Griffith UNIVERSITY
Queensland, Australia
Centre for Applied Health Economics

Evidence Evaluation for the Diagnostic Accuracy of Iridology: Systematic Review Research Protocol

FINAL

Version 2.0

Date 28 March 2022

Prepared for

National Health and Medical Research Council

Prepared by

Centre for Applied Health Economics

Griffith University

Contents

Authors

McFadden, K. [1], Maujean, A.[1], Downes, M.[1]

1. Centre for Applied Health Economics, Griffith University, Australia

Contact person

Associate Professor Martin Downes

Room 2.36, Sir Samuel Griffiths Building (N78)

Kessels Road, Nathan, 4111

QLD, Australia

Email: m.downes@griffith.edu.au

Ph: +61 7 3735 9106

1 Background

1.1 Description of condition and setting

Iridology is a method which examines the patterns, colours and structure of a person's iris in order to determine information about the wider health of the body (1). The practice of analysing the iris dates back centuries, but modern iridology was popularised by Dr Ignatz von Peczely (1). Classified as a complementary and alternative medicine (CAM) in Australia, iridology is generally performed at holistic therapy practices and there are options for non-medical training certifications for practitioners.

1.2 Description of intervention

Iridology is a diagnostic system based on the premise that every organ has corresponding location(s) within the iris of the eye, in which structural and pigmentation components can serve as indicators for condition(s) and/or problem(s) in the human body. Practitioners of iridology examine the iris and capture images of the iris to identify the indicators for conditions. The iridologist then compares observations of an individual's iris to iris charts, which are "maps" that divide the iris into regions linked to specific organs or body parts (1). Typically there are 80-90 areas identified on topographic charts of the iris, with minor variations based on different schools of thought (1).

There are multiple options for capturing images of the iris for the purpose of iridology. These include physical observation, or images/scans obtained via digital cameras, integrated and/or adapted iridoscopes (which are purpose-built cameras for iris photography), other types of illumination and image recording, and image editing software (e.g., Adobe Photoshop, or specific software for images of irises) (2-4). Interpretation of iris images are completed by a practitioner of iridology. Alternatively, it has been proposed that interpretation can be done by a computer via the use of machine learning algorithms (MLAs), though MLAs are not currently used in practice in Australia.

With recent advancements in technology, machine learning is being increasingly used in diagnostic medicine, with successful implementation in areas like skin cancer diagnosis (5, 6). MLAs use experience and exposure to data to "learn" patterns in order to improve predictive capabilities (6). For iridology, MLAs are developed so that a computer is presented with example inputs (iris images) and outputs (diagnosis), with the goal being to develop a general, pattern-based rule that maps inputs to outputs. There are multiple types of MLAs (also called a classifier or classification system), including: Cubic Support Vector Machine (CSVM), Median Gaussian Support Vector Machine (MGSVM), Quadratic Support Vector Machine (QSVM), Fine Gaussian Support Vector Machine (FGSVM), Boosted Tree (EBOT), Bagged Tree (EBAT), Subspace K-Nearest Neighbour (SKNN), Complex Tree (CT), Median Tree (MT), Naive Bayes Classifier Algorithm and Random Forest (RF). In addition, classifiers can be combined using stack learning (ensemble methods).

1.3 Description of how the intervention might work

Proponents of iridology believe that changes in the intricate tissue structure of the iris can indicate a current or future clinical manifestation of a disease, and as such iridology can be used as a diagnostic tool (1). Because the iris is connected to hundreds of thousands of nerve endings, blood vessels and other tissues, it is thought to correspond to the body's internal function. Using iris charts as a guide, practitioners use observations of the iris, or more recently computer-aided pattern recognition via machine learning, for diagnosis.

Iridology is practiced primarily by “iridologists” (certification is available, though this is non-medical). Naturopaths and other holistic health practitioners also use iridology as part of their wider practice to assist in decision making.

1.4 Why it is important to do this review

An overview review conducted in 2015 did not find clear systematic review evidence for the efficacy of iridology as a diagnostic tool. The purpose of this review is to inform the Australian Government’s Natural Therapies Review 2019-20, which is evaluating evidence of the clinical effectiveness of 16 therapies, including iridology as a diagnostic tool. This review will differ from the 2015 publication by including primary research including additional studies published in the last five years, or relevant studies not identified in the 2015 review.

