



Extended Research Article

Tele-ophthalmology-enabled and artificial intelligence-ready referral pathway for community optometry referrals of retinal disease: HERMES cluster randomised trial with a diagnostic accuracy study

Anitta Sharma,^{1,2,3,4} Rima Hussain,^{1,3} Annastazia E Learoyd,⁵ Angela Aristidou,^{6,7} Taha Soomro,¹ Ann Blandford,⁸ John G Lawrenson,⁹ Gabriela Grimaldi,^{1,3,10} Abdel Douiri,⁵ Ashleigh Kernohan,¹¹ Tomos Robinson,¹¹ Najmeh Moradi,¹¹ Christiana Dinah,¹² Evangelos Minos,¹³ Dawn Sim,³ Tariq Aslam,¹⁴ Avinash Manna,¹⁵ Alastair K Denniston,¹⁵ Praveen J Patel,^{1,3,10} Pearse A Keane,^{3,4,10} Catey Bunce,¹⁶ Luke Vale^{11,17} and Konstantinos Balaskas^{1,3,4,10*}

¹Moorfields Ophthalmic Reading Centre and Artificial Intelligence Lab, Moorfields Eye Hospital NHS Foundation Trust, London, UK

²Optometry, Moorfields Eye Hospital NHS Foundation Trust, London, UK

³NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust, London, UK

⁴Institute of Ophthalmology, University College London, London, UK

⁵School of Population Health and Environmental Sciences, King's College London, London, UK

⁶School of Management, University College London, London, UK

⁷Digital Economy Lab, Stanford University, Stanford, CA, USA

⁸UCLIC, University College London, London, UK

⁹School of Health and Medical Sciences, City St George's, University of London, London, UK

¹⁰Medical Retina Service, Moorfields Eye Hospital NHS Foundation Trust, London, UK

¹¹Health Economics Group, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

¹²Ophthalmology, London North West University Healthcare NHS Trust, London, UK

¹³North West Anglia NHS Foundation Trust, Peterborough, UK

¹⁴School of Biomedical Sciences, University of Manchester, Manchester, UK

¹⁵NIHR Birmingham Biomedical Research Centre, London, UK

¹⁶Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK

¹⁷Global Health Economics Centre, London School of Hygiene and Tropical Medicine, London, UK

*Corresponding author konstantinos.balaskas@gmail.com

Published December 2025

DOI: 10.3310/QNDF3325

Scientific summary

Tele-ophthalmology-enabled and artificial intelligence-ready referral pathway for community optometry referrals of retinal disease: HERMES cluster randomised trial with a diagnostic accuracy study

Health Technology Assessment 2025; Vol. 29: No. 69

DOI: 10.3310/QNDF3325

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

Scientific summary

Background

Ophthalmic services are challenged by an ageing demographic and the associated rise in common retinal diseases. Despite an increasing demand, there is an anticipated global shortfall of ophthalmologists. Community-based optometrists (COs) are essential in managing and referring people to hospital eye services (HES). Optical coherence tomography (OCT) devices have been increasingly installed in community practices. Although OCT expanded optometrists' scope of practice, its widespread use caused an increase in overall and unnecessary referrals. Efforts to streamline incoming referrals are vital to help alleviate pressures to HES and to improve patients' access to timely diagnosis and treatment.

This could be achieved with teleophthalmology, which is the use of medical information exchanged from one site to another via electronic communications to improve a patient's eye health. A systematic review revealed the evidence gap from well-designed, randomised controlled trials (RCTs) for validation of teleophthalmology referrals. The HERMES study assesses the clinical- and cost-effectiveness of teleophthalmology referrals compared to standard care, in reducing unnecessary hospital visits and misclassified 'urgent' referrals through a cluster RCT (cRCT).

Additionally, there have been significant developments in applying machine learning, specifically deep learning, in medical imaging for disease diagnosis. The HERMES study included a prospective observational study, reviewing artificial intelligence decision support systems (AI DSS), specifically the Moorfields-Google DeepMind model ('Octane') (Google DeepMind, DeepMind Technologies Limited, London), compared to human assessors in terms of its accuracy for retinal diagnosis and referral decisions. Its use as an AI DSS was tested for its generalisability in a diverse clinical care environment.

