



World Health
Organization

TDR  For research on
diseases of poverty
UNICEF • UNDP • World Bank • WHO

Determining the local calibration of computer-assisted detection (CAD) thresholds and other parameters

A toolkit to support the effective use of CAD for TB screening



FOREWORD

Chest radiography (CXR) plays a key role in the detection of pulmonary tuberculosis (TB). While current World Health Organization (WHO) guidelines recommend diagnosis on the basis of bacteriological or molecular findings, the integration of CXR into screening and triaging algorithms can assist with early detection of TB, as CXR is a highly sensitive tool for detecting TB disease, and can guide the effective use of diagnostic testing to improve case detection and cost-efficiency.

Computer-aided detection (CAD) products use artificial intelligence (AI) to analyse CXR images for the presence of abnormalities suggestive of pulmonary TB, producing an abnormality score that can be used to determine the need for follow-on diagnostic testing for TB relative to a selected threshold. CAD technology can improve the feasibility and performance of CXR for screening and triage for TB disease and may benefit TB programmes by enhancing capacity for TB screening. Such technology can replace or augment human expert interpretation of plain CXR when screening for TB and can avoid inter-reader variability and reduce delays in reading radiographs when skilled personnel are scarce.

Following a review of the evidence, a new recommendation that supports the use of CAD for adults over the age of 15 years in place of human readers for interpretation of digital chest radiography in both screening and triage for pulmonary TB disease was included in the updated WHO TB screening guidelines and recommendations. This new recommendation is expected to increase the consideration of CAD by NTPs and other partners as a tool to enhance the use of CXR in screening and triage contexts.

Effectively integrating CAD products into routine programming requires determining an appropriate CAD threshold that will be used to signal probable TB cases and trigger further TB diagnostic evaluation. Identifying the ideal threshold for each implementation requires calibration of CAD products based on the local context and intended use case, as well as decision making around the goals for screening and acceptable costs.

This toolkit has been developed by the Special Programme for Research and Training in Tropical Diseases (TDR) and the Global TB Programme (GTB) at WHO, in collaboration with partners and experts working on CAD and TB screening. This toolkit aims to support the implementation of CAD for TB screening and detection and is intended for use by national TB programmes (NTPs) and other implementers who have decided to use CAD.

The specific objectives of this toolkit are to:

- Support new users of CAD to understand threshold scores and their programmatic implications within the context of a TB screening or triage programme;
- Describe a simplified operational research protocol to collect and analyse the required data necessary to determine the appropriate CAD threshold for a specific CAD implementation;
- Support the analysis, interpretation and application of the resulting CAD diagnostic performance data and cost estimates to determine the most appropriate CAD thresholds based on local context and use case.

These objectives are addressed through the following three components that make up this toolkit:

- **Part A** is a background document that introduces CAD calibration studies, including an overview of proposed study designs, procedures and outcomes of interest.
- **Part B** is a generic study protocol that describes the proposed research methodology and procedures for a calibration study. The protocol can be adapted by users and used to seek ethics approval. Adapting the protocol will be driven largely by the specific study design and intended use case selected by users.
- **Part C** is a user guide to support the use of the [CAD for TB detection calibration tool](#), an online tool which has been developed to support data analysis for calibration studies. The online tool can help users to determine an appropriate CAD threshold by calculating the accuracy (sensitivity and specificity) of a range of possible threshold values, and demonstrating the practical implications of these thresholds on factors including the number of over- and under-diagnosed TB cases and costs incurred related to follow up confirmatory testing. Part C also includes guidance to help users interpret the results generated by the online tool and how to practically apply them to their local context.

By providing a standardised methodology, it is envisaged that this toolkit may support the collection and analysis of standardised data on CAD performance across various settings that can be used to better inform the use of CAD globally.

TABLE OF CONTENTS

FOREWORD	2
TABLE OF CONTENTS	4
LIST OF TABLES	7
LIST OF FIGURES	7
LIST OF BOXES	7
ACRONYMS	8
ACKNOWLEDGEMENTS.....	9
PART A: STUDY OVERVIEW	11
1. Introduction.....	11
1.1 CAD technologies: an overview	11
1.2 Consideration of the use and approval of CAD by WHO	12
1.3 Integration of CAD into screening or triage algorithms.....	13
1.4 Calibrating CAD	13
2. Implications of different settings for CAD use	14
3. CAD calibration studies	15
3.1 Proposed study designs and procedures	16
3.2 Selecting the most appropriate study design for calibration studies	18
3.3 Bacteriological reference standard.....	21
3.4 Number of participants required for calibration studies.....	22
3.5 Sub-analysis to determine unique CAD thresholds for specific populations.....	24
3.6 Study outcomes of interest	25
3.7 Data collection and analysis	27
3.8 Study duration	29
4. Countries with future plans to use CAD.....	30
5. Calibration study costs.....	30
PART B: GENERIC CAD CALIBRATION STUDY PROTOCOL	33
1. Background.....	38