In Australia, complementary medicine and therapies are often used in conjunction with conventional medicine. Iridology is mostly used by some natural health practitioners as a diagnostic tool in conjunction with other tools, particularly when diagnosis achieved through other methods is unclear. However, some practitioners treat patients based solely on results obtained from examination of the iris. For this reason, it is important to have relevant, up-to-date synthesis of the evidence for the effectiveness of iridology as a diagnostic tool. This will enable consumers, health care providers and policy makers to make informed decisions about care.

2 Objectives

The objective is to collate, synthesise and critically appraise available evidence on the effectiveness of iridology as a diagnostic tool for any described injury, disease, medical condition, or preclinical condition commonly seen by iridologists.

3 Methods

The methodologies for this systematic review are based on those reported in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (7). EndNote20 (endnote.com) and Covidence (covidence.org) will be used for screening, managing citations, and data extraction. Where appropriate, OpenMeta (cebm.brown.edu/openmeta) will be used for any meta-analyses. GRADEpro GDT software (gradepro.org) will be used to record decisions and derive an overall assessment of the certainty of evidence for each outcome guided by GRADE methodology. The final approved systematic review protocol will be registered on the International Prospective Register of Systematic Reviews (PROSPERO).

3.1 Criteria for considering studies for this review

3.1.1 Study designs

This review will include evidence for the diagnostic accuracy of iridology from primary studies. Eligible study designs will include any study which measures the diagnostic accuracy of iridology for a condition via comparison to a valid reference standard (diagnosis by a medical practitioner). Diagnostic case control studies will be eligible for inclusion. Both prospective and retrospective studies will be eligible for inclusion. Studies will be limited to human studies and there will be no restrictions on the recruitment of participants, though it is expected most will be in a clinical setting. As a study type, systematic reviews will be excluded.

All studies will be presented regardless of quality; however, only studies with a moderate to low risk of bias will be included in the interpretation of the diagnostic accuracy.

For studies exploring machine learning for iridology, studies will only be included if results are conducted on a "test" set. That is, when developing machine learning algorithms, a "training" set of images is used to help the algorithm learn pattern identification. Results which report on the diagnostic accuracy of this "training" set will be excluded, and only studies which explicitly use a separate set of images for testing will be included. Often, the process of a separate training and testing set is called "cross-validation", so, any studies including this as their methodology will be included.

3.1.2 Publication date

There are no limitations on publication date, however, studies published after the systematic review literature search date will not be included. Studies published (or submitted to the Department) after the literature search date will be listed in the "Studies Awaiting Classification" table of the Evidence Evaluation Report. These studies will not undergo formal evidence review, however, a brief statement about the study and its potential impact on the overall conclusions of the evidence review will be included.

3.1.3 Studies published in languages other than English

Database searches, as well as the Department's call for evidence, will not exclude studies based on language of publication. Databases in languages other than English will not be searched, however, studies in languages other than English may be identified via the English-language databases. For practicality, potentially eligible studies will not undergo full-text translation or data extraction but will be documented via a process outlined in the "Studies published in languages other than English" section.

3.1.4 Participants

People of any age with any injury, disease, medical condition, or preclinical condition are eligible for inclusion.

3.1.5 Index test(s)

All studies which evaluate the practice of iridology as a diagnostic tool will be included, that is, any activity named as iridology which involves interpreting observations or images of an iris with reference to an iris chart to diagnose a pre-specified condition(s) and/or problem(s). Studies are to be included irrespective of whether diagnosis is completed by a certified iridologist. Any method of observing or capturing images of the iris, and interpreting iris images, will be included.

Studies which include MLA as part of the index test will only be included where MLA is specified as part of iridology's diagnostic process, or as a comparator. Studies must use MLA as a diagnostic tool for a pre-specified hypothesis about the relationships between a condition and a region/pattern of the iris; that is, data-driven studies "fishing" for correlations between diagnosis and iris patterns will not be included.

3.1.6 Target conditions

Iridology is used to diagnose many target conditions, including diabetes, chronic liver disease, and kidney disease (as identified via a scoping review conducted to inform this Research Protocol). Target conditions will not be pre-specified to include the breadth of iridology practice. Searches will be limited to any human clinical condition. The scoping review has identified a wide range of target clinical conditions, including diabetes, ulcerative colitis, asthma, coronary heart disease, and kidney disease.