Objectives

To assess the effectiveness and cost-effectiveness of a digital referral pathway between community optometry and HES for referral of retinal disease enabled by a teleophthalmology platform. To measure the diagnostic and referral accuracy of the Octane AI DSS in the context of referral pathways between community optometry and HES.

Methods

Design

A multicentre, superiority cRCT to assess the clinical- and cost-effectiveness of a teleophthalmology referral pathway for retinal disease. A prospective, observational diagnostic accuracy (validation) study to measure the performance of Octane AI DSS for diagnostic and referral support. A human-computer interaction (HCI) assessment via a theoretically informed, qualitative study to explore participants' and healthcare professionals' perspectives on teleophthalmology and AI DSS. A small-scale exploratory post-implementation observational study of real-life teleophthalmology.

Settings

Community optometry practices with an OCT device and HES based in the UK.

Participants

Adults (≥ 18 years) attending an eye examination with a macular OCT scan at the participating optometry practice were recruited if there was a suspicion of retinal disease in the opinion of CO. Conditions included neovascular age-related macular degeneration (nAMD), dry age-related macular degeneration, diabetic retinopathy, macular oedema, central serous chorioretinopathy, vitreoretinal interface abnormalities, genetic eye disease and any other retinal condition not requiring an emergency referral. Participants were required to give consent. Individuals with known

retinal comorbidities in either eye triggering a referral and cases where acquisition of a good-quality OCT scan was not possible were excluded.

Interventions

Twenty-six community optometry practices (clusters) in the catchment areas of four HES sites in the UK were randomised 1 : 1 to standard care or teleophthalmology for the cRCT. Practices randomised to the control arm continued to refer participants with suspicion of retinal disease to HES using their standard method. Upon receiving informed written consent, COs uploaded their clinical findings, imaging (OCT scans), diagnosis and referral decision to the teleophthalmology referral platform.

Practices randomised to the intervention arm referred participants with suspicion of retinal disease to HES using the teleophthalmology referral platform. Human experts based at the corresponding HES reviewed every case remotely. A referral decision was provided to participants and COs within 48 hours. A hospital appointment was arranged if required. The efficiency of teleophthalmology was assessed by comparing CO and teleophthalmology recommendations against a reference standard (RS).

For the AI study, a subset of OCT scans was processed by the Octane AI DSS. The Octane AI only used the OCT scans as input without any other clinical information. Its diagnosis and referral recommendations were compared against an independent RS.

Additionally, a model-based economic evaluation was conducted, including a model-based cost-effectiveness analysis (CEA), a discrete choice experiment (DCE), and by using the results of both CEA and DCE for a cost-benefit analysis (CBA).

A HCI evaluation was conducted; this involved a combination of situated observations and semistructured interviews with healthcare professionals and participants to investigate their perspectives on teleophthalmology models of care and AI DSS.

Outcome measures

The cRCT primary outcome was the proportion of false-positive referrals (unnecessary HES visits or incorrect referral urgency) in each arm in the overall enrolled and the referred participants against the RS. Secondary outcomes included the proportion of wrong diagnosis, wrong referral urgency, false-negative referrals, uncommon referrals (rare disease) which could be safely triaged, the time from referral to consultation and to treatment and an economic evaluation to assess the costs and benefits from the NHS perspective of the teleophthalmology referral triaging pathway.

The primary outcome for the observational study was the diagnostic accuracy of the referral decision made by AI DSS when compared with the RS. Secondary measures included the diagnostic accuracy of retinal disease and referral urgency, the proportion of false-positive referrals, wrong diagnosis, the wrong referral urgency in each arm if human assessors were replaced by AI DSS, the time required by the AI DSS to process an OCT scan and its average time of end-to-end referral recommendation.

A HCI analysis using qualitative methods identified factors that facilitate the successful implementation of a digital referral platform to ensure acceptability and acceptance.

Statistical analysis

A hierarchical two-level mixed-effects model was used to calculate the required sample size. Twenty-six clusters split between arms in a 1 : 1 ratio needed to recruit 306 participants, with an average of 10 participants per cluster, to achieve 89.27% power to detect a difference in the proportion of false-positive referrals of 30%. To calculate estimates of diagnostic accuracy for the AI study, a higher total sample size of 370 participants (accounting for the drop-out rate), with 351 participants being correctly diagnosed, would produce estimates of diagnostic accuracy with a two-sided 95% confidence interval (CI).

