

Protocol

Target Product Profile for a Machine Learning–Automated Retinal Imaging Analysis Software for Use in English Diabetic Eye Screening: Protocol for a Mixed Methods Study

Trystan Macdonald^{1,2,3}, MBChB; Jacqueline Dinnes³, MSc, PhD; Gregory Maniatopoulos⁴, PhD; Sian Taylor-Phillips⁵, PhD; Bethany Shinkins⁵, PhD; Jeffrey Hogg⁶, MBBS, MRes; John Kevin Dunbar⁷, MBChB; Ameenat Lola Solebo^{8,9}, PhD; Hannah Sutton¹⁰, BA; John Attwood¹¹, MA; Michael Pogose¹², MSc; Rosalind Given-Wilson¹³, MBBS; Felix Greaves^{14,15}, PhD; Carl Macrae¹⁶, PhD; Russell Pearson¹⁷, PhD; Daniel Bamford⁷, MA; Adnan Tufail^{9,18}, MD; Xiaoxuan Liu^{1,2,3*}, PhD; Alastair K Denniston^{1,2,3,19,20*}, PhD

¹Ophthalmology Department, Queen Elizabeth Hospital Birmingham, University Hospitals Birmingham National Health Service Foundation Trust, Birmingham, United Kingdom

²Academic Unit of Ophthalmology, Institute of Inflammation and Aging, College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom

³National Institute for Health and Care Research Birmingham Biomedical Research Centre, Birmingham, United Kingdom

⁴School of Business, University of Leicester, Leicester, United Kingdom

⁵Warwick Medical School, University of Warwick, Coventry, United Kingdom

⁶Population Health Sciences Institute, Faculty of Medical Sciences, The University of Newcastle upon Tyne, Newcastle, United Kingdom

⁷NHS England, Leeds, United Kingdom

⁸Population Policy and Practice, University College London Great Ormond Street Institute of Child Health, London, United Kingdom

⁹Moorfields Eye Hospital NHS Foundation Trust, London, United Kingdom

¹⁰Lay member, Oxford, United Kingdom

¹¹Alder Hey Children's Hospital, Alder Hey Children's Hospital NHS Foundation Trust, Liverpool, United Kingdom

¹²Hardian Health, London, United Kingdom

¹³St. George's University Hospitals National Health Service Foundation Trust, London, United Kingdom

¹⁴National Institute for Health and Care Excellence, London, United Kingdom

¹⁵Faculty of Medicine, School of Public Health, Imperial College London, London, United Kingdom

¹⁶Nottingham University Business School, University of Nottingham, Nottingham, United Kingdom

¹⁷Medicines and Healthcare Products Regulatory Agency, London, United Kingdom

¹⁸Institute of Ophthalmology, University College London, London, United Kingdom

¹⁹Centre for Regulatory Science and Innovation, Birmingham Health Partners, Birmingham, United Kingdom

²⁰National Institute for Health and Care Research Biomedical Research Centre at Moorfields and University College London Institute of Ophthalmology, London, United Kingdom

* these authors contributed equally

Corresponding Author:

Alastair K Denniston, PhD

Ophthalmology Department

Queen Elizabeth Hospital Birmingham

University Hospitals Birmingham National Health Service Foundation Trust

Mindelsohn Way

Edgbaston

Birmingham, B15 2TH

United Kingdom

Phone: 44 1213716905

Email: a.denniston@bham.ac.uk

Abstract

Background: Diabetic eye screening (DES) represents a significant opportunity for the application of machine learning (ML) technologies, which may improve clinical and service outcomes. However, successful integration of ML into DES requires careful product development, evaluation, and implementation. Target product profiles (TPPs) summarize the requirements necessary for successful implementation so these can guide product development and evaluation.

Objective: This study aims to produce a TPP for an ML-automated retinal imaging analysis software (ML-ARIAS) system for use in DES in England.

Methods: This work will consist of 3 phases. Phase 1 will establish the characteristics to be addressed in the TPP. A list of candidate characteristics will be generated from the following sources: an overview of systematic reviews of diagnostic test TPPs; a systematic review of digital health TPPs; and the National Institute for Health and Care Excellence's Evidence Standards Framework for Digital Health Technologies. The list of characteristics will be refined and validated by a study advisory group (SAG) made up of representatives from key stakeholders in DES. This includes people with diabetes; health care professionals; health care managers and leaders; and regulators and policy makers. In phase 2, specifications for these characteristics will be drafted following a series of semistructured interviews with participants from these stakeholder groups. Data collected from these interviews will be analyzed using the shortlist of characteristics as a framework, after which specifications will be drafted to create a draft TPP. Following approval by the SAG, in phase 3, the draft will enter an internet-based Delphi consensus study with participants sought from the groups previously identified, as well as ML-ARIAS developers, to ensure feasibility. Participants will be invited to score characteristic and specification pairs on a scale from "definitely exclude" to "definitely include," and suggest edits. The document will be iterated between rounds based on participants' feedback. Feedback on the draft document will be sought from a group of ML-ARIAS developers before its final contents are agreed upon in an in-person consensus meeting. At this meeting, representatives from the stakeholder groups previously identified (minus ML-ARIAS developers, to avoid bias) will be presented with the Delphi results and feedback of the user group and asked to agree on the final contents by vote.

Results: Phase 1 was completed in November 2023. Phase 2 is underway and expected to finish in March 2024. Phase 3 is expected to be complete in July 2024.

Conclusions: The multistakeholder development of a TPP for an ML-ARIAS for use in DES in England will help developers produce tools that serve the needs of patients, health care providers, and their staff. The TPP development process will also provide methods and a template to produce similar documents in other disease areas.

International Registered Report Identifier (IRRID): DERR1-10.2196/50568

(*JMIR Res Protoc* 2024;13:e50568) doi: [10.2196/50568](https://doi.org/10.2196/50568)

KEYWORDS

artificial intelligence; design; developers; diabetes mellitus; diabetic eye screening; diabetic retinopathy; diabetic; DM; England; eye screening; imaging analysis software; implementation; machine learning; retinal imaging; study protocol; target product profile

Introduction

Diabetic Eye Screening and Its Delivery in the United Kingdom

Diabetic eye screening (DES) aims to prevent vision loss from diabetic retinopathy (DR), a sight-threatening microvascular complication of diabetes [1]. In the United Kingdom, people with diabetes aged 12 years or older undergo yearly fundus photography with images interpreted by human graders to detect and assess the severity of DR. This allows detection of DR in its early asymptomatic stages, facilitating early referral to hospital eye services where treatment is more effective [2,3]. DES is associated with improved clinical [4] and economic [5] outcomes and a reduction in blindness from DR since its introduction in 2003 [6]. However, DES is both labor and resource intensive [7], and costs are expected to increase with a projected rise in diabetes prevalence [8]. In the United Kingdom, grading is undertaken by experienced professionals in a quality-assured and controlled multilevel system [9] with a high sensitivity for sight-threatening diseases [10]. However, grading represents a significant cost despite the fact that the majority of people with diabetes have no evidence of DR or

only mild disease [11] and are at low risk of visual loss in the near term [12].

Automated Retinal Imaging Analysis Software

DES grading is considered a promising use case for automation, and indeed, DES in Scotland has used automated retinal imaging analysis software (ARIAS) since 2012 [13,14]. This software uses "symbolic" artificial intelligence (AI) with grading "rules" programmed by humans [15] and is used in a primary grader role. The Scottish "autograder" has demonstrated a sensitivity of 90.5% (95% CI 89.3%-91.6%) and specificity of 67.4% (95% CI 66%-68.8%) for referable disease or ungradable images [13], with 50% of images it graded in 2018 requiring no further human review [16]. It is, however, unable to process the optic disc-centered images acquired in screening protocols elsewhere in the United Kingdom and has not been adopted outside Scotland [17].

