# Automated grading in the Diabetic Eye Screening Programme: protocol for an evidence summary


Review group: Julia Geppert
Pooja Suresh Kumar
Samantha A Johnson
Pranshu Mundada
Hannah Williams
Sian Taylor-Phillips
Chris Stinton
On behalf of WeBS-ESG

Correspondence to: Chris Stinton
Warwick Screening
Warwick Applied Health
Warwick Medical School
University of Warwick
Coventry CV4 7AL

Tel: 024 7657 4701
Email: c.stinton@warwick.ac.uk



Version: v3.3

Date completed: 21 July 2025

This project is funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis Programme (NIHR168057). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. *Conflicts of interest of authors are listed at the end of the document.*

**Abbreviations**

| | |
|---|---|
| AI | Artificial intelligence |
| ARIAS | Automated Retinal Image Analysis System |
| DESP | Diabetic eye screening programme |
| DM | Diabetes Mellitus |
| DL | Deep Learning |
| DR | Diabetic retinopathy |
| EDESP | English Diabetic Eye Screening Programme |
| ICTRP | International Clinical Trials Registry Platform |
| ML | Machine Learning |
| MODY | Maturity-onset diabetes of the young |
| NA | Not applicable |
| NIHR | National Institute for Health and Care Research |
| RCT | Randomised controlled trial |
| ROC | Receiver operating characteristic |
| UK NSC | UK National Screening Programme |
| WHO | World Health Organization |

# 1 Background

## 1.1 Condition

**Diabetes**

Diabetes mellitus (DM) is a chronic, metabolic condition characterised by elevated levels of blood glucose (hyperglycaemia). In people with DM, the body cannot produce enough of a hormone called insulin (which regulates the blood glucose levels), or the produced insulin is not effective. As a result, glucose cannot enter the cells and accumulates in the blood. There are two main types of DM: In 'Type 1 DM', the body's immune system attacks and destroys the cells in the pancreas that produce insulin. In 'Type 2 DM', the body cannot produce enough insulin, or it does not use the insulin properly. Other, less common, types include gestational diabetes (which develops during pregnancy), type 3c (due to damage to the pancreas), maturity-onset diabetes of the young (MODY), neonatal diabetes, and latent autoimmune diabetes in adults.[1] Over time, hyperglycaemia can lead to serious damage of the heart, blood vessels, eyes, kidneys and nerves.

According to the World Health Organization, the number of people living with DM globally rose from 200 million in 1990 to 830 million in 2020.[2] In 2022, the DM prevalence was 14% among adults aged 18 years and older, an increase from 7% in 1990. In 2021, DM and kidney disease due to DM caused over 2 million deaths worldwide. Additionally, around 11% of cardiovascular deaths were caused by hyperglycaemia.[2]

In the UK, the number of adults (20 – 79 years) with diabetes rose from 1.5 million in 2000 to 4.5 million in 2024, and numbers are expected to increase to 4.9 million by 2050.[3]

**Diabetic eye disease**

Diabetic eye disease encompasses a group of eye problems that can affect people with DM. Diabetic retinopathy (DR) is the most common form of eye problem in DM. Hyperglycaemia caused by DM can damage the blood vessels in the back of the eyes (the retina). As a result, they can become blocked, leak, or grow incorrectly. If left undiagnosed and untreated, DR can cause progressive damage to the retina and lead to vision loss. DR develops in stages:[4 5]

- *Stage 1 - Background DR:* Tiny bulges (microaneurysms) are visible in the blood vessels of the retina, which may leak small amounts of blood. Background DR usually does not affect eyesight.
- *Stage 2 - Pre-proliferative DR:* More severe and widespread changes in the blood vessels of the retina are present, including more significant bleeding into the eye. There is a high risk that the vision can eventually be compromised.
- *Stage 3 - Proliferative DR*: At this stage, many of the retinal blood vessels become damaged or blocked, which can lead to insufficient blood supply (ischaemia). To make up for the lack of oxygen, the body grows new, abnormal blood vessels on the retina or into the vitreous gel. These new vessels are weak and prone to bleeding. Scar tissue can pull the retina away from the back of the eye (retinal detachment). At this stage, there is a very high risk of irreversible vision loss.
- *Diabetic maculopathy:* Here, DR affects the macula, the middle part of the retina which provides central vision. This happens when the blood vessels near the macula are leaky or blocked, and fluid can build up and cause macular swelling (macular oedema). This causes the vision to be blurred and distorted and can also make colours appear washed out.

In 2017 in the UK, the prevalence of DR, sight-threatening retinopathy and diabetic maculopathy among people with DM aged 12 years and above were 33.8%, 12.3% and 7.9%, respectively.[6] DR or diabetic maculopathy was the second leading cause of blindness among working age adults in England and Wales in 2009/2010, causing 14.4% cases of blindness during that time period.[7]

## 1.2 Diabetic eye screening in the UK

Screening for DR aims to detect people with stage 2 (pre-proliferative) DR, stage 3 (proliferative) DR or with diabetic maculopathy during its common asymptomatic stage, so that treatment can be administered when it is most effective and the risk of vision loss can be reduced.[8] Diabetic Eye Screening programmes (DESPs) were introduced in England, Scotland and Wales in 2003[9] and in Northern Ireland in 2008. Screening is offered to individuals with DM aged 12 years and over (excluding women with gestational diabetes). The differences between DESPs across the UK 4 nations are summarised in **Table 1**.

**Table 1. Differences between DESPs across the 4 UK nations** (modified from Zhelev et al. 2021[10])

| | **England** | **Wales** | **Scotland** | **Northern Ireland** |
|---|---|---|---|---|
| **Software** | No national system, 2 suppliers | Single commissioned National system | Single commissioned National system | Single commissioned National system |
| **Automation** | None | None | Automated primary grading | None |
| **Images** | 2-field | 2-field | 1-field | 2-field |
| **Mydriasis** | Yes | Yes | Some cases | Yes |
| **Extended screening intervals for people at lowest risk** | Yes | Yes | Yes | Yes |

In the English DESP (EDESP), the screening test involves taking two 45° digital retinal photographs per eye (one macula-centred and one disc-centred) taken with pupil dilation (mydriasis) that are manually graded. Possible grading results are reported in **Table 2**.[11]

**Table 2. Retinal image grading in the English Diabetic Eye Screening Programme**

| **Level of DR or maculopathy** | **Grade** | **Possible interpretation** | |
|---|---|---|---|
| No observable retinopathy | R0 | **No DR** | **No DR** |
| No maculopathy | M0 | | |
| Background retinopathy | R1 | **Non-referable DR** | **Any DR** |
| Maculopathy – low risk or high risk | M1 | **Referable DR** | |
| Pre-proliferative retinopathy - low risk (R2L) or high risk (R2H) | R2 | | |
| Proliferative retinopathy - active (R3A) or treated and stable (R3S) | R3 | | |
| Ungradable | U | | |

DR, Diabetic retinopathy.