3.1.7 Comparators

As this review will examine the diagnostic accuracy of iridology for any clinical condition, it is not possible to specify at the Research Protocol stage which comparator tests will be considered. We will consider studies that compare iridology to any other comparator (diagnostic test).

3.1.8 Reference standards

Reference standards are required to confirm the presence or absence of a condition within a population. These vary depending by target condition but must include diagnosis by a medical practitioner. Diagnoses by iridologists or other non-medical practitioners will not be considered as a reference standard. Studies in which the reference standard is not directly reported as diagnosis by a medical practitioner, or where this is unclear, will be assessed for appropriateness on a case-by-case basis.

3.1.9 Outcome measures

The outcome of interest in this study is diagnostic accuracy (including true positives, false positives, true negatives, false negatives, sensitivity, specificity, negative predictive value, positive predictive value, receiver operating curve or accuracy). Prognostic accuracy outcomes will not be included as it is difficult to measure diagnostic accuracy in these studies, and studies and outcomes are lacking in this space (i.e. they don't follow up people to determine if the progress to the disease, and there is generally no or poor comparison) from our scoping review. Patient-reported measures of experience (e.g., satisfaction), safety, quality or economic outcomes will be excluded. Studies will not be assessed on outcome measures at the title-abstract review phase in the interest of comprehensive screening.

3.2 Search methods for identification of studies

3.2.1 Electronic searches

We will search the following electronic databases, from inception until present:

- AMED
- CINAHL
- Cochrane Library
- Embase
- Emcare
- JBI Database of Systematic Reviews and Implementation Reports
- MANTIS
- MEDLINE
- PsycINFO
- Systematic Review Data Repository (SRDR)
- Natural Medicines Comprehensive Database
- OVID
- Scopus
- Web of Science
- relevant databases in the PAHO Virtual Health Library.

The following search strategy was developed for MEDLINE (based on a recent published scoping search for iridology (8)):

- iridology.mp *OR* iridodiagnos*.mp. *OR* (Iris/ AND Complementary Therapies/)

This search string will be modified to suit the required syntax for other databases.

Retracted studies will be flagged and excluded using EndNote20.

3.2.2 Search restrictions

Searches will be limited to human research. No date, language or geographic limitations will be applied when conducting the search of English language databases. However, non-English databases will not be searched. If non-English studies are found because of English language database searches, the process outlined in “Studies published in languages other than English” will be followed.

3.2.3 Other sources

Reference lists of all included studies will be reviewed for potential eligible studies (ancestry search). In addition, studies citing the included studies will also be reviewed for inclusion (forwards citation search). The Natural Therapies Review Expert Advisory Panel (NTREAP) or NTWC members may recommend studies to be considered in the review. The Department of Health has invited the public and key stakeholders to provide published research evidence. Publicly submitted evidence will be provided to evidence reviewers once the Protocol is finalised. Potential studies from both sources will be considered and assessed against the predetermined inclusion criteria (see Data collection section below). Grey literature will be considered out of scope.

3.3 Data collection

3.3.1 Inclusion decisions – title/abstract screening

Citations (title/abstract/year/journal) retrieved by the literature searches will be imported into Covidence or EndNote and duplicates removed. Two reviewers will independently screen the titles and abstracts identified in the database searches, citation searches, and those provided by the Department for eligibility against the inclusion criteria. Any discrepancies will be resolved via discussion with reference to the inclusion criteria; any unresolved items will be checked with a third reviewer if required. Citations that are in a language other than English will be tagged and managed as described in the below under “Studies published in languages other than English”.

3.3.2 Inclusion decisions – full text screening

The lead reviewer will retrieve full-text copies of eligible articles, and two reviewers will independently screen the studies for inclusion. Any disagreements will be resolved by discussion, or with reference to a third reviewer. Ineligible studies will be marked with a reason for exclusion and listed in a table in the Evaluation Report under “Characteristics of excluded studies”. The selection process will be recorded in sufficient detail to complete a PRISMA flow diagram. If a study does not contain the required PICO information for a decision to be made regarding its eligibility, the information will be sought from the study’s authors. Eligible studies that are not available in English will be noted and managed as described in the below under “Studies published in languages other than English”.

3.3.3 Studies published in languages other than English

Studies published in languages other than English will undergo title and abstract translation using Google translate (or an equivalent tool). If online translation does not facilitate understanding of the title and abstract, then these studies will be listed in a table as “Studies unable to be translated or interpreted at the title/abstract stage”. Translated titles and abstracts will be screened during the title/abstract screening stage and reported in the PRISMA flow diagram.

