Artificial intelligence for detecting keratoconus

Cochrane Library

Cochrane Database of Systematic Reviews

Artificial intelligence for detecting keratoconus (Protocol)

Vandevenne MMS, Favuzza E, Veta M, Lucenteforte E, Berendschot T, Mencucci R, Nuijts RMMA, Virgili G, Dickman MM

Vandevenne MMS, Favuzza E, Veta M, Lucenteforte E, Berendschot T, Mencucci R, Nuijts RMMA, Virgili G, Dickman MM.
Artificial intelligence for detecting keratoconus (Protocol).
Cochrane Database of Systematic Reviews 2021, Issue 12. Art. No.: CD014911.
DOI: 10.1002/14651858.CD014911.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Protocol]

Artificial intelligence for detecting keratoconus

Magali MS Vandevenne[1], Eleonora Favuzza[2], Mitko Veta[3], Ersilia Lucenteforte[4], Tos Berendschot[1], Rita Mencucci[2], Rudy MMA Nuijts[1], Gianni Virgili[2,5], Mor M Dickman[1]

[1]University Eye Clinic Maastricht, Maastricht University Medical Center (MUMC+), Maastricht, Netherlands. [2]Department of Neurosciences, Psychology, Pharmacology and Child Health, University of Florence, Florence, Italy. [3]Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands. [4]Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy. [5]Queen's University Belfast, Belfast, UK

Contact: Mor M Dickman, mor.dickman@mumc.nl.

Editorial group: Cochrane Eyes and Vision Group.
Publication status and date: New, published in Issue 12, 2021.

Citation: Vandevenne MMS, Favuzza E, Veta M, Lucenteforte E, Berendschot T, Mencucci R, Nuijts RMMA, Virgili G, Dickman MM. Artificial intelligence for detecting keratoconus (Protocol). *Cochrane Database of Systematic Reviews* 2021, Issue 12. Art. No.: CD014911. DOI: 10.1002/14651858.CD014911.

ABSTRACT

Objectives

This is a protocol for a Cochrane Review (diagnostic). The objectives are as follows:

The primary objective is to assess the diagnostic accuracy of AI algorithms in the detection of keratoconus in patients presenting with refractive errors, especially those whose vision can no longer be corrected fully with glasses, patients seeking corneal refractive surgery or those suspected of having keratoconus. AI could help ophthalmologists, optometrists and other eye-care professionals to make decisions on referral to cornea specialists for these patients.

Secondary objectives

To compare different AI algorithms, e.g. neural networks, decision trees, support vector machines.

To assess potential causes of heterogeneity in diagnostic performance across studies, according to the following:

- index test methodology: pre-processing techniques, core AI method and postprocessing techniques;
- sources of input to train algorithms: topography and tomography images from Placido-disc system or Scheimpflug system or slit-scanning system or OCT, number of training and testing cases/images, label/endpoint variable used for training;
- study setting;
- study design, retrospective or prospective studies;
- ethnicity, or geographic area as its proxy;
- different index test positivity criteria provided by topography or tomography device;
- reference standard used, topography or tomography, one or two cornea-specialists;
- definition of keratoconus used;
- mean age;
- patient recruitment;
- severity of keratoconus:
  - clinically manifest keratoconus
  - subclinical keratoconus

BACKGROUND

Target condition being diagnosed

Keratoconus is an ectatic degenerative disorder of the cornea, usually affecting both eyes. Ultra-structural examination of the human cornea *ex vivo* has revealed disruption and loss of the native collagen network, leading to biomechanical instability and severe corneal thinning (Hayes 2012; Meek 2005). The disease is generally progressive in nature, resulting in the cornea taking a typical cone shape. This causes myopia and irregular astigmatism, impairing visual acuity.

The usual onset of keratoconus is during puberty and tends to gradually progress until patients are in their 30s. Keratoconus tends to progress more rapidly in patients younger than 17 years (Ferdi 2019). The disease usually stabilizes when patients get older (Ferdi 2019). The prevalence of keratoconus varies among studies (Hashemi 2020). This may be due to several reasons e.g. different diagnostic criteria, different diagnostic methods, change in testing rates over time, genetic variation and/or environmental differences.

The pathophysiology of keratoconus is not well understood. However, both environmental and genetic factors seem to play a role (Rabinowitz 2021). One of the risk factors that has been investigated extensively is eye rubbing; others include the wearing of contact lenses and allergic disease. Research on the genetic contribution to keratoconus suggests a possible association (Rabinowitz 2021). However, genetic testing as a diagnostic tool is not currently available.

Some patients who undergo refractive surgery may be at risk of developing iatrogenic keratoectasia. The reported incidence of this is low (Giri 2017). However, the consequences can be sight-threatening. It is therefore important to detect those corneas at risk of developing this condition. Some risk factors are suggested, e.g. irregular topography and thin corneal pachymetry (Giri 2017).

The treatment for keratoconus depends on the severity of the disease. In the initial stage, treatment of keratoconus typically aims at improving visual acuity with the use of glasses and specialised contact lenses. These treatments, however, do not cure keratoconus. As the disease progresses, visual acuity often can no longer be corrected with glasses. Corneal cross-linking has been used since 2003 to stop the progression of keratoconus (Sykakis 2015). However, this treatment cannot reverse the visual impairment. Before corneal cross-linking, the only treatment to cure keratoconus was corneal transplantation. Despite the development of cross-linking, keratoconus is still one of the most common reasons for corneal transplantation (Kelly 2011; Röck 2018). Thus, the diagnosis of keratoconus and especially its early diagnosis may help to avoid poor visual outcomes and possible corneal transplantation.