Subsequent advances in AI, most notably in machine learning (ML), where AI "learns" patterns from data it is presented as opposed to being programmed by humans, have given rise to a newer generation of machine learning-automated retinal imaging analysis software (ML-ARIAS). These have demonstrated better performance than symbolic tools on

routinely collected UK screening images, including disc-centered photos. Heydon et al [18] found ML-ARIAS was able to achieve sensitivity and specificity of 95.7% (95% CI 94.8%-96.5%) and 54% (95% CI 53.4%-54.5%), respectively, for referable disease or ungradable images in a large sample taken from routine screening in 3 UK programs. A health technology assessment has also estimated that ML-ARIAS in a triage or primary grader role could be cost-effective and cost-saving compared to purely manual grading [17]. Consequently, there have been calls for ML-ARIAS's adoption into DES in the United Kingdom [19].

Bridging the Implementation Gap

In response to these calls, the UK National Screening Committee (NSC), an independent scientific advisory group that advises UK governments on screening programs, commissioned a review to ascertain whether ARIAS should be more widely adopted in 2021 [19]. While the review found evidence suggesting ML-ARIAS to be safe, accurate, and cost-effective, it also highlighted evidence gaps that prevented a recommendation for their use. The review highlighted that the acceptability of ML-ARIAS with patients and health care professionals was unknown, as was their real-world performance post implementation, recognizing that this may be affected by a range of clinical and technical factors. While research to address these evidence gaps is currently ongoing [20], the review's conclusions reflect an awareness in the wider field of health care AI that evidence of good diagnostic accuracy cannot predict the impact tools will have post deployment. This awareness contributes toward a growing "implementation gap" [21,22] or "chasm" [23] in health care AI, with increasing numbers of tools failing to be adopted into clinical use following promising research. This situation has a number of negative consequences, such as significant opportunity cost, with tools' development requiring significant public [24] or private investment [25], as well as a failure for patients, health care professionals, or systems to realize the benefits tools may offer for a lack of appropriate evidence generation.

A similar situation has existed in the pharmaceutical and medical device industries for decades. Despite significant investment in research and development, few biomarkers, drugs, or devices are ever adopted into routine clinical care, often failing to achieve regulatory approval or demonstrate sufficient clinical or economic benefits to health care providers [26,27]. To combat this situation, pharmaceutical companies began to outline the characteristics necessary for a successful product in the earliest stages of development, summarizing these in "target product profiles" (TPPs). These documents often incorporate considerations essential to successful clinical use but not covered in traditional research, such as interaction with existing care pathways and resource constraints. TPPs are used to guide research development as part of a broader "quality by design" approach that has been encouraged by regulatory agencies [28,29] and widely adopted in industry with notable successes [30,31]. TPPs have since been used by governments [32], health care providers [33], and nonprofit organizations [34] for drugs [35,36] and medical devices [37], particularly in the context of infectious diseases [38,39] and the developing world [40,41].

Aim

This study aims to produce a TPP for an ML-ARIAS for use in English DES (E-DES), enabling the development of ML tools that match the needs of people with diabetes, health care professionals, regulators, and providers. This work will focus on the English context, as variations in the commissioning and organization of DES between UK nations may affect the stakeholders needed as well as the TPP's final contents. However, we hope this work will be valuable for screening programs in other nations in the United Kingdom and beyond.

Objectives

The objectives of this study are to (1) identify which characteristics should be addressed in a TPP for an ML-ARIAS (this will be achieved through an overview of systematic reviews of TPPs, a systematic review of digital health TPPs, and by extracting standard's from the National Institute of Health and Care Excellence's (NICE) Evidence Standards Framework for Digital Health Technologies (ESF); (2) draft specifications for an ML-ARIAS TPP through semistructured interviews with members of key stakeholder groups; and (3) validate the contents of a TPP for an ML-ARIAS through a modified Delphi consensus study.

Methods

Ethical Considerations

Ethical approval for this study has been granted by the University of Birmingham Institutional Research Board (ERN_2023-0620). The study will be undertaken in accordance with the Helsinki Declaration of 1975, as revised in 2000. Informed consent will be sought from participants. The interviews and consensus meeting will be recorded and transcribed, with participants given the opportunity to opt out with deletion of their responses up to a week after their interview or meeting date, after which their responses will be anonymized and therefore no longer attributable to them. Recordings of the interview and consensus meeting will be deleted following transcription, with the anonymized transcripts stored on secure institutional research servers for up to 10 years. Delphi participants will be informed that their responses are anonymized and therefore cannot be deleted following submission. Contact details they provide in order to be able to participate in the consensus meeting will be stored on secure servers until the end of the study, after which they will be deleted.

People with diabetes participating in the study advisory group (SAG) and consensus meeting will have their travel costs reimbursed as well as their time reviewing documents or attending meetings at £25 (US \$31.57) per hour in line with National Institute for Health and Care Research (NIHR) guidance [42,43].

Identifying Key Stakeholder Groups

The successful use of ML-ARIAS in E-DES will require the approval, collaboration, and consent of a diverse range of stakeholder groups. To ensure our TPP is both comprehensive

and authoritative it is therefore important to incorporate the knowledge and opinions of all these groups in its development.

A recent qualitative systematic review on stakeholder perspectives of clinical AI implementation identified the following key stakeholder groups influencing implementation: patients, carers, and the public; health care professionals; health care managers and leaders; regulators and policy makers; and developers [44]. Mapped to our use case and context, these broad stakeholder groups can be further divided as follows: people with diabetes, carers, and the public; E-DES technicians and graders, as well as ophthalmologists (health care professionals); managers of E-DES services provided by either the NHS or independent providers (health care managers and leaders); the Medicines and Healthcare Products Regulatory Agency (MHRA), NICE, the Care Quality Commission (CQC), and NSC (regulators and policy makers); and ML-ARIAS developers. Our study will aim to encompass the opinions of all these groups as outlined below.

Study Delivery and Oversight

Study Management Group

A study management group (SMG) led by the chief investigator AKD will be responsible for the design and delivery of the study.

Recruitment

SMG members will be recruited from the host institution to facilitate ease of communication and regular meetings.

Sampling

The SMG will include experts in the relevant clinical, scientific, and methodological domains, project management support, and a person with diabetes.

Study Advisory Group

The SMG will be supported by a SAG, which will bring additional breadth of expertise from a range of key stakeholder groups. The SAG will provide critical feedback and advice

throughout the project in quarterly project review meetings and at specific junctures outlined in the protocol and ad hoc.

Recruitment

SAG recruitment strategies will be tailored to individual stakeholder groups. People with diabetes will be recruited through diabetes charities. Clinical and methodological experts with previous experience developing TPPs will be identified through their publications in the field. Representatives of relevant regulatory and policy bodies will be recruited through their organizations (NSC, MHRA, NICE, and CQC).

Sampling

The SAG will aim to recruit at least 2 representatives from each of the broad stakeholder groups previously identified, with the exception of ML-ARIAS developers, omitted to avoid bias arising from conflict of interest. Experts in TPP development and health economics will also be invited to the SAG to add a greater breadth of expertise.