For studies not published in English but which are eligible for full-text review and are likely to meet the inclusion criteria, or if there is any uncertainty, the full-text report will not be translated to determine the studies’ compliance with eligibility criteria. They will be recorded in a “Studies Awaiting Classification” table and this information will be reflected in the PRISMA flow diagram.

Once the “Studies Awaiting Classification” table is finalised, a copy will be provided to NHMRC and included in the Evidence Evaluation Report.

The potential risk of language bias and its implications for the evaluation will be discussed in relevant sections of the Evidence Evaluation Report (such as “Overall completeness and applicability of evidence” and “Agreements and disagreements with other studies or reviews”). Appropriate qualifying statements will be made throughout the Research Protocol and Evidence Evaluation Report to acknowledge that only evidence published in English was reviewed. In relevant sections of the Evidence Evaluation Report, any potential limitations due to language bias that might influence the conclusions of the review will be discussed.

3.3.4 Evidence provided through the Department’s public call for evidence

Evidence provided through the Department’s public call for evidence (or provided by any other key stakeholders) will be assessed according to the inclusion criteria. Evidence not meeting the inclusion criteria will be considered out of scope, and a rationale for exclusion will be provided for such studies. Eligible studies that have not been identified in database searches and other search processes will be incorporated into the review.

3.3.5 Data collection process

Two reviewers will independently extract data from reports of included studies using data extraction forms (Appendix 1). The extraction form will be piloted to test practicality and reliability – during piloting, two reviewers will jointly extract the data from two studies into the extraction forms to ensure consistent understanding and suitability of the forms. The remaining included studies will be extracted by two authors independently. Data extractions will be compared by a third reviewer to identify discrepancies in extractions, and discrepancies will be reconciled by discussion.

3.3.6 Requests for data

If key information is missing from reports of the included studies, the corresponding authors will be contacted.

3.4 Data analysis

3.4.1 Data items for extraction

The following characteristics of included studies will be extracted:

- Author, year, study design, clinical setting, and location;
- Journal
- Number of participants, participant characteristics (including demographics, diagnosis, health status);
- Index test (including target condition, type of equipment used to capture iris image, equipment specifications, analysis technique, classifier (if using machine learning algorithm) and reference standard (if provided);
- Outcomes (true positives, false positives, true negatives, false negatives, inconclusive results, accuracy, sensitivity, specificity, receiver operating curve).

3.4.2 Dealing with missing data

If numerical outcome data are missing from reports of the included studies, and they cannot be obtained from the authors, where feasible, we will calculate them from other available statistics according to methods described in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (7). If missing data have been calculated, this will be noted in the Evaluation Report for transparency. Studies where missing data cannot be reasonably calculated

will be included as narrative (non-quantitative) synthesis of results. Consequences of missing data will be discussed and considered when assessing the risk of bias.

3.4.3 Assessment of methodological quality

Risk of bias assessment will be completed by one reviewer, and the other reviewer will independently check and confirm assessments made. Disagreements will be resolved by discussion, with reference to a third reviewer if necessary. Risk of bias assessments for each study will be made on a scale (e.g., low, moderate, high, critical). A supporting rationale with reference to the study will be provided in separate risk of bias forms. An overall risk of bias judgement for each study will be presented in the main Evaluation Report.

3.4.4 Tools to assess risk of bias in individual studies

Risk of bias of included studies will be assessed using QUADAS-2 tool for any studies measuring diagnostic accuracy (9), as advised by the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (7). For studies where machine learning is included as part of the index test, the PROBAST risk of bias assessment tool (designed for predictive modelling studies) will be used (10).

3.4.5 Statistical analysis and data synthesis

Our primary analysis of interest is accuracy of the iris assessment against the dichotomous outcome variable 'disease/no disease'. To explore this, we will apply the recommended Cochrane framework treating each test separately. For each test, we will extract the relevant diagnostic accuracy outcome (e.g., sensitivity/specificity). Where these are not available and the data allows, we will extract data to populate a standard 2x2 data table of binary test results against the reference standard. From this table we will calculate sensitivities and specificities, with 95% confidence intervals, at individual test level. Where certain tests require a threshold level (e.g., HbA1c for type 2 diabetes) primary thresholds of interest will be based on the thresholds proposed in the original paper describing the study or on agreed upon clinical threshold used in practice guidelines.