The diagnosis of keratoconus is based primarily on corneal topographic and tomographic analysis. This is performed in patients presenting with refractive errors, especially those whose vision can no longer be corrected fully with glasses, and in patients seeking corneal refractive surgery. A global consensus committee of ophthalmology experts concluded that "abnormal posterior ectasia, abnormal corneal thickness distribution, and clinical non-inflammatory corneal thinning are mandatory findings to diagnose keratoconus" (Gomes 2015). However, applying this definition in practice is not straightforward, since the consensus doesn't mention cut-offs or even parameters. The definition is open to the interpretation of the specialist. The ocular findings that may indicate subclinical keratoconus include abnormal keratometry readings and a distorted red reflex when using an ophthalmoscope, both indicating an irregular cornea. Detecting keratoconus at an early stage may be challenging, as patients are often asymptomatic, and there are few or no clinical signs. In later stages of the disease, clinical signs are visible during slit lamp examination, e.g. stromal thinning, conical protrusion of the cornea at the apex, Fleischer ring or Vogt striae (Zadnik 1996). As mentioned above, early diagnosis is crucial to monitor the disease and to treat it on time. A missed or delayed diagnosis could compromise the visual prognosis and may lead to a corneal transplant. Another challenge in the diagnosis of keratoconus is detecting an at-risk cornea or subclinical keratoconus in patients seeking corneal refractive surgery. Iatrogenic keratoectasia due to biomechanical decompensation may occur in these patients, if the disease is not detected (Giri 2017).

Currently, there is no accurate and objective method to detect keratoconus. An artificial intelligence (AI)-based tool for keratoconus detection would be helpful for ophthalmologists, optometrists and other eye-care professionals to make decisions on referral to cornea specialists for these patients.

AI is a growing field within ophthalmology, and is expected to play an important role in the diagnosis and characterisation of eye diseases. Recently, there has been an increasing interest in application of AI methods for diseases of the anterior segment (Ting 2020). This review will seek to determine if AI is a valid tool to diagnose keratoconus, as an aid for ophthalmologists.

Index test(s)

This review will evaluate the application of AI in the diagnosis of keratoconus. AI methods are already contributing to many aspects of human life and society, ranging from home automation, smart assistants (e.g. 'Siri', 'Google Assistant') and self-driving cars to facial recognition and automatic detection of “fake news” on social media. Notable progress with the use of AI has already been made in the field of medical image analysis, including applications in ophthalmology (Ting 2019).

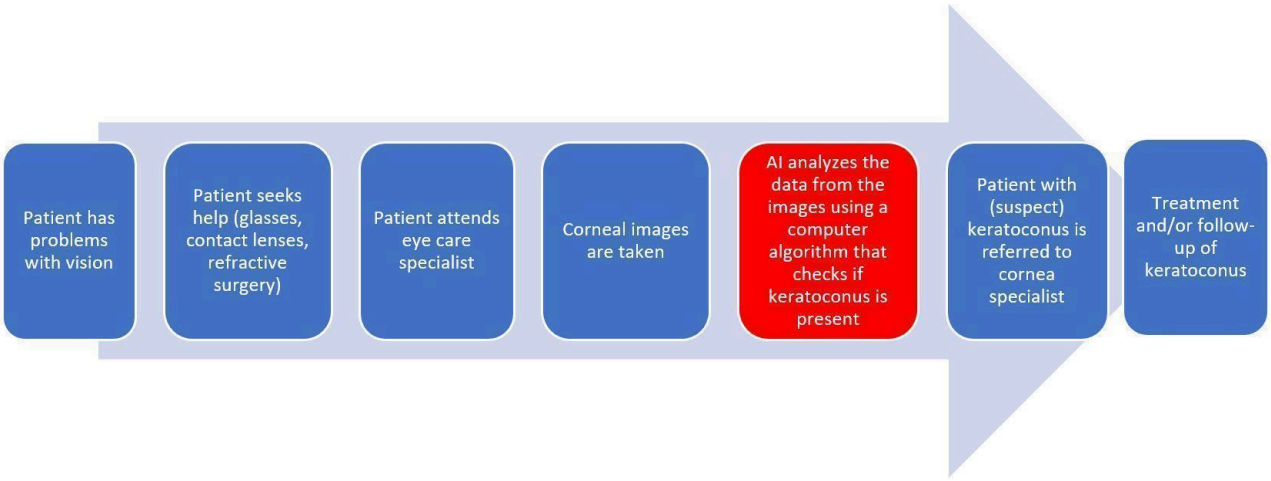
AI provides machines with the capability to adapt, reason and find solutions. Machine learning is a sub discipline of AI; It provides a machine to be able to learn from data and experience through algorithms. Examples of machine learning algorithms are support vector machine, random forest and decision tree. Deep learning is a sub discipline of machine learning, it uses neural networks much like how our own brain works. It has the ability to learn through pattern recognition and even to improve itself (LeCun 2015).

Initially, most of the AI research in ophthalmology was focused on the posterior segment. Multiple deep learning applications have been investigated for several common ophthalmic diseases, including diabetic retinopathy (Abràmoff 2016; Gargeya 2017; Gulshan 2016; Ting 2017), age-related macular degeneration (Grassmann 2018; Ting 2017), glaucoma (Shibata 2018) and retinopathy of prematurity (Brown 2018). Recently, more research has been conducted on the development of deep learning

applications for the anterior segment, in particular keratoconus (Ting 2020).

In keratoconus the AI algorithm will analyse images of the cornea using a computer and determine if keratoconus is present or not (see Figure 1). The images are acquired during the patient's visit. There are different devices that can take these corneal images, which are called topography or tomography images. Most devices such as Scheimpflug-based device or optical coherence tomography take both tomography and topography images. However, some devices only take topography images such as a placido disk device. The image is uploaded on to the computer, where the algorithm performs a series of analyses to come to a decision whether keratoconus is present or not.

Figure 1. Clinical pathway



The first step in developing an algorithm is collecting a representative dataset for keratoconus, which includes topographic or tomographic images of both keratoconus and healthy eyes. The dataset is then divided into training, validation, and test sets. The training set is used to determine the parameters or features of keratoconus via an optimisation procedure. The validation set is used for model selection (e.g. determining the best neural network architecture) and monitoring for overfitting, i.e. the algorithm is only applicable to the data on which it was trained. The independent test set is used for evaluation of the model, i.e. determining the performance of the model on unseen data. In principle, the test set should only be used once, after the model is developed and trained. When these three phases are completed, the algorithm will in theory be able to differentiate keratoconus eyes from healthy eyes.

We will systematically organize the included studies based on the main characteristics of the AI methodology (pre-processing techniques, core AI method and post-processing techniques), data that was used to train the model (patient inclusion criteria, number of training and testing cases/images, label/endpoint variable used for training) and evaluation (evaluation metric, reported performance on the independent test set).

