# Teleophthalmology versus standard of care for community optometry referrals of retinal disease (HERMES): a cluster randomised controlled trial with linked prospective diagnostic accuracy assessment of artificial intelligence support

*Anitta Sharma, Rima Hussain, Annastazia E Learoyd, Angela Aristidou, Taha Soomro, Ann Blandford, John G Lawrenson, Gabriela Grimaldi, Abdel Douiri, Ashleigh Kernohan, Tomos Robinson, Najmeh Moradi, Christiana Dinah, Evangelos Minos, Tariq Aslam, Avinash Manna, Alastair K Denniston, Praveen J Patel, Pearse A Keane, Catey Bunce, Luke Vale, Konstantinos Balaskas*

## Summary

**Background** Telemedicine can support integrated care between community and hospital services for disease management, potentially enhanced by medical artificial intelligence (AI). Common retinal disease exemplifies this opportunity. Optometry-to-hospital referral pathways generate unnecessary visits, exacerbating hospital pressures. Evidence for the effectiveness of teleophthalmology and AI in optimising these pathways is scarce.

**Methods** This cluster randomised controlled trial was done at optometry practices that referred to four hospital sites in the UK and that were randomly assigned (1:1) to standard care or teleophthalmology using random permuted blocks of varying sizes stratified by hospital site. Statisticians were masked to group assignment for the primary analysis. Individuals aged 18 years or older with suspected macular disease and good quality optical coherence tomography (OCT) scans were included; same-day emergency cases were excluded. In the teleophthalmology group, OCT scans were remotely reviewed by hospital clinicians; the standard care group followed standard referral pathways. The primary outcome was false-positive referral rate (unnecessary overall or urgent [ie, <2 weeks] referrals) against an independent reference standard, analysed in the enrolled and referred populations (superiority margin 30%) using a modified intention-to-treat analysis. A parallel prospective, observational, diagnostic accuracy (validation) study evaluated automated referral recommendations by the Moorfields-DeepMind-AI model, assessing sensitivity and specificity. This trial is registered with ISRCTN.com (ISRCTN18106677) and is closed.

**Findings** Between Jan 26, 2021, and Dec 14, 2022, 71 optometry sites were assessed for eligibility and 26 recruited and randomly assigned to standard care (13 sites) or teleophthalmology (13 sites). Between July 5, 2021, and March 31, 2023, 304 participants were recruited, of whom 294 were included in the analysis (136 in the standard care group and 158 in the teleophthalmology group; 127 [43%] of 294 participants were male, and 167 [57%] were female). Among all enrolled participants, false-positive referrals occurred in ten (7%) of 136 participants in the standard care group versus two (1%) of 158 in the teleophthalmology group (absolute difference 6% [95% CI –5 to 17]; odds ratio [OR] 6·16 [95% CI 1·28 to 58·80]; p=0·018); urgent false-positive referrals occurred in 24 (18%) of 136 versus one (1%) of 158 (17% [11 to 24]; OR 33·37 [5·28 to 1392·05]; p=0·0004). In referred participants only, false-positive referrals occurred in ten (8%) of 125 versus two (2%) of 124 (6% [–5 to 18]; OR 5·27 [1·09 to 50·52]; p=0·035); urgent false-positive referrals occurred in 24 (63%) of 38 versus one (4%) of 27 (59% [41 to 78]; OR 42·00 [5·69–1901·06]; p=0·0001). The superiority margin of a 30% reduction in false-positive referrals was met only for urgent referred cases. Moorfields-DeepMind-AI could process images from 204 (52%) of 396 participants with sensitivity and specificity of 96% (95% CI 92–99) and 20% (8–37), respectively, for all referrals, and 74% (54–89) and 90% (85–94), respectively, for urgent referrals. No serious adverse events occurred.

**Interpretation** Teleophthalmology significantly reduced unnecessary urgent hospital referrals, the main source of capacity pressure in retinal care, demonstrating superiority among referred participants and supporting timely, safe care. Its impact on overall unnecessary referrals was inconclusive due to the low number of non-referred participants. When evaluable (in ∼50% participants), Moorfields-DeepMind-AI recommended more unnecessary referrals overall than clinically indicated. Simulating human experts, AI performed worse than hospital specialists in the teleophthalmology pathway, and similarly to community optometrists in standard care, probably reflecting differences between clinician judgement and fixed AI rules.

**Funding** National Institute for Health and Care Research.



*Lancet Prim Care* 2025; 1: 100041

Published Online October 21, 2025
https://doi.org/10.1016/j.lanprc.2025.100041

**Moorfields Ophthalmic Reading Centre and Artificial Intelligence Lab, Moorfields Eye Hospital NHS Foundation Trust, London, UK** (A Sharma MSc, R Hussain MSc, T Soomro MD, G Grimaldi MD, Prof K Balaskas MD); **Optometry, Moorfields Eye Hospital NHS Foundation Trust, London, UK** (A Sharma); **NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK** (A Sharma, R Hussain, G Grimaldi, P J Patel MD[Res], Prof P A Keane MD, Prof K Balaskas); **Institute of Ophthalmology, University College London, London, UK** (A Sharma, Prof P A Keane, Prof K Balaskas); **Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK** (C Bunce DSc); **UCL Interaction Centre, University College London, London, UK** (A Blandford PhD); **School of Health and Medical Sciences, City St George's, University of London, London, UK** (Prof J G Lawrenson PhD); **Medical Retina Service, Moorfields Eye Hospital NHS Foundation Trust, London, UK** (G Grimaldi, P J Patel, Prof P A Keane, Prof K Balaskas); **School of Population Health and Environmental Sciences, King's College London, London, UK** (A E Learoyd PhD, Prof A Douiri PhD); **Health Economics Group, Population Health Sciences Institute, Newcastle University,**

**Newcastle upon Tyne, Tyne and Wear, UK** (A Kernohan PhD, T Robinson PhD, N Moradi PhD, Prof L Vale PhD)**; Global Health Economics Centre, London School of Hygiene & Tropical Medicine, London, UK** (Prof L Vale)**; Ophthalmology, London North West University Healthcare NHS Trust, London, UK** (C Dinah MRes)**; North West Anglia NHS Foundation Trust, Peterborough, Cambridgeshire, UK** (E Minos MD)**; University of Manchester, Manchester, UK** (Prof T Aslam PhD)**; NIHR Birmingham Biomedical Research Centre, UK** (A Manna FRCOphth, Prof A K Denniston PhD)**; School of Management, University College London, London, UK** (A Aristidou PhD)**; Stanford Digital Economy Lab, Stanford University, Stanford, CA, USA** (A Aristidou)

Correspondence to: Prof Konstantinos Balaskas, Institute of Ophthalmology, University College London, London EC1V 9EL, UK **konstantinos.balaskas@gmail.com**

## Introduction

Hospital eye services are under increasing strain from the ageing population and rising prevalence of chronic, treatment-intensive retinal diseases.[1] The number of new cases of late age-related macular degeneration (AMD) is projected to increase globally from 5·24 million in 2020 to 6·41 million by 2050.[2] Yet, the expansion of the ophthalmology workforce continues to lag behind population ageing.[3] These pressures are compounded by inefficiencies in the referral process from community-based optometrists to hospital care, delaying access to treatment and worsening outcomes.[4]

Community optometry practices are the primary providers of eye care in the UK and many other countries, including Germany, Denmark, the USA, and Canada,[5,6] and serve as the main source of hospital referrals.[7] Increasingly, these practices are equipped with optical coherence tomography (OCT) devices, which are non-invasive imaging tools that rapidly generate high-resolution 3D retinal scans.[8–10] OCT has transformed the diagnosis and management of retinal diseases such as AMD and diabetic retinopathy, but its wider use in primary care also raises the risk of over-referral, particularly for benign findings.[11] Notably, a high rate of unnecessary urgent referrals (ie, <2 weeks) for suspected neovascular AMD has been reported.[11]

Digital health innovations, such as teleophthalmology and artificial intelligence (AI), offer opportunities to improve referral accuracy while maintaining safety. Although cloud computing has enhanced medical data exchange, a recent systematic review identified a scarcity of well powered randomised controlled trials evaluating the clinical effectiveness of teleophthalmology in eye referral pathways.[12,13] In parallel, advances in deep learning have enabled medical AI systems trained on retinal images to reach expert-level diagnostic accuracy.[14] However, most AI validation studies are retrospective and in silico, with performance often declining in prospective, real-world settings.[14]

