Yes

continued

TABLE 64 Quality assessment using the Drummond checklist for the two economic evaluations of sensor CPTs for ADHD (*continued*)

Drummond criteria	AQUA trial[18]	AHSN study[69]
24. The choice of rate(s) is justified	Yes	No
25. An explanation is given if costs or benefits are not discounted	Yes	NA
26. Details of statistical tests and CIs are given for stochastic data	NA	No
27. The approach to sensitivity analysis is given	N/A	Yes
28. The choice of variables for sensitivity analysis is justified	N/A	No
29. The ranges over which the variables are varied are stated	N/A	Yes
30. Relevant alternatives are compared	Yes	Yes
31. Incremental analysis is reported	ICER is reported	No
32. Major outcomes are presented in a disaggregated as well as aggregated form	No	Yes
33. The answer to the study question is given	Yes	Yes
34. Conclusions follow from the data reported	No, because QbTest cost excluded from both arms	Yes
35. Conclusions are accompanied by the appropriate caveats	Yes	Yes

N/A, not applicable.

EME
HSDR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library