Each AI algorithm has its own grading system to classify keratoconus and healthy eyes. Depending on the goal of the AI tool, screening or diagnosis, the thresholds of sensitivity and specificity will differ.

According to current guidelines, corneal tomography is the 'gold standard' to diagnose keratoconus (Gomes 2015). Previously, topography was considered to be the gold standard. However, topography only analyses the anterior corneal surface. Tomography analyses both anterior and posterior corneal surfaces and is able to create three-dimensional images. It is therefore more accurate than topography. In clinical practice, tomography and topography parameters are used to diagnose keratoconus, e.g. maximum keratometry, minimal pachymetry, astigmatism and asphericity. These show only a moderate correlation with keratoconus (Kanellopoulos 2013a; Kanellopoulos 2013b; Lopes 2012; Sedghipour 2012). Most devices also provide objective indices to help with the diagnosis of keratoconus, e.g. the keratoconus index, the index of surface variance and the inferior-superior index. However, these parameters and indices individually do not give enough information (Martínez-Abad 2017). The parameters and indices need to be combined and interpreted together. Unfortunately, not all ophthalmologists, optometrists or eye care professionals have these diagnostic skills. A second issue is the intra- and inter-observer variability in the diagnosis of keratoconus (Brunner 2018; Flynn 2016). AI can be a solution for both problems, since it will aid in the diagnosis of keratoconus by combining tomography and topography parameters and indices based on an enormous amount of data with ease, and reduce diagnostic variability. In other words, AI is a support tool to help with the interpretation of the topography and tomography images. It can help young ophthalmologists, ophthalmologists in non-academic centres, optometrists and other eye-care specialists to be able to diagnose the disease and refer to a cornea specialist. With the help of AI, keratoconus could be detected sooner, so follow-up can start earlier and possible progression can be detected before visual loss. Thus, patients may be treated on time, which in turn would lead to a better visual outcome.

Clinical pathway

The clinical pathway to diagnose keratoconus is based on clinical examination including visual acuity testing, slit lamp examination of the anterior segment and corneal imaging. Corneal imaging is performed in patients presenting with refractive errors, especially those whose vision can no longer be corrected fully with glasses,

patients seeking corneal refractive surgery or those suspected of having keratoconus, referred by ophthalmologists, optometrists and other eye-care professionals.

There are different devices for corneal imaging e.g., Placido topography, Scheimpflug tomography, or slit-scanning tomography. The interpretation of the images can be challenging, and the signs of keratoconus can be subtle for general ophthalmologists, optometrists and other eye-care professionals. In current practice, the ophthalmologist will analyze the corneal images. They will look for patterns and evaluate certain parameters depending on the device available, such as keratometry, elevation and pachymetry parameters. Since the global consensus mentions no cut-offs in the definition of keratoconus, specialists need to rely on their knowledge and experience to diagnose keratoconus, making decisions more subjective.

When the diagnosis of keratoconus is made, the patient will need regular follow-up to check if the disease progresses. The global consensus document states that treatment should be done when there is documented clinical progression (Gomes 2015). Clinical progression is defined in Gomes 2015 as follows: a consistent change in at least two of the following parameters where the magnitude of the change is above the normal noise of the testing system:

1. steepening of the anterior corneal surface;
2. steepening of the posterior corneal surface;
3. thinning and/or an increase in the rate of corneal thickness change from the periphery to the thinnest point.

As with the definition of keratoconus, this definition is open to interpretation. There are no cut-offs, time intervals or specific parameters mentioned.

A missed diagnosis of keratoconus could lead to a delayed treatment, poor visual outcome and a greater risk for corneal transplant. All of this in turn has an impact on quality of life of the patients because it affects young people who are active and in their primary income-earning years.

The same corneal images that are analyzed by clinicians will be uploaded in a computer and analyzed by the AI algorithm. AI based on a large ophthalmic dataset can achieve high accuracy in distinguishing a normal cornea from a keratoconus cornea by analyzing the topography or tomography images (Lin 2019 ; Lopes 2019). Since the global consensus does not give an accurate definition of keratoconus, or of the progression of the disease, AI could be helpful in making this decision. It could help with early diagnosis of keratoconus, so patients could be monitored. As a result progression, could be detected sooner. Once progression is detected and confirmed by the specialist, the patient would receive corneal cross-linking to halt the deterioration of the disease, which in turn would lead to better visual prognosis and less risk of corneal transplants. In other words, AI is a support tool to help with the interpretation of topography and tomography images. Since the cornea specialist is still responsible for the diagnosis, the first role of AI will be as triage to make decisions on referral.

The conditions to implement an AI algorithm in clinical practice are as follows: firstly, the algorithm needs to be efficient and the analysis of an image should not take very much time; secondly, it should give one clear indication, and the output should be clear in terms of diagnosis. In conclusion, the AI algorithm should be able to analyze the topography or tomography image in a few seconds, and give a clear indication whether keratoconus is present or absent.

Devices that measure biomechanical properties, such as the Corvis ST or the ORA, will not be included in this review.

Rationale

AI is a rapidly growing field in ophthalmology with numerous new developments in the detection of keratoconus (Ting 2020). It is important that we have reliable evidence regarding the accuracy of these developments. This review will give a clear overview of the different AI detection tools and their accuracy.

Corneal imaging devices are becoming increasingly sophisticated, and with the help of AI algorithms they are able to detect keratoconus earlier. AI uses a vast amount of data to learn characteristic features of keratoconus. It is able to process thousands of images in a short amount of time to learn how to detect the disease, in comparison to an ophthalmologist, who needs years of practice. AI will help ophthalmologists, optometrists and eye-care professionals in the diagnosis of keratoconus and potentially help to diagnose it in an early stage. This is beneficial for the patient, because they may have a better visual outcome, which in turn has an effect on the quality of life. Also the financial consequences are important, not only healthcare costs but also personal costs.

There are, however, limitations to AI. The accuracy of the algorithms relies on the generalizability of the training sets. If training sets did not contain enough data or enough variability in data, the algorithms will miss a diagnosis, due to insufficient learning (LeCun 2015).

