



Computer-aided reading of tuberculosis chest radiography: moving the research agenda forward to inform policy

Faiz Ahmad Khan^{1,2}, Tripti Pande¹, Belay Tessema^{3,4}, Rinn Song^{3,5,6}, Andrea Benedetti^{1,2,7}, Madhukar Pai^{1,7}, Knut Lönnroth^{8,9} and Claudia M. Denkinger³

Affiliations: ¹McGill International Tuberculosis Centre, McGill University, Montreal, QC, Canada. ²Respiratory Epidemiology and Clinical Research Unit, Research Institute of the McGill University Health Centre, Montreal, QC, Canada. ³FIND, Geneva, Switzerland. ⁴Dept of Medical Microbiology, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia. ⁵Division of Infectious Diseases, Boston Children's Hospital, Boston, MA, USA. ⁶Dept of Pediatrics, Harvard Medical School, Boston, MA, USA. ⁷Depts of Medicine and of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada. ⁸Dept of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden. ⁹Global TB Programme, WHO, Geneva, Switzerland.

Correspondence: Faiz Ahmad Khan, 1001 boulevard Décarie, Room D05.2510, Montréal, Québec H4A 3J1 Canada. E-mail: faiz.ahmadkhan@mcgill.ca

 @ERSpublications

Key gaps limit the evidence base for computer-aided reading of TB on CXR. We describe a research agenda to fill them <http://ow.ly/JkmQ30cynWR>

Cite this article as: Ahmad Khan F, Pande T, Tessema B, *et al.* Computer-aided reading of tuberculosis chest radiography: moving the research agenda forward to inform policy. *Eur Respir J* 2017; 50: 1700953 [<https://doi.org/10.1183/13993003.00953-2017>].

Introduction

Current issues in tuberculosis detection and the potential of computer-aided chest radiography interpretation

In 2015, there were an estimated 10.4 million new tuberculosis (TB) cases, but only 6.1 million (59%) were detected, and notified to national TB programmes (NTPs) [1]. The World Health Organization (WHO) emphasises that more proactive efforts are needed to close this case detection gap to move towards TB elimination [2, 3]. Progress has been made in improving laboratory services in recent decades. New tests for TB diagnosis have become available, and their use is being scaled up [4]. Efforts have been made to improve the evaluation of people who seek care and have symptoms consistent with TB. However, many people with TB remain undiagnosed or are diagnosed and treated only after long delays [1].

A large proportion of persons with active TB do not have classical TB symptoms, while TB abnormalities can be detected early in the disease course with the help of chest radiography (CXR) [5]. However, access to

Received: May 09 2017 | Accepted: May 12 2017

Support statement: This work was supported by the Dutch government through a grant to FIND (PDP15CH14) and through a USAID grant to WHO. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. R. Song was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (K23HD072802). F. Ahmad Khan receives salary support from the Fonds de Recherche Québec Santé.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com

K. Lönnroth is a staff member of the WHO. The author alone is responsible for the views expressed in this publication and they do not necessarily represent the decisions or policies of the WHO.

The content of this work is copyright of the authors or their employers. Design and branding are copyright ©ERS 2017.