Moorfields-DeepMind-AI is a diagnostic AI model that analyses OCT scans alone, without clinical input, to

### Research in context

#### Evidence before this study

We searched PubMed, Scopus, Science Direct, and IEEE Explore from database inception to Jan 20, 2020, and Medline Ovid (1946–2020), Embase Ovid (1947–2020), ISRCTN, ClinicalTrials.gov, and the Cochrane database between Feb 1, 2019, and Jan 20, 2020, for published or ongoing trials involving telemedicine and medical artificial intelligence (AI) in ophthalmology and other health applications. English language manuscripts or abstracts were considered. A combination of the following search terms was used: "telemedicine", "teleophthalmology", "telehealth", "community optometry", "optical coherence tomography", "age-related macular degeneration", "medical Artificial Intelligence", "computer-assisted clinical decision support", "AI validation", and "randomised controlled trial". We identified more than 1500 articles, of which 125 were deemed relevant to our study. 12 telemedicine randomised controlled trials (RCTs) were identified but none involved teleophthalmology. Findings from case series suggest that teleophthalmology might help reduce unnecessary referrals. Of 86 AI RCTs found, none were related to the use of AI in ophthalmology. A concern regarding publication bias in favour of positive results in AI RCTs has been documented. In ophthalmology, 36 AI systems with regulatory approval were identified, of which seven involved optical coherence tomography (OCT) scans. Ten prospective observational or interventional validation studies and one RCT were found, all involving colour fundus images in diabetic retinopathy. All identified OCT AI validation studies were retrospective.

#### Added value of this study

To our knowledge, this is the first multisite RCT to assess whether teleophthalmology could reduce unnecessary referrals and downgrade unnecessary urgent referrals for retinal disease from primary care (community optometry) to hospital eye services compared with standard care. Our study shows that integrating telemedical technologies with advanced ophthalmic image viewing functionalities—specifically 3D-imaging with OCT—could reduce unnecessary hospital visits while maintaining safety standards of hospital-based care, as evidenced by similar low observed rates of missed required referrals in both groups. This approach could liberate specialist capacity, increase availability for early treatment of patients at risk of preventable vision loss, and reduce referral times to hospital review and treatment, thus improving patients' care experience and supporting more appropriate use of services. HERMES adds further value through the parallel observational, real-life diagnostic accuracy study of a medical AI system for diagnostic and referral support (ie, Moorfields-DeepMind-AI). To our knowledge, this is the first prospective observational validation study of an AI clinical support system for OCT scans. Moorfields-DeepMind-AI processing was feasible for only specific, commonly used OCT devices. When processing was possible, the Moorfields-DeepMind-AI generated more unnecessary referrals overall than community optometrists and hospital clinicians and generated a similar number of unnecessary urgent referrals as community optometrists.

#### Implications of all the available evidence

If widely implemented, teleophthalmology has the potential to reduce unnecessary urgent referrals, which could improve patient care experience and issues around hospital capacity. For medical AI, our findings illustrate the value of prospective, real-life diagnostic accuracy and validation studies to guide model refinements for replicating performance observed in retrospective validation studies and accelerate meaningful clinical use.

generate retinal disease diagnoses and referral recommendations.[15] This model was developed using Topcon OCT scans (Topcon 3D-OCT 2000, Topcon, Tokyo, Japan) and a smaller set from Heidelberg Spectralis (Heidelberg Engineering, Heidelberg, Germany). Although both scans are common in hospital settings, community practices use a broader range of OCT devices. The model's retrospective internal validation reported high accuracy (area under the curve [AUC] 0·96), which is similar to that of leading retinal specialists.[16] However, unlike clinicians, AI does not consider symptoms, history, or visual acuity. Moorfields-DeepMind-AI remains a research use-only model.[17] More recently, other AI models have been licensed for patient use (eg, the RetinSight model obtained a CE mark under MDR 2017/745 in May 2022), supporting wider device compatibility.[18] Although developed for different purposes, comparative benchmarking of such models would inform their clinical use in detecting, referring, and managing chronic retinal diseases such as AMD.

The HERMES study aimed to evaluate two innovations in parallel: remote triage of optometry referrals via a custom-built teleophthalmology platform supporting full 3D OCT scan transfer and review by hospital-based specialists, and prospective, real-world validation of Moorfields-DeepMind-AI for referral recommendation accuracy in retinal disease.

## Methods

### Study design

We conducted a cluster randomised controlled trial involving community optometry practices (clusters; appendix pp 22–23) that referred to one of four hospital sites in the UK: Moorfields Eye Hospital, London; University Hospitals Birmingham, Birmingham; Central Middlesex Hospital, London; and Hinchingbrooke Hospital, Huntingdon, and Peterborough City Hospital, Peterborough. Hinchingbrooke Hospital and Peterborough City Hospital constituted a single referral centre with extended geographical reach. In a parallel, observational medical AI prospective diagnostic accuracy (validation) study, OCT scans from recruited participants (including, but not limited to, those who were enrolled in the trial) were processed with the Moorfields-DeepMind-AI system. Up to an additional five clusters could be included in the study to recruit participants contributing data for the AI study only. An exploratory post-implementation substudy of teleophthalmology was conducted across three community optometry practices in Manchester, UK (appendix pp 1–2).

The study protocol has been published online.[19] Protocol amendments are listed in the appendix (pp 3–4). Amendments to the protocol were mainly driven by the COVID-19 pandemic. These included the optional increase of the number of clusters from 24 to 26, if acceleration of recruitment was required, with an average of ten participants recruited per cluster, leading to a recruitment target of 306 participants, accounting for 15% loss to follow-up, while retaining statistical power. The Manchester University National Health Service (NHS) Trust was removed as a recruitment site due to region-wide teleophthalmology commissioning during the pandemic (appendix pp 3–4).

Patient champions, members of the public, and eye charities were involved in study design, conduct, and reporting through forum meetings and Trial Steering Committee representation. Ethics approval was granted by the London–Bromley Research Ethics Committee (REC 20/LO/1299).

This trial is registered with ISRCTN.com (ISRCTN18106677) and is closed.

### Community optometry practices and participants

Participating community optometry practices were required to have an OCT device and refer within the catchment area of one of the four hospital sites. To reflect real-world variability, no minimum experience or post-graduate credentials were required for community optometrists; supervised pre-registration optometrists were eligible. Optometrists verbally consented and uploaded their clinical assessments to the study database. Each hospital site had an average of two clinicians (ophthalmologists or specialist optometrists) reviewing referrals in the teleophthalmology group, with second opinions sought through local escalation policies as needed. All hospital site clinicians had substantive medical retina expertise and routine involvement in retinal care.

All community and hospital clinicians underwent study-specific training, including site initiation and refresher meetings led by the chief investigator (KB), with appropriate Good Clinical Practice training. All provided care under their professional codes.

See Online for appendix

Patients aged 18 years or older undergoing an eye exam including a macular OCT at participating community practices were eligible if the optometrist suspected retinal disease (new or pre-existing patients). Included macular-affecting or OCT-detectable conditions were dry AMD, neovascular AMD, macular oedema, central serous chorioretinopathy, macular holes, epiretinal membranes, genetic retinal disease, and any non-emergency retinal condition. Individuals with poor-quality OCT scans preventing diagnosis (eg, due to media opacity or fixation issues) or those receiving hospital-based retinal care were excluded. Patients previously discharged from hospital care but re-presenting to community optometry with new visual symptoms were eligible.

Each community site could recruit up to 16 participants to the cluster randomised controlled trial. Any additional participants were included only in the AI study (appendix pp 22–23). The same eligibility criteria were applied to the AI study, with the additional requirement of Moorfields-DeepMind-AI-compatible scans (Topcon or Heidelberg Spectralis).

Data on sex were self-reported; options were male or female. Ethnicity data were not collected due to sensitivities raised by optometry councils. Recruitment efforts were guided by National Institutes for Health and Care Research equality, diversity, and inclusion advisors (appendix p 13). Optometrists recruited eligible participants after providing study information and obtaining written consent in person

during the community optometry visit or written consent by telephone following a specified and REC-approved standard operating procedure with a witness cosignatory.