User Group

A TPP user group (UG) consisting of ML-ARIAS developers will be assembled to provide feedback on the TPP as outlined below to ensure feasibility is factored into the development process.

Sampling

The UG will consist of only ML-ARIAS developers.

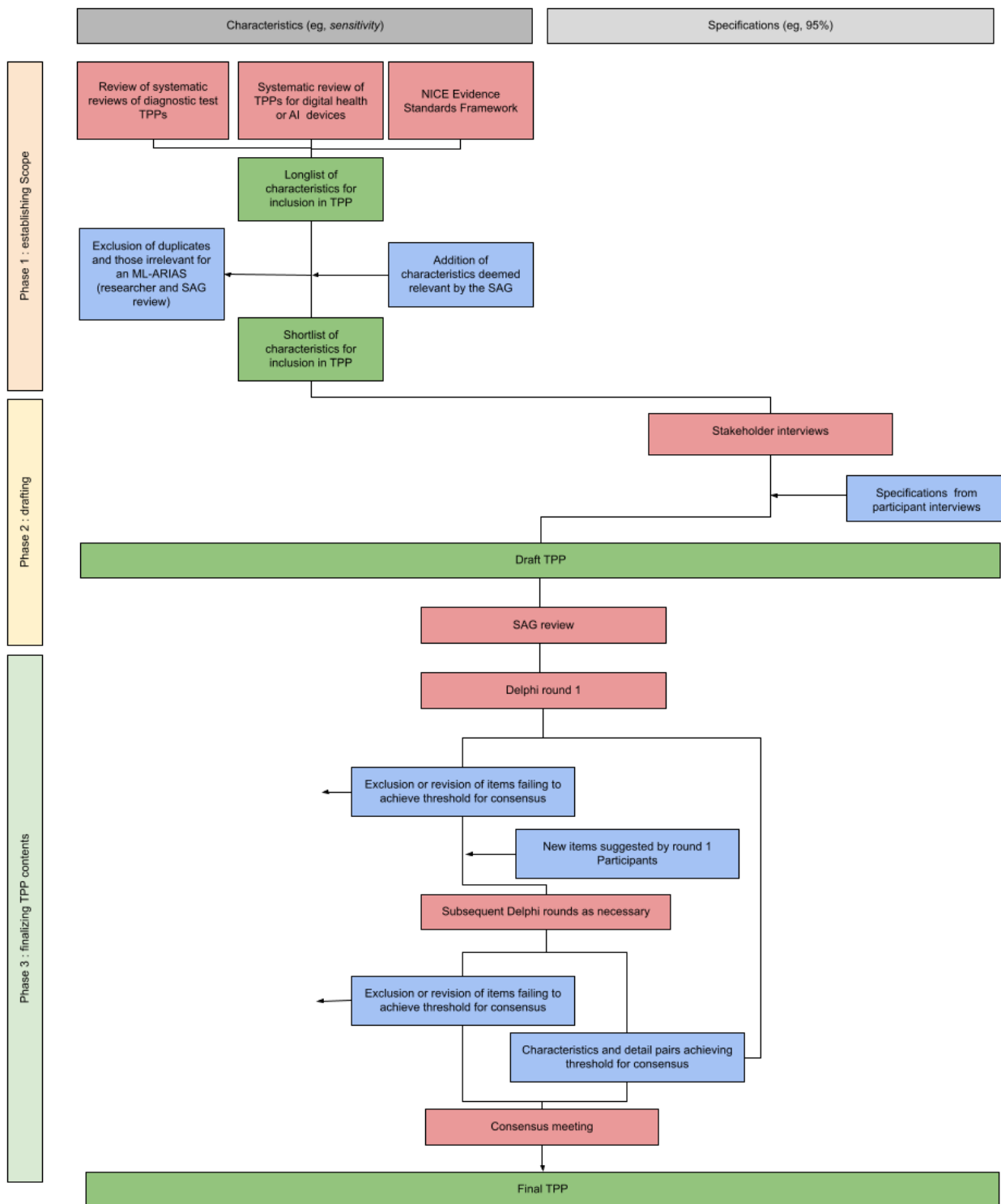
Recruitment

Developers with ML-ARIAS approved by British, European, Australian, or US medical device regulators will be invited to participate, along with those with technologies included in the UK NSC's 2021 review.

Phase 1: Literature Reviews to Establish the Scope of an ML-ARIAS TPP

See [Figure 1](#) for an overview of the ML-ARIAS TPP development process.

Figure 1. Diagram of the machine learning–automated retinal imaging analysis software (ML-ARIAS) target product profile (TPP) development process. AI: artificial intelligence; NICE: National Institute for Health and Care Excellence; SAG: study advisory group.



Rationale

This phase will seek to establish what characteristics a TPP for an ML-ARIAS should address in order to be both relevant and comprehensive. A long list of potential characteristics for inclusion in an ML-ARIAS TPP will be generated through 2 reviews. These will be built on the work of Cocco et al [45], who published a list of all characteristics previously reported in diagnostic test TPPs in a 2020 systematic review.

Overview of Systematic Reviews of Target Product Profiles for Diagnostic Tests

This will seek to identify all systematic reviews of diagnostic test TPPs with the aim of extracting the characteristics reported in each review or the TPPs they capture. Based on a scoping search, we anticipate this review will include the systematic review by Cocco et al [45] and any new resources published since.

Systematic Review of TPPs for Digital Health Tools

As Cocco et al's [45] review did not include any TPPs for digital health tools, our long list of potential characteristics may therefore omit characteristics unique to these tests. The purpose of this review is to identify all TPPs published for digital health tools (including AI) and extract the characteristics they report. This will include TPPs for therapeutic as well as diagnostic devices.

[Multimedia Appendix 1](#) includes MEDLINE search strategies for both reviews, developed in consultation with an information specialist and using validated search filters where possible [45-47]. Strategies for other bibliographic databases will be adapted from these, as will a web search strategy using a systematic method previously reported [48]. The 2 reviewers (TM and JH) will perform a title and abstract screening, then a full-text review to assess eligibility. Disagreements will be arbitrated by the senior investigator (AKD).

Generating AI-Specific Characteristics from NICE's Evidence Standards Framework for Digital Health Technologies

Rationale

As scoping searches for digital health TPPs have returned very few TPPs for AI and specifically ML tools, relying on the above reviews alone to determine the scope of our TPP may risk omitting characteristics unique to these devices and essential for their successful deployment. To generate AI-specific characteristics, we will therefore adapt standards from NICE's ESF for digital health technologies, a gold standard tool designed to evaluate digital health technologies for UK commissioning decisions. This was revised in 2022 to encompass AI devices, with the new AI standards being generated through a multistakeholder Delphi consensus process [49,50]. The 2 reviewers (TM and JH) will review the ESF, extracting relevant standards, with disagreements resolved by discussion or arbitration by the senior investigator (AKD). The extracted standards will be adapted into characteristics and added to the long list from the earlier reviews.

Shortlisting Characteristics for Inclusion in a Draft TPP

The list of candidate characteristics will be reviewed by the SMG to remove any irrelevant to the use case or proposed technology and consolidate the remainder, removing any duplicates. Disagreements between SMG members will be resolved through discussion and arbitration by the senior investigator (AKD) if necessary.

The remaining characteristics will then be mapped to the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework [51], a tool outlining the interacting complexity of factors at the policy, organizational, and practice level that may influence the successful implementation of digital health technology [52]. The NASSS will help group the characteristics into broader themes and help identify factors affecting successful implementation that they do not cover.

