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

Scientific summary

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

Health Technology Assessment 2025; Vol. 29: No. 58
DOI: 10.3310/DRDR7171

NIHR Journals Library www.journalslibrary.nihr.ac.uk

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.

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).