A simplified version of the multi-level manual grading pathway of retinal photographs[12] is shown in **Error! Reference source not found.**. All retinal images are assessed by primary graders. Images graded as R3, U or with other non-DR diseases identified go straight to the referral outcome grader. Secondary graders review all images graded as R1, R2 or M1 by primary graders. For quality control, secondary graders also review a random 10% of images that were graded as R0M0 by primary graders. Anyone receiving a R3, U or agreed R2 or M1 between primary and secondary graders is referred to a Digital Surveillance clinic for a rescreen after 3-6 months or the hospital eye service. Disagreements between primary and secondary graders (other than R3 and U) are assessed by arbitration graders. Based on the final grades - the specific level of DR (R0, R1, R2, or R3) and the presence/absence of diabetic maculopathy (M0 or M1) – the frequency of screening and the need for referral, respectively, are determined.[12]

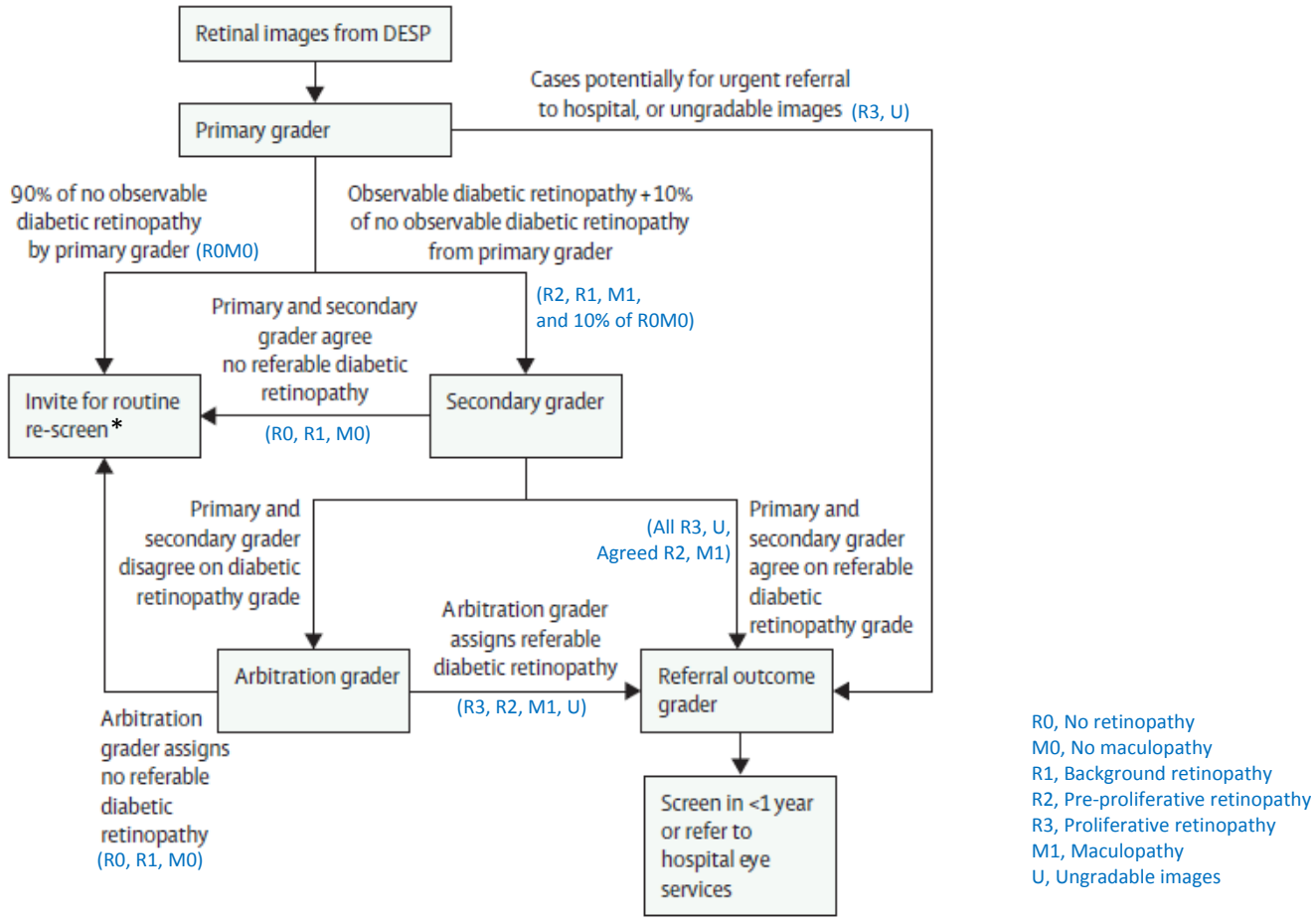


**Figure 1. Current grading practice in the English DESP** (modified from MacDonald et al. 2025[13]).

* Annual rescreen after a single negative screening result (R0M0 or R1M0); biennial rescreen after two consecutive negative screening results (R0M0 and R1M0).

Patients graded R0M0 (no DR) or R1M0 (non-referable DR) in the worse eye are invited to return for rescreening 12 months after a single negative result. After two consecutive negative screening results (R0M0 or R1M0), patients are invited for a rescreen after 24 months. Patients graded R2L, R3S and/or low risk M1 are referred to a Digital Surveillance clinic and are kept under surveillance and screened more frequently (every 3-6 months). Patients graded R3A or those who have higher risk maculopathy (M1) are referred to the Hospital Eye Service for management and treatment of DR. Patients with ungradable images (U) are referred to undergo a second test called slit lamp biomicroscopy.

A total of 1.49 million people were offered diabetic eye screening in 2020/2021 in England, among whom the uptake rate was 67.9%.[14] As each screen creates at least 4 retinal images that require assessing by up to 3 trained human graders, the current EDESP is very labour intensive. This represents a major challenge for the NHS, especially as the numbers of retinal photographs that need grading are expected to escalate in the future given that both prevalence and incidence of DM are expected to increase markedly.[15]

## 1.3 Automated Retinal Image Analysis Systems

Automated retinal image analysis systems (ARIASs) are artificial intelligence (AI)-based algorithms designed to read digital retinal images and classify them into 'disease', 'no disease' and 'ungradable images'. AI broadly refers to "machines that perform tasks normally performed by human intelligence, especially when the machines learn from data how to do those tasks".[16] Machine learning (ML) is a subset of AI, and deep learning (DL) is a subfield of ML (see **Figure 2**).