2.	Study objectives.....	39
3.	Study design	39
3.1	Study settings	40
3.2	Study population.....	40
3.3	Participant selection and sample size	42
4.	Outcomes of interest	43
5.	Study procedures	44
5.1	Bacteriological testing.....	45
6.	Data collection and management	46
7.	Data analysis.....	47
8.	Ethical considerations	48
8.1	Ethics approval.....	48
8.2	Protection of patient and confidentiality	48
8.3	Informed consent.....	49
9.	Study timelines	50
	PART C: CAD FOR TB DETECTION CALIBRATION TOOL USER GUIDE.....	53
1.	Introduction.....	53
2.	Overview of the CAD for TB detection calibration tool.....	53
3.	Preparing the data set.....	54
4.	Uploading the data set.....	54
4.1	Setting the parameters	55
5.	Selecting variables for ROC curve and outcomes table	55
6.	Selecting variables for sub - analyses	57
7.	Interpreting calibration analysis.....	58
	ANNEX.....	61
	REFERENCES	63

Draft

LIST OF TABLES

Table 1: Inclusion criteria for cross-sectional calibration studies.....	16
Table 2: Definitions of eligibility for case-control calibration studies.....	17
Table 3: Potential sources of pre-existing data for calibration studies.....	17
Table 4: Comparison of study designs.....	19
Table 5: Sample size estimations by study design and level of sensitivity (based on $\pm 5\%$ precision)	22
Table 6: Sample size estimations by TB prevalence and level of sensitivity (based on $\pm 5\%$ precision)	24
Table 7: Key outcome definitions.....	26
Table 8: Variables for data collection.....	27
Table 9: Suggested budget lines and estimated expenses for a CAD study.....	31
Table 10: Key outcome definitions.....	43
Table 11: Data collection variables for CAD calibration study.....	46

LIST OF FIGURES

Figure 1: Sensitivity vs. specificity across CAD threshold spectrum	14
Figure 2: Comparison of CAD readings with bacteriological test results	21
Figure 3: Illustrative receiver operator characteristic (ROC) curve	29
Figure 4: Comparison of CAD reading with bacteriological test results.....	47
Figure 5: Landing page of online tool	54
Figure 6: Uploading the dataset.....	54
Figure 7: Selecting variables for ROC curve and outcomes table	56
Figure 8: ROC curve display.....	56
Figure 9: Selecting variables for sub-analysis	57
Figure 10: Selecting variables for sub-analysis	57

LIST OF BOXES

Box 1: Illustrative example of a cross-sectional calibration study within a TB screening setting	17
Box 2: Illustrative example of a case-control calibration study within a TB triage setting	18
Box 3: Illustrative example of sample size calculation for a combined case-control study	23
Box 4: Illustrative example of a sample size calculation for a cross-sectional study	23
Box 5: Illustrative example of a sample size calculation for a cross-sectional study	24
Box 6: Illustrative example of a sample size calculation for sub-analyses	25
Box 7: Illustrative example of a country with future plans to use CAD	30
Box 8: Case study (based on a fictional example)	58

ACRONYMS

AI	Artificial intelligence
CAD	Computer-aided detection
CXR	Chest radiography (X-Ray)
DICOM	Digital imaging and communications in medicine
FN	False negative
FP	False positive
GDG	Guideline development group
mWRDs	Molecular WHO-recommended rapid diagnostic tests
NTP	National Tuberculosis Programme
NPV	Negative predictive value
PPV	Positive predictive value
Se	Sensitivity
Sp	Specificity
TN	True negative
TP	True positive

ACKNOWLEDGEMENTS

This toolkit was developed by Corinne Merle, Vanessa Veronese and Debora Pedrazzoli from the Special Programme for Research and Training in Tropical Diseases (TDR) and Cecily Miller and Dennis Falzon from the Global TB Programme (GTB) at the World Health Organization (WHO) in collaboration with the following group of experts: Morten Ruhwald and Sandra Kik (The Foundation for Innovative New Diagnostics; FIND), Jacob Creswell (Stop TB Partnership), Katherine Fielding (London School of Hygiene and Tropical Medicine), Nazir Ismail (WHO/GTB). The protocol was reviewed by the following external peer reviewers: Monde Muyoyeta (Centre for Infectious Disease Research, Zambia), Andrew Codlin and Luan Vo (Friends for International TB Relief, Vietnam), and Shifa S Habib (Community Health Solutions, Pakistan). Funding was provided by grants provided to WHO by USAID, the Government of the Netherlands and the Russian Federation.