### Randomisation and masking

Community optometry practices were randomly assigned (1:1) to standard care or teleophthalmology. A senior data manager generated the allocation list using random permuted blocks of varying size, stratified by hospital site, with simple randomisation (Stata MP, version 16). The list was provided to the trial project manager (RH), who assigned recruited practices consecutively. Allocation was concealed at the cluster level. Statisticians were masked to allocation until the primary analysis was completed.

### Procedures

Community optometry practices referred participants per random assignment. Optometrists conducted routine examinations and imaging, including pupil dilation if clinically required. In the standard care group, referrals were submitted to hospitals through regional hospital eye services pathways without imaging (appendix p 1). All optometrists uploaded clinical data and full-volume 3D OCT scans to the HERMES platform (appendix pp 5–12) with eye-level diagnoses and referral decisions, including decisions not to refer. These decisions were implemented and used to determine each patient's clinical reference standard. Routine retinal referrals generally required 2–18 weeks; patients with neovascular AMD required less than 2 weeks.[20] Optometrists noted if they planned to monitor non-referred cases. In the teleophthalmology group, optometrists uploaded clinical data and OCT scans to the HERMES platform, which were reviewed remotely by hospital clinicians within 48 h. The hospital's diagnoses and referral decisions were implemented with prompt appointment scheduling by site coordinators. Optometrists and participants were notified directly. Optometrists' own suggested diagnoses and referrals were also recorded but not acted on. Clinical data from referred participants' first hospital visit were collected when available. Participants who were not referred, or whose optometrist's referral suggestion differed from teleophthalmology decisions, received detailed safety netting instructions on symptoms indicating possible disease activity, including contact details for the local eye emergency hospital.

All clinical and imaging data, including previous hospital records and any additional imaging, were reviewed by two senior expert graders at the Moorfields Reading Centre for all study participants and were adjudicated by the Reading Centre director (KB), in cases of discordant graders' opinions, to provide the clinical reference standard for correct referral decision and retinal diagnosis. This process was isolated from other study-related activities, particularly clinical examination and referral recommendations made by hospital clinicians for participants in the teleophthalmology group referred to Moorfields as a study site. Data from the first in-person hospital visit, when available, were considered for the reference standard. Referral decisions by clinicians in each case were evaluated against the clinical reference standard. Clinicians doing teleophthalmology reviews and referral decisions did not contribute to the decision of the reference standard.

Compatible OCT scans (good quality Topcon scans of specific size and density), uploaded in a full-volume open-source format, were processed as part of the AI study.[16] The Moorfields-DeepMind-AI model provided an eye-level diagnosis and a participant-level referral recommendation. Specifically, OCT scans from participants recruited by community optometrists in the AI study were uploaded to the dedicated HERMES platform. OCT scans were then downloaded and processed by the Moorfields-DeepMind-AI model within the Moorfields Information Technology environment using local Graphic Processing Unit infrastructure. For each processed OCT, it provided a referral recommendation and retinal diagnosis. No other clinical or demographic participant data were input to the AI model. Diagnoses included choroidal neovascularisation, dry AMD (drusen and macular atrophy), central serous chorioretinopathy, macular oedema, vitreoretinal interface abnormalities, or normal. Choroidal neovascularisation indicates diagnosis of neovascular AMD in nearly all cases. Referral recommendations were urgent (for choroidal neovascularisation), routine, or no referral (for cases without pathology).

### Outcomes

The primary outcome for the cluster randomised controlled trial was the proportion of false-positive referrals per group at the participant level, in enrolled participants overall, and in referred participants, compared with the clinical reference standard. The statistical analysis plan prespecified that this outcome should also be reported by urgency. False-positive rates are reported for all referrals (ie, when not needed) and urgent referrals (ie, when only routine or no referral was needed). Secondary outcomes included the proportion of incorrect diagnoses (eye level); incorrect referral urgency (participant level); false-negative referrals (participant level); sensitivity and specificity (participant level); identification and the number of uncommon referrals (for rare disease) safely triaged via teleophthalmology (participant level); and referral-to-consultation and referral-to-treatment times (routine and urgent). Within-trial cost-effectiveness and cost-consequences, as well as modelled economic evaluations of referral strategies, will be reported separately.

The primary outcome for the AI study was diagnostic accuracy (ie, sensitivity and specificity) of AI referral recommendations (referral *vs* no referral) against the clinical reference standard in the full AI study cohort. A prespecified sensitivity analysis assessed sensitivity and specificity of AI urgent referral recommendations. Secondary outcomes were AI sensitivity and specificity of referral urgency (participant level); AI diagnostic accuracy for retinal diseases (eye level); proportion of false-positive referrals (participant level), incorrect referral urgency

(participant level), and incorrect diagnoses (eye level) if AI replaced human assessors (in the cluster randomised controlled trial); AI uptime and inference speed in seconds (OCT); time from referral to AI output in minutes (participant level); and cost consequences and net benefit of AI-enabled referral pathway (not reported).

Given the low-risk nature of the intervention, safety concerns were minimal. All adverse events were recorded per group by study teams at each hospital site. Safety was not a prespecified endpoint. Standard NHS harm risk assessment and adverse event reporting processes were implemented.

### Statistical analysis

The sample size was calculated using nQuery version 8.3.10 with a two-level hierarchical mixed effects model. 24 clusters (optionally 26) needed to recruit 12 patients each (ten if 26 clusters) to provide 89% power to detect a 30% absolute difference in false-positive referrals, assuming an intracluster correlation of 0·15 on the basis of previous ophthalmology data.[21] The same effect size assumptions applied to urgent false-positive referrals. For the AI study, 370 recruited participants (with 351 correctly diagnosed) would yield 95% CIs of 0·046 (appendix pp 15–19).

Analyses were conducted using Stata MP, version 17. Outcomes relating to referrals were analysed at the participant level and diagnosis-related outcomes were analysed at the eye level. For both components, key parameters were reported with 95% CIs. Continuous data were summarised using means (SDs) if normally distributed or medians (IQRs) if non-normally distributed. Categorical data were presented as proportions.

Analyses of the cluster randomised controlled trial used two-sided hypothesis testing with 95% CIs. The primary outcome (false-positive referrals) was assessed using two denominators: all enrolled participants and referred participants only using a modified intention-to-treat approach. Although the statistical analysis plan did not prespecify a preferred denominator, the false-positive rate used in the sample size was derived from a referred-only hospital dataset. Both denominators are valid; results are reported for each. Differences in proportions were estimated with 95% CIs using the exact binomial method, adjusting for clustering via group-specific intraclass correlation coefficients. A superiority margin of 30% was applied.[11] An exact logistic regression model was used to estimate unadjusted odds ratios (ORs) for false-positive referrals between groups. Due to convergence issues, clustering was not adjusted for, diverging from the statistical analysis plan (appendix pp 15–16), to improve accuracy given low event rates. False-positive urgent referrals were analysed similarly. In a post-hoc analysis, we excluded practices that were outliers with respect to their disproportionate contribution to false-positive referrals, as observed during data analysis.

Secondary outcomes involving proportions used the same methods as for the primary analysis. Time to consultation and treatment outcomes (median time to event [IQR]) were compared between groups using Cox regression models, following verification of the proportional hazards assumption via log-log plots and Kaplan–Meier curves. Mean (95% CI) and median (IQR) times to consultation and treatment are also reported for all referrals and separately for routine and urgent referrals in the appendix (p 17). A post-hoc analysis of time to event outcomes excluded cases with unrelated administrative delays (eg, late referral notifications or booking issues). Time to event is reported as mean (95% CI) and hazard ratio.

Diagnostic accuracy metrics for referral decisions were reported with exact binomial 95% CIs (appendix p 17). Post-hoc analyses included accuracy of neovascular AMD (choroidal neovascularisation) diagnosis (eye level) and referral (participant level); stratified referral accuracy (participant level) per practice with corresponding intraclass correlation coefficients (all and urgent referrals); and rates of recurrent diagnoses (reactivation or new diagnosis in eyes with known previous retinal disease; eye level) per group.