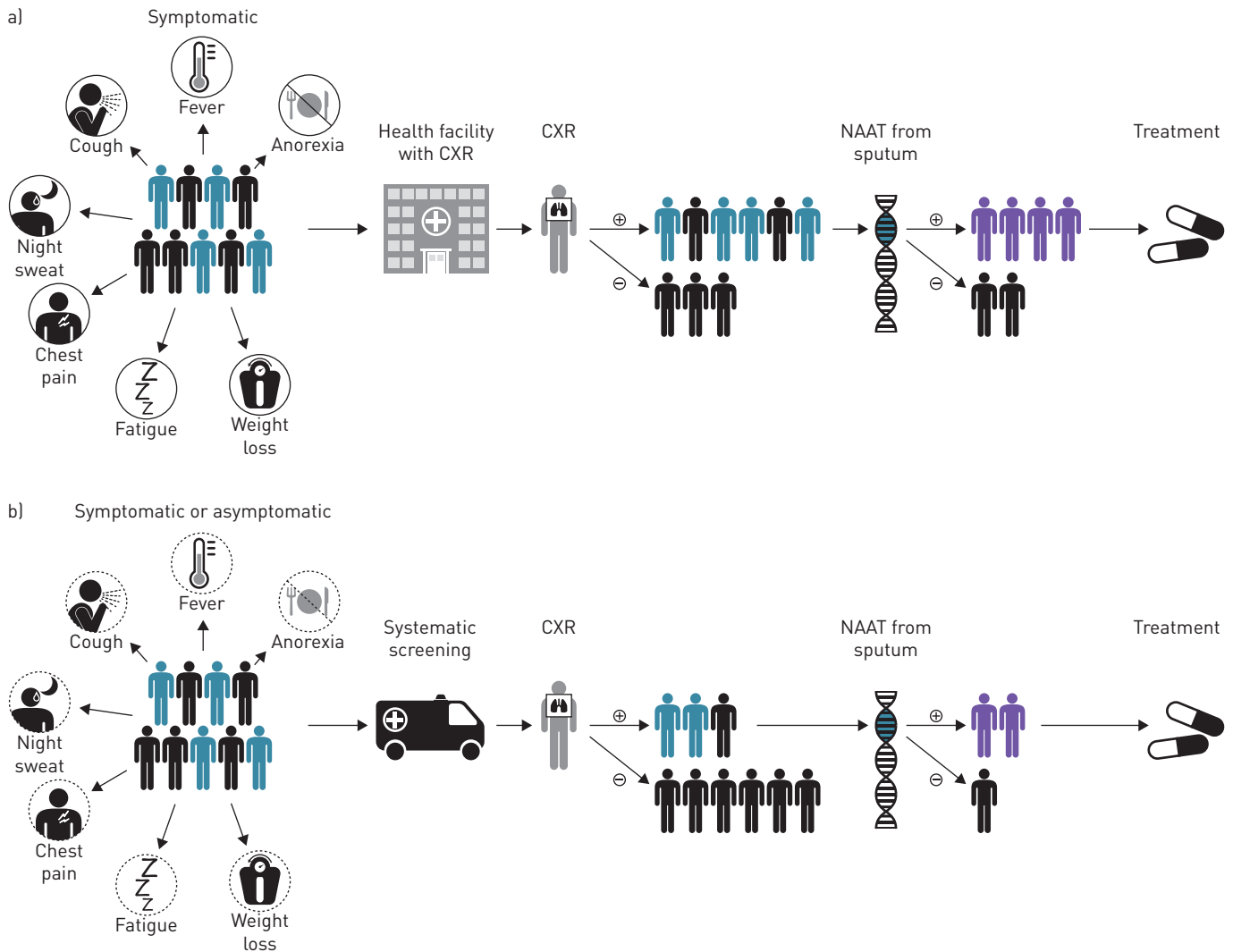


FIGURE 1 Chest radiography (CXR) with computer-aided detection (CAD) used as a triage test (a) or for systematic screening (b).

high-quality radiography with expert interpretation is limited in many settings, and high hardware costs as well as infrastructure requirements make it challenging to decentralise the technology [6]. For this reason, as well as concerns due to low specificity of CXR-based TB diagnosis and high intra- and inter-reader variability, previous WHO recommendations for resource-limited settings emphasised CXR to be used primarily when pulmonary TB cannot be confirmed bacteriologically, thus at the end of diagnostic algorithms.

Recently, however, CXR has been promoted as a useful tool that can be placed early in systematic screening (*i.e.* for patients who do not actively seek medical care but are at risk of having TB), and in triaging algorithms (*i.e.* among patients who present to care with symptoms) to identify those who need confirmatory laboratory testing (figure 1) [7]. An important reason for re-evaluating the role of CXR is the increased availability of digital radiography which presents numerous advantages over conventional radiography, such as lower running costs, better image quality, and better safety [8].

Radiography access has improved, but human resources for reading radiographs are still limited, especially in rural areas in resource-limited countries [9]. One potential solution is computer-aided detection (CAD), which is the use of computer programmes that analyse CXR for the presence of TB-compatible abnormalities. CAD could aid or replace radiologists or other health care staff reading radiographic films, and thus help realise the potential of rapidly scaling up digital radiology for TB systematic screening, and triaging [7].

Current landscape of radiography and computer-aided detection

To date, the only commercially available CAD programme evaluating TB on CXR is CAD4TB, a proprietary software owned by Delft Imaging Systems (Veenendaal, the Netherlands). Image interpretation



FIGURE 2 a) Inside a mobile tuberculosis screening truck fitted with a digital radiography machine in Karachi, Pakistan. Image courtesy of Interactive Research and Development, Aamir J. Khan. b) Screenshot of CAD4TB CXR analysis. Image courtesy of Interactive Research and Development, Arman Majidullah.

with CAD4TB is based on machine-learning methods. The programme analyses radiography image characteristics, and, using distributions developed from a set of training images, produces an abnormality score ranging from 0 to 100 (higher scores suggest greater likelihood of TB) (figure 2).