Where possible we will present the results by plotting their sensitivity and specificity (and their 95% confidence intervals) both in forest plots and in a scatter plot in receiver operating characteristic (ROC) space. For the meta-analysis of diagnostic accuracy measures, we will use the bivariate model (Reitsma 2005). For studies with a common threshold, this model takes into account within-study variation and between-study variation and focuses on estimating a summary operating point (i.e. a summary value for sensitivity and specificity). In addition, we will estimate the 95% confidence region and the 95% prediction region around the summary operating point. We will perform these analyses using the command xtmelogit in STATA, according to the licenses available.

3.4.6 Investigations of heterogeneity

Forest plots will be used to look for evidence of heterogeneity within both sensitivity and specificity. We will also use receiving operating characteristics (ROC) plots to look for evidence of a threshold effect and heterogeneity due to differences in accuracy. Where possible the effects of covariates thought to be potential sources of heterogeneity on sensitivity or specificity, will be examined.

3.4.7 Sensitivity analysis

Where appropriate (i.e., if not already explored in our analyses of heterogeneity) and as data allow, we will explore the effect of methodological aspects of the included studies. In the first instance we will run a sensitivity analysis limited to studies considered to be at low risk of bias.

3.4.8 Risk of reporting bias across studies

Risk of bias for individual studies will be assessed using the appropriate tool as outlined in Section 3.4.4. Assessments for each study will be made on a scale (e.g., low, moderate, high, critical), and any studies with high or critical risk of bias will be excluded. Individual assessments will be used to appraise the risk of reporting bias across studies based on the study limitations, imprecision, inconsistency of results, indirectness of evidence, and the likelihood of publication bias.

3.4.9 Addressing risk of bias

Initially, only studies with a low or moderate risk of bias will be included in the interpretation of evidence, though all studies will be presented. Concerns relating to the bias or applicability of the evidence will be discussed. Where appropriate, sensitivity analysis will be conducted based on the level of bias.

3.4.10 Subgroup analyses

Subgroup analyses are not planned, however if there is inconsistency between effect estimates, subgroup analysis may be used to explore possible sources of heterogeneity or modelled using the bivariate model (Reitsma 2005).

3.4.11 Certainty of the evidence

The GRADE approach will be used to assess the certainty of the body of evidence. Using this approach, certainty will be rated as:

- Very low (⨁⊖⊖⊖): the true effect is probably markedly different from the estimated effect.
- Low (⨁⨁⊖⊖): the true effect might be markedly different from the estimated effect.
- Moderate (⨁⨁⨁⊖): further research is likely to have an important impact in the confidence in the estimate of effect.
- High (⨁⨁⨁⨁): further research is very unlikely to change the confidence in the estimate of effect.

The GRADE framework will determine the certainty of the evidence based on consideration of five factors (11):

- Risk of bias –assessed using relevant tools for the study design (i.e., QUADAS-2 or PROBAST). Given studies with high or critical risk of bias will be excluded from evidence interpretation, it is not expected that certainty of evidence will be impacted by risk of bias concerns. Risk of bias may be higher if: consecutive patients are not recruited as a single cohort and classified by disease state; the selection or referral process is not clearly described; evaluators are not blind to results of index test and reference standard.
- Inconsistency – assessed by examining consistency in sensitivity, specificity, or likelihood ratios. Unexplained inconsistency may reduce quality of evidence.
- Imprecision – This will be assessed by examining confidence intervals for estimates of test accuracy or true and false positive and negative rates. Wide confidence intervals may reduce quality of evidence.
- Indirectness – assessed by examining the relevance of the outcome(s) and population(s) of the studies to the intended outcome(s) and population(s).
- Publication bias – This will be based on the extent to which the evidence is available. Publication bias would be suspected when the evidence is limited to a small number of small trials.

3.4.12 'Summary of Findings' tables and evidence statements

Findings will be reported in a “Summary of Findings” tables in the Evidence Evaluation Report. This will include all reported results on diagnostic accuracy, grouped by condition of interest and/or index test. The Summary of Findings tables will provide a synthesis of the body of evidence, key numerical results, and a summary judgment about the certainty of the underlying evidence for each outcome. Plain language statements will be used to describe findings, including the size of effects, and concerns relating to false positives and false negatives.