A recent narrative review suggests that AI may be a reliable tool; however, it only gave a summation of the different included articles and did not compare the AI algorithm accuracies (Lin 2019). Another review discusses AI in the anterior segment, and also mentions the detection of keratoconus (Ting 2020). However, this review also did not compare the accuracies of the algorithms.

There is a need for a reliable overview of current knowledge and accuracy of the different existing AI algorithms.

OBJECTIVES

The primary objective is to assess the diagnostic accuracy of AI algorithms in the detection of keratoconus in patients presenting with refractive errors, especially those whose vision can no longer be corrected fully with glasses, patients seeking corneal refractive surgery or those suspected of having keratoconus. AI could help ophthalmologists, optometrists and other eye-care professionals to make decisions on referral to cornea specialists for these patients.

Secondary objectives

To compare different AI algorithms, e.g. neural networks, decision trees, support vector machines.

To assess potential causes of heterogeneity in diagnostic performance across studies, according to the following:

- index test methodology: pre-processing techniques, core AI method and postprocessing techniques;
- sources of input to train algorithms: topography and tomography images from Placido-disc system or Scheimpflug system or slit-scanning system or OCT, number of training and testing cases/images, label/endpoint variable used for training;
- study setting;
- study design, retrospective or prospective studies;
- ethnicity, or geographic area as its proxy;
- different index test positivity criteria provided by topography or tomography device;
- reference standard used, topography or tomography, one or two cornea-specialists;
- definition of keratoconus used;
- mean age;
- patient recruitment;
- severity of keratoconus:
  - clinically manifest keratoconus
  - subclinical keratoconus

METHODS

Criteria for considering studies for this review

Types of studies

The following study designs will be included:

- cross-sectional studies;
- diagnostic case-control studies, including both prospective and retrospective studies.

Included studies will be organized based on the main characteristics of the AI methodology (pre-processing techniques, core AI method and post-processing techniques), data that was used to train the model (patient inclusion criteria, number of training and testing cases/images, label/endpoint variable used for training) and evaluation (evaluation metric, reported performance on the independent test set).

Participants

We aim to include patients with refractive errors, whose vision cannot be fully corrected with glasses, patients seeking refractive surgery or patients suspected of keratoconus, for whom a decision is to be made on referral to cornea specialists. However, research in this field is still in its early stages and we will accept studies which do not satisfy this optimal definition of participants, including case-control studies, in which keratoconus patients and healthy controls are included based on different sets of criteria.

Since keratoconus can progress until the fourth decade of life, patients up to the age of 50 years will be included.

Index tests

We will include studies reporting accuracy data for tests using automated diagnosis. All AI algorithms that are developed for the analysis of corneal topography or tomography to detect keratoconus will be included.

Target conditions

The target condition that the AI algorithms will need to diagnose is keratoconus of any stage. When studies report accuracy for multiple severity levels, we will preferably extract data for at least mild severity. In fact, 'fruste' keratoconus is generally non-progressive, or very slowly progressive.

Reference standards

The reference standard for keratoconus is topography or tomography. These examinations are routinely done in patients who come for refractive surgery or patients referred to an ophthalmologist for suspected keratoconus. The corneal images should be analysed and interpreted independently by two or more cornea specialists. We will accept studies using only one cornea specialist for diagnosis as a low quality reference standard.

Topography examines the anterior corneal surface. The Placido disc system is a device that uses topography. Concentric rings of light are projected on the cornea. Thousands of points along these concentric rings are analysed, which in turn is translated in the curvature of the anterior corneal surface (Fan 2018). The main parameters measured by Placido systems are the maximum keratometry, steep keratometry, flat keratometry and astigmatism.

Tomography examines both the anterior and posterior corneal surfaces. The Scheimpflug system uses a single rotating Scheimpflug camera, e.g. Pentacam (Oculus GmbH, Wetzlar, Germany), a single rotating Scheimpflug camera combined with Placido disk topography (Sirius, CSO, Italy) or uses a dual-Scheimpflug camera with Placido disc technology incorporated to improve curvature information on the central cornea, e.g. the Galilei (Ziemer, Biel, Switzerland). Another device that uses tomography is the slit-scanning system; this is an elevation-based method for assessment of topography. Multiple complimentary slits are used to perform an assessment of the corneal surface, e.g. the Orbscan IIz (Bausch & Lomb, Rochester, NY). Next to the keratometry, which is also measured by the Placido systems, the Scheimpflug system and slit-scanning system measure the corneal elevation and pachymetry.

Optical Coherence Tomography (OCT) also examines both the anterior and posterior corneal surfaces. Anterior Segment OCT (AS-OCT) uses low-coherence interferometry to assess the cornea and the anterior segment. The low-coherent light is emitted and split by a inferometer in a reference beam and a probe beam (Wojtkowski 2010). The latter is backscattered from the different corneal layers. The echo time delay is measured and transformed into two- or three- dimensional images by the OCT (Subhash 2013). Recently, a new instrument, MS-39 (CSO, Italy), which combines Placido disc corneal topography with high resolution OCT-based anterior segment tomography, has been developed. This AS-OCT measures the keratometry, elevation and pachymetry and other parameters.

Search methods for identification of studies

Electronic searches

The Cochrane Eyes and Vision Information Specialist will search the following electronic databases. There will be no restrictions to language or date of publication.

- Cochrane Central Register of Controlled Trials (CENTRAL; latest issue) (which contains the Cochrane Eyes and Vision Trials Register) in the Cochrane Library (Appendix 1).
- MEDLINE Ovid (1946 to present) (Appendix 2).
- Embase Ovid (1980 to present) (Appendix 3).
- System for Information on Grey Literature in Europe (OpenGrey) (1995 to present) (Appendix 4).
- ISRCTN registry (www.isrctn.com/editAdvancedSearch) (Appendix 5).
- US National Institutes of Health Ongoing Trials Register - ClinicalTrials.gov (www.clinicaltrials.gov) (Appendix 6).
- World Health Organization International Clinical Trials Registry Platform (www.who.int/ictrp) (Appendix 7).

Currently, the Aggressive Research Intelligence Facility database (ARIF) is unavailable and is not being updated. If this resource becomes available again it will be searched on subsequent updates of this review.