Clinical decision making requires categorical interpretation of radiographic abnormalities; hence, to operationalise the continuous abnormality score in a clinically meaningful way, a “threshold score” needs to be chosen: TB is ruled out if a radiograph’s abnormality score is below this threshold, whereas TB remains possible if the score is equal to or greater than this threshold. Currently, the developer does not recommend a threshold score, instead users are advised to choose their own based on available data from the intended use population. The challenges and potential limitations of this approach are discussed below.

Other CAD solutions in development include CADx (Advenio, India). CADx, similar to CAD4TB, is hardware neutral. Another group has recently evaluated software programmes that use deep learning algorithms, GoogleNet and AlexNet, with promising results in a case–control design diagnostic study [10]. All solutions, in their current form, have been developed to assess the likelihood of TB but do not explicitly offer information about other lung findings and conditions.

Evidence on currently available computer-aided detection solutions

To inform WHO guidance on CXR in TB detection, we recently conducted two systematic reviews of the evidence base as of July 2016 addressing the diagnostic accuracy of CAD4TB for microbiologically confirmed pulmonary TB, with the second review also including unpublished studies [7, 11]. No published studies are available for CADx from Advenio to date. We identified nine studies that provided data on diagnostic accuracy compared to a microbiological reference standard of culture or nucleic acid amplification test (NAAT). CXR was used to evaluate persons seeking care with symptoms of presumptive TB, *i.e.* a triage use-case, in six studies, and for systematic screening regardless of symptoms (*e.g.* in active case finding or prevalence surveys) in three studies. Hereafter, we will use the terms triage *versus* systematic screening to separate these two main use cases.

The review suggested the CAD4TB software is capable of achieving a high sensitivity but at the cost of a low specificity for the triage use-case, *i.e.* sensitivities of >85% were associated with specificities ranging from 23% to 69%; and that there were too few screening use-case studies to confidently estimate diagnostic accuracy in that scenario. Importantly, the review identified a number of limitations that could have resulted in a biased assessment of diagnostic accuracy, and limited the generalisability of the findings (table 1). One overarching concern was that several versions of CAD4TB exist, and that the evidence base for recent versions, in particular, consists only of a small number of studies. Another broad concern is that most studies have not evaluated the performance of CAD4TB when it is operationalised the way it will most likely be used in the field. For example, in two out of the three systematic screening studies, the accuracy of CAD4TB was retrospectively evaluated only on a subset of individuals who were symptomatic or in whom human readers had identified radiographs as abnormal; hence, these studies are not applicable to situations where the software will be applied for systematic screening.

TABLE 1 Limitations observed in CAD4TB evidence-base

QUADAS-2 Domain	Limitations observed in CAD4TB evidence base	Published and unpublished studies affected by limitation (number of studies with major concern/total)
Patient selection	Exclusion of participants	Systematic screening (2/3)
	Potential biased assessment of diagnostic accuracy Limited generalisability to how software will be used in the field	Triage (1/6)
Index test (CAD4TB)	CXR for training, and evaluation of accuracy originate from same dataset	Systematic screening (1/3)
	Potential biased assessment of diagnostic accuracy Limited generalisability to how software will be used in the field	Triage (2/6)
	Threshold score not pre-specified	Systematic screening (2/3)
Reference standard for TB	Potential biased assessment of diagnostic accuracy Limited generalisability to how software will be used in the field	Triage (4/6)
	Composite reference standard	Systematic screening (0/3) [#]
	Potential biased assessment of diagnostic accuracy	Triage (1/6)
	Use of NAAT as reference standard	Systematic screening (0/3) [#]
	Potential biased assessment of diagnostic accuracy due to limited sensitivity for smear-negative TB	Triage (3/6)

TB: tuberculosis; CXR: chest radiography; NAAT: nucleic acid amplification test. [#]: microbiological tests were used to diagnose TB in all persons with symptoms or human-reader identified radiographic abnormalities; asymptomatic persons with radiographs interpreted as “normal” were assumed to not have TB, and did not undergo microbiological testing.