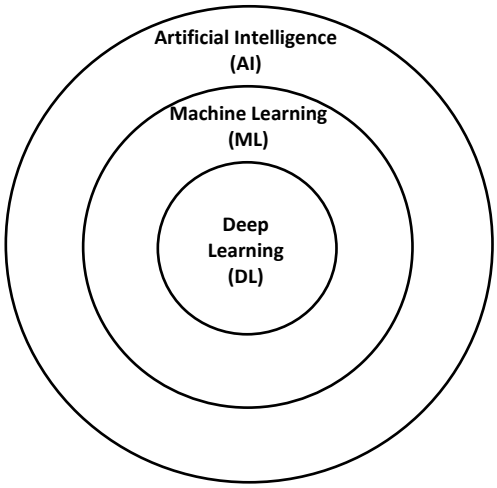


**Figure 2. Hierarchical organisation of artificial intelligence, machine learning, and deep learning**

Non-ML-based algorithms (also called symbolic AI) extract pre-specified 'hand-crafted' features (e.g. microaneurysms) from the retinal photographs and use them to classify the image sets. Most of the recently developed ARIASs use ML-based algorithms that do not depend on pre-specified features; instead, these algorithms use labelled input data and learn to distinguish 'disease' from 'no disease' images on their own. DL is a subset of ML that uses artificial neural networks with multiple layers to analyse data. DL algorithms are trained using large amounts of data and can learn and improve over time, becoming more accurate as they process more data.

The development of ML-based algorithms typically comprises 3 stages: training, fine-tuning (internal validation) and testing (external validation).[17] During internal validation, the AI algorithm is evaluated in the same dataset that was used for development and training, e.g. using cross validation or split sample validation. This can substantially overestimate the performance of the algorithm.[17] External validation studies, on the other hand, assess the AI performance with separate and independent data not used for model development. Here, studies using images collected from different sites (geographical validation) should be preferred as they allow an assessment of the generalisability of the ARIAS performance across potentially different technical parameters (such as different machines) and operating personnel.[13 17]

ARIASs have already been introduced in the DESPs in Scotland[18] and Portugal[19] and are now considered for clinical use in other countries.[20] In 2024, the UK National Screening Committee (UK NSC) received a proposal to modify the DESP by using an ARIAS for primary grading to triage patients into those with sight-threatening DR or other retinal abnormalities, from those at low risk of sight-threatening DR (**Error! Reference source not found.**). The proposal indicates that using ARIAS for diabetic eye screening would reduce human grader requirements, thereby reducing screening costs.

Under the proposed screening pathway with ARIAS as primary grader, any patient images denoted as ARIAS test-positive (or considered as technical failures) and a random 10% of those classed as ARIAS test-negative would proceed to secondary grading (**Error! Reference source not found.**). The pathway would then work exactly the same as for the manual grading arm.

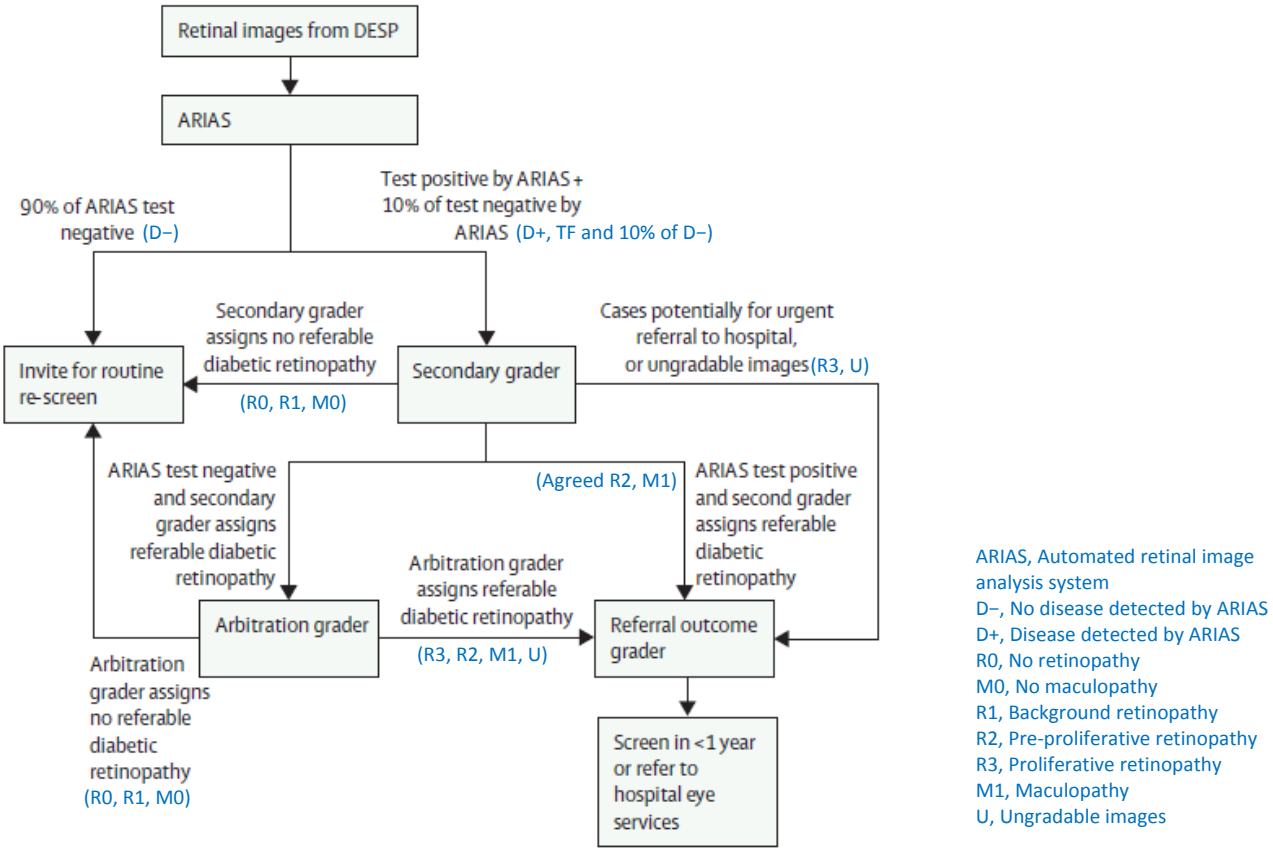


**Figure 3. Proposed grading pathway replacing the Level 1 grader with an automated retinal image analysis system (ARIAS)**

(modified from MacDonald et al. 2025[13])

* Annual rescreen after a single negative screening result (R0M0 or R1M0); biennial rescreening after two consecutive negative screening results (R0M0 and R1M0).

## 1.4 UK NSC recommendation

DESPs were implemented in England, Scotland and Wales in 2003,[9] and in Northern Ireland in 2008. The UK NSC currently does not recommend the incorporation of automated grading into the UK DESPs.