For the parallel AI study, a higher recruitment target was determined to produce estimates of diagnostic accuracy with narrow 95% CIs. Primary (sensitivity and specificity) and secondary (positive predictive value, negative predictive value, and AUC) outcomes for AI referral accuracy were assessed in all participants enrolled in the AI study against the clinical reference standard. A prespecified sensitivity analysis examined AI referral accuracy for urgent referral recommendations, due to their clinical importance. A post-hoc analysis compared AI performance against the rule-based reference standard, where Moorfields-DeepMind-AI model's preset referral rules (which often differed from clinical practice) were applied, to better assess the performance of the intrinsic model (appendix p 14). Secondary outcomes from the simulated cluster randomised controlled trial (if AI replaced community optometrists in the standard care group and hospital-based specialists in the teleophthalmology group) were reported with 95% exact binomial CIs. Clinical significance of between-group differences can be interpreted through 95% CIs and their overlap. Technical performance metrics were reported as discrete values. Additional post-hoc analyses included neovascular AMD (choroidal neovascularisation)-specific outcomes and AI performance using Topcon OCT scans only.

The study was overseen by an independent data monitoring committee.

### Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, interpretation, or writing of the report.

## Results

Between Jan 26, 2021, and Dec 14, 2022, 71 optometry sites were assessed for eligibility, and 33 sites were randomly assigned (18 to the standard care group and 15 to the teleophthalmology group; figure). Seven sites withdrew, resulting in 26 clusters included in the study, recruiting

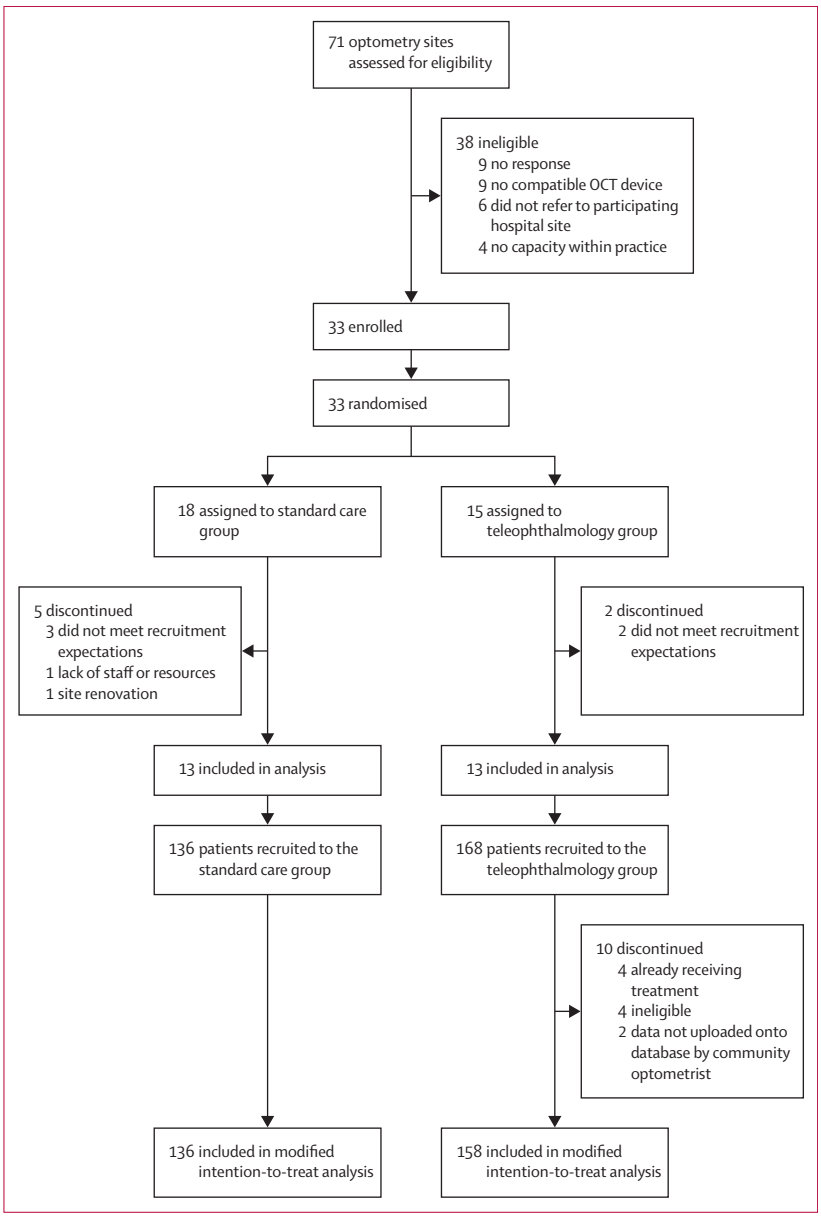


***Figure:*** **Trial profile**
OCT=optical coherence tomography.

participants for the cluster randomised controlled trial. Three additional sites were included for recruitment of participants into the AI study only and are not among the seven sites that withdrew. Between July 5, 2021, and March 31, 2023, 304 participants were recruited (136 in the standard care group and 168 in the teleophthalmology group). No participants discontinued in the standard care group; however, ten discontinued in the teleophthalmology group, such that 136 participants in the standard care group and 158 in the teleophthalmology group were included in the modified intention-to-treat analyses. The data cutoff was on Sept 6, 2023.

Participant characteristics are presented in table 1. Overall, 127 (43%) of 294 participants were male, and 167 (57%) were female. At the participant level, median age was 73 years (IQR 61·0–78·0). Groups were well matched, except for a history of diabetes (21 [15%] of 136 *vs* 39 [25%] of 158) and diabetic macular oedema (one [1%] *vs* nine [6%]) in the in the standard care group and teleophthalmology group, respectively (additional participant-level and cluster-level baseline data are provided in the appendix [pp 19–21]). No data were missing.

Of all enrolled participants, false-positive referrals occurred in ten (7%) of 136 in the standard care group versus two (1%) of 158 in the teleophthalmology group (absolute difference 6% [95% CI –5 to 17]; OR 6·16 [95% CI 1·28 to 58·80]; p=0·018; table 2). The event rate for decisions of no referral was unexpectedly low in both groups (11 [8%] of 136 participants in the standard care group and 34 [22%] of 158 participants in the teleophthalmology group), thus, introducing some potential statistical uncertainty. Urgent false-positive referrals were statistically significantly higher with standard care (24 [18%] of 136 participants) than with teleophthalmology (one [1%] of 158; absolute difference 17% [95% CI 11 to 24]; OR 33·37 [95% CI 5·28–1392·05]; p=0·0004). Among referred participants only, false-positive referrals occurred in ten (8%) of 125 participants in the standard care group and two (2%) of 124 in the teleophthalmology group (absolute difference 6% [95% CI –5 to 18; OR 5·27 [95% CI 1·09 to 50·52]; p=0·035), whereas urgent false-positive referrals occurred in 24 (63%) of 38 versus one (4%) of 27 (59% [41 to 78]; 42·00 [5·69 to 1901·06]; p=0·0002; table 2). The prespecified superiority margin of 30% for the primary outcome was met for false-positive urgent referrals among referred participants.

Among all enrolled participants, false-negative referrals were similarly low between the groups: seven cases of missed (false-negative) referrals overall and six missed urgent referrals, with 1% (95% CI –3 to 5) more false-negative referrals and 3% (–3 to 9) more false-negative urgent referrals in the standard care group than in the teleophthalmology group (appendix pp 23–24).

Diagnostic accuracy for correctly identifying necessary referrals was higher in the teleophthalmology group than in the standard care group. For urgent referrals, accuracy was also higher with teleophthalmology than with standard care (table 3).

Incorrect retinal disease diagnoses were significantly lower in the teleophthalmology group than in the standard care group: 18% (26 of 146 participants with available data) for right eyes and 17% (25 of 146) for left eyes compared with 38% (42 of 112) for right eyes and 36% (42 of 117) for left eyes, respectively (appendix pp 24–25).

Hospital consultations were offered to 222 (76%) of 294 participants, and 188 (85%) attended. Median time to consultation (all referrals combined, routine, and urgent) was 70 days (IQR 30–127) in the standard care group and 69 days (24–107) in the teleophthalmology group (p=0·84).