The [online tool that accompanies this toolkit](#) was developed by Takuya Yamanaka, from the London School of Hygiene & Tropical Medicine, UK and Nagasaki University, School of Tropical Medicine & Global Health, Japan.

PART A: STUDY OVERVIEW



TB screening in Vietnam
Andrew Codlin, Friends for International Relief

PART A: STUDY OVERVIEW

1. Introduction

This section provides an introduction to CAD technology and calibration and an overview of the proposed approach to conducting a calibration study. The template for the study protocol is included in Part B.

The calibration study has been designed based on the following assumptions:

- Users are not seeking to validate CAD by comparing performance to a human reader; rather users have already decided to integrate CAD into their TB screening and/or triage algorithms and require assistance in determining the most appropriate thresholds based on their local settings and use case;
- Users have, or are preparing to have, a functional set up for CAD in place that includes:
 - Digital radiography equipment, computed radiography (CR) or digital radiography (DR) retrofit radiography equipment that generates a digital CXR images (DICOM¹)
 - A CAD product ([for an overview of available products, click here](#))
 - Staff who are trained on digital radiography and CAD
 - Capacity to undertake bacteriological confirmation (culture or molecular WHO recommended rapid diagnostic testing such as Xpert MTB/RIF^R)
 - Infrastructure to host and run the technologies (i.e., electricity, internet)

1.1 CAD technologies: an overview

Chest radiography (CXR) plays a key role in the detection of pulmonary TB. While current World Health Organization (WHO) guidelines recommend diagnosis on the basis of bacteriological or molecular findings, the integration of CXR into screening and triaging algorithms can assist with early detection of TB through screening, triage, and as part of comprehensive clinical evaluation. CXR is a highly-sensitive screening tool for detecting people who require diagnostic testing for TB disease, and can guide the effective allocation of molecular WHO-recommended rapid diagnostic tests (mWRDs) to improve case detection and cost-efficiency(1). CXR can also assist in the enhanced detection of low-grade and/or sub-clinical TB, thereby facilitating early detection and treatment initiation, and reducing the likelihood of onward transmission(2-4). However, factors

¹ DICOM – Digital imaging and communications in medicine – is a worldwide standard for the storage and transmission of medical images.

such as high variability in human interpretation of CXR, suboptimal specificity and limited availability of appropriately trained personnel to interpret CXR images may limit the potential utility of CXR in some settings.

CAD software products are increasingly used as a tool to enhance the feasibility and accuracy of CXR interpretation. CAD uses artificial intelligence (AI) to analyse digital CXR images for abnormalities and provides a quantitative scoring for TB-related abnormalities. Such technology can be used to replace or augment human expert interpretation of plain CXR when screening for pulmonary TB and can address inter-reader variability and reduce delays in reading radiographs when skilled personnel are scarce. Some CAD products express their reading of an abnormality as a numerical score (either between 0 - 100 or 0 - 1). A score above a determined threshold can be used to trigger a referral for further diagnostic evaluation for TB.

1.2 Consideration of the use and approval of CAD by WHO

CAD technologies are a new advancement in the field of TB detection and care and were first reviewed by WHO in 2016. Due to limited evidence available, a recommendation on their use was not issued at that time. In 2020 the WHO's Global TB Programme held a Guideline Development Meeting to update the guidelines for TB screening which included a series of questions focused on the diagnostic accuracy and performance of screening tools. As part of this meeting, the Guideline Development Group (GDG) evaluated the performance of CAD software for automated reading of digital CXR for the detection of pulmonary TB disease, compared to human readers. The GDG considered the performance of CAD software separately in the screening and triage use cases, where screening refers to the systematic identification of people with possible TB disease in a predetermined target group, regardless of the presence of signs or symptoms of TB (often referred to as active case finding when conducted in the community setting), while triage refers to the process of deciding the diagnostic and care pathways for people based on their signs, symptoms, risk makers and/or test results within the health care setting. Individuals may be referred to TB triage testing through various pathways, including: self-presentation/reporting on the basis of symptoms suggestive of TB; via referral for further evaluation for TB based on provider suspicion of TB, or; outcomes of an initial TB screening tool (such as symptom screening or CXR) conducted as part of a systematic screening programme screening necessitating further investigation.