## 1.5 Previous reviews

The UK NSC commissioned an evidence summary in 2021 to evaluate the use of AI in the DESP, specifically focusing on ARIAS.[10] The review concluded that, although AI technologies show promise, the evidence base was insufficient to support widespread implementation of ARIAS in the DESP at that time.

A key finding of the previous review was that only limited evidence of ARIAS had been assessed in large-scale clinical studies that provided robust and reliable data on their diagnostic accuracy in real world practice. Moreover, the available evidence was highly context specific; study results could not be generalised beyond the settings in which the evaluations were conducted. This highlighted the need for future research to be carried out in settings that closely reflect the environments where ARIAS are intended to be deployed, ensuring external validity and relevance to routine screening programmes.

The review also identified a significant gap in high-quality comparative research. No randomised controlled trials (RCTs) or prospective cohort studies were found that directly compared clinical outcomes or other impact measures between DESPs employing fully manual primary grading and those using ARIAS based primary grading. Consequently, the clinical impact of integrating AI systems into screening pathways remains uncertain.

Economic evaluations included in previous reviews were limited and insufficient for definitive conclusions about cost effectiveness. Although these studies provided a useful foundation, they lacked long-term data and did not fully account for the evolving nature of AI technologies. Future cost effectiveness assessments should be longitudinal, incorporating software updates and real-world performance changes over time.

Additionally, concerns about the social and ethical implications of AI deployment, including acceptance by health professionals and patients, were noted but insufficiently explored. Overall, the 2021 evidence summary emphasised the necessity of further independent, high-quality research that includes diagnostic accuracy in relevant clinical contexts, clinical outcome comparisons, comprehensive economic analyses, and considerations of social acceptability before ARIAS can be confidently implemented in the DESP.

## 2 Decisions questions

Decision questions and the UK NSC criteria to which they relate are shown in **Table 3**.

**Table 3. Key questions for the evidence summary, and relationship to UK NSC screening criteria**

| Key question | UK NSC criterion |
|---|---|
| **Question 1.** What is the diagnostic accuracy* of Automated Retinal Image Analysis Systems at detecting diabetic eye disease in patients with diabetes mellitus? | 4. There should be a simple, safe, precise and validated screening test.<br><br>5. The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed. |
| **Question 2.** What is the wider clinical impact of diabetic eye screening programmes with the use of Automated Retinal Image Analysis Systems for primary grading compared with diabetic eye screening programmes with fully manual grading? | 11. There should be evidence from high quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity. Where screening is aimed solely at providing information to allow the person being screened to make an 'informed choice' (such as Down's syndrome or cystic fibrosis carrier screening), there must be evidence from high quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened. |
| **Question 3.** What is the cost-effectiveness of Automated Retinal Image Analysis Systems in diabetic eye screening programmes compared with diabetic eye screening programmes with manual grading? | 14. The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (value for money). Assessment against this criteria should have regard to evidence from cost benefit and/or cost effectiveness analyses and have regard to the effective use of available resource. |

* The term 'diagnostic accuracy' does not imply that the system is used to diagnose diabetic retinopathy or any other condition; the only use of automated retinal image systems investigated in the current review is as a first line screening test designed to identify patients with 'no disease' or 'non-referable disease', as part of a multi-level screening programme, such as the English Diabetic Eye Screening Programme.

# 3 Methods

The review will be undertaken using the UK NSC assessment approach for Evidence Summaries,[21] and be guided by the methods employed in the previous review on this topic for the UK NSC.[10]

## 3.1 Identification and selection of studies

### 3.1.1 Search strategy

Systematic literature searches will be undertaken using terms for diabetic retinopathy, screening and ARIAS which will identify evidence for all review questions. The search strategy from the previous UK NSC review[10] was used as a starting point, with the addition of MeSH headings for Early Diagnosis and Mass Screening. The search will be adapted to include 25 product and manufacturer names of commercially available ARIAS with regulatory approval as identified by a 2025 scoping review.[22] The search strategies were developed in MEDLINE (Ovid).

The search will be adapted for Embase (OVID) and the Cochrane Library (Wiley). Embase and the Cochrane Library now include the clinical trials harvested from ClinicalTrials.gov. Searches will also be undertaken in the International Clinical Trials Registry Platform (ICTRP) database from the WHO. The search strategies that are likely to be used in each of the databases are provided in **Appendix 1**.

The search strategy will comprise the following elements:

1) Searching of electronic bibliographic databases,
2) Searching of the ICTRP database,
2) Scrutiny of the references of included studies and relevant systematic reviews.

We will run alerts of the main search in the electronic databases until 30th September 2025.

### 3.1.2 Study eligibility criteria

Studies that satisfy the criteria listed in **Table 4** will be included.

## Table 4. Inclusion and exclusion criteria for the key questions

| Key question | Inclusion criteria | | | | | | | Exclusion criteria* |
|---|---|---|---|---|---|---|---|---|
| | **Population** | **Target condition** | **Intervention** | **Reference standard** | **Comparator** | **Outcome** | **Study type** | |
| 1. What is the diagnostic accuracy** of ARIAS at detecting diabetic eye disease in patients with DM? | People with type 1 and type 2 DM ≥12 years, including rarer forms of DM such as MODY, who underwent standard fundus photography for DR screening | 'Any DR' (R1 or higher and/or M1) or Referable DR' (R2, R3) and/or maculopathy (M1) | ARIAS, alone or as part of the workflow | Any | Manual grading, no comparator, head-to-head comparison of ARIAS | Overall accuracy measures for 'Any DR' (R1 or higher and/or M1; prioritised) or 'Referable DR' (R2 or higher and/or M1; deprioritised) and, if reported, for each grade of retinopathy | RCTs, prospective or retrospective single-gate studies, two-gate studies, systematic reviews*** and meta-analyses (studies will be prioritised by design, see below) | Non-English language, published before 2020, conference abstracts, ARIAS in development, internal validations of ARIAS |
| 2. What is the wider clinical impact of DESPs with the use of ARIAS for primary grading compared with DESPs with fully manual grading? | People with type 1 and type 2 DM ≥12 years, including rarer forms of DM such as MODY | NA | DESP for detecting DR and/or maculopathy using ARIAS on fundus photographs for primary grading followed by manual grading for secondary and arbitration grading | NA | DESP for detecting DR and/or maculopathy using human manual grading on fundus photographs at all levels of grading | Any clinical utility outcomes, such as: Vision loss, health-related quality of life.<br><br>Any patient management and practical implications outcomes, such as: Workforce (e.g. workload), inequalities. | RCTs, comparative prospective and retrospective cohort studies, and systematic reviews*** and meta-analyses of these (studies will be prioritised by design, see below) | Non-English language, published before 2020, conference abstracts |
| 3. What is the cost-effectiveness of an ARIAS in DESP compared with DESP with manual grading? | People with type 1 and type 2 DM ≥12 years, including rarer forms of DM such as MODY | NA | DESP for detecting DR and/or maculopathy using ARIAS on fundus photographs for primary grading followed by manual grading for secondary and arbitration grading | NA | DESP for detecting DR and/or maculopathy using human manual grading on fundus photographs at all levels of grading | Any cost-effectiveness or modelled clinical outcomes | Economic evaluations (of any type) and reviews of these | Non-English language, published before 2020, conference abstracts, non-UK-based evaluations |

ARIAS, Automated Retinal Image Analysis System; DESP, Diabetic eye screening programme; DM, Diabetic mellitus; DR, Diabetic retinopathy; MODY, Maturity-onset diabetes of the young; NA, Not applicable; RCT, Randomised controlled trial.