Searching other resources

We will search the reference lists of the review's included studies.

Data collection and analysis

Selection of studies

Two review authors (MV and EF) will independently evaluate the articles from the search. We will review the titles of the studies and eliminate irrelevant articles. After this, the review authors will assess the full texts of the remaining articles to check if they meet the inclusion criteria. When the review authors disagree, they will discuss this to come to an agreement. If necessary, the other review authors will be asked to join the discussion.

Data extraction and management

The two review authors will independently extract data from the selected articles with a standardized collection form. The following data will be extracted from each included study: study design, study population, definition of keratoconus, reference standard, the index tests, description of architecture and training mechanisms, the ground truth (one observer versus multiple observers), the size of datasets used and data required to fill in a 2 x 2 diagnostic contingency table for each index test.

We will compare the data collected independently by the two review authors, and resolve any discrepancies through discussion and consensus. If we need to obtain further data from a paper or if there is missing data, we will try to contact the study author for further clarification.

When there are multiple AI algorithms in an article we will select the one with the highest YOUDEN index. We are aware that this selection may inflate accuracy, especially in smaller studies, and we will highlight this as a limitation, which may on the other hand reduce redundancy and be acceptable in this early research stage. Examples of algorithms are random forest, support vector machine, decision tree and neural network.

Assessment of methodological quality

Two authors (MV and EF) will independently assess the included studies for bias using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool, as described in chapter 9 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Reitsma 2009). The QUADAS-2 tool has four assessment domains: patient selection, index test, reference test and flow and timing. Each domain has signalling questions to assess the risk of bias. The first three domains are also assessed on applicability.

Regarding the direct comparison of different AI tests we added two signalling questions in our QUADAS-2 guidance (Appendix 8).

Statistical analysis and data synthesis

Statistical analysis and data synthesis will be conducted in accordance with Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2010).

Initially, we will present data in a 2 x 2 table, showing cross-classification of the index test result versus the reference standard outcome. For each index test, in all studies, we will calculate the sensitivity and specificity with a 95% confidence interval. In order to visually evaluate the variation in calculations of sensitivity and specificity we will use Review Manager 5 (RevMan 5) (Review Manager 2014) to generate coupled forest plots and ROC (receiver operating characteristic) plots.

Since a definite threshold that is comparable across studies is unlikely to be available in AI studies, we will use a hierarchical summary ROC (HSROC) model (Macaskill 2010) and we will estimate average sensitivity at fixed specificity values according to cut-offs for tertiles of specificity.

We will consider direct comparison between the index tests, (different types or data sources for AI) if sufficient data are available. We will conduct these analyses with a test covariate in the HSROC model. If few studies provide data for comparisons between tests, we will compute the relative diagnostic odds ratio (DOR) using a simplified HSROC model assuming symmetrical underlying summary ROC (SROC) curves. If data are very sparse we will use a fixed-effects model.

We will conduct analyses using the 'metadas' user-written command in SAS software (SAS software) and we will make predictions at fixed specificities using NLMIXED procedure post-estimation commands.

Investigations of heterogeneity

In order to investigate heterogeneity, where data are available, we will add covariates in a meta-regression, using the sources presented in the Objectives , following guidance provided in Macaskill 2010 . All covariates will be used as categorical variables.

Sensitivity analyses

We will create a summary table to report pre-specified sensitivity analysis, specifically regarding QUADAS 2 domains with most studies at high risk of bias. These will cover the case-control studies, which have a high risk of bias.

Additional sensitivity analyses identified during the review will be added if needed and reported as changes to the review protocol

Assessment of reporting bias

We will assess reporting biases if a study protocol is available for the included studies. We will attempt to maximise data collection

through comprehensive search methods and contact with study authors to request further information or data that are needed for a study to meet our inclusion criteria.

ACKNOWLEDGEMENTS

Cochrane Eyes and Vision (CEV) will create and run the electronic searches. We thank Vito Romano and Roy Schwartz for their comments on this protocol and Anupa Shah, Managing Editor for CEV for her assistance throughout the review process.

REFERENCES

Additional references

Abràmoff 2016

Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative Ophthalmology and Visual Science* 2016;**57**(13):5200-6.

Brown 2018

Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmology* 2018;**136**(7):803-10.

Brunner 2018

Brunner M, Czanner G, Vinciguerra R, Romano V, Ahmad S, Batterbury M, et al. Improving precision for detecting change in the shape of the cornea in patients with keratoconus. *Scientific Reports* 2018;**8**(1):1-7.

Fan 2018

Fan R, Chan TC, Prakash G, Jhanji V. Applications of corneal topography and tomography: a review. *Clinical and Experimental Ophthalmology* 2018;**42**(2):133-46.

Ferdi 2019

Ferdi AC, Nguyen V, Gore DM, Allan BD, Rozema JJ, Watson SL. Keratoconus natural progression: a systematic review and meta-analysis of 11 529 eyes. *Ophthalmology* 2019;**126**(7):935-45.

Flynn 2016

Flynn TH, Sharma DP, Bunce C, Wilkins MR. Differential precision of corneal Pentacam HR measurements in early and advanced keratoconus. *British Journal of Ophthalmology* 2016;**100**(9):1183-7.

Gargeya 2017

Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 2017;**124**(7):962-9.

Giri 2017

Giri P, Azar DT. Risk profiles of ectasia after keratorefractive surgery. *Current Opinion in Ophthalmology* 2017;**28**(4):337-42.

Gomes 2015

Gomes JA, Tan D, Rapuano CJ, Belin MW, Ambrósio RJ, Guell JL, et al. Global consensus on keratoconus and ectatic diseases. *Cornea* 2015;**34**(4):359-69.

Grassmann 2018

Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology* 2018;**125**(9):1410-20.

Gulshan 2016

Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Journal of the American Medical Association* 2016;**316**(22):2402-10.

Hashemi 2020

Hashemi H, Heydarian S, Hooshmand E, Saatchi M, Yekta A, Aghamirsalim M, et al. The Prevalence and Risk Factors for Keratoconus: A Systematic Review and Meta-Analysis. *Cornea* 2020;**39**(2):263-70.