4 Protocol

4.1 Protocol registration

On final approval by the NHMRC, the protocol will be registered on the International Prospective Register of Systematic Reviews (PROSPERO).

4.2 Differences between the protocol and systematic review

Any differences between the protocol and the systematic review will be documented in a specified “Differences between the protocol and systematic review” section of the Evidence Evaluation Report, together with reasons for deviation from the original registered protocol.

Acknowledgements

Thank you to the members of the Department of Health's Natural Therapies Review Expert Advisory Panel and the National Health and Medical Research Council's Natural Therapies Working Committee for their advice and comments throughout the creation of this document.

Membership and other details of the Panel and Committee can be found at:

https://www.health.gov.au/committees-and-groups/natural-therapies-review-expert-advisory-panel

https://www.nhmrc.gov.au/about-us/leadership-and-governance/committees/natural-therapies-working-committee

Declarations of interest

All authors report no known conflicts of interest.

Funding

This review is funded by the Australian Government Department of Health via the NHMRC, under contract 2021-22C007.

References

1. Jensen B. Iridology simplified. 5th ed. California, USA: Iridologists International; 1980.
2. Hareva DH, Lukas S, Suharta NO, editors. The smart device for healthcare service: Iris diagnosis application. 2013 Eleventh International Conference on ICT and Knowledge Engineering; 2013 20-22 Nov. 2013.
3. Ma L, Zhang D, Li N, Cai Y, Zuo W, Wang K. Iris-based medical analysis by geometric deformation features. IEEE J Biomed Health Inform. 2013;17(1):223-31.
4. Othman Z, Prabuwono AS, editors. Preliminary study on iris recognition system: Tissues of body organs in iridology. 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES); 2010 30 Nov.-2 Dec. 2010.
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-8.
6. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Medical Research Methodology. 2019;19(1):64.
7. Deeks J. W, S., Davenport, C. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy: Cochrane; 2013. Available from: https://methods.cochrane.org/sdt/handbook-dta-reviews.
8. Esteves R, Morero J, Pereira S, Mendes K, Hegadoren K, Cardoso L. Parameters to increase the quality of iridology studies: A scoping review. European Journal of Integrative Medicine. 2021;43:101311.
9. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.
10. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ. 2021;375:n2281.
11. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ (Clinical research ed). 2008;336(7653):1106-10.

Appendix 1: Proposed data extraction form

General Information							Participants						Study						
Author	Year	Title	Country	Clinical setting	Funding sources	Study design	Number	Age	Gender	Health status	Diagnosis	Ethnicity	Target condition	Equipment details	Image segmentation method	Feature extraction method	Analysis method	Classifier/s (MLA)	Reference standard

Results																		
Diagnostic accuracy results										Iridology				Control				
TP	FP	FN	TN	Accuracy	Specificity	Sensitivity	PPV	NPV	Area under curve	Number tested	Number positive	Number negative	Number inconclusive	Number tested	Number positive	Number negative	Number inconclusive	Interpretation of reference test(s)

Appendix 2: Proposed Search Terms

Database	Search terms
MEDLINE/OVID	iridology.mp *OR* iridodiagnos*.mp. *OR* (Iris*/ AND Complementary Therap*/)
AMED	iridology OR iridodiagnos*
Campbell Systematic Reviews	iridology OR iridodiagnos*
CINAHL	iridology *OR* iridodiagnos* *OR* (Iris* Complementary Therap*/)
Cochrane Library	iridology
Embase	iridology OR iridodiagnos* OR (iris* AND 'alternative medicine')
Emcare (OviD)	(iridology or iridodiagnos*).mp. or (iris*.mp. and exp alternative medicine/)
JBI Database of Systematic Reviews and Implementation Reports (Ovid)	iridology OR iridodiagnos*
MANTIS	iridology OR iridodiagnos*
PsycINFO	iridology.mp *OR* iridodiagnos*.mp. *OR* (Iris*/ AND Complementary Therap*/)
Systematic Review Data Repository (SRDR)	iridology
Natural Medicines Comprehensive Database	iridology
Scopus	TITLE-ABS-KEY (iridology OR iridodiagnos* OR (iris* AND complementary AND therap*))
Web of Science	((TS=(iridology)) OR TS=(iridodiagnos*)) OR TS=("Iris" AND "complementary")
PAHO Virtual Health Library	iridology OR iridodiagnosis