A strength of the evidence base is that most triage studies used either culture or NAAT as the reference standard to diagnose pulmonary TB. In one triage study, a composite of microbiological tests and clinical follow-up data was used as the reference standard for ruling out TB, this was identified as an important source of potential bias as it resulted in a large proportion of persons being excluded from analyses of diagnostic accuracy because they did not present for follow-up [12]. In systematic screening studies, microbiological tests were used to diagnose TB in all persons with symptoms or human-reader identified radiographic abnormalities, but asymptomatic persons with radiographs interpreted as “normal” were assumed to not have TB without undergoing microbiological tests. We did not consider this to be an important source of bias as the likelihood of false-negative classification of the untested subset was considered quite low in the context of screening.

There has been substantial variability in how threshold scores were selected. In the majority of systematic screening, and triage studies, investigators either reported the diagnostic accuracy across a range of threshold scores, or retrospectively selected a score that matched the performance of a human reader. Other studies chose to define the threshold using a training set of CXRs from the larger validation set. For research use, such approaches are valid, however, for a diagnostic test, pre-specified thresholds are needed. The requirement that each CAD4TB user identify an adequate threshold using a training set of local radiographs prior to use of the software will be difficult to operationalise while avoiding potential biases arising from the selection of radiographs for the training set, choice and use of reference standards, and insufficient power to ensure precision of estimates of diagnostic accuracy measures.

Pre-specified threshold scores for the different use cases will need to be determined on sets of CXRs that adequately represent the patient populations, chest pathology and differential diagnoses found, and that have not been used for training of the software. The currently published and unpublished evidence does not meet these criteria as shown by the high prevalence of TB (*i.e.* 15–60% in triage studies and in systematic screening studies 2–4%), and of HIV (*e.g.* in four of the six triage studies 33–68%), thus results cannot be generalised across populations. Finally, little data are available on whether patient characteristics modify the software’s accuracy, most notably, age, gender, HIV status, previous history of active TB, co-morbidities and smear status.

Based on the limited number of studies, and the methodological issues outlined above, the WHO has not yet started a process to develop guidelines on CAD for TB detection. The 2016 document summarising current WHO recommendations on CXR suggested for CAD to be used for research, ideally following a protocol that contributes to the required evidence base for guideline development [7].

The ideal computer-aided detection solution

The characteristics of an ideal CAD solution have been identified with input from experts in the field. The ideal CAD solution should not only indicate possible TB but also identify other abnormalities that require further clinical workup (e.g. cardiomegaly, lung nodule, interstitial patterns). A categorisation of “abnormal – likely TB” *versus* “abnormal – unlikely TB” *versus* “normal” should be possible at a minimum. This would overcome ethical concerns of the use of CAD4TB instead of a human reader, in patients who present with symptoms, and are identified not to have TB on CAD4TB systematic screening, but require further assessment for pathologies present on radiography. Furthermore, thresholds of CAD should be predefined to optimise performance across the two use cases.

A CAD solution should be easy to use, ideally by a minimally trained healthcare worker (with less than 1 day of training). Performance characteristics for the detection of pulmonary TB (TB) in adults and children should be similar to those currently reported with human readers across the two use cases [13].

The CAD solution should be able to interpret images from different CXR devices, and different populations without changes in accuracy. The software should either be able to identify quality issues with the CXR, and indicate them to the user or at a minimum the developer should offer an assessment of radiograph quality, and suitability for use with the CAD solution prior to use for clinical care. The CAD solution should be able to transmit images for telemedicine consultations with experts if the need arises, and do so through secure means that protect patient confidentiality.

The throughput should be at least one radiograph per minute, and optimally faster. The interpretation should be less than 1 USD per image, which would amount to a monthly cost of around 480USD if the software were utilised to evaluate, on average, 20 patients per day, 6 days per week.

Research questions to be addressed for policy guidance

Several key questions need to be addressed for CAD to be used as a clinical aid.

In order for CAD programmes that produce continuous abnormality scores, such as CAD4TB, to be used as a diagnostic test, and not for research only, a threshold needs to be predefined. Thresholds might differ across use cases as it is expected that in the context of systematic screening, CXRs would have minimal pathology, while in the context of triage testing, more substantial lung findings would be present. Thresholds can be specified on sets of CXRs that adequately represent the populations for the different use cases. The sets of CXRs need to contain key populations as characterised by HIV status, age, gender, prior TB, co-morbidities (e.g. obstructive lung disease, silicosis), and smear status in order to guarantee generalisability of the data within these groups.