Median time to treatment was 60 days (26–131) and 41 days (22–81) in the same groups, respectively (p=0·67; appendix p 26). In a post-hoc analysis, after excluding external delays, mean time to consultation significantly reduced for overall and urgent referrals in the teleophthalmology group compared with the standard care group. Mean time to treatment also reduced in the teleophthalmology group compared with the standard care group, although not significantly (appendix pp 26–28).

Uncommon (rare disease) diagnoses occurred in eight (3%) of 294 participants: six of 136 in the standard care group and two of 158 in the teleophthalmology group (both of whom were safely triaged as requiring routine referral; appendix pp 28–29).

A post-hoc analysis on neovascular AMD referrals found 10% (95% CI 1–18) more false-positive urgent referrals in the standard care group than in the teleophthalmology group, with corresponding diagnostic accuracy results (appendix pp 29–31). Referral accuracy varied across community optometry practices. One standard care group practice notably produced many false-positive referrals (seven [44%] of 16) yet was not an isolated outlier among all study-participating practices. In the teleophthalmology group, five optometry practices would have made multiple false-positive referrals based on the optometrist's recorded recommendation (13–44% of referrals) but hospital-based expert review via teleophthalmology averted it (appendix pp 32–35). A post-hoc analysis excluding the standard care group practice with a disproportionately high false positive rate reduced false-positive referral differences to 1% (95% CI –3 to 5; appendix pp 36–37). Analysis of select secondary outcomes (false-negative referral rate, wrong referral urgency, and sensitivity and specificity for correctly identifying necessary referrals) following exclusion of this standard care group are presented in the appendix (pp 37–39).

Community optometrists selected active monitoring without referral for 11 participants (nine in the teleophthalmology group and two in the standard of care group), correctly in all but one participant in each group (clinical reference standard). Of the eyes referred for neovascular AMD (51 eyes) and diabetic macular oedema (26 eyes), 11 and 17 had a known history of the disease, respectively (appendix pp 40–44).

A total of 396 participants were enrolled in the AI study (including the 294 participants in the cluster randomised controlled trial), of whom 179 (45%) were excluded because their OCT scans were unsuitable for Moorfields-DeepMind-AI processing (eg, non-compatible OCT device, non-compatible OCT scan size or format, other technical limitations of the AI model or AI developer; appendix pp 47–48) and 13 (3%) for other reasons (appendix p 45). Patient characteristics and reasons for non-processed OCT scans are provided in the appendix (pp 47–48). Of five OCT devices found in HERMES optometry practices, Moorfields-DeepMind-AI could process Topcon (Maestro or 3D-OCT-1000/2000) and Nidek (RS-330 or other; Nidek, Japan) scans.

| | Standard care group (n=136) | Teleophthalmology group (n=158) |
|---|---|---|
| **Participant level** | | |
| Age, years | 73·0 (62·0–78·0) | 73·0 (61·0–78·0) |
| Sex | | |
| Male | 63 (46%) | 64 (41%) |
| Female | 73 (54%) | 94 (59%) |
| Smoking history (previous or current) | 9 (7%) | 18 (11%) |
| Medical history of (previous or ongoing) systemic disease | 91 (67%) | 112 (71%) |
| Ocular history of known (previous or ongoing) eye disease | 62 (46%) | 70 (44%) |
| Specified conditions | | |
| Wet AMD | 11 (8%) | 12 (8%) |
| Dry AMD | 25 (18%) | 29 (18%) |
| Retinal vein occlusion | 2 (1%) | 4 (3%) |
| Diabetic macular oedema | 1 (1%) | 9 (6%) |
| Previous eye procedures | 44 (32%) | 48 (30%) |
| Specified procedure | | |
| Cataract surgery | 39 (29%) | 42 (27%) |
| **Cluster level** | | |
| Age, years | 69·6 (68·1–72·8) | 71·8 (69·9–74·0) |
| Sex | | |
| Male | 44·4% (38·5–50·0) | 37·5% (31·3–50·0) |
| Female | 55·6% (50·0–61·5) | 62·5% (50·0–68·8) |
| Smoking history (previous or current) | 0·0 (0·0–11·1) | 6·2% (0·0–12·5) |
| Medical history of (previous or ongoing) systemic disease | 66·7% (55·6–77·8) | 68·8% (56·3–83·3) |
| Ocular history of known (previous or ongoing) eye disease | 45·5% (38·5–50·0) | 40·0% (25·0–62·5) |
| Specified conditions | | |
| Wet AMD | 9·1% (0·0–12·5) | 0·0 (0·0–12·5) |
| Dry AMD | 15·4% (9·1–37·5) | 12·5% (6·2–25·0) |
| Retinal vein occlusion | 0 | 0 |
| Diabetic macular oedema | 0 | 6·2% (0·0–6·2) |
| Previous eye procedures | 36·4% (23·1–41·7) | 37·5% (16·7–43·8) |
| Specified procedure | | |
| Cataract surgery | 30·8% (22·2–40·0) | 25·0% (16·7–40·0) |

Data are median (IQR), n (%), or median (IQR) as a percentage per cluster. AMD=age-related macular degeneration.

*Table 1:* **Participant characteristics in the cluster randomised trial, at the participant and cluster level**

Compared with the clinical reference standard, AI sensitivity and specificity for referral recommendations was 96% (95% CI 92–99; high) and 20% (8–37; low), respectively, and for urgent referrals 74% (54–89; modest) and 90% (85–94; high), respectively. In a post-hoc analysis versus the rule-based reference standard, sensitivity and specificity were high for all referrals, and modest and high, respectively, for urgent referrals (table 4). The positive predictive value and negative predictive value analysis is provided in the appendix (p 48).

Moorfields-DeepMind-AI showed moderate sensitivity (68% [95% CI 63–73]) and specificity (67% [55–78]) for detecting the correct retinal diagnosis (clinical reference standard); both remained modest, despite improvement in specificity, in the post-hoc analyses versus the rule-based reference standard (appendix p 49). The AI diagnostic accuracy of choroidal neovascularisation only (post hoc) is provided in the appendix (p 49). A total of 130 participants were in both the cluster randomised controlled trial and the

| | Standard care group (n=136) | Teleophthalmology group (n=158) | Difference in proportions (95% CI) | Odds ratio (95% CI) | Unadjusted p value for OR |
|---|---|---|---|---|---|
| Referral when not needed | | | | | |
| Proportion of all participants | 10/136 (7%) | 2/158 (1%) | 6% (–5 to 17) | 6·16 (1·28 to 58·80) | 0·018 |
| Intraclass correlation coefficient | 0·49 | 0·0005 | NA | NA | NA |
| Proportion of referred participants | 10/125 (8%) | 2/124 (2%) | 6% (–5 to 18) | 5·27 (1·09 to 50·52) | 0·035 |
| Intraclass correlation coefficient | 0·47 | 0·0003 | NA | NA | NA |
| Urgent referral when routine or no referral needed | | | | | |
| Proportion of all participants | 24/136 (18%) | 1/158 (1%) | 17% (11 to 24) | 33·37 (5·28 to 1392·05) | 0·0003 |
| Intraclass correlation coefficient | 0·0005 | 0·0002 | NA | NA | NA |
| Proportion of referred participants | 24/38 (63%) | 1/27 (4%) | 59% (41 to 78) | 42·00 (5·69 to 1901·06) | 0·0002 |
| Intraclass correlation coefficient | 0·07 | 0·0003 | NA | NA | NA |

NA=not applicable.

***Table 2:*** **Proportion of false-positive referrals per study group**

| | Standard care group (n=136) | Teleophthalmology group (n=158) |
|---|---|---|
| Definition of false positive: referred when not needed | | |
| Sensitivity | 97% (92–99) | 98% (93–100) |
| Specificity | 41% (18–67) | 94% (80–99) |
| Positive predictive value | 92% (86–96) | 98% (94–100) |
| Negative predictive value | 64% (31–89) | 91% (76–98) |
| Area under ROC curve | 0·69 (0·57–0·81) | 0·96 (0·91–1·00) |
| Definition of false positive: urgent referral when not needed | | |
| Sensitivity | 74% (49–91) | 96% (81–100) |
| Specificity | 80% (71–86) | 99% (96–100) |
| Positive predictive value | 37% (22–54) | 96% (81–100) |
| Negative predictive value | 95% (89–98) | 99% (96–100) |
| Area under ROC curve | 0·77 (0·66–0·87) | 0·98 (0·94–1·00) |

Indicators come from a diagnostic test that does not account for clusters. All indicators are presented with 95% CIs in parentheses. ROC=receiver operating characteristic.