The included characteristics will then be presented to the SAG, along with those excluded and the NASSS domains not covered by the shortlist. SAG members will be asked to identify any

characteristics incorrectly excluded and approve a final shortlist of characteristics for subsequent research phases through discussion or vote.

Phase 2: Drafting the TPP

Scoping Interviews

A TPP typically includes a table of characteristics and their specifications. Characteristics might include "target population" with "people with diabetes aged 12 years or older" as its associated specification. The drafting of specifications for each of the characteristics shortlisted in phase 1 will be informed by semistructured interviews with participants from each key stakeholder group previously identified. The final output of phase 2 will be a full draft TPP for entry into a modified Delphi consensus study in phase 3.

Sampling

Purposive sampling will initially be used with the aim of interviewing at least 1 individual from all the stakeholder groups relevant to the use case previously identified, with the exception of ML-ARIAS developers. As the interviews progress, theoretical sampling will be used to better explore areas of disagreement or characteristics that have not been addressed in previous interviews. Snowball sampling will also be used, with participants invited to suggest potential interviewees. An approximate target of 20 participants has been set; however, the scale of recruitment will be refined by the SMG as the interviews progress.

Recruitment

People with diabetes will be recruited through diabetes charities and offered compensation for their time in accordance with NIHR guidelines [42,43]. Clinical and methodological experts, as well as industry representatives with previous knowledge or experience of AI in DES or TPPs, will be identified through their publications in the field. Representatives of relevant regulatory and policy bodies will be recruited by contacting their organizations (NSC, MHRA, and NICE). It is anticipated that members of the SAG will facilitate the recruitment of key stakeholders in their respective stakeholder groups.

Data Collection

Ahead of the interviews, participants will be sent information in plain English on E-DES, AI, and TPPs.

The interviews will be conducted over video teleconferencing software (Zoom; Zoom Video Communications, Inc). Interview participants will initially be asked to provide demographic data, including stakeholder groups, location, age, sex, gender, and ethnicity. A preprepared topic guide informed by the shortlist of characteristics produced in phase 1 will be developed and used to encourage a free-flowing conversation between interviewer and interviewee focused on the latter's opinions, priorities, and concerns regarding the adoption of ML-ARIAS in E-DES. The topic guide will be iterated as the interviews progress to ensure all implementation considerations are covered and adapted to different stakeholder groups to facilitate meaningful participation.

All the interviews will be conducted by 2 clinicians (TM and JH) with previous training and experience in qualitative interviewing. They will present themselves as researchers during the interviews so as not to introduce bias to the data collected [53]; however, as their status may become apparent during the interview and can be discovered on the internet, they will use reflective journaling to identify and mitigate against any effects this may have on data collection [54]. Both interviewers will take field notes during and after the interviews to provide context for the data analysis.

Data Analysis

All the interviews will be digitally recorded and transcribed verbatim. Data analysis will begin after the first interview and proceed in parallel with the remainder. One researcher (TM) will code the interview transcripts using qualitative analysis software (NVivo, Lumivero). The coded data will then be mapped to the shortlist of characteristics produced from phase 1, using this as a framework to aid analysis [55].

The data collected will be reviewed throughout the collection period by the SMG to identify and manage biases in sampling, data collection, analysis, and interpretation. The data will be used to iterate the topic guide, recruitment, and sampling strategies to allow the exploration of problematic issues, divergent views between or within stakeholder groups, and implementation considerations not previously explored.

On completion of the data analysis, the SMG will draft characteristics' specifications based on the analyzed data. An initial draft TPP will be presented to the SAG for review and approval before entry into phase 3.

Phase 3: Finalizing the TPP

Internet-Based Delphi Consensus Rounds

The draft TPP will be entered into an internet-based Delphi study using purpose-made software (Qualtrics XM; Qualtrics). The contents and form of the Delphi survey will be coproduced by the SMG and SAG with the close involvement of both groups' patient representatives. It will also be piloted to ensure it is broadly comprehensible and accessible.

At the start of each Delphi round, participants will be presented with information in plain English on E-DES, AI, and TPPs. In subsequent rounds after round 1, participants will also be provided with a report of the previous round (see Data Analysis section below). Each round will remain open for a minimum of 2-4 weeks, and it is expected that at least 2 rounds will be conducted. Extensions to rounds' time frames and further rounds may be considered in consultation with the SAG.

Sampling

In line with similar studies successfully conducted in other contexts [56-58], we aim to recruit more than 100 respondents overall, with at least 10 from each broader stakeholder group previously outlined, including ML-ARIAS developers. Completed returns will be monitored by the stakeholder group, and further measures taken targeted at individual groups to boost recruitment if necessary. These may include targeted email or newsletter reminders.

Recruitment

People with diabetes will be recruited by asking diabetes charities to disseminate a link to the Delphi survey among their members. Similarly, the British Association of Retinal Screeners will be asked to disseminate a link among their members to recruit DES professionals. The SMG and SAG will be asked to disseminate a survey link to their wider networks to recruit regulators, clinicians, and academics. ARIAS developers with technologies previously identified by the UK NSC's Rapid Review will be invited to participate, as will those receiving funding from the NHS AI Health and Care Award.

Data Collection

Participants will be asked which stakeholder groups they belong to, their geographic location, age, sex, gender, and ethnicity. In each round, participants will be asked to score each characteristic and specification on a Likert scale ranging from 1 to 5, with 1 being "definitely exclude" and 5 being "definitely include." Recognizing that some stakeholder groups may not have the expertise to confidently score characteristic and specification pairs in some areas, participants will be given the option to omit pairs or identify other stakeholder groups they would be comfortable making this decision on their behalf. Free-text boxes will be provided to comment on characteristic and specification pairs or suggest edits. Another box at the end of the survey will provide an opportunity for participants to suggest entirely new characteristic and specification pairs.

Data Analysis

After each round, the responses will be aggregated and analyzed. A report will be produced summarizing response rates and characteristic and specification pairs scores descriptively, with scores broken down by subgroup. Free-text responses will be imported into qualitative analysis software and coded.

Consensus will be defined as >70% of respondents scoring a characteristic and specification pair 4-5 and <15% 1 in any stakeholder group. Characteristic and specification pairs exceeding this threshold will exit the survey process and proceed to the consensus meeting for discussion. Those not meeting this threshold will be revised by the SMG, taking into account their scores in previous rounds by different stakeholder groups and their coded free-text responses. The revised characteristic and specification pairs will then enter subsequent rounds, along with any new items suggested by participants in the previous round.

Consensus Meeting

The Delphi results will be reviewed, and the final TPP contents will be agreed upon at an in-person consensus meeting.

Before the consensus meeting, feedback will be sought from members of the UG. This feedback, along with the Delphi round results broken down by stakeholder group, will be sent to consensus meeting participants ahead of the meeting.

The consensus meeting itself will be led by an experienced facilitator. At the meeting, participants will be presented with each characteristic and specification pair in turn, and an opportunity will be provided for discussion and edits. The Delphi round results broken down by stakeholder group as well as the UG's feedback on the characteristic and specification

pairs will be made available to aid discussion. Meeting participants will be asked to vote on the inclusion of each characteristic and specification pair, with a supermajority of >70% among voting participants needed for inclusion in the final TPP. An abstention will be permitted.

Recruitment

Participants in the Delphi study will be asked whether they wish to be considered for participation. Meeting attendees will be agreed upon with the SAG and invited to the meeting by email.