The performance of CAD against human readers is important for implementation decisions, even though it is challenging to measure due to variability in the characteristics of human readers. Ideally, a comparison is done against well-trained radiologists; however, implementation decisions need to consider that availability of radiologists is limited, and often much less trained providers end up reading a CXR. Depending on the use case, one can also expect different performance of readers (in the field for systematic screening *versus* in the hospital for a triage use case). A comparison to human readers therefore should specify the training level of the reader, and ideally include both well-, and less trained readers. With predefined thresholds, CAD will be able to mimic a binary or categorical interpretation of results, similar to what human readers use in clinical settings.

As a software solution, version updates need to be expected more frequently than is the case for *in vitro* diagnostics, and a rapid review of comparative performance needs to be feasible to make recommendations on the suitability of a new version. Regulatory bodies such as the Food and Drug Administration have provided guidance on this topic, and allow for reporting of performance on a predefined CXR image bank that is housed with the manufacturer [14]. In order to allow for comparability of data across different CAD solutions, and different versions of one solution, a predefined CXR image bank that allows for the assessment across key populations would better be housed with an independent, neutral organisation (such as WHO or a WHO collaborating centre).

Crucially important for the implementation of CAD is the recognition that CXRs are used for non-TB lung diseases. Thus, the clinical impact of a solution that is TB focused needs to be considered. In the context of the systematic screening use case, a CXR that identifies TB early can help prevent transmission

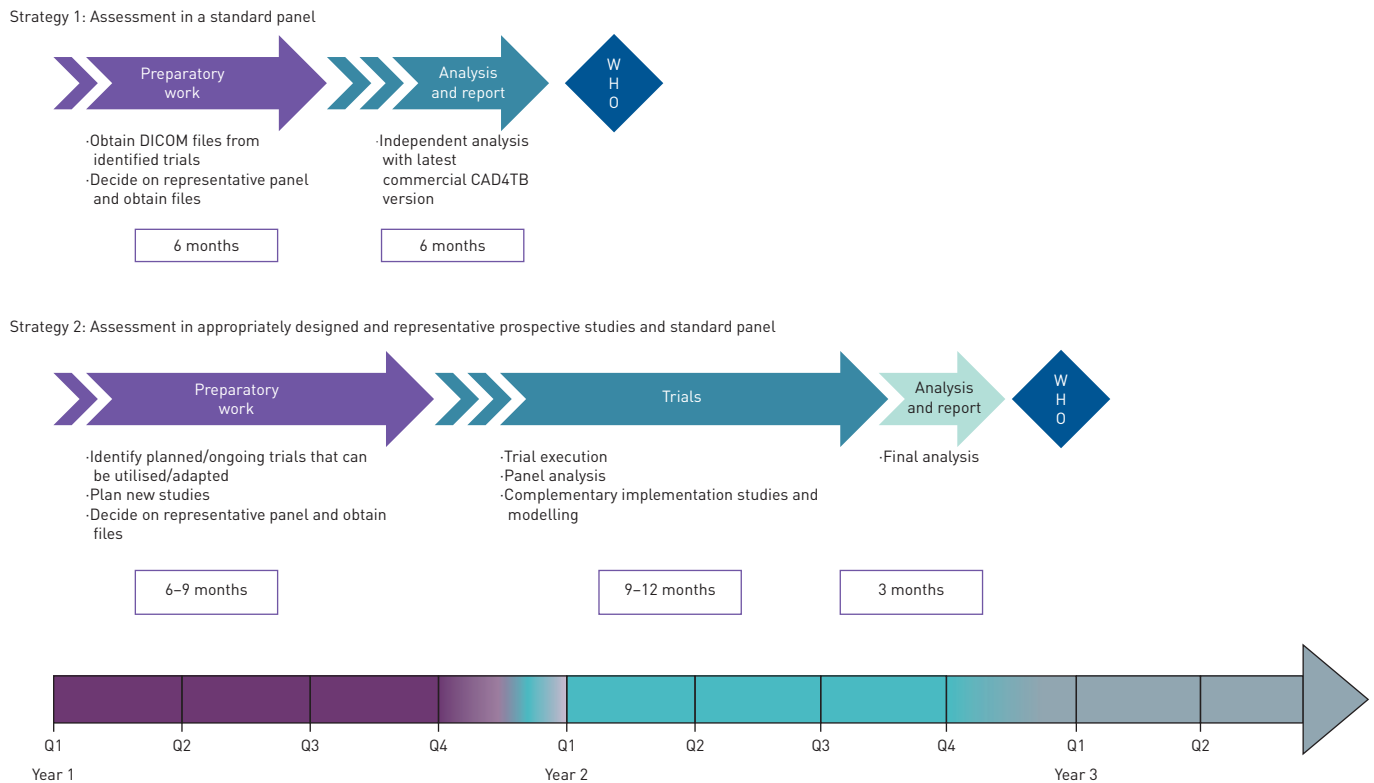


FIGURE 3 Strategies to inform guidance development and implementation of computer-aided detection for pulmonary tuberculosis.