Hayes 2012

Hayes S, Khan S, Boote C, Kamma-Lorger CS, Dooley E, Lewis J, et al. Depth profile study of abnormal collagen orientation in keratoconus corneas. *Archives of Ophthalmology* 2012;**130**(2):251-2.

Kanellopoulos 2013a

Kanellopoulos AJ, Asimellis G. Revisiting keratoconus diagnosis and progression classification based on evaluation of corneal asymmetry indices, derived from Scheimpflug imaging in keratoconic and suspect cases. *Clinical Ophtholmology* 2013;**7**:1539-48. [DOI: 10.2147/opth.S44741]

Kanellopoulos 2013b

Kanellopoulos AJ, Moustou V, Asimellis G. Evaluation of visual acuity, pachymetry and anterior-surface irregularity in keratoconus and crosslinking intervention follow-up in 737 cases. *International Journal of Keratoconus and Ectatic Corneal Diseases* 2013;**2**(3):95.

Kelly 2011

Kelly T, Williams KA, Coster DJ. Corneal transplantation for keratoconus: a registry study. *Archives of Ophthalmology* 2011;**129**(6):691-7.

LeCun 2015

LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;**521**(7553):436-44.

Lin 2019

Lin SR, Ladas JG, Bahadur GG, Al-Hashimi S, Pineda R. A review of machine learning techniques for keratoconus detection and refractive surgery screening. *Seminars in Ophthalmology* 2019;**34**(4):317-26.

Lopes 2012

Lopes BT, Ramos IC, Faria-Correia F, Luz A, Freitas Valbon B, Belin MW, et al. Correlation of topometric and tomographic indices with visual acuity in patients with keratoconus. *International Journal of Keratoconus and Ectatic Corneal Diseases* 2012;**1**(3):167-72.

Lopes 2019

Lopes BT, Eliasy A, Ambrosio R. Artificial Intelligence in corneal diagnosis: Where are we? *Current Ophthalmology Reports* 2019;**7**(3):204-11.

Macaskill 2010

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

Martínez-Abad 2017

Martínez-Abad A, Pinero DP. New perspectives on the detection and progression of keratoconus. *Journal of Cataract and Refractive Surgery* 2017;**43**(9):1213-27.

Meek 2005

Meek KM, Tuft SJ, Huang Y, Gill PS, Hayes S, Newton RH, et al. Changes in collagen orientation and distribution in keratoconus corneas. *Investigative Ophthalmology and Visual Science* 2005;**46**(6):1948-56.

Rabinowitz 2021

Rabinowitz YS, Galvis V, Tello A, Rueda D, García JD. Genetics vs chronic corneal mechanical trauma in the etiology of keratoconus. *Experimental Eye Research* 2021;**202**:108328.

Reitsma 2009

Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2009. Available from: http://srdta.cochrane.org/.

Review Manager 2014 [Computer program]

The Cochrane Collaboration Review Manager 5 (RevMan 5). Version 5.4. Copenhagen: Nordic Cochrane Centre: The Cochrane Collaboration, 2014.

Röck 2018

Röck T, Bartz-Schmidt KU, Röck D. Trends in corneal transplantation at the University Eye Hospital in Tübingen, Germany over the last 12 years: 2004–2015. *Plos One* 2018;**13**(6):e0198793.

SAS software [Computer program]

SAS Institute Inc SAS software. Version 9.2. North Carolina, USA: SAS Institute Inc, Accessed prior to 28 September 2021.

Sedghipour 2012

Sedghipour MR, Sadigh AL, Motlagh BF. Revisiting corneal topography for the diagnosis of keratoconus: use of Rabinowitz's KISA% index. *Clinical Ophthalmology* 2012;**6**:181. [DOI: 10.2147/OPTH.S24219]

Shibata 2018

Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Scientific Reports* 2018;**8**(1):1-9.

Subhash 2013

Subhash HM, Wang RK. Optical coherence tomography: technical aspects. In: Liang R, editors(s). Biomedical Optical Imaging Technologies. Biological and Medical Physics, Biomedical Engineering. Berlin, Heidelberg: Springer, 2013:163-212. [DOI: https://doi.org/10.1007/978-3-642-28391-8_5]

Sykakis 2015

Sykakis E, Karim R, Evans JR, Bunce C, Amissah-Arthur KN, Patwary S, et al. Corneal collagen cross-linking for treating keratoconus. *Cochrane Database of Systematic Reviews* 2015, Issue 3. Art. No: CD010621. [DOI: 10.1002/14651858.CD010621.pub2]

Ting 2017

Ting DS, Cheung CY, Lim G, Tan GS, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *Journal of the American Medical Association* 2017;**318**(22):2211-33.

Ting 2019

Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, Wong TY. Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology* 2019;**103**(2):167-75.

Ting 2020

Ting DS, Foo VH, Yang LW, Sia ST, Ang M, Lin H, et al. Artificial intelligence for anterior segment diseases: emerging applications in ophthalmology. *British Journal of Ophthalmology* 2020;**105**(2):158-68.

Wojtkowski 2010

Wojtkowski M. High-speed optical coherence tomography: basics and applications. *Applied Optics* 2010;**49**(16):D30-61.

Zadnik 1996

Zadnik K, Barr JT, Gordon MO, Edrington TB. Biomicroscopic signs and disease severity in keratoconus. Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study Group. *Cornea* 1996;**15**(2):139-46.