Following a review of the evidence, a new recommendation was included in the updated TB screening guidelines (5, 6), that supports the use of CAD for individuals aged 15 years and older

in place of human readers for interpretation of digital chest radiography in both screening and triage for TB disease.

1.3 Integration of CAD into screening or triage algorithms

CXR can be combined with other screening tests in various possible algorithms, leading to referral for confirmatory diagnostic testing to establish or rule out a diagnosis. Due to its high sensitivity, it is advisable to place CXR early in a screening or triage algorithm to maximize case detection, particularly among people with asymptomatic or subclinical TB disease. In addition, CAD can be integrated into TB screening or triage algorithms where human interpretation is not available or to be used alongside trained readers to reduce workload. The following are potential use cases:

- CAD can be used as an initial screen, with any abnormal result by CAD (i.e., an abnormality score above the determined threshold) referred to a human reader for final interpretation, before being referred for diagnostic evaluation (confirmatory testing with an mWRD);
- CAD can be used as an initial screen and all individuals with an abnormality score above the determined threshold are referred for diagnostic evaluation, with a portion of all CXR images reread by a human reader for verification (e.g., all images with abnormal CAD reading and 10% of images with normal CAD reading);
- CAD can be used to entirely replace a human reader, with all abnormal results by CAD referred for diagnostic evaluation;
- CAD and human reading can be done in parallel, with an abnormal reading from either reading referred for diagnostic evaluation.

1.4 Calibrating CAD

Many CAD software products do not come with a pre-set manufacturer-recommended threshold that defines an “abnormal CXR” to be referred for diagnostic evaluation for TB. Some manufacturers do recommend thresholds for defining images as abnormal, however the evaluations reviewed by the GDG demonstrated substantial variation in the diagnostic accuracy (sensitivity and specificity) of CAD programmes across settings, even within the same technology set to the same threshold. The guidelines note that calibration of CAD thresholds will be essential for a given software in each setting and population in which it will be used, to ensure the accuracy, predictive values, overall yield, and requirements for further diagnostic testing are as expected.

Calibration of CAD requires an estimation of its diagnostic accuracy against a validated means of detecting TB that can act as a reference standard, such as bacteriological testing, to enable a user

to determine the most appropriate abnormality threshold score for their setting and use case, above which a confirmatory diagnostic test will be conducted. However, there is an inherent trade-off in the selection of the threshold score; a lower threshold score will maximize sensitivity of the tool to detect true TB cases among the population being screened but will incur additional costs related to unnecessary follow-on diagnostic testing due to reduced specificity. On the other hand, a higher threshold score will reduce the volume, and thus costs, of follow-on diagnostic testing and will likely identify more severe cases, but its reduced sensitivity will result in missed cases (Figure 1).