prior to presentation to care. One can argue that missing other diseases in this context is not as harmful as the patient has not been seeking care. Nevertheless, a solution that does not pick up, for example, a bone mass or a lung cancer nodule, might miss an opportunity for early intervention and give a patient false reassurance. For patients who seek care because of symptoms (*i.e.* triage use case) the limitation of only identifying TB *versus* non-TB could weigh more heavily. It would be paramount for a CAD solution to trigger further workup by identifying an abnormality even if it is unlikely to be TB. In the absence of the ability to make the call “abnormal – unlikely TB”, all CXRs would have to be read by a radiologist in addition to CAD. This could have implications on cost-effectiveness. Similarly, because some CXRs lesions that are compatible with pulmonary TB could in fact be due to cancers or infections other than TB (*e.g.* cavities, nodules, pleural effusions), guidance will be needed about what clinicians should do for patients flagged by CAD4TB to have a CXR that is “abnormal – likely TB” but in whom microbiological tests do not identify pulmonary TB. Hence, safety, and ethical concerns need to be considered if issuing recommendations about use of TB-centric CAD if results are not verified by of human readers.

A number of implementation, and economic issues will also need to be considered to ensure acceptability, as well as equitable, and sustainable usage: these include costs of the software, updates, and different payment models.

Different strategies to address the research questions

We propose two strategies that could be used to generate the necessary evidence to answer the above research questions (figure 3).

Strategy 1: assessment in a standard panel

As in the case of standard sample panels that are available for test validation, we propose to assemble a bank of digital CXR files that has 1) representative spectrum of TB pathology [15]; 2) representative pathology of differential diagnoses [16]; 3) representative distribution of patients according to gender, HIV-status, age, and geographic origin; and 4) representative patient groups for the two different use cases of triage and systematic screening. This “standard panel” will be housed independent of CAD manufacturers/developers at WHO or a WHO collaborating centre. It will serve the purpose of evaluating available CAD solutions, and future versions of the same software, as well as allow for comparative assessments of novel CAD programmes. This assessment on a standard panel will be complimentary to an

TABLE 2 Study designs to evaluate diagnostic accuracy of computer-aided detection (CAD) programmes

Study characteristic	Ideal	Acceptable
Study population	Broad eligibility criteria to enhance generalisability. Sufficient sample size to permit stratified analyses by key characteristics. Should not have contributed CXR for developing/training CAD software that is under study. Prevalence of TB in study population should reflect that expected in similar use-case settings.	Limited to specific high-risk groups, or key populations.
Participant selection	Minimise exclusion of participants. Report number of persons screened for eligibility, ineligible for enrolment, and enrolled. Report reasons for ineligibility.	Minimise exclusion of participants. Report number of persons screened for eligibility, ineligible for enrolment, and enrolled.
Design	Prospective or retrospective study of diagnostic accuracy. CAD programme should be used how it will be used in the clinical context (including additional procedures, e.g. verification procedure prior to use of the software).	Cross-sectional study of diagnostic accuracy. If data from other studies are retrospectively used to study a CAD solution, investigators should report how data were collected in the original study, the criteria for selecting data for the CAD evaluation, and the number of participants from the original study whose data were excluded. CAD programme should be used how it will be used in the clinical context (including additional procedures, e.g. verification procedure prior to use of the software).
Reference standard	Microbiological reference standard of solid, and/or liquid cultures performed on all participants regardless of CXR result. [#] A combined clinical and microbiological reference standard can be considered. Here it is important that the final classification of patients should be well standardised or performed by an expert group that is blind to the CAD result. A microbiological reference standard should always be reported in addition.	Microbiological reference standard of NAAT performed on all participants regardless of CXR result. For systematic screening use-case only: liquid culture or NAAT performed on participants with a CXR classified as abnormal by a human reader, whereas pulmonary TB is assumed absent in CXRs classified as normal by a human reader. [¶]
Comparison to human readers	A comparison to a human reader should be reported. Human readers should: Be blind to results of CAD, and microbiological tests when interpreting CXR; Use simple two or three category readings [15]. Study should report training, and years of experience with radiographic reading of the human readers. For each category of experience, there should be more than one human reader (e.g. at least two experts, at least two non-experts). Inter-reader reliability should be reported.	A comparison to a human reader should be reported. Human readers should: Be blind to results of CAD, and microbiological tests when interpreting CXR; Use simple two or three category readings [15]. Study should report training, and years of experience with radiographic reading of the human readers. Inter-reader reliability should be reported.
Index test	CAD solution should not have been developed/trained with CXRs from population being studied.	Report if study site contributed CXRs for developing or training CAD programme, the number of CXR used, and whether this could affect generalisability of study results to sites where CXR have not contributed to software development or training.
Test performance and interpretation	Evaluate a pre-specified threshold score that is recommended for use in similar settings. Explore diagnostic heterogeneity arising from age, gender, HIV-infection and CD4 count, and smear status.	Report accuracy measures across a range of pre-specified threshold scores. Explore diagnostic heterogeneity arising from HIV-infection and smear status.