APPENDICES

Appendix 1. CENTRAL search strategy

#1 MeSH descriptor: [Keratoconus] this term only
#2 keratoconus*
#3 cornea* near/5 ectatic*
#4 cornea* near/5 ectasia
#5 conical near/2 cornea*

#6 cornea* near/2 thinning
#7 #1 OR #2 OR #3 OR #4 OR #5 OR #6
#8 MeSH descriptor: [Artificial Intelligence] this term only
#9 MeSH descriptor: [Deep Learning] this term only
#10 MeSH descriptor: [Machine Learning] explode all trees
#11 MeSH descriptor: [Neural Networks, Computer] this term only
#12 MeSH descriptor: [Algorithms] this term only
#13 MeSH descriptor: [Decision Trees] this term only
#14 MeSH descriptor: [Automation] this term only
#15 MeSH descriptor: [Databases, Factual] this term only
#16 MeSH descriptor: [Electronic Data Processing] this term only
#17 artificial NEAR/1 intelligence
#18 (deep or machine) NEAR/2 learning
#19 vector NEAR/3 machine
#20 AI or DL or DLS
#21 (deep or convolutional or neural) NEAR/3 network*
#22 automat* NEAR/2 (screen* or detect* or diagnos* or algorithm* or identif* or grading or graded or method*)
#23 Bagging
#24 Naive NEAR/1 Bayes
#25 Multilayer NEAR/1 Perceptron
#26 (multi-layer NEAR/1 perceptron) or MLP
#27 Radial NEAR/1 Basis NEAR/1 Function
#28 Random NEAR/1 Forest
#29 Ensemble NEAR/1 Selection
#30 (Ada or gradient) NEAR/1 boost*
#31 LASSO
#32 Elastic NEAR/1 Net
#33 genetic NEAR/1 algorithm*
#34 (decision or classification or regression or probability or model*) NEAR/3 tree*
#35 logistic* NEAR/2 regression
#36 augment* NEAR/1 clinical NEAR/1 decision* NEAR/1 mak*
#37 nearest NEAR/1 (neighbor or neighbour)
#38 fuzzy NEAR/3 (logit or logic or logistic)
#39 kernel
#40 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39
#41 #7 AND #40

Appendix 2. MEDLINE Ovid search strategy

1. Keratoconus/
2. keratoconus$.tw.
3. (cornea$ adj5 ectatic$).tw.
4. (cornea$ adj5 ectasia).tw.
5. (conical adj2 cornea$).tw.
6. (cornea$ adj2 thinning).tw.
7. or/1-6
8. artificial intelligence/
9. deep learning/
10. exp machine learning/
11. "neural networks (computer)"/
12. fuzzy logic/
13. algorithms/
14. decision tree/
15. automation/
16. databases, factual/
17. information processing/
18. (artificial adj1 intelligence).tw.
19. ((deep or machine) adj2 learning).tw.
20. (vector adj3 machine).tw.
21. (AI or DL or DLS).tw.
22. ((deep or convolutional or neural) adj3 network$).tw.

23. (automat$ adj2 (screen$ or detect$ or diagnos$ or algorithm$ or identif$ or grading or graded or method$)).tw.
24. Bagging.tw.
25. (Naive adj1 Bayes).tw.
26. (Multilayer adj1 Perceptron).tw.
27. ((multi-layer adj1 perceptron) or MLP).tw.
28. (Radial adj1 Basis adj1 Function).tw.
29. (Random adj1 Forest).tw.
30. (Ensemble adj1 Selection).tw.
31. ((Ada or gradient) adj1 boost$).tw.
32. LASSO.tw.
33. (Elastic adj1 Net).tw.
34. (genetic adj1 algorithm$).tw.
35. ((decision or classification or regression or probability or model$) adj3 tree$).tw.
36. (logistic$ adj2 regression).tw.
37. (augment$ adj1 clinical adj1 decision$ adj1 mak$).tw.
38. (nearest adj1 (neighbor or neighbour)).tw.
39. (fuzzy adj3 (logit or logic or logistic)).tw.
40. kernel.tw.
41. or/8-40
42. 7 and 41

Appendix 3. Embase Ovid search strategy

1. keratoconus/
2. keratoconus$.tw.
3. (cornea$ adj5 ectatic$).tw.
4. (cornea$ adj5 ectasia).tw.
5. (conical adj2 cornea$).tw.
6. (cornea$ adj2 thinning).tw.
7. or/1-6
8. artificial intelligence/
9. deep learning/
10. machine learning/
11. supervised machine learning/ or support vector machine/ or unsupervised machine learning/
12. perceptron/
13. artificial neural network/
14. convolutional neural network/
15. deep neural network/
16. automated pattern recognition/
17. decision tree/
18. detection algorithm/
19. learning algorithm/
20. classification algorithm/
21. data classification/
22. disease classification/
23. disease simulation/
24. automation/
25. information processing/
26. feature extraction/
27. bayesian learning/
28. fuzzy system/
29. k nearest neighbor/
30. kernel method/
31. random forest/
32. (artificial adj1 intelligence).tw.
33. ((deep or machine) adj2 learning).tw.
34. (vector adj3 machine).tw.
35. (AI or DL or DLS).tw.
36. ((deep or convolutional or neural) adj3 network$).tw.
37. (automat$ adj2 (screen$ or detect$ or diagnos$ or algorithm$ or identif$ or grading or graded or method$)).tw.
38. Bagging.tw.
39. (Naive adj1 Bayes).tw.

40. (Multilayer adj1 Perceptron).tw.
41. ((multi-layer adj1 perceptron) or MLP).tw.
42. (Radial adj1 Basis adj1 Function).tw.
43. (Random adj1 Forest).tw.
44. (Ensemble adj1 Selection).tw.
45. ((Ada or gradient) adj1 boost$).tw.
46. LASSO.tw.
47. (Elastic adj1 Net).tw.
48. (genetic adj1 algorithm$).tw.
49. ((decision or classification or regression or probability or model$) adj3 tree$).tw.
50. (logistic$ adj2 regression).tw.
51. (augment$ adj1 clinical adj1 decision$ adj1 mak$).tw.
52. (nearest adj1 (neighbor or neighbour)).tw.
53. (fuzzy adj3 (logit or logic or logistic)).tw.
54. kernel.tw.
55. or/8-54
56. 7 and 55

Appendix 4. OpenGrey search strategy

keratoconus AND (Artificial intelligence OR deep learning OR machine learning)

Appendix 5. ISRCTN search strategy

keratoconus AND (Artificial intelligence OR deep learning OR machine learning)

Appendix 6. ClinicalTrials.gov search strategy

keratoconus AND (Artificial intelligence OR deep learning OR machine learning)

Appendix 7. WHO ICTRP search strategy

keratoconus AND Artificial intelligence OR keratoconus AND deep learning OR keratoconus AND machine learning