* See below for more detailed exclusion criteria.

** The term 'diagnostic accuracy' does not imply that the system is used to diagnose diabetic retinopathy or any other condition; the only use of automated retinal image systems investigated in the current review is as a first line screening test designed to identify patients with 'no disease' or 'non-referable disease', as part of a multi-level screening programme, such as the English Diabetic Eye Screening Programme.

*** Systematic reviews will be defined as per Centre for Reviews and Dissemination (CDR) Database of Abstract of Reviews of Effects (DARE) criteria.[23]

Papers that fulfil the following criteria will be excluded:

Qualitative studies, studies without relevant outcomes, studies where more than 10% of the population do not meet our inclusion criteria and are not reported separately, any algorithms in development, internal validations of ARIAS, articles not available in the English language, articles published prior to 2020, single case studies (one patient or one family), letters, reviews, editorials, communications, commentaries, conference abstracts and other grey literature, studies that have been retracted, studies where it cannot be established if the inclusion criteria are met.

## 3.2 Review strategy

Study selection will be managed through the Rayyan platform.[24] All inclusion and exclusion decisions will be made by human reviewers.

Titles and abstracts of records identified by the searches will be screened by one reviewer. A second reviewer will independently assess a random 20% sample of the titles/abstracts. We will retrieve the full publication of records considered potentially relevant by either reviewer. Full text articles will be assessed against the eligibility criteria by one reviewer, with a random 20% sample assessed independently by a second reviewer. Disagreements will be resolved by consensus, or through discussion with a third reviewer. Records rejected at full text stage will be documented (including reasons for exclusion).

## 3.3 Data extraction strategy

Data of all prioritised (see section 3.5) studies will be extracted by one reviewer, with a random 20% checked by a second reviewer. All data extraction will be entered into a piloted electronic data collection form in Excel. The following information will be abstracted: general information on the trial design and methods (e.g., eligibility criteria, study flow, country, setting, population / dataset, photographic protocol, intervention, comparator, reference standard, vendor involvement), accuracy and effectiveness results, and conclusions. Disagreements will be resolved by discussion and, if necessary, by involving a third reviewer.

No formal data extraction will be performed for deprioritised studies, but study characteristics and main results will be reported in a table in the appendix.

## 3.4 Assessment of study quality

The quality appraisal tool used for each study design are reported in **Table 5**. For question 1, quality appraisal of test accuracy studies will be conducted using a modified QUADAS-2 tool[25] with signalling questions tailored for critical appraisal of machine learning studies.[26] Quality appraisal for all prioritised studies will be conducted by one reviewer, with a random 20% independently assessed by a second reviewer. Disagreements will be resolved by discussion and, if necessary, by involving a third reviewer. No quality assessment will be performed for deprioritised studies.

**Table 5. Risk of bias/quality appraisal tools**

| Study design | Tool |
|---|---|
| Systematic reviews with/without meta-analysis | A MeaSurement Tool to Assess systematic Reviews (AMSTAR 2)[27] |
| Randomised controlled trials | Cochrane Risk of Bias 2 tool (RoB 2)[28] |
| Test accuracy studies | QUADAS-2[25]<br>QUADAS-C[29] |
| Cohort studies | Joanna Briggs Institute Checklist for Cohort Studies[30] |
| Economic evaluations | Joanna Briggs Institute Checklist for Economic Evaluations[31] |

## 3.5 Methods for reporting

Studies for questions 1 and 2 will be prioritised. Only the highest priority studies will be extracted and data synthesised. A similar approach as used in the 2021 UK NSC review[10] will be taken to prioritise studies for extraction.

Question 1

The highest priority study would be a comparative UK-based study that evaluates the performance of commercially available, CE-marked and/or FDA approved ARIAS(s) compared to human primary graders to detect 'Any retinopathy' (R1 or higher and/or M1) in dilated eyes (with mydriasis) using an external, geographically separate study population.

In the absence, or minimal volume, of such studies, we will prioritise studies using a combination of the following criteria in discussion with the UK NSC:

- Studies from comparable countries (i.e. North-Western European, America, Australia);
- Studies assessing the performance of ARIAS in undilated eyes;
- Studies reporting 'Referable retinopathy' (R2 or higher and/or M1) only;
- Studies using an external, temporal validation;
- Studies without a human comparator in the following (descending) order: RCTs, prospective test accuracy studies with single-gate design, retrospective test accuracy studies with single-gate design;
- Studies using ARIAS that is not CE-marked and/or FDA approved;
- Studies using ARIAS that is not commercially available.

Question 2

The ideal study would be a systematic review of UK studies. In the absence of a suitable systematic review, primary UK-based RCTs or comparative prospective cohort studies will be prioritised next.

In the absence, or minimal volume, of such studies, we will prioritise studies as follows:

- Studies from comparable countries (i.e. North-Western European, America, Australia);
- Retrospective cohort studies.

We will report study characteristics and main results of deprioritised studies in an appendix table.

## 3.6 Methods for analysis/synthesis

### 3.6.1 Overall approach

Only the highest priority studies will be synthesised. Information on their study design, population, setting, the used ARIAS(s), comparator, reference standard and outcome measures will be summarised in text and tables.

For question 1, original data extracted from the studies will be used to construct 2x2 tables. The resulting pairs of sensitivities and specificities will be plotted on a receiver operator characteristic (ROC) curve. Pairs of sensitivities and specificities will also be displayed in a paired forest plot to demonstrate scatter and uncertainty, grouped by type of ARIAS (non-ML-based vs ML-based). Test accuracy results will further be stratified by ARIAS vendor and product.

Meta-analysis will be considered if sufficient data from reasonably homogeneous studies are available (e.g., at least 5 clinically and methodologically similar studies[32]).

Where data permit, we will additionally present subgroup data and may undertake subgroup analyses by grade of retinopathy, age group and ethnicity.

For question 2, we will summarise study characteristics and findings, grouped by type of outcome.

We will use the following effect measures for the final health outcomes:

- Hazard ratio and median time for time-to-event data (e.g. time to vision loss);
- Risk ratio for dichotomous outcomes (e.g. incidence of vision loss);
- Mean difference between arms for continuous outcomes (e.g. health-related quality of life).