For the cRCT primary outcome, the difference in the proportion of false-positive referrals between study arms was calculated with 95% CI, calculated using the exact binomial method adjusting for cluster. A superiority margin of 30% was used. A regression model estimating the difference in odds of false-positive referral between study arms as an odds ratio (control arm/intervention arm) was also measured. As differing referral urgency levels are of interest, false-positive urgent referrals (clinician urgently referring the patient when RS deemed an urgent referral was unnecessary) were examined using the same methods.

Diagnostic accuracy for the referral decision is reported using the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), where exact binomial CIs were provided. The mean time from referral to consultation and to treatment was generated with 95% CIs calculated using the exact method for a Poisson distribution. Cox model proportional hazards regression adjusting for clustering was used to compare the time from referral to consultation between arms. A sensitivity analysis was performed following the exclusion of cases delayed by external factors from the intervention arm. The remaining secondary outcomes were examined using the same methods for the primary outcome. An additional exploratory analysis focusing on the detection and referral of nAMD was conducted.

For the AI study primary outcome, sensitivity, specificity, PPV and NPV are reported, where exact binomial CIs are provided with the exact binomial method. Summary statistics for the diagnostic accuracy of the referral decision dichotomised as urgent referral versus non-urgent and of referral urgency (urgent vs. routine referral) were computed as the primary outcome methods. Proportions for all human-assessor-related outcomes were analysed with 95% CIs calculated using the exact binomial method. The technical infrastructure details were reported in terms of time per OCT scan.

Results

One hundred and thirty-six participants were recruited to the control arm, and 158 participants were recruited to the intervention arm. Asynchronous teleophthalmology with clinician triage significantly reduced the proportion of false-positive urgent referrals by 59.5% as compared to standard care in referred participants. It lowered the proportion of false-positive referrals (for urgent/routine referrals) and of false-negative referrals. In terms of wrong referral urgency (reported as means with CIs), the standard pathway had 25.7% (14.1% to 37.3%) more incorrect referral decisions than the teleophthalmology pathway, a significant difference at the 5% level.

Both arms showed high sensitivity for detecting the need for a referral, however, specificity was significantly lower in the control arm [41.2% (18.4% to 67.1%)] versus the intervention arm [93.9% (79.8% to 99.3%)]. For urgent referrals only, there was greater accuracy in the intervention arm [sensitivity: 96.3% (81.0% to 99.9%), specificity: 99.2% (95.8% to 100%)] versus the control arm [sensitivity: 73.7% (48.8% to 90.9%), specificity: 79.5% (71.0% to 86.4%)]. There was also a significant increase in the proportion of wrong diagnoses in the standard pathway versus the teleophthalmology pathway [right eye: 19.7% (6.1% to 33.3%); left eye: 18.8% (4.3% to 33.2%)].

Focusing on nAMD, the control arm had 54.2% (30.3% to 78.0%) more false-positive referrals than the intervention arm. Additionally, the intervention arm had greater sensitivity [90.9% (70.8% to 98.9%)] and specificity [100% (97.3% to 100%)] when making a nAMD-related referral than the control arm sensitivity [57.9% (33.5% to 79.7%)] and specificity [88.9% (81.7% to 93.9%)]. The standard pathway also led to more false-positive nAMD diagnoses [53.6% (33.8% to 72.5%)] versus the teleophthalmology pathway [0, (0%)].

Following a sensitivity analysis, the mean time to consultation was significantly lower in the intervention arm [53 (51 to 55) days] versus the control arm [89 (87 to 91) days] ($p = 0.039$). The mean time to treatment was lower in the intervention arm [55 (52 to 57) days] versus the control arm [90 (87 to 94)] ($p = 0.151$). Assessment of safe triage of rare diseases showed each arm referred all suspected cases to HES.

For the AI study, 204 out of 396 participants from 17 of the 29 participating CO practices were included, where the majority of participants were from 14 sites, 201 of the 204 participants (98.5%). Two OCT device manufacturers found

in CO practices were supported, of which 30% of images were not suitable for processing by Octane due to image size or format.