Appendix 8. QUADAS 2 guidance

DOMAIN	Low risk/concern	Unclear	High risk/concern
PATIENT SELECTION	Describe methods of patient selection: Describe included patients (prior testing, presentation, intended use of index test and setting):		
Was a consecutive or random sample of patients enrolled?	Consecutive sampling or random sampling seeking refractive error correction or refractive surgery in eye services	Unclear whether consecutive or random sampling used	Selection of non-consecutive subjects
Was a case-control design avoided?	No selective recruitment of people with or without keratoconus	Unclear selection mechanism	Selection of either cases or control in a predetermined, non-random fashion; or enrichment of the cases from a selected population
Did the study avoid inappropriate exclusions?	Exclusions are detailed and felt to be appropriate, e.g. people already diagnosed with keratoconus or with other corneal diseases	Exclusions are not detailed (pending contact with study authors)	Inappropriate exclusions are reported, e.g. of people with borderline index test results.
Risk of bias: Could the selection of patients have introduced bias?	'no' for any of the above		

(Continued)

Concerns regarding applicability: Are there concerns that the included patients do not match the review question?	Inclusion of patients seeking refractive error correction or refractive surgery in primary or secondary care eye services	Unclear inclusion criteria	Inclusion of patients attending cornea services for known disease, population-based studies, registry based studies
INDEX TEST	Describe the index test and how it was conducted and interpreted:		
Were the index test results interpreted without knowledge of the results of the reference standard?	Test performed "blinded" or "independently and without knowledge of" reference standard results are sufficient and full details of the blinding procedure are not required; or clear temporal pattern to the order of testing that precludes the need for formal blinding	Unclear whether results are interpreted independently	Reference standard results available to those who conducted or interpreted the index tests
If a threshold was used, was it pre-specified?	The study authors declare that the selected cut-off used to dichotomise data was specified a priori, or a protocol is available with this information.	No information on preselection of index test cut-off values.	A study is classified at higher risk of bias if the authors define the optimal cut-off post-hoc based on their own study data.
Risk of bias: Could the conduct or interpretation of the index test have introduced bias?	'no' for any of the above		
Concerns regarding applicability: Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Tests used and testing procedure clearly reported and tests executed by personnel with sufficient training.	Unclear execution of the tests or unclear study personnel profile, background and training.	Tests used are not validated or study personnel is insufficiently trained.
REFERENCE STANDARD	Describe the reference standard and how it was conducted and interpreted:		
Is the reference standard likely to correctly classify the target condition?	Topography and/or tomography interpreted independently by 2 or more cornea specialists.	Topography and/or tomography interpreted by cornea specialists, but not enough details to adjudicate 'yes' or 'no'	Topography and/or tomography interpreted by only one cornea specialist.
Were the reference standard results interpreted without knowledge of the results of the index test?	Reference standard performed "blinded" or "independently and without knowledge of" index test results are sufficient and full details of the blinding procedure are not required; or clear temporal pattern to the order of testing that precludes the need for formal blinding.	Unclear whether results are interpreted independently	Index test results available to those who conducted the reference standard
Risk of bias: Could the reference standard, its conduct, or its interpretation have introduced bias?	'no' for any of the above		
Concerns regarding applicability: Are there concerns	Not applicable for this review		

(Continued)

that the target condition as defined by the reference standard does not match the review question?		
FLOW AND TIMING	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram): Describe the time interval and any interventions between index test(s) and reference standard	
Was there an appropriate interval between index test(s) and reference standard?	No more than three months between index and reference test execution	More than three months between index and reference test execution.
Did all patients receive a reference standard?	All participants receiving the index test are verified with the reference standard.	Not all participants receving the index test are verified with the reference standard.
Did all patients receive the same reference standard?	Not applicable for this review	
Were all patients included in the analysis?	The number of participants included in the study match the number in analyses or participants with undefined or borderline test results are excluded.	The number of participants included in the study does not match the number in analyses or participants with undefined or borderline test results are excluded.
Risk of bias: Could the patient flow have introduced bias?	'no' for any of the above	
ADDITIONAL QUESTIONS	These question concern the direct comparisons between AI tests	
Were different AI tests developed and interpreted without knowledge of each other?	Different AI tests were developed and interpreted “blinded” or “independently and without knowledge of” the results of each other.	Different AI tests were developed or their results interpreted with knowledge of the results of each other.
Are the proportions and reasons for missing data similar for all index tests?	Missing data and their causes were similar for each AI test.	The amount of missing data or their causes differed between AI tests.

CONTRIBUTIONS OF AUTHORS

MMSV: Development of the protocol
EF: Critical review on clinical sections of the protocol
MV: Critical review on artificial intelligence sections of the protocol, providing advice on artificial intelligence
EL: Critical review of statistical section, providing advice on statistics
TB: Critical review on the protocol
RM: Critical review on clinical sections of the protocol
RMMAN: Critical review on clinical sections of the protocol
GV: Development of statistical section and critical review of all protocol sections
MMD: Critical review on all protocol sections

DECLARATIONS OF INTEREST

MMSV, EF, MV, EL, TB, RM, GV, MMD: have no conflicts of interest.

RMMAN is involved in clinical trials of medical devices, e.g.. intraocular lenses in several companies.

SOURCES OF SUPPORT

Internal sources

- University Eye Clinic Maastricht, Maastricht University Medical Center (MUMC+), Maastricht, Netherlands

  Authors' place of employment (MMSV, TB, RMMAN, MMD)
- Department of Neurosciences, Psychology, Drug Research and Child Health (NEUROFARBA), University of Florence, Florence, Italy

  Authors' place of employment (RM, GV)
- Department of Neurosciences, Psychology, Pharmacology and Child Health, University of Florence, Florence, Italy

  Authors' place of employment (EF)
- Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

  Authors' place of employment (MV)
- Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

  Authors' place of employment (EL)
- Department of Neuroscience, School for Mental Health and Neuroscience (MHeNS), Maastricht University, Maastricht, Netherlands

  Authors' place of employment (MMSV)
- Centre for Public Health, Queen’s University Belfast, UK

  Authors' place of employment (GV)

External sources

- Public Health Agency, UK

  This protocol was supported by the HSC Research and Development (R&D) Division of the Public Health Agency which funds the Cochrane Eyes and Vision editorial base at Queen's University Belfast.