Where treatment effect estimates (e.g., risk ratios, mean differences) are not explicitly reported, but sufficient data are available (e.g., event counts, mean values for arms), we will calculate these estimates and their corresponding confidence intervals using standard methods.[33]

If data permits and studies are clinically similar, we will pool the results in meta-analyses stratified by type of ARIAS. If data do not permit a pooled analysis, then we will conduct a narrative synthesis. Where data permits, we will additionally present subgroup data and may undertake subgroup analyses by age group and ethnicity.

For question 3, the findings from individual studies will be compared narratively due to the expected heterogeneity in the economic analyses (e.g. different aims/objectives, study designs, populations, and methods).

# 4 Timescale

| | Jun | | | | | Jul | | | | Aug | | | | Sep | | | | | Oct | | | | Nov | | | | Dec | Jan | Feb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Task name** | **2nd** | **9th** | **16th** | **23rd** | **30th** | **7th** | **14th** | **21st** | **28th** | **4th** | **11th** | **18th** | **25th** | **1st** | **8th** | **15th** | **22nd** | **29th** | **6th** | **13th** | **20th** | **27th** | **3rd** | **10th** | **17th** | **24th** | | | |
| Kick off meeting with UK NSC | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Develop search strategy | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Draft protocol | | | 16th | | | | | | | | | | | | | | | | | | | | | | | | | |
| Final protocol | | | | | | 11th | | | | | | | | | | | | | | | | | | | | | | |
| Run searches | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Develop data extraction sheet | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Tailoring of QUADAS-2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Title / abstract sift | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Retrieval of full texts | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Full text assessment | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Final selection of included studies | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data extraction | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Quality appraisal | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Synthesis | | | | | | | | | | | | | | | | | | | | | 24th | | | | | | | |
| Report writing | | | | | | | | | | | | | | | | | | | | | 24th | | | | | | | |
| QA of draft report | | | | | | | | | | | | | | | | | | | | | | | 7th | | | | | |
| Address QA feedback | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Draft report to NSC | | | | | | | | | | | | | | | | | | | | | | | | 14th | | | | |
| Address NSC feedback (several rounds) | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Final report to NSC | | | | | | | | | | | | | | | | | | | | | | | | | | | 31st | |
| ARG Meeting | | | | | | | | | | | | | | | | | | | | | | | | | | | | 26th |

**Figure 4. Project timeline**

## 5 Project team contributions

Name: Julia Geppert

Job title: Research Fellow

Address: Warwick Screening, Warwick Medical School, University of Warwick, Coventry, CV4 7AL

Email: J.Geppert@warwick.ac.uk

Contribution: Protocol development, study selection, reviewer 1 for key questions 1 and 2 (data extraction, quality appraisal), reviewer 2 for key question 3 (data extraction, quality appraisal), data synthesis, report writing

Name: Samantha A Johnson

Job title: Information Specialist

Address: Warwick Screening, Warwick Medical School, University of Warwick, Coventry, CV4 7AL

Email: Samantha.A.Johnson@warwick.ac.uk

Contribution: Development of search strategies, undertaking searches, reference management

Name: Pooja Suresh Kumar

Job title: Research Assistant

Address: Warwick Screening, Warwick Medical School, University of Warwick, Coventry, CV4 7AL

Email: pooja.suresh-kumar@warwick.ac.uk

Contribution: Protocol development, study selection, reviewer 2 for key questions 1 and 2 (data extraction, quality appraisal), report writing

Name: Pranshu Mundada

Job title: Research Associate

Address: Warwick Screening, Warwick Medical School, University of Warwick, Coventry, CV4 7AL

Email: pranshu.r.mundada@warwick.ac.uk

Contribution: Protocol development, study selection, reviewer 1 for key question 3 (data extraction, quality appraisal), report writing

Name: Hannah Williams

Job title: Project Manager

Address: Warwick Screening, Warwick Medical School, University of Warwick, Coventry, CV4 7AL

Email: H.Fraser@warwick.ac.uk

Contribution: Project management

Name: Sian Taylor-Phillips

Job title: Professor of Population Health

Address: Warwick Screening, Warwick Medical School, University of Warwick, Coventry, CV4 7AL

Email: s.taylor-phillips@warwick.ac.uk

Contribution: Joint project lead, protocol development, report writing

Name: Chris Stinton

Job title: Senior Research Fellow

Address: Warwick Screening, Warwick Medical School, University of Warwick, Coventry, CV4 7AL

Email: c.stinton@warwick.ac.uk

Contribution: Joint project lead, protocol development, report writing

## 6 Competing interests of authors

Sian Taylor-Phillips is a member of the UK NSC and Chair of the of the UK NSC's Research and Methodology subgroup. Chris Stinton is a member of the UK NSC's Adult Reference Group. The other authors declare no conflicts of interest. No funding from the industry to self or institution. Public funding to the institution from the NIHR Evidence Synthesis Programme (NIHR168057) and an NIHR award (NIHR302434).

# 7 References


1. Diabetes UK. About diabetes: Diabetes UK; [Available from: https://www.diabetes.org.uk/about-diabetes?gad_campaignid=15219795550&gad_source=1 accessed 10 June 2025.

2. World Health Organization. Diabetes: WHO; 2024 [Available from: https://www.who.int/news-room/fact-sheets/detail/diabetes accessed 6 June 2025.

3. International Diabetes Federation. United Kingdom - Diabetes country report 2000 — 2050: International Diabetes Federation; 2025 [Available from: https://diabetesatlas.org/data-by-location/country/united-kingdom/ accessed 4 July 2025.

4. Royal National Institute of Blind People. Diabetes related eye conditions including Diabetic Retinopathy: RNIB; 2023 [updated 1 June 2023. accessed 6 June 2025.

5. NHS. Stages - Diabetic retinopathy [updated 16 December 2021. Available from: https://www.nhs.uk/conditions/diabetic-retinopathy/stages/#:~:text=Stage%201%3A%20background%20retinopathy,common%20in%20people%20with%20diabetes. accessed 9 June 2025.

6. Haider S, Thayakaran R, Subramanian A, et al. Disease burden of diabetes, diabetic retinopathy and their future projections in the UK: cross-sectional analyses of a primary care database. *BMJ Open* 2021;11(7):e050058. doi: 10.1136/bmjopen-2021-050058 [published Online First: 20210712]

7. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. *BMJ Open* 2014;4(2):e004015. doi: 10.1136/bmjopen-2013-004015 [published Online First: 20140212]

8. Amoaku WM, Ghanchi F, Bailey C, et al. Diabetic retinopathy and diabetic macular oedema pathways and management: UK Consensus Working Group. *Eye (Lond)* 2020;34(Suppl 1):1-51. doi: 10.1038/s41433-020-0961-6

9. Scanlon PH. The English National Screening Programme for diabetic retinopathy 2003-2016. *Acta Diabetol* 2017;54(6):515-25. doi: 10.1007/s00592-017-0974-1 [published Online First: 20170222]

10. Zhelev Z, Peters J, Rogers M, et al. Automated grading in the Diabetic Eye Screening Programme - External review against programme appraisal criteria for the UK National Screening Committee. 2021. https://assets.publishing.service.gov.uk/media/619e80088fa8f5037e8ccb2f/Evidence_summary_AI_in_DESP_2021.pdf (accessed 7 June 2025).