***Table 3:*** **Diagnostic accuracy of each group for referral decisions against the clinical-based reference standard**

AI study, of whom 108 were referred to hospital in both groups (90 routine and 18 urgent). In these participants, the false-positive referral rate of Moorfields-DeepMind-AI was higher at 15% (95% CI 9–22) than of all clinicians at 3% (1–8), with non-overlapping CIs (appendix pp 49–51). AI's incorrect referral-urgency was higher than human assessors in the teleophthalmology group at 9% (95% CI 4–19) versus 2% (0–8), respectively, but matched that of human assessors in the standard care group at 23% (12–38; appendix p 52). Regarding incorrect diagnoses of retinal disease, the rate of AI false-positive diagnoses (ie, AI diagnoses different to clinician diagnoses) was similar to clinicians' false-positive diagnoses of 12% (18 of 153 participants) for Moorfields-DeepMind-AI versus 11% (19 of 176 participants) for clinicians, and the rate of AI false-negative diagnoses (AI normal or unclassified disease diagnoses *vs* named disease diagnoses by clinicians) was higher than clinicians' false-negative diagnoses of 65% (56 of 86) versus 43% (26 of 61), respectively, and remained (appendix p 53).

The mean AI processing time was 40 s per OCT and AI-enabled referral triaging was 8 min per participant.

Study-related adverse events are reported in the appendix (p 54).

A post-hoc Topcon-only analysis yielded similar performance for Moorfields-DeepMind AI to the full cohort (appendix pp 54–57). The research team's experience with primary care is provided in the appendix (p 57).

## Discussion

In HERMES, teleophthalmology significantly reduced unnecessary urgent hospital referrals by 17% compared with standard care, which did not meet the prespecified superiority margin. Among referred participants only, teleophthalmology reduced unnecessary urgent referrals by 59%, meeting the prespecified 30% superiority margin. Although unnecessary referrals overall were reduced by 6% in all participants and referred participants, the total number of such cases was small (n=12), limiting interpretability. These findings show the potential of teleophthalmology to reduce unnecessary urgent referrals—particularly for suspected neovascular AMD, a major pressure point for hospital eye services. By contrast, evidence for reducing overall unnecessary referrals was inconclusive.

The observed event rate of false-positive (unnecessary) referrals made by participating community optometry practices overall was lower than anticipated across both groups. Early recruitment patterns suggested optometrists across both groups adopted greater caution with borderline cases, probably due to awareness of scrutiny (a Hawthorne effect), indicating the trial environment promoted more considered referrals.[22] Although beneficial clinically, this behaviour might reflect a temporary behavioural shift. Additionally, post-COVID-19 upskilling initiatives in community optometry probably improved practitioner competence, contributing to enhanced referral accuracy.

The low event rate of unnecessary referrals by community optometrists overall irrespective of urgency level led to few non-referred cases, with two implications. First, it limited

| | Clinical reference standard (n=204) | Rule-based reference standard (n=204)* |
|---|---|---|
| Routine or urgent referral *vs* no referral | | |
| Sensitivity | 96% (92–99) | 98% (95–99) |
| Specificity | 20% (8–37) | 100% (66–100) |
| Positive predictive value | 85% (80–90) | 100% (98–100) |
| Negative predictive value | 54% (25–81) | 69% (39–91) |
| Area under ROC curve | 0·58 (0·51–0·65) | 0·99 (0·98–1·00) |
| Urgent referral *vs* routine or no referral | | |
| Sensitivity | 74% (54–89) | 88% (72–97) |
| Specificity | 90% (85–94) | 95% (91–98) |
| Positive predictive value | 54% (37–71) | 78% (62–90) |
| Negative predictive value | 96% (92–98) | 98% (94–99) |
| Area under ROC curve | 0·82 (0·74–0·91) | 0·92 (0·86–0·98) |

Data are indicators with 95% CIs in parentheses. AI=artificial intelligence. ROC=receiver operating characteristic. *Post-hoc analysis.

***Table 4:*** **Diagnostic accuracy of the AI model for referral decisions against the clinical reference standard and the rule-based reference standard**

statistical power to detect overall referral differences, causing absolute and relative measures (ORs) to diverge; therefore, we prioritised absolute differences in our primary interpretation. Second, for urgent referrals—when optometrists are more cautious due to the high risk of missed diagnoses—the Hawthorne effect is less likely to occur, allowing a more robust assessment of teleophthalmology. In this context, the reduction in urgent referrals was significant and met the superiority threshold. Because urgent referrals for conditions such as neovascular AMD are less affected by behavioural bias, they better reflect the true potential effect of teleophthalmology. Reducing these referrals could meaningfully relieve pressure on overstretched specialist services.

Randomised controlled trials evaluating integrated care models, rather than products or devices, are inherently affected by human factors (eg, practitioner performance)[23] and system-level changes (eg, policy-driven upskilling), which can rapidly shift and undermine baseline assumptions. Despite these challenges, such trials are essential to generate robust, translational evidence for health-care policy and practice.

False-negative referral rates were low across both groups, with no statistically significant between-group differences, indicating similar safety. Previous research has shown community optometrists tend to be cautious when making referral decisions,[11] which was corroborated in HERMES, with no difference in missed diagnoses or urgency misclassification between standard care and teleophthalmology. Standard care produced more incorrect referral decisions (combined false-positives and false-negatives) than teleophthalmology. Over-referral can increase patient anxiety and hospital burden; under-referral risks delayed diagnoses and irreversible vision loss.[7,12]

For retinal diagnosis—particularly neovascular AMD—teleophthalmology showed high diagnostic accuracy (AUC 0·96) versus standard care (AUC 0·73), an important distinction given the severe consequences of missed neovascular AMD diagnosis. Service delays in neovascular AMD have been associated with permanent visual loss.[4]

Reduced diagnostic accuracy in community referrals might reflect variation in training or experience.[23] One prospective study found only 37% of urgent community referrals for neovascular AMD were correctly diagnosed, largely due to challenges interpreting OCT.[24] Previous studies have also reported low feedback rates (13–16%) from hospital eye services to community optometrists.[7,25] In HERMES, the teleophthalmology group received structured referral feedback, which might help to improve diagnostic and referral accuracy.

The HERMES teleophthalmology pathway enabled two-way communication between community optometry and hospital eye services. External administrative delays (eg, capacity constraints and booking issues) were recorded and excluded from time-to-care analyses. Teleophthalmology reduced overall mean time to consultation by 36 days and time to treatment by 35 days. Trial co-ordinators supported streamlined scheduling and direct communication, which probably minimised common delays associated with multi-step standard care pathways. As such, teleophthalmology could offer a route to faster access and improved patient outcomes.

Only approximately half of recruited participants in the AI study had eligible scans for inclusion in the AI study analysis. Community care reflected the real-world diversity in OCT device types and formats. Moorfields-DeepMind-AI performed best with its primary training domain (mainly Topcon), but reduced compatibility with other devices led to a smaller sample size and wider 95% CIs, affecting precision but not statistically significantly affecting diagnostic accuracy.

Non-processing of OCTs within the training domain of Moorfields-DeepMind-AI was largely due to incompatible OCT dimensions, sparse scan volumes, or unsupported file formats. Approximately 40% of Topcon scans could not be processed because of wide-field or sparse acquisition protocols (appendix pp 47–48). Although partly trained on Heidelberg OCTs, the AI could not process 75 participants' scans due to late discovery by the developers of incompatibility in converting proprietary (.e2e) to open-source (.dicom) OCT formats (appendix pp 47–48). Unexpectedly, Nidek OCTs, despite not being part of training data, were processed successfully, probably due to format similarity with Topcon scans. These scans were included in the analysis, although further model finetuning is needed to optimise Nidek performance.