Sampling

Purposive sampling will be used to ensure there are at least 2 members from all the key stakeholder groups previously identified, with the exception of ML-ARIAS developers, to avoid conflicts of interest. Travel expenses will be offered to all attendees.

Results

This project has received funding from the NIHR Birmingham Biomedical Research Centre since February 2022 and an NIHR incubator grant for regulatory science awarded to the University Hospitals Birmingham NHS Foundation Trust in June 2023.

Both the SMG and SAG have been assembled and have convened their first meetings. Phase 1 began in April 2023; database searches were performed in May 2023 and the phase was completed in November 2023. Phase 2 began in November 2023 and is expected to be completed in March 2024. As of March 2024, 21 interviews have been performed. Phase 3 is expected to begin in April 2024 and be completed in July 2024. The final TPP and its methods will be submitted to a peer-reviewed journal for publication and reported using the Consolidated Criteria for Reporting Qualitative Research (COREQ) criteria [59].

Discussion

Benefits of Establishing a TPP for an ML-ARIAS for Use in E-DES

ML-ARIAS offer an innovative approach for E-DES to meet rising demand and improve clinical, economic, or service outcomes. However, there is a risk that any or all of these could be negatively impacted by the introduction of an ML-ARIAS due to failures to design, evaluate, or deploy such tools appropriately. For example, placing an ML-ARIAS designed to be highly sensitive but poorly specific as a primary grader may necessitate additional decision arbitration in the screening service, or risk increasing unnecessary referrals to hospital eye services. Either eventuality may significantly reduce (or even negate) any cost benefits, or negatively impact the experience of people with diabetes or E-DES professionals. Alternatively, an ML-ARIAS's economic value could be significantly reduced by any requirement to update existing IT infrastructure, making it economically unviable.

While it is impossible to forecast all such potential pitfalls, some are predictable and could be avoided by outlining the NHS' requirements with regard to an ML-ARIAS within a TPP. This

can then be used to guide product development and evaluation. Depending on who is involved in TPP development, these documents can also reflect the priorities of all stakeholder groups, including those often neglected in traditional procurement processes, such as patients and delivery staff.

The knowledge of what should go into a TPP for an ML-ARIAS likely already exists, albeit spread over members of different stakeholder groups and sometimes not formally recorded. Where desired characteristics are not clearly understood, for example, when trade-offs exist between competing priorities, our proposed consensus methodology provides a means to better establish requirements or areas where future research is needed.

Given that product requirements can be hard to identify, it is unsurprising that innovation frequently aligns poorly with the actual needs of health care professionals, health services, and their users. Creating a TPP to gather this knowledge in a concise format can accelerate the development of products that meet these stakeholders' needs at a cost much less than that of late-stage product failures or unsuccessful deployments. As well as improving developers' efficiency in product development or testing, other stakeholders stand to gain from TPPs' development. Knowing the essential characteristics of a tool, commissioners can make more assured commissioning decisions, drawing on the collective knowledge of all stakeholders contributing to the TPP. In an ML-ARIAS context, people with diabetes will also have a unique opportunity to influence product design and implementation strategies, increasing the likelihood that tools are acceptable to them and good DES uptake is maintained.

Comparison to Previous Work

To our knowledge, this project represents the first public use of TPPs for an AI health care technology. Our final TPP and the learning accrued through its development will provide a strong basis for the development of TPPs for other disease areas. As no best practice currently exists on TPP development [45], we will use a 3-phase multistakeholder-modified Delphi-consensus method similar to that used to develop reporting guidelines [56,60], core outcome sets [57,61] and previous TPPs [35,40,62]. This method aims to foster values such as inclusivity, patient empowerment, and consensus, and we intend to transfer these values into best practices in a new field.

Limitations

This project has a number of potential risks that we have sought to mitigate, most notably a failure to recruit and retain participants in the Delphi process, the omission of critical characteristics, and a failure to adequately represent and balance the needs of stakeholder groups.

With regard to securing good recruitment and retention for the Delphi process, we have secured the strong engagement of all relevant stakeholder groups with representatives of each on the SAG. Their involvement in the design of the Delphi survey aims to ensure it is broadly acceptable and achievable by all, as well as mitigate against a significant "drop-off" between Delphi rounds, a common issue that can introduce bias. To further mitigate this, we will also monitor returns collectively and at a

stakeholder group level during each round, allowing interventions such as reminders, further engagement activities, or the extension of round durations, if necessary.

The risk of omitting an essential characteristic in our TPP is increased as this will be the first developed for an AI diagnostic test, and characteristics cannot simply be copied from those previously published in the field. Additionally, some characteristics may only be identified and prioritized by single stakeholder groups. It is therefore important that the TPP development process provide sufficient opportunity for these novel or underappreciated characteristics to be identified. To do this, our reviews will specifically target resources in the field of digital and AI health care to establish which characteristics to report, and our method will provide opportunities for members of all stakeholder groups to put forward additional requirements at multiple stages.

There is also the risk that our methods may generate characteristics that put undue emphasis on the priorities of 1 stakeholder group over another. One may hypothesize that a developer might wish to set TPP requirements at a low, easily achievable level that would be unacceptable to patients and health service providers on the grounds of safety or quality. Conversely, patients and health professionals may have an unrealistic idea of what is possible and argue for TPP requirements to be set near or at perfection (eg, “must not miss any cases”). Our TPP development process seeks to both surface

and balance competing priorities to produce a common, achievable target where possible, including being open regarding which stakeholder groups hold different views, potential conflicts of interest, and the actual consequences of decisions and compromises. To ensure that the requirements of the TPP remain achievable, ML-ARIAS developers will be invited to participate in the internet-based Delphi rounds as well as the UG, feedback from which will be presented at the meeting where the final TPP will be agreed. Bidirectional thresholds will also be used in the Delphi phase, such that characteristic and specification pairs must have a minimum level of support as well as a maximum level of dissent allowed to progress, increasing the likelihood that these are achievable as well as acceptable to all.

Conclusions

In developing a TPP for an ML-ARIAS, we will for the first time bring patients, health care professionals, commissioners, methodologists, clinical AI experts, and developers together to provide a target for AI developers to work toward. It is our aim that this will increase the likelihood of the development of ML-ARIAS that are fit-for-purpose for the NHS, improve screening outcomes and of benefit all stakeholders. In addition, we hope that this first public use of TPPs in health care AI and our open sharing of our methods will enable others to develop TPPs for a range of unmet needs, providing clarity to developers and accelerating innovation toward products that will be welcomed by patients and providers alike.

Acknowledgments

All authors were involved in the conception and design of the work. TM, XL, AKD, JD, JH, ST-P, GM, and BS drafted the manuscript. All authors reviewed the manuscript and approved the final version. TM is a doctoral fellow funded by the University Hospitals Birmingham NHS Foundation Trust.

Data Availability

Data sharing is not applicable to this article, as no data sets were generated or analyzed during this study.

Conflicts of Interest

CM acts as the national professional advisor for patient safety at the Care Quality Commission. RGW is a member of the UK National Screening Committee and a nonexecutive director of Moorfields Eye Hospital. FG is employed by NICE and the Department of Health and Social Care in England.

Multimedia Appendix 1

Review search strategies.