11. York and Scarborough Teaching Hospitals - NHS Foundation Trust. Understanding national grading results and outcomes - Screening for diabetic retinopathy 2025 [Available from: https://www.yorkhospitals.nhs.uk/our-services/organdonation/a-z-of-services/diabetic-eye-screening/understanding-national-grading-results-and-outcomes/ accessed 6 June 2025.

12. NHS England. NHS England Diabetic eye screening: Patient grading referral surveillance. 2024. https://assets.publishing.service.gov.uk/media/668423b903469096e184059d/20230725_Diabetic_eye_screening_pathways_-_patient_grading_referral_surveillance_.pdf#:~:text=Screening%20is%20offered%20once%20every%2024%20months%20for,positive%2C%20non-referrable%20result%20or%20a%20single%20negative%20result. (accessed 6 June 2025).

13. Macdonald T, Zhelev Z, Liu X, et al. Generating evidence to support the role of AI in diabetic eye screening: considerations from the UK National Screening Committee. *Lancet Digit Health* 2025;7(5):100840. doi: 10.1016/j.landig.2024.12.004 [published Online First: 20250403]

14. Office for Health Improvement & Disparity. Corporate report - NHS screening programmes in England: 2020 to 2021: GOV.UK; [updated 16 February 2023. Available from: https://www.gov.uk/government/publications/nhs-screening-programmes-annual-report/nhs-screening-programmes-in-england-2020-to-2021#nhs-diabetic-eye-screening-des-programme accessed 7 June 2025.

15. Public Health England. Diabetes Prevalence Model. 2016. https://assets.publishing.service.gov.uk/media/5a82c07340f0b6230269c82d/Diabetesprevalencemodelbriefing.pdf (accessed 6 June 2025).

16. HM Government. National AI Strategy. 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf (accessed 7 June 2025).

17. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology* 2018;286(3):800-09. doi: 10.1148/radiol.2017171920 [published Online First: 20180108]

18. Philip S, Lee N, Black M, et al. Impact of introducing automated grading into the Scottish national diabetic retinopathy screening programme. *Diabetic Medicine* 2017;34:172.

19. Ribeiro L, Oliveira CM, Neves C, et al. Screening for Diabetic Retinopathy in the Central Region of Portugal. Added Value of Automated 'Disease/No Disease' Grading. *Ophthalmologica* 2014 doi: 10.1159/000368426 [published Online First: 20141126]

20. Bellemo V, Lim G, Rim TH, et al. Artificial Intelligence Screening for Diabetic Retinopathy: the Real-World Emerging Application. *Curr Diab Rep* 2019;19(9):72. doi: 10.1007/s11892-019-1189-3 [published Online First: 20190731]

21. Public Health England. Guidance - UK NSC: evidence review process [updated 26 July 2024. Available from: https://www.gov.uk/government/publications/uk-nsc-evidence-review-process/uk-nsc-evidence-review-process accessed 6 June 2025.

22. 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(1):323. doi: 10.1038/s41746-025-01726-8 [published Online First: 20250529]

23. Centre for Reviews and Dissemination. Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews York (UK): Centre for Reviews and Dissemination; 1995-2025; [Available from: https://www.crd.york.ac.uk/CRDWeb/ accessed 6 June 2025.

24. Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5(1):210. doi: 10.1186/s13643-016-0384-4 [published Online First: 20161205]

25. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009

26. Faes L, Liu X, Wagner SK, et al. A Clinician's Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies. *Transl Vis Sci Technol* 2020;9(2):7. doi: 10.1167/tvst.9.2.7 [published Online First: 20200212]

27. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008. doi: 10.1136/bmj.j4008 [published Online First: 20170921]

28. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. doi: 10.1136/bmj.l4898 [published Online First: 20190828]

29. Yang B, Mallett S, Takwoingi Y, et al. QUADAS-C: A Tool for Assessing Risk of Bias in Comparative Diagnostic Accuracy Studies. *Ann Intern Med* 2021;174(11):1592-99. doi: 10.7326/M21-2234 [published Online First: 20211026]

30. Joanna Briggs Institute. Checklist for Cohort Studies - Critical Appraisal tools for use in JBI Systematic Reviews. 2020. https://jbi.global/sites/default/files/2021-10/Checklist_for_Cohort_Studies.docx (accessed 6 June 2025).

31. Joanna Briggs Institute. Checklist for Economic Evaluations. 2020. https://jbi.global/sites/default/files/2021-10/Checklist_for_Economic_Evaluations.docx (accessed 6 June 2025).

32. Cochrane. Chapter 10: Analysing data and undertaking meta-analyses: Cochrane; 2024 [Version 6.5:[Available from: https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-10#section-1 accessed 4 July 2025.

33. Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions (current version): Cochrane; 2024 [Version 6.5:[Available from: https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current accessed 4 July 2025.

# 8 Appendices

**Appendix 1. Search strategies**

**Ovid MEDLINE(R) ALL <1946 to July 16, 2025>**

| | | |
|---|---|---|
| 1 | exp eye diseases/ | 675229 |
| 2 | retinopathy.ti,ab. | 60893 |
| 3 | eye pathology.ti,ab. | 371 |
| 4 | maculopathy.ti,ab. | 5852 |
| 5 | diabetic eye.ti,ab. | 1082 |
| 6 | diabetic macular.ti,ab. | 6435 |
| 7 | retinal fundus.ti,ab. | 681 |
| 8 | fundus photograph*.ti,ab. | 8422 |
| 9 | fundus camera*.ti,ab. | 1248 |
| 10 | retinal photograph*.ti,ab. | 1425 |
| 11 | retinal camera*.ti,ab. | 331 |
| 12 | 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 | 699475 |
| 13 | exp Diagnostic Techniques, Ophthalmological/ | 204203 |
| 14 | exp Diagnosis, Computer-Assisted/ | 91432 |
| 15 | "Sensitivity and Specificity"/ | 380834 |
| 16 | diagnostic test*.ti,ab. | 64670 |
| 17 | diagnostic accuracy.ti,ab. | 74475 |
| 18 | diagnostic performance.ti,ab. | 34884 |
| 19 | screening.ti,ab. | 768195 |
| 20 | imaging.ti,ab. | 1178551 |
| 21 | (Sensitivity or specificity).ti,ab. | 1403535 |
| 22 | reference standard.ab. | 22002 |
| 23 | optical coherence tomography.ti,ab. | 57612 |
| 24 | early diagnosis/ | 32178 |
| 25 | Mass Screening/ | 122073 |
| 26 | or/13-25 | 3640230 |