Continued

analysis of existing published data in a systematic review (ideally an individual patient data meta-analysis utilising the latest version of the CAD software).

The digital CXR files for this standard panel could be sourced from: 1) completed or ongoing studies that utilise digital CXR either in a triage or systematic screening use-case, and perform the required reference standard as defined below (but are not utilising an existing CAD software to avoid that developers access the files for software training), 2) TB prevalence surveys or 3) an imaging bank of existing digital CXR

TABLE 2 Continued

Study characteristic	Ideal	Acceptable
Detection of important non-TB abnormalities	<p>Include an assessment of the sensitivity of the CAD solution (compared to human readers) for identifying important non-TB abnormalities, defined as a radiological finding that could require treatment (e.g. bacterial pneumonia) or further clinical, and radiological follow-up (e.g. lung nodule/mass, pleural effusion).</p> <p>The reference standard for important non-TB abnormalities should be a panel composed of at least a radiologist and a clinician blinded to CAD results.</p>	<p>Include an assessment of the sensitivity of the CAD solution (compared to human readers) for identifying important non-TB abnormalities, defined as a radiological finding that could require treatment (e.g. bacterial pneumonia) or further clinical, and radiological follow-up (e.g. lung nodule/mass, unfolded aorta, pleural effusion).</p> <p>The reference standard for important non-TB abnormalities is a single specialist physician.</p>

CXR: chest radiography; TB: tuberculosis; NAAT: nucleic acid amplification test. [#]: a microbiological reference standard is preferred over human-interpretation because the role of CXR in TB diagnosis is to select persons to undergo microbiological testing. Studies assessing the accuracy of CAD for detecting microbiologically confirmed TB provide evidence that is directly applicable to its role in the field, whereas studies using human reading as a reference standard provide evidence that is indirect, and less generalisable. [¶]: due to large sample sizes of systematic screening studies, it may be unfeasible to perform a microbiological test on all participants; hence it would be reasonable to use a human-reader interpretation of a CXR as normal as the reference standard for absence of pulmonary TB. In such studies, the human reader should have been instructed to “over read” CXR to lower the false-negative rate.

files with appropriate pathology of TB, and differential diagnoses (e.g. International Organization for Migration database).

Strategy 2: assessment in appropriately designed, and representative prospective studies

The second strategy would involve identification of planned or ongoing trials to integrate an evaluation of CAD, or undertaking new studies. The design considerations should follow those suggested below. The second strategy could be complimented by an analysis of a standard panel.

An expert group commissioned by the WHO is currently in the process of evaluating the feasibility, and timeline of these strategies. While the first strategy is conceivable possible within months, the second would require at least 1.5 years (figure 3).

Suggested design elements for studies evaluating computer-aided detection

For all future studies of CAD solutions, investigators should consider certain design characteristics that we have enumerated in table 2, that will help to ensure that data generated are most informative for implementation purposes. We have focused on features that will minimise potential for bias, and increase the generalisability of the evidence-base. Reporting should follow guidelines for reporting diagnostic accuracy studies [17]. Table 2 is limited to studies of diagnostic accuracy; however, additional studies, and modelling are needed to inform guidance for best scale-up. Implementation studies and transmission modelling should inform best strategies to ensure maximal impact, for example, in the context of systematic screening within high-risk groups or infection control in the hospital setting through early triage of high-risk patients. Furthermore, studies should assess the cost-effectiveness of CAD in isolation but also within implementation strategies [18, 19]. Ideally, implementation studies would compare CAD against alternative solutions (e.g. tele-radiology), and against the status quo in different contexts [20].