[[DOCX File , 19 KB-Multimedia Appendix 1](#)]

References

1. Diabetic eye screening: programme overview. Gov.UK. URL: <https://www.gov.uk/guidance/diabetic-eye-screening-programme-overview> [accessed 2022-11-15]
2. Photocoagulation treatment of proliferative diabetic retinopathy. Clinical Application of Diabetic Retinopathy Study (DRS) findings, DRS Report Number 8. The Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1981;88(7):583-600. [Medline: [7196564](#)]
3. Savolainen EA, Lee QP. Diabetic retinopathy—need and demand for photocoagulation and its cost-effectiveness: evaluation based on services in the United Kingdom. *Diabetologia*. 1982;23(2):138-140. [doi: [10.1007/BF01271176](https://doi.org/10.1007/BF01271176)] [Medline: [6813169](#)]

4. Fendrick AM, Javitt JC, Chiang YP. Cost-effectiveness of the screening and treatment of diabetic retinopathy. What are the costs of underutilization? *Int J Technol Assess Health Care*. 1992;8(4):694-707. [doi: [10.1017/s0266462300002385](https://doi.org/10.1017/s0266462300002385)] [Medline: [1464489](https://pubmed.ncbi.nlm.nih.gov/1464489/)]
5. Javitt JC, Aiello LP. Cost-effectiveness of detecting and treating diabetic retinopathy. *Ann Intern Med*. 1996;124(1 Pt 2):164-169. [doi: [10.7326/0003-4819-124-1_part_2-199601011-00017](https://doi.org/10.7326/0003-4819-124-1_part_2-199601011-00017)] [Medline: [8554212](https://pubmed.ncbi.nlm.nih.gov/8554212/)]
6. 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. [FREE Full text] [doi: [10.1136/bmjopen-2013-004015](https://doi.org/10.1136/bmjopen-2013-004015)] [Medline: [24525390](https://pubmed.ncbi.nlm.nih.gov/24525390/)]
7. Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of type 1 and type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabet Med*. 2012;29(7):855-862. [doi: [10.1111/j.1464-5491.2012.03698.x](https://doi.org/10.1111/j.1464-5491.2012.03698.x)] [Medline: [22537247](https://pubmed.ncbi.nlm.nih.gov/22537247/)]
8. Haider S, Thayakaran R, Subramanian A, Toulis KA, Moore D, Price MJ, 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. [FREE Full text] [doi: [10.1136/bmjopen-2021-050058](https://doi.org/10.1136/bmjopen-2021-050058)] [Medline: [34253675](https://pubmed.ncbi.nlm.nih.gov/34253675/)]
9. Scanlon PH. The English national screening programme for diabetic retinopathy 2003-2016. *Acta Diabetol*. 2017;54(6):515-525. [FREE Full text] [doi: [10.1007/s00592-017-0974-1](https://doi.org/10.1007/s00592-017-0974-1)] [Medline: [28224275](https://pubmed.ncbi.nlm.nih.gov/28224275/)]
10. Oke JL, Stratton IM, Aldington SJ, Stevens RJ, Scanlon PH. The use of statistical methodology to determine the accuracy of grading within a diabetic retinopathy screening programme. *Diabet Med*. 2016;33(7):896-903. [FREE Full text] [doi: [10.1111/dme.13053](https://doi.org/10.1111/dme.13053)] [Medline: [26666463](https://pubmed.ncbi.nlm.nih.gov/26666463/)]
11. Scanlon PH, Nevill CR, Stratton IM, Maruti SS, Massó-González EL, Sivaprasad S, et al. Prevalence and incidence of diabetic retinopathy (DR) in the UK population of Gloucestershire. *Acta Ophthalmol*. 2022;100(2):e560-e570. [FREE Full text] [doi: [10.1111/aos.14927](https://doi.org/10.1111/aos.14927)] [Medline: [34180581](https://pubmed.ncbi.nlm.nih.gov/34180581/)]
12. Scanlon PH, Aldington SJ, Leal J, Luengo-Fernandez R, Oke J, Sivaprasad S, et al. Development of a cost-effectiveness model for optimisation of the screening interval in diabetic retinopathy screening. *Health Technol Assess*. 2015;19(74):1-116. [FREE Full text] [doi: [10.3310/hta19740](https://doi.org/10.3310/hta19740)] [Medline: [26384314](https://pubmed.ncbi.nlm.nih.gov/26384314/)]
13. Philip S, Fleming AD, Goatman KA, Fonseca S, McNamee P, Scotland GS, et al. The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme. *Br J Ophthalmol*. 2007;91(11):1512-1517. [FREE Full text] [doi: [10.1136/bjo.2007.119453](https://doi.org/10.1136/bjo.2007.119453)] [Medline: [17504851](https://pubmed.ncbi.nlm.nih.gov/17504851/)]
14. Fleming AD, Goatman KA, Philip S, Prescott GJ, Sharp PF, Olson JA. Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts. *Br J Ophthalmol*. 2010;94(12):1606-1610. [doi: [10.1136/bjo.2009.176784](https://doi.org/10.1136/bjo.2009.176784)] [Medline: [20858722](https://pubmed.ncbi.nlm.nih.gov/20858722/)]
15. Confalonieri R, Coba L, Wagner B, Besold TR. A historical perspective of explainable artificial intelligence. *Wiley Interdiscip Rev Data Min Knowl Discov Wiley*. 2021;11(1):e1391. [FREE Full text] [doi: [10.1002/widm.1391](https://doi.org/10.1002/widm.1391)]
16. Styles CJ. Introducing automated diabetic retinopathy systems: it's not just about sensitivity and specificity. *Eye (Lond)*. 2019;33(9):1357-1358. [FREE Full text] [doi: [10.1038/s41433-019-0535-7](https://doi.org/10.1038/s41433-019-0535-7)] [Medline: [31358924](https://pubmed.ncbi.nlm.nih.gov/31358924/)]
17. Tufail A, Kapetanakis VV, Salas-Vega S, Egan C, Rudisill C, Owen CG, et al. An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. *Health Technol Assess*. 2016;20(92):1-72. [FREE Full text] [doi: [10.3310/hta20920](https://doi.org/10.3310/hta20920)] [Medline: [27981917](https://pubmed.ncbi.nlm.nih.gov/27981917/)]
18. Heydon P, Egan C, Bolter L, Chambers R, Anderson J, Aldington S, et al. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30000 patients. *Br J Ophthalmol*. 2021;105(5):723-728. [FREE Full text] [doi: [10.1136/bjophthalmol-2020-316594](https://doi.org/10.1136/bjophthalmol-2020-316594)] [Medline: [32606081](https://pubmed.ncbi.nlm.nih.gov/32606081/)]
19. Zhivko Z, Jaime P, Morwenna R, Michael A, Jenny L, Goda K, et al. Automated grading in the diabetic eye screening programme. UK National Screening Committee. 2021. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1035903/Evidence_summary_AI_in_DESP_2021.pdf [accessed 2024-02-27]
20. Ethnic differences in performance and perceptions of artificial intelligence retinal image analysis systems for the detection of diabetic retinopathy in the NHS Diabetic Eye Screening Programme. National Institute for Health and Care Research. URL: https://fundingawards.nihr.ac.uk/award/AI_HI200008 [accessed 2023-08-16]
21. DECIDE-AI Steering Group. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nat Med*. 2021;27(2):186-187. [doi: [10.1038/s41591-021-01229-5](https://doi.org/10.1038/s41591-021-01229-5)] [Medline: [33526932](https://pubmed.ncbi.nlm.nih.gov/33526932/)]
22. Seneviratne MG, Shah NH, Chu L. Bridging the implementation gap of machine learning in healthcare. *BMJ Innov*. 2019;6(2):45-47. [FREE Full text] [doi: [10.1136/bmjinnov-2019-000359](https://doi.org/10.1136/bmjinnov-2019-000359)]
23. Aristidou A, Jena R, Topol EJ. Bridging the chasm between AI and clinical implementation. *Lancet*. 2022;399(10325):620. [doi: [10.1016/S0140-6736\(22\)00235-5](https://doi.org/10.1016/S0140-6736(22)00235-5)] [Medline: [35151388](https://pubmed.ncbi.nlm.nih.gov/35151388/)]
24. Artificial intelligence funding. National Institute for Health and Care Research. URL: <https://www.nihr.ac.uk/explore-nihr/funding-programmes/ai-award.htm> [accessed 2023-02-22]
25. Krasniansky A, Evans B, Zweig M, from: W, Shah P, Hawks C, et al. 2021 year-end digital health funding: seismic shifts beneath the surface. *Rock Health*. 2022. URL: <https://rockhealth.com/insights/2021-year-end-digital-health-funding-seis-mic-shifts-beneath-the-surface/> [accessed 2023-02-22]