When compared to human assessors (clinical RS), the sensitivity of the Octane AI model was 96.4% (92.4% to 98.7%) and the specificity was 20.0% (8.4% to 36.9%). The AI model was less accurate when deciding referral urgency for required referrals (routine vs. urgent), with wrong referral urgency by AI of 14.8% (9.0% to 22.3%) and by human assessors 2.8% (0.6% to 7.8%). Of note, human assessors considered information from OCT scans, clinical history and patient preferences when making referral decisions, whereas the AI model considered information solely from OCT scans. When recommending urgent referrals, the Octane model is less accurate than hospital-based experts in the teleophthalmology arm, with wrong referral urgency by AI of 25% (5.5% to 57.2%) and by hospital-based experts of 0%, yet the overall absolute numbers of wrong urgent referrals are small; the model is of a similar accuracy as community optometrists for urgent referrals, with wrong referral urgency by AI of 66.7% (34.9% to 90.1%) and by CO of 62.5% (35.4% to 84.8%).

In a post-hoc analysis, the Octane model showed good referral accuracy when applying its own preset referral rules (rule-based RS) for recommending a referral versus no referral. If the AI model offers a referral to a patient, there is a 100% (98.1% to 100%) probability that a referral is truly needed (PPV). If a referral is not offered by the AI, however, there is a 69.2% (38.6% to 90.9%) probability that a referral is truly not needed (NPV). Absolute numbers of no referrals being small in both pathways, however, PPV is increased and NPV is decreased due to high prevalence of participants requiring a referral.

In a post-hoc analysis, the Octane model shows good referral accuracy against the rule-based RS for recommending urgent referral, with modest reduction in referral efficiency (PPV); that is, if the Octane model offers an urgent referral to a patient, there is a 78.4% (61.8% to 90.2%) probability that an urgent referral is truly needed. For referral urgency, the Octane model chose the wrong referral urgency a comparable number of times to CO [AI: 22.7% (11.5% to 37.8%) vs. CO: 22.7% (11.5% to 37.8%)] and a greater number compared to hospital-based experts [AI: 9.4% (3.5% to 19.3%) vs. HES: 1.6% (0.0% to 8.4%)]. In terms of diagnostic accuracy for the diagnosis of retinal disease, the Octane model exhibits moderate sensitivity [68.1% (62.5% to 73.4%)] when detecting the same diagnosis as the clinical RS, which improves to 82.6% (72.9% to 89.9%) when compared to the rule-based RS.

The post-implementation substudy recruited 17 patients overall. Barriers to implementation of the Manchester Electronic eye Referral System included training gaps, inadequate communication channels among primary care, secondary care and technology suppliers and an insufficient support network.

The CEA of the economic decision model showed the HERMES pathway to have a greater effect at a lower cost than the standard pathway, meaning it was the dominant intervention. The DCE results showed the public had a greater preference for a more effective intervention that would be delivered more quickly. Importantly, they had a greater preference for obtaining a correct diagnosis compared to a reduction in the waiting time. A CBA, based on the DCE results, demonstrated a net benefit for the HERMES pathway compared with the standard pathway of £992 for every patient seen. Deterministic and probabilistic sensitivity analyses were carried out to assess the robustness of the conclusions, and in these analyses the conclusions were found not to change. These results provide strong evidence for the efficiency of the HERMES pathway compared with standard care. These data are significant as previous studies reviewing different referral interventions, including asynchronous teleophthalmology, provide limited information on this key area.

The HCI-related in-depth interview study with participants, optometrists and ophthalmologists showed that they generally attribute value in implementing teleophthalmology through improving efficiency and the ability to provide and receive feedback. The normalisation process theory analysis highlighted the need to consider multiple factors when developing and implementing teleophthalmology platforms, especially if aiming to have it adopted and normalised at a large scale.

Conclusions

Asynchronous teleophthalmology can reduce the number of unnecessary referrals, provide more appropriate referral timescales and is more cost-effective compared to standard care. Its role in reducing unnecessary referrals overall was inconclusive. When comparing the Octane AI system for referral decisions to hospital-based human expert assessors, it was less accurate for making routine and urgent referral decisions and was of similar accuracy to CO for making urgent referral decisions. The AI model presented challenges relating to its generalisability in a real-world evaluation context.

Trial registration

This trial is registered as ISRCTN18106677.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: NIHR127773) and is published in full in *Health Technology Assessment*; Vol. 29, No. 69. See the NIHR Funding and Awards website for further award information.

Health Technology Assessment

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 HTA programme as award number NIHR127773. The contractual start date was in March 2020. The draft manuscript began editorial review in September 2023 and was accepted for publication in March 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.

Copyright © 2025 Sharma *et al.* This work was produced by Sharma *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

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