Conclusions

CAD solutions might offer an opportunity to expand the use of CXR to improve case finding, infection control, and reduce the cost of case detection within triage algorithms. Further evidence is needed to guide the use of CAD in clinical care to assess its performance, ensure it meets ethical standards, and ensure the highest impact.

Acknowledgements

Author contributions: F. Ahmad Khan, T. Pande and B. Tessema reviewed the literature; figures were compiled by F. Ahmad Khan, C.M. Denkinger and M. Pai; F. Ahmad Khan, T. Pande, B. Tessema, R. Song, A. Benedetti, M. Pai, K. Lönnroth and C.M. Denkinger conducted the data collection, analysis, and interpretation. They also wrote and reviewed the article.

References

- 1 World Health Organization. Global Tuberculosis Control: WHO report 2016. Report No.: WHO/HTM/TB/2016.13. Geneva, World Health Organization, 2016.
- 2 Lönnroth K, Migliori GB, Abubakar I, *et al.* Towards tuberculosis elimination: an action framework for low-incidence countries. *Eur Respir J* 2015; 45: 928–952.
- 3 Uplekar M, Weil D, Lönnroth K, *et al.* WHO's new end TB strategy. *Lancet* 2015; 385: 1799–1801.
- 4 World Health Organization. Xpert MTB/RIF implementation manual: technical and operational 'how-to'; practical considerations; 2014. Report No.: WHO/HTM/TB/2014.1. Geneva, World Health Organization, 2014.
- 5 World Health Organization. Tuberculosis prevalence surveys: a handbook. Report No.: WHO/HTM/TB/2010.17. Geneva, World Health Organization, 2011.
- 6 Pedrazzoli D, Lalli M, Boccia D, *et al.* Can tuberculosis patients in resource-constrained settings afford chest radiography? *Eur Respir J* 2017; 49: 1601877.
- 7 World Health Organization. Chest radiography in tuberculosis detection - Summary of current WHO recommendations and guidance on programmatic approaches. Report No.: WHO/HTM/TB/2016.20. Geneva, World Health Organization, 2016.
- 8 Bansal GJ. Digital radiography. A comparison with modern conventional imaging. *Postgrad Med J* 2006; 82: 425–428.
- 9 Pande T, Pai M, Khan FA, *et al.* Use of chest radiography in the 22 highest tuberculosis burden countries. *Eur Respir J* 2015; 46: 1816–1819.
- 10 Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017; 162326.
- 11 Pande T, Cohen C, Pai M, *et al.* Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. *Int J Tuberc Lung Dis* 2016; 20: 1226–1230.
- 12 Breuninger M, van Ginneken B, Philipsen RH, *et al.* Diagnostic accuracy of computer-aided detection of pulmonary tuberculosis in chest radiographs: a validation study from sub-Saharan Africa. *PLoS One* 2014; 9: e106381.
- 13 Van't Hoog AH, Onozaki I, Lönnroth K. Choosing algorithms for TB screening: a modelling study to compare yield, predictive value and diagnostic burden. *BMC Infect Dis* 2014; 14: 532.
- 14 Food and Drug Administration (FDA). Guidance for Industry and Food and Drug Administration Staff - Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions. Report No.: 1697. Washington, FDA, 2012.
- 15 Pinto L, Dheda K, Theron G, *et al.* Development of a Simple Reliable Radiographic Scoring System to Aid the Diagnosis of Pulmonary Tuberculosis. *PLoS One* 2013; 8: e54235.
- 16 Rabinovitch B, Pai M. Radiology of tuberculosis: case studies for the GP. *GP Clinics* 2013; 4: 21–29.
- 17 Bossuyt PM, Reitsma JB, Bruns DE, *et al.* STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; 351: h5527.
- 18 Murray M, Cattamanchi A, Denkinger CM, *et al.* Cost-effectiveness of triage testing for facility-based systematic screening of tuberculosis among Ugandan adults. *BMJ Glob Health* 2016; 1: e000064.
- 19 van't Hoog AH, Cobelens F, Vassall A, *et al.* Optimal triage test characteristics to improve the cost-effectiveness of the Xpert MTB/RIF assay for TB diagnosis: a decision analysis. *PLoS One* 2013; 8: e82786.
- 20 Arinaminpathy N, Dowdy D. Understanding the incremental value of novel diagnostic tests for tuberculosis. *Nature* 2015; 528: S60–S67.