26. Ioannidis JPA, Bossuyt PMM. Waste, leaks, and failures in the biomarker pipeline. *Clin Chem*. 2017;63(5):963-972. [FREE Full text] [doi: [10.1373/clinchem.2016.254649](https://doi.org/10.1373/clinchem.2016.254649)] [Medline: [28270433](https://pubmed.ncbi.nlm.nih.gov/28270433/)]
27. Wouters OJ, McKee M, Luyten J. Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA*. 2020;323(9):844-853. [FREE Full text] [doi: [10.1001/jama.2020.1166](https://doi.org/10.1001/jama.2020.1166)] [Medline: [32125404](https://pubmed.ncbi.nlm.nih.gov/32125404/)]
28. Guidance for industry Q8(R2) pharmaceutical development. Federal Drug Administration. 2009. URL: <https://www.fda.gov/media/71535/download> [accessed 2024-02-27]
29. Quality by design. European Medicines Agency. 2018. URL: <https://www.ema.europa.eu/en/human-regulatory/research-development/quality-design> [accessed 2023-02-24]
30. U.S. Food and Drug Administration, Centre for Drug Evaluation and Research. Januvia NDA 21-995. U.S. Food and Drug Administration. 2006. URL: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2006/021995s000_ChemR.pdf [accessed 2024-02-27]
31. Gavyzia NDA 125486Orig1s000. U.S. Food and Drug Administration. 2013. URL: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/125486Orig1s000ChemR.pdf [accessed 2024-02-27]
32. Target product profile: point of care SARS-CoV-2 detection tests. Gov.UK. URL: <https://www.gov.uk/government/publications/how-tests-and-testing-kits-for-coronavirus-covid-19-work/target-product-profile-point-of-care-sars-cov-2-detection-tests> [accessed 2022-11-15]
33. Woodhall SC, Gorwitz RJ, Migchelsen SJ, Gottlieb SL, Horner PJ, Geisler WM, et al. Advancing the public health applications of chlamydia trachomatis serology. *Lancet Infect Dis*. 2018;18(12):e399-e407. [FREE Full text] [doi: [10.1016/S1473-3099\(18\)30159-2](https://doi.org/10.1016/S1473-3099(18)30159-2)] [Medline: [29983342](https://pubmed.ncbi.nlm.nih.gov/29983342/)]
34. About us. FIND. 2016. URL: <https://www.finddx.org/about/> [accessed 2022-11-24]
35. Lewin SR, Attoye T, Bansbach C, Doehle B, Dubé K, Dybul M, et al. Multi-stakeholder consensus on a target product profile for an HIV cure. *Lancet HIV*. 2021;8(1):e42-e50. [FREE Full text] [doi: [10.1016/S2352-3018\(20\)30234-4](https://doi.org/10.1016/S2352-3018(20)30234-4)] [Medline: [33271125](https://pubmed.ncbi.nlm.nih.gov/33271125/)]
36. Target product profile for sleeping sickness. Drugs for Neglected Diseases initiative (DNDi). 2020. URL: <https://dndi.org/diseases/sleeping-sickness/target-product-profile/> [accessed 2022-11-15]
37. Preferred Product Characteristics for Personal Protective Equipment for the Health Worker on the Frontline Responding to Viral Hemorrhagic Fevers in Tropical Climates. Geneva, Switzerland. World Health Organisation; 2018.
38. Malaria Vaccines: Preferred Product Characteristics and Clinical Development Considerations. Geneva, Switzerland. World Health Organization; 2022.
39. Measles-Rubella Microarray Patch (MR-MAP) target product profile. In: World Health Organization. Geneva, Switzerland. World Health Organization; 2020.
40. Pellé KG, Rambaud-Althaus C, D'Acromont V, Moran G, Sampath R, Katz Z, et al. Electronic clinical decision support algorithms incorporating point-of-care diagnostic tests in low-resource settings: a target product profile. *BMJ Glob Health*. 2020;5(2):e002067. [FREE Full text] [doi: [10.1136/bmjgh-2019-002067](https://doi.org/10.1136/bmjgh-2019-002067)] [Medline: [32181003](https://pubmed.ncbi.nlm.nih.gov/32181003/)]
41. High-priority target product profile for hepatitis C diagnosis in decentralized settings: report of a consensus meeting. FIND. 2015. URL: <https://forumresearch.org/projects/hcv-drug-development-advisory-group/additional-projects-of-the-hcv-drag/1167-high-priority-tpp-for-hepatitis-c-diagnosis-in-decentralized-settings-report-of-a-consensus-meeting> [accessed 2024-02-27]
42. Payment guidance for researchers and professionals, version 1.4. National Institute for Health and Care Research. 2023. URL: <https://www.nihr.ac.uk/documents/payment-guidance-for-researchers-and-professionals/27392> [accessed 2023-02-22]
43. UK standards for public involvement. National Institute for Health Research. 2019. URL: <https://www.invo.org.uk/wp-content/uploads/2019/11/UK-standards-for-public-involvement-v6.pdf> [accessed 2024-02-27]
44. Hogg HDJ, Al-Zubaidy M, Talks J, Denniston AK, Kelly CJ, Malawana J, et al. Stakeholder perspectives of clinical artificial intelligence implementation: systematic review of qualitative evidence. *J Med Internet Res*. 2023;25:e39742. [FREE Full text] [doi: [10.2196/39742](https://doi.org/10.2196/39742)] [Medline: [36626192](https://pubmed.ncbi.nlm.nih.gov/36626192/)]
45. Cocco P, Ayaz-Shah A, Messenger MP, West RM, Shinkins B. Target product profiles for medical tests: a systematic review of current methods. *BMC Med*. 2020;18(1):119. [FREE Full text] [doi: [10.1186/s12916-020-01582-1](https://doi.org/10.1186/s12916-020-01582-1)] [Medline: [32389127](https://pubmed.ncbi.nlm.nih.gov/32389127/)]
46. Search strategies. *BMJ*. 2017. URL: <https://bestpractice.bmj.com/info/toolkit/learn-ebm/study-design-search-filters/> [accessed 2023-05-12]
47. Wu L. Research guides: electronic health records: EHR literature. University of Tennessee Health Science Center. 2017. URL: <https://libguides.uthsc.edu/c.php?g=657371&p=4615105> [accessed 2023-05-12]
48. Godin K, Stapleton J, Kirkpatrick SI, Hanning RM, Leatherdale ST. Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. *Syst Rev*. 2015;4:138. [FREE Full text] [doi: [10.1186/s13643-015-0125-0](https://doi.org/10.1186/s13643-015-0125-0)] [Medline: [26494010](https://pubmed.ncbi.nlm.nih.gov/26494010/)]
49. Unsworth H, Wolfram V, Dillon B, Salmon M, Greaves F, Liu X, et al. Building an evidence standards framework for artificial intelligence-enabled digital health technologies. *Lancet Digit Health*. 2022;4(4):e216-e217. [FREE Full text] [doi: [10.1016/S2589-7500\(22\)00030-9](https://doi.org/10.1016/S2589-7500(22)00030-9)] [Medline: [35337640](https://pubmed.ncbi.nlm.nih.gov/35337640/)]