Moorfields-DeepMind-AI showed low specificity relative to the clinical reference standard, reflecting over-referral due to conservative referral thresholds. Although the threshold ensured nearly all necessary referrals were captured, it also led to unnecessary ones. Unlike human clinicians, who integrate imaging with clinical context, the AI uses rigid referral rules. For example, it refers all dry AMD cases (drusen or atrophy), despite UK clinical

guidance advising against routine hospital referrals of early AMD cases. Alignment with clinical practice would require threshold adjustment. The model's reliance on imaging alone might affect precision in real-world use. In the 2018 retrospective internal validation of Moorfields-DeepMind-AI, human experts applied the model's preset referral rules, which diverge from standard clinical practice, leading to reported human-level or superior AI performance.[16] These artificial rules differ from real-world decision making, reinforcing the need for prospective evaluations to assess actual clinical use. Notably, few licensed ophthalmic AI tools have undergone prospective validation and none using OCT, despite its centrality in retinal diagnostics.[26]

Ground truth reliability (the reference standard) depends on expert knowledge. HERMES used reference standards generated by two senior graders with adjudication by a retinal specialist, aligning with best practice recommendations for high-quality reference standards.[27]

In a simulated scenario in which AI replaced human assessors in the cluster randomised controlled trial, Moorfields-DeepMind-AI was less accurate than hospital-based specialists (via teleophthalmology). Compared with standard care community optometrists, Moorfields-DeepMind-AI did not improve referral accuracy. These findings suggest that decision support with AI at the community level would probably not reduce unnecessary hospital referrals.

Diagnostic accuracy for some retinal conditions (eg, vitreoretinal interface abnormalities, macular oedema, and central serous chorioretinopathy) should be interpreted cautiously, as validated probability thresholds were unavailable. However, validated thresholds were recently reported for choroidal neovascularisation (neovascular AMD) and dry AMD (drusen or atrophy), which are both clinically significant for hospital workload.[28]

Ethnicity data were not collected, as community optometrists found this challenging. However, study sites were geographically diverse across England, and recruitment strategies were informed by NIHR Equality, Diversity and Inclusion advisors (appendix p 13). In the cluster randomised controlled trial, behavioural shifts in referral decisions, particularly for mild or non-urgent disease, probably attenuated differences in overall referral rates between groups. The trial evaluated a specific pathway configuration; future research could explore modified pathways, such as direct-to-treatment tele-referral models, especially for urgent conditions like neovascular AMD.

In the AI validation study, performance was limited by device compatibility. Moorfields-DeepMind-AI, in its current form, is not suitable for direct care. Rebuilding the system using modern machine learning libraries and fine-tuning across a wider range of OCT devices is needed for regulatory approval and clinical use. Integrating clinical findings and patient history with image analysis would better reflect real-life decision making. Combining OCT data with structured clinical inputs and natural language processing might enhance AI reliability and use. Of note, Moorfields-DeepMind AI is a research model, and this study is not intended for licensing purposes.

Teleophthalmology was shown to be effective for triaging community optometry referrals, reducing unnecessary urgent referrals, primarily for neovascular AMD, a key driver of hospital demand. It has the potential to streamline referral pathways, reduce system burden, and improve patient care for vision-threatening retinal disease. In the prespecified safety analysis (false negatives or missed referrals), teleophthalmology was shown to be as safe as standard practice. As OCT adoption in community optometry grows, teleophthalmology's impact will increase.[29]

The AI validation study identified opportunities to refine Moorfields-DeepMind-AI, particularly its referral thresholds, integration of clinical inputs, and compatibility with diverse OCT devices. Additional training and fine-tuning are needed, ideally incorporating newer AI architectures.[30] Our study underscores the importance of prospective evaluation in medical AI development and regulation, while signposting developers towards specific improvements for clinical readiness. Our findings are contextualised through cited findings of linked human–computer interaction analyses, with respect to patients' positive experience, attitudes, and perception of teleophthalmology-related benefits, clinicians' positive perceptions and attitudes, as well as study-related behavioural adjustments (full analyses and findings are reported separately).[13]

#### Contributors

KB (chief investigator) was responsible for the trial concept, design, research application, protocol development, oversight of trial delivery, validation of clinical artificial intelligence (AI) methodological expertise, data analysis, and interpretation of trial findings, drafting of this paper followed by critical revision for important intellectual content. RH and AS were involved in protocol development, trial management, recruitment, data collection, data analysis, interpretation of trial findings, oversight of trial delivery, and contributed to manuscript writing. TS was involved in interpretation of trial findings and contributed to manuscript writing. CB provided critical input in methodological design and principles of applied statistics for data analysis. AB designed and supervised the qualitative HCI study and contributed to manuscript writing. JGL led community optometry engagement and provided intellectual contribution to study design. GG provided critical support for data collection, data analysis, and critical review of the report. AEL performed the statistical analyses, interpretation of data, and contributed to manuscript writing. AD supervised the statistical analyses and contributed to manuscript writing, study, data analysis, interpretation of results, and contributed to manuscript writing. AK contributed to the design of study, data analysis, interpretation of results, and manuscript writing. TR contributed to the design and analysis of the experiment, data interpretation, and manuscript writing. LV and AK led the health economics components. NM contributed to the design and the analysis of the economic decision model, data interpretation, and manuscript writing. CD, EM, AM, and TA contributed to data collection and critical review of the final report. AA contributed to process evaluation, cultural and organisational impact of medical AI. PJP contributed to data collection and data analysis, trial oversight, and critical review of the final paper. AKD provided intellectual contribution to study design. PAK provided intellectual contribution to study design and AI data analysis. KB, AS, and RH directly accessed and verified the underlying data reported in the manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

#### Declaration of interests

LV reports a grant from the National Institute for Health and Care Research (NIHR) HTA programme for this study. TA reports an NIHR grant for trial activities. AA reports funding from UK Research and Innovation Future Leaders Fellowship scheme (MR/Y034430/1). AB reports a grant from the NIHR HTA programme for this study and a grant for EPSRC CDT in medical imaging. EM reports consulting fees for retinal diseases drugs and management from AbbVie; honoraria for lectures, presentations, manuscript writing, or education events for retinal diseases and their management and drugs technology from AbbVie, Alimera, and Bayer; fees for a video testimony for drug use in retinal disease from AbbVie; support for attending conference meetings, events and/or travel for retinal disease, and AI use in them from AbbVie, Alimera, Bayer, Novartis, and Roche; and participation in an advisory board for drugs' use in retinal diseases from AbbVie and Roche. PJP reports consulting fees from Bayer, Genentech, and Roche; payment from Bayer, Boehringer Ingelheim, and Roche; and support for attending meetings and/or travel from Bayer and Roche. CD reports grants or contracts from NIHR, Apellis Pharmaceuticals, Roche Products, and Topcon; consulting fees from Boehringer Ingelheim, F Hoffman-La Roche, Bayer, AbbVie, EyePoint, Alimera Sciences, Ocular Therapeutix, Jansen, Astellas, and Apellis; payment or honoraria from Boehringer Ingelheim, F Hoffman-La Roche, Bayer, Apellis, AbbVie, and Topcon; support for attending meetings and/or travel from Roche UK; and a leadership or fiduciary role in Oxford Ophthalmology Congress, NIHR Ophthalmology Speciality Group, NICE Diabetic Retinopathy Expert Committee and NIHR Commercial Research Delivery Centres network. JGL reports a grant from the NIHR HTA programme for this study. GG reports grants or contracts from Bayer and Roche; consulting fees from Apellis, Bayer, and Roche; payment or honoraria from AbbVie, Apellis, Bayer, and Roche; support for attending meetings and/or travel from Roche and Bayer; and participation on a data safety monitoring board or advisory board for Apellis, Bayer, and Roche. PAK reports grants or contracts by a UK Research and Innovation Future Leaders Fellowship (MR/R019050/1) and The Rubin Foundation Charitable Trust; consulting fees from Retina Consultants of America, Topcon, Roche, and Boehringer-Ingelheim; payment or honoraria from Zeiss, Topcon, Novartis, Boehringer-Ingelheim, Apellis, Roche, and AbbVie; support for attending meetings and/or travel from Bayer, Topcon, and Roche; active patents in generalisable medical image analysis using segmentation and classification neural networks and pending Patents in predicting disease progression from tissue images and tissue segmentation maps; participation on a data safety monitoring board or advisory board for Topcon, Bayer, Boehringer-Ingelheim, RetinAI, and Novartis; stock options with Bitfount and stock with Big Picture Medical; and consultancy work for Google DeepMind. KB reports speaker fees from Novartis, Bayer, and Roche; meeting or travel fees from Novartis and Bayer; consulting fees from Novartis, Roche, Bayer, and Boehringer-Ingelheim; research support from Apellis, Novartis, and Bayer; and stock with Read AI. All other authors declare no competing interests.