27 exp Artificial Intelligence/ 243216
28 artificial intelligence.ti,ab. 66235
29 deep learning.ti,ab. 77052
30 neural network*.ti,ab. 128064
31 automated retinal image analysis system.ti,ab. 7
32 automated grading.ti,ab. 160
33 automated level.ti,ab. 13
34 (automated adj (tool* or technique* or identification or detection)).ti,ab. 7468
35 ARIAS.ti,ab. 377
36 iGradingM.ti,ab. 3
37 EyeArt.ti,ab. 25
38 IDx-DR.ti,ab. 28
39 DR-RACS.ti,ab. 0
40 RetinaLyze.ti,ab. 3
41 RetmarkerSR DR.ti,ab. 0
42 Singapore Eye Lesion Analyzer.ti,ab.0
43 RetinaVue.ti,ab. 6
44 TRIAD network.ti,ab. 2
45 LumineticsCore.ti,ab. 3
46 Retmarker.ti,ab. 7
47 DAIRET.ti,ab. 2
48 eyRIS.ti,ab. 0
49 (Automated Disease Assessment or ARDA).ti,ab. 93
50 Medios.ti,ab. 840
51 OphtAI.ti,ab. 1
52 RetCAD.ti,ab. 8
53 DeepDee.ti,ab. 0
54 MONA DR.ti,ab. 1
55 AEYE-DS.ti,ab. 1
56 EyeCheckup.ti,ab. 1
57 Reti-EyeReti.ti,ab. 0

58 DrNoon.ti,ab. 0

59 ITOS Mass Screening.ti,ab. 0

60 Nexy AI.ti,ab. 0

61 EyeWisdom.ti,ab. 3

62 iPredict System.ti,ab. 0

63 TeleMedC DR grader.ti,ab. 0

64 Eyetelligence system.ti,ab. 0

65 DRISTi.ti,ab. 9

66 VUNO Med.ti,ab. 5

67 AutoGrader.ti,ab. 2

68 UMI DR.ti,ab. 0

69 or/27-68 370017

70 12 and 26 and 69 4166

71 limit 70 to ed=20200625-20250717 2638

72 limit 70 to dt=20200625-20250717 2944

73 limit 70 to ez=20200625-20250717 2933

74 limit 70 to ep=20200625-20250717 2362

75 71 or 72 or 73 or 74 3174

76 limit 75 to english language 3122

## Protocol for Public and Patient Involvement and Engagement (PPIE) in Rapid Review: Automated grading in the Diabetic Eye Screening Programme

The structure of this protocol is guided by the ACTIVE framework (Pollock, 2019).

### Recruitment of PPIE group members

Recruitment will be closed (by invitation) using purposeful sampling.

We will set up a group of 5-6 public participants. We will recruit a mix of people (age, gender, ethnicity, socioeconomic background), with and without experience of diabetes and sight-loss, and including individuals who have and have not experienced diabetic eye screening.

Given the timescale of the rapid review, recruitment will be targeted through known contacts, including ophthalmology departments. If necessary, we will use third party organisations Diabetes UK or Breakthrough T1D, formerly JDRF, who specialise in contacting underserved populations, provided they can recruit without public-facing open calls such as via social media. While potential participants may be suggested by these contacts or organisations, we will undergo an interview process with them to determine their views, background, any conflicts of interest, and how they might interact in a group-setting, to ensure their suitability. PPIE group members will also be required to complete a conflict-of-interest form.

### Compensation of time

Participation will be compensated at NIHR rates for involvement. Additional preparatory or post-meeting work will also be compensated. We will fund other costs that might preclude involvement, such as carers/childcare and any additional accessibility costs (for example, where participants have vision impairment).

**Mode and level of Involvement**

PPIE involvement will be through direct interaction (online meetings), with the same group meeting on 5 occasions. Accessible preparatory information will be sent prior to meetings to make best use of meeting time.

We will create a deliberative knowledge space in meetings, which allows for the integration of technical forms of knowledge in an accessible format to enable detailed exploration of meanings and facilitate the contribution of all members.

The level of involvement will vary with the stage of the review; we envisage members will contribute to the review processes of sifting, integration, and interpretation by providing feedback for the research team. The introduction and final meeting will focus on information provision. The group will be directly involved in discussing the communication/dissemination process, sense-checking the report, and providing content for the plain English summary. The group will not have a role in independently communicating/disseminating results.

**Proposed meeting schedule**

The meetings will run alongside the rapid review process, with input at key stages, and without impacting on the review timeline**.**

(1) Introduction to the project (including the topic and concept of rapid reviews), introduction to each other and the research team. Discussion of objectives, purpose, research plan, and discussion around ways of working / expectations (from research team and PPIE group)
(2) Discussion around searching/sifting strategies, data extraction and emerging findings
(3) Discussion around results and integration, with interpretation of clinical impact and practical implications
(4) Discussion around reporting, including highlighting key messages
(5) Online meeting and/or email correspondence (indirect interaction), following feedback from the NSC and any stakeholder consultation

**Evaluation and reporting**

Post-meeting surveys will be used to gather feedback and capture any reflections since the meeting. They will also be used as an opportunity for PPI contributors to highlight impact from meeting discussions and any additional support they may need to access the content of the meetings. The perspectives of the review team will be gathered, for example, to determine where and how PPIE impacted the review process, what worked well, and whether there are areas for further development. Reflections on the process will also be sought from the UK NSC.

We will follow GRIPP2 guidelines (Staniszewska et al, 2017) for reporting patient and public involvement in health and social care research. Any arising publications would only come at the end of the review cycle, and with UK NSC approval to ensure the messaging is appropriate. Where third-party organisations are involved in PPIE recruitment, they may also want their own report on how successful PPIE recruitment was and how this involvement impacted the work. This would be opted-out of where possible or, where not, based on the GRIPP2 reporting guidelines (which detail the nature of the PPIE involvement, stages of contribution, and any PPIE impact) and would similarly be reviewed by the UK NSC for approval before being sent to the organisation.

## References

Pollock A, Campbell P, Struthers C, et al. Development of the ACTIVE framework to describe stakeholder involvement in systematic reviews. *Journal of Health Services Research & Policy*. 2019;24(4):245-255. doi:10.1177/1355819619841647

Staniszewska, S., Brett, J., Simera, I. *et al.* GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. *Res Involv Engagem* **3**, 13 (2017). https://doi.org/10.1186/s40900-017-0062-2