50. Evidence standards framework for digital health technologies. National Institute of Health and Care Excellence. 2022. URL: <https://www.nice.org.uk/corporate/ecd7> [accessed 2024-02-27]
51. Greenhalgh T, Wherton J, Papoutsis C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J Med Internet Res*. 2017;19(11):e367. [FREE Full text] [doi: [10.2196/jmir.8775](https://doi.org/10.2196/jmir.8775)] [Medline: [29092808](https://pubmed.ncbi.nlm.nih.gov/29092808/)]
52. Maniatopoulos G, Procter R, Llewellyn S, Harvey G, Boyd A. Moving beyond local practice: reconfiguring the adoption of a breast cancer diagnostic technology. *Soc Sci Med*. 2015;131:98-106. [FREE Full text] [doi: [10.1016/j.socscimed.2015.02.036](https://doi.org/10.1016/j.socscimed.2015.02.036)] [Medline: [25769108](https://pubmed.ncbi.nlm.nih.gov/25769108/)]
53. Richards H, Emslie C. The 'doctor' or the 'girl from the university'? Considering the influence of professional roles on qualitative interviewing. *Fam Pract*. 2000;17(1):71-75. [FREE Full text] [doi: [10.1093/fampra/17.1.71](https://doi.org/10.1093/fampra/17.1.71)] [Medline: [10673494](https://pubmed.ncbi.nlm.nih.gov/10673494/)]
54. Orlipp M. Keeping and using reflective journals in the qualitative research process. *TQR*. 2008;13(4):695-705. [FREE Full text] [doi: [10.46743/2160-3715/2008.1579](https://doi.org/10.46743/2160-3715/2008.1579)]
55. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)] [Medline: [32100154](https://pubmed.ncbi.nlm.nih.gov/32100154/)]
56. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *BMJ*. 2020;370:m3164. [FREE Full text] [doi: [10.1136/bmj.m3164](https://doi.org/10.1136/bmj.m3164)] [Medline: [32909959](https://pubmed.ncbi.nlm.nih.gov/32909959/)]
57. Tallouzi MO, Mathers JM, Moore DJ, Bucknall N, Calvert MJ, Murray PI, et al. COSUMO Working Group. Development of a core outcome set for clinical trials in non-infectious uveitis of the posterior segment. *Ophthalmology*. 2021;128(8):1209-1221. [FREE Full text] [doi: [10.1016/j.ophtha.2021.01.022](https://doi.org/10.1016/j.ophtha.2021.01.022)] [Medline: [33515595](https://pubmed.ncbi.nlm.nih.gov/33515595/)]
58. Denkinger CM, Dolinger D, Schito M, Wells W, Cobelens F, Pai M, et al. Target product profile of a molecular drug-susceptibility test for use in microscopy centers. *J Infect Dis*. 2015;211(Suppl 2):S39-S49. [FREE Full text] [doi: [10.1093/infdis/jiu682](https://doi.org/10.1093/infdis/jiu682)] [Medline: [25765105](https://pubmed.ncbi.nlm.nih.gov/25765105/)]
59. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19(6):349-357. [FREE Full text] [doi: [10.1093/intqhc/mzm042](https://doi.org/10.1093/intqhc/mzm042)] [Medline: [17872937](https://pubmed.ncbi.nlm.nih.gov/17872937/)]
60. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2):e1000217. [FREE Full text] [doi: [10.1371/journal.pmed.1000217](https://doi.org/10.1371/journal.pmed.1000217)] [Medline: [20169112](https://pubmed.ncbi.nlm.nih.gov/20169112/)]
61. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET handbook: version 1.0. *Trials*. 2017;18(Suppl 3):280. [FREE Full text] [doi: [10.1186/s13063-017-1978-4](https://doi.org/10.1186/s13063-017-1978-4)] [Medline: [28681707](https://pubmed.ncbi.nlm.nih.gov/28681707/)]
62. Kadam R, White W, Banks N, Katz Z, Dittrich S, Kelly-Cirino C. Target product profile for a mobile app to read rapid diagnostic tests to strengthen infectious disease surveillance. *PLoS One*. 2020;15(1):e0228311. [FREE Full text] [doi: [10.1371/journal.pone.0228311](https://doi.org/10.1371/journal.pone.0228311)] [Medline: [31995628](https://pubmed.ncbi.nlm.nih.gov/31995628/)]

Abbreviations

- AI:** artificial intelligence
- ARIAS:** automated retinal imaging analysis software
- COREQ:** Consolidated Criteria for Reporting Qualitative Research
- CQC:** Care Quality Commission
- DES:** diabetic eye screening
- DR:** diabetic retinopathy
- E-DES:** English diabetic eye screening
- ESF:** Evidence Standards Framework
- MHRA:** Medicines and Healthcare Products Regulatory Agency
- ML:** machine learning
- ML-ARIAS:** machine learning-automated retinal imaging analysis software
- NASSS:** nonadoption, abandonment, scale-up, spread, and sustainability
- NICE:** National Institute for Health and Care Excellence
- NIHR:** National Institute for Health and Care Research
- NSC:** National Screening Committee
- SAG:** study advisory group
- SMG:** study management group
- TPP:** target product profile
- UG:** user group

Edited by A Mavragani; submitted 14.09.23; peer-reviewed by G Lim, A Hidki; comments to author 26.12.23; revised version received 02.02.24; accepted 13.02.24; published 27.03.24

Please cite as:

Macdonald T, Dinnes J, Maniatopoulos G, Taylor-Phillips S, Shinkins B, Hogg J, Dunbar JK, Solebo AL, Sutton H, Attwood J, Pogose M, Given-Wilson R, Greaves F, Macrae C, Pearson R, Bamford D, Tufail A, Liu X, Denniston AK

Target Product Profile for a Machine Learning–Automated Retinal Imaging Analysis Software for Use in English Diabetic Eye Screening: Protocol for a Mixed Methods Study

JMIR Res Protoc 2024;13:e50568

URL: <https://www.researchprotocols.org/2024/1/e50568>

doi: [10.2196/50568](https://doi.org/10.2196/50568)

PMID: [38536234](https://pubmed.ncbi.nlm.nih.gov/38536234/)

©Trystan Macdonald, Jacqueline Dinnes, Gregory Maniatopoulos, Sian Taylor-Phillips, Bethany Shinkins, Jeffrey Hogg, John Kevin Dunbar, Ameenat Lola Solebo, Hannah Sutton, John Attwood, Michael Pogose, Rosalind Given-Wilson, Felix Greaves, Carl Macrae, Russell Pearson, Daniel Bamford, Adnan Tufail, Xiaoxuan Liu, Alastair K Denniston. Originally published in JMIR Research Protocols (<https://www.researchprotocols.org>), 27.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Research Protocols, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.researchprotocols.org>, as well as this copyright and license information must be included.