#### Data sharing

All data requests should be submitted to the corresponding author for consideration. Access to anonymised data might be granted following review. The code base for the deep-learning framework, Moorfields-DeepMind-AI, makes use of proprietary components and we are unable to publicly release the full code base. The three-dimensional augmentation code (using the caffe framework) is available as part of the three-dimensional U-net source code at https://lmb.informatik.uni-freiburg.de/resources/opensource/unet.en.html. Additionally, although we are unable to make all the Google proprietary components available, we are in the process of making the augmentation operations for TensorFlow available in the official TensorFlow code.

#### Acknowledgments

We thank the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme for providing funding (NIHR127773) and Moorfields Eye Hospital National Health Service (NHS) Foundation Trust for sponsoring the trial. We thank the NIHR Biomedical Research Centre based at Moorfields Eye Hospital NHS Foundation Trust, the NIHR North Thames Clinical Research Network, and University College London Institute of Ophthalmology for their support. The HERMES Research Team members also wish to extend their gratitude to the following members of the HERMES trial advisory committees, Patient and Public Involvement and Equality, Diversity and Inclusion advisors' group for their expert advice and guidance to study planning and delivery: Richard Gale Irene Stratton, Bushra Mushtaq, Gabriella De Salvo, and Geraldine Hoad (trial steering committee), Louise Downey, Gabriela Czanner, and Markus Groppe (data monitoring committee); and Jocelyn Cammack, Helen Baker, Subhojit Chakraborty, and Geraldine Hoad (Moorfields Patient and Public Involvement and Biomedical Research Centre Equality, Diversity and Inclusion). The authors would like to specially thank all HERMES trial participants. We also wish to express sincere gratitude to all recruiting community optometry sites and their corresponding NHS sites for their valuable contribution.

#### References

1 Wong WL, Su X, Li X, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health* 2014; **2:** e106–16.

2 Wang Y, Zhong Y, Zhang L, et al. Global incidence, progression, and risk factors of age- related macular degeneration and projection of disease statistics in 30 years: a modeling study. *Gerontology* 2022; **68:** 721–35.

3 Resnikoff S, Felch W, Gauthier T-M, Spivey B. The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200,000 practitioners. *Br J Ophthalmol* 2012; **96:** 783–87.

4 Foot B, MacEwen C. Surveillance of sight loss due to delay in ophthalmic treatment or review: frequency, cause and outcome. *Eye (Lond)* 2017; **31:** 771–75.

5 Thomas D, Weegen L, Walendzik A, Wasem J, Jahn R. Comparative analysis of delivery of primary eye care in three European countries. Institut für Betriebswirtschaft und Volkswirtschaft, 2011.

6 Muttuvelu DV, Buchholt H, Nygaard M, Rasmussen MLR, Sim D. Danish teleophthalmology platform reduces optometry referrals into the national eye care system. *BMJ Open Ophthalmol* 2021; **6:** e000671.

7 Evans BJW, Edgar DF, Jessa Z, et al. Referrals from community optometrists to the hospital eye service in England. *Ophthalmic Physiol Opt* 2021; **41:** 365–77.

8 Young K. Specsavers reaches OCT installation milestone of 1000 devices. Sept 29, 2022. https://www.aop.org.uk/ot/news/2022/09/29/specsavers-reaches-oct-installation-milestone-of-1000-devices (accessed June 10, 2025).

9 Dabasia PL, Edgar DF, Garway-Heath DF, Lawrenson JG. A survey of current and anticipated use of standard and specialist equipment by UK optometrists. *Ophthalmic Physiol Opt* 2014; **34:** 592–613.

10 Toth CA. Optical coherence tomography and eye care. *N Engl J Med* 2023; **389:** 1526–29.

11 Kern C, Fu DJ, Kortuem K, et al. Implementation of a cloud-based referral platform in ophthalmology: making telemedicine services a reality in eye care. *Br J Ophthalmol* 2020; **104:** 312–17.

12 Carmichael J, Abdi S, Balaskas K, Costanza E, Blandford A. The effectiveness of interventions for optometric referrals into the hospital eye service: a review. *Ophthalmic Physiol Opt* 2023; **43:** 1510–23.

13 Blandford A, Abdi S, Aristidou A, et al. Protocol for a qualitative study to explore acceptability, barriers and facilitators of the implementation of new teleophthalmology technologies between community optometry practices and hospital eye services. *BMJ Open* 2022; **12:** e060810.

14 Esteva A, Chou K, Yeung S, et al. Deep learning-enabled medical computer vision. *NPJ Digit Med* 2021; **4:** 5.

15 Bazvand F, Ghassemi F. Artefacts in macular optical coherence tomography. *J Curr Ophthalmol* 2020; **32:** 123–31.

16 De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018; **24:** 1342–50.

17 Wilson M, Chopra R, Wilson MZ, et al. Validation and clinical applicability of whole-volume automated segmentation of optical coherence tomography in retinal disease using deep learning. *JAMA Ophthalmol* 2021; **139:** 964–73.

18 Pawloff M, Gerendas BS, Deak G, Bogunovic H, Gruber A, Schmidt-Erfurth U. Performance of retinal fluid monitoring in OCT imaging by automated deep learning versus human expert grading in neovascular AMD. *Eye (Lond)* 2023; **37:** 3793–800.

19 Han JED, Liu X, Bunce C, et al. Teleophthalmology-enabled and artificial intelligence-ready referral pathway for community optometry referrals of retinal disease (HERMES): a cluster randomised superiority trial with a linked diagnostic accuracy study-HERMES study report 1-study protocol. *BMJ Open* 2022; **12:** e055845.

20 Parkin E. NHS maximum waiting time standards. March 31, 2020. House of Commons Library, 2020.

21 Day AC, Burr JM, Bunce C, et al, and the FACT Group. Randomised, single-masked non-inferiority trial of femtosecond laser-assisted versus manual phacoemulsification cataract surgery for adults with visually significant cataract: the FACT trial protocol. *BMJ Open* 2015; **5:** e010381.

22 Shah R, Edgar D, Evans BJW. Measuring clinical practice. *Ophthalmic Physiol Opt* 2007; **27:** 113–25.

23 Carmichael J, Abdi S, Balaskas K, Costanza E, Blandford A. Assessment of optometrists' referral accuracy and contributing factors: a review. *Ophthalmic Physiol Opt* 2023; **43:** 1255–77.

24 Muen WJ, Hewick SA. Quality of optometry referrals to neovascular age-related macular degeneration clinic: a prospective study. *JRSM Short Rep* 2011; **2:** 64.

25 Harvey K, Edgar DF, Agarwal R, Benwell MJ, Evans BJ. Referrals from community optometrists in England and their replies: a mixed methods study. *Ophthalmic Physiol Opt* 2022; **42:** 454–70.

26 Ong AY, Taribagil P, Sevgi M, et al. A scoping review of artificial intelligence as a medical device for ophthalmic image analysis in Europe, Australia and America. *NPJ Digit Med* 2025; **8:** 323.

27 Lebovitz S, Levina N, Lifshitz-Assaf H. Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. *Manage Inf Syst Q* 2021; **45:** 1501–26.

28 Williamson DJ, Struyven RR, Antaki F, et al. Artificial Intelligence to facilitate clinical trial recruitment in age-related macular degeneration. *Ophthalmol Sci* 2024; **4:** 100566.

29 Powell S. NHS to deliver OCT scans in the community. Dec 19, 2024. https://www.aop.org.uk/ot/news/2024/12/19/nhs-to-deliver-oct-scans-in-the-community (accessed July 14, 2025).

30 Zhou Y, Chia MA, Wagner SK, et al, and the UK Biobank Eye & Vision Consortium. A foundation model for generalizable disease detection from retinal images. *Nature* 2023; **622:** 156–63.