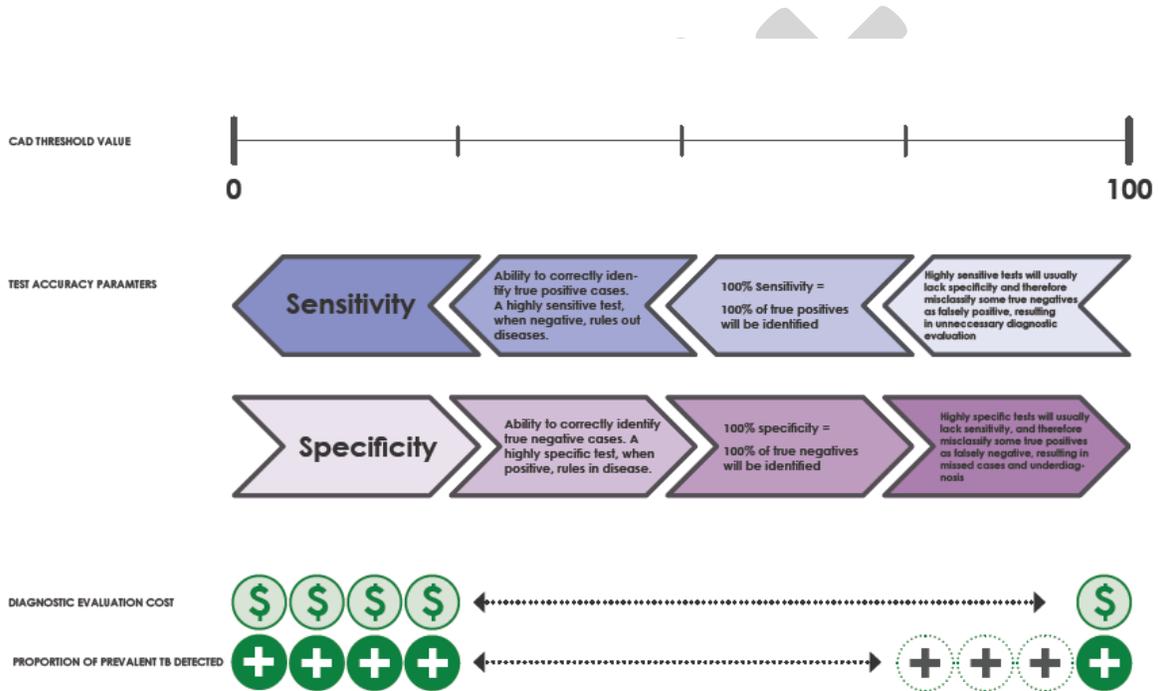


Figure 1: Sensitivity vs. specificity across CAD threshold spectrum

2. Implications of different settings for CAD use

Threshold scores for detecting TB will be influenced by the setting in which CAD is intended for use due to differences in population characteristics, such as:

- The severity of TB disease. For example, TB cases diagnosed in the community are more likely to be asymptomatic and/or subclinical – and therefore less severe – compared to cases diagnosed in a facility setting which are typically prompted by the presence of symptoms.

- The underlying TB prevalence. The prevalence of TB will typically be much lower in screening populations (<5%) than in triage populations (10-20%), which will impact a test's predictive values and the numbers of correctly and incorrectly diagnosed individuals.
- The presentation of TB in individuals with co-morbidities. For example, TB among people living with HIV is typically less cavitary in presentation compared to TB in individuals without HIV(7), and therefore, abnormalities due to TB are less evident than in non-HIV, increasing the risk of false negatives.
- The proportion of lung diseases other than TB and the prevalence of risk factors for TB in certain populations. For example, smoking may be more common in urban vs rural populations which can affect the texture of the lung.

3. CAD calibration studies

A CAD calibration study is an operational research study, designed to gather the requisite data to allow users to evaluate the diagnostic accuracy of CAD and calibrate the threshold for its implementation in a specific setting and population. CAD calibration studies are designed to be conducted in a relatively rapid and efficient way using data on the following points collected from all eligible individuals:

- Key demographic and clinical data
- Results from a bacteriological reference standard to confirm or rule out TB diagnosis for each individual, regardless of suspicion of TB and/or CAD abnormality score
- A CAD abnormality score based on the reading of a digital CXR (for all individuals, regardless of suspicion of TB and/or reference standard test results)

Before embarking on a CAD calibration study, users must first determine the specific settings and population in which CAD will be used. A calibration study may involve several study sites, provided the target population and settings are the same across all locations. Conversely, multiple calibration projects may be required in the same country or region if operational conditions and/or target populations differ significantly (e.g., triage in a hospital setting in a big city, health center screening for miners, community-wide screening in a rural area). Individuals who are recruited into the calibration study/studies must be representative of the population in which CAD will be implemented and used.

3.1 Proposed study designs and procedures

Required data for calibration studies may be prospectively collected or extracted from pre-existing data sources using either a *cross-sectional* or *case-control* study design.

Cross-sectional study

A cross-sectional study will prospectively and consecutively select all individuals meeting the eligibility criteria described in Table 1 in the setting where CAD is to be implemented. Note that individuals must be eligible for screening based on local TB case finding definitions/programme and according to the specific use case.

This design offers a relatively straight-forward approach to conducting calibration studies which can be integrated into new or routine practice and therefore run in parallel with ongoing community-screening or TB triaging activities.

Table 1: Inclusion criteria for cross-sectional calibration studies

Inclusion criteria	<ul style="list-style-type: none">• Aged 15 and over• Willing and able to provide consent• Willing and able to receive CXR and bacteriological testing• Eligible for community-based <u>or</u> triage screening based on local case finding definitions <p><i>For cross-sectional studies conducted in TB triage settings only:</i></p> <ul style="list-style-type: none">• Signs and symptoms suggestive of TB or provider suspicion or referral for TB evaluation
Exclusion criteria	<ul style="list-style-type: none">• Pregnant women• Medical conditions not compatible with performing CXR

Once recruited, each eligible participant will undergo:

1. Collection of key demographic and clinical patient information (see Annex for data collection template)
2. Digital CXR and reading with CAD product
3. Collection of 1-2 x sputum samples for testing using a bacteriological reference standard (defined as smear microscopy, culture or WRD, such as Xpert MTB/RIF which employ molecular or biomarker-based techniques for the diagnosis of TB)(8)

Box 1: Illustrative example of a cross-sectional calibration study within a TB screening setting

An NTP is running an active case finding intervention using mobile vans where everyone in an urban poor area is invited for CXR screening and participants with an abnormal CXR are offered follow-up testing by Xpert MTB/RIF. In order to gather data for the calibration study, in this scenario everyone targeted by the mobile van is invited to receive a CXR screen as well as the Xpert MTB/RIF for verification of TB status (regardless of CXR result)

Case-control study

In a case-control study design, individuals are selected separately and intentionally on the basis of their TB status (Table 2). A case-control study offers the possibility of using pre-existing patient data, provided it is of high quality, meets the requirements of the calibration study, and is representative of the population to be targeted for CAD use.

Table 2: Definitions of eligibility for case-control calibration studies

Cases	Individuals aged 15 years and above for whom a CXR and a bacteriological confirmatory diagnosis of pulmonary TB exists.
Controls	Individuals aged 15 years and above for whom a CXR was conducted and TB diagnosis ruled out based on the negative bacteriological test results.

Table 3 illustrates potential sources of pre-existing data that may be used in a case-control study based on use case. Available data must include key demographic and clinical patient information, digital CXR images for CAD reading, and results of bacteriological testing for all eligible individuals. Where multiple CXR and/or test results are available for an eligible individual, the earliest records on file - i.e., those taken at the point of initial examination or diagnosis – should be used.

Table 3: Potential sources of pre-existing data for calibration studies

	Typical characteristics of sample	Potential sources of pre-existing data
For CAD use in a TB triage setting	Population of more sick individuals seeking care for TB-related or other symptoms Symptomatic / more advanced TB presentation	Outpatient department (OPD) records TB clinic records

	High underlying level of TB prevalence (10 – 20%) and other lung conditions.	
For CAD use in a TB community-based screening setting	<p>A largely healthy population, usually not actively seeking care at the time of screening</p> <p>Sub-clinical TB (e.g., asymptomatic or subclinical)</p> <p>Low underlying level of TB prevalence (<5%)</p>	<p>Recent TB prevalence surveys</p> <p>Data from community screening-based research studies</p>

In many contexts, suitable pre-existing data may be available only for known, previously-diagnosed TB cases and not for non-TB cases. In these situations, a *combined retrospective/prospective* approach to participant selection may be used, allowing for pre-existing data on known TB cases to be used, while, at the same time, prospectively recruiting additional participants until the required number of TB and non-TB cases have been identified (sample size requirements are discussed in Section 3.6).

Box 2: Illustrative example of a case-control calibration study within a TB triage setting

An NTP has decided to run the calibration study within a major, urban TB clinic. Known, previously-diagnosed TB cases are retrospectively selected from the clinic records, from which CXR taken at the time of diagnosis are selected for CAD reading, and results of bacteriological outcomes and a basic set of patient data captured in the patient files at the point of diagnosis are extracted. In parallel, all eligible prospective cases attending the TB clinic are prospectively selected and invited for CXR and bacteriological testing, until the estimated number of TB and non-TB cases have been identified.

3.2 Selecting the most appropriate study design for calibration studies

Each study type is associated with different benefits, disadvantages and requirements. The following table outlines the key differences between the two study types and can help users determine which design may be most appropriate.

Table 4: Comparison of study designs

	Cross-sectional study	Case-control study
<i>General conditions for selecting study design</i>	<p>There is no pre-existing data source that can be used for either TB or non-TB cases.</p> <p>Existing data does not resemble the characteristics of the population intended for CAD use (e.g., patient data from TB clinics cannot be used in a calibration study intended for community-based screening).</p>	<p>Pre-existing, appropriate data which is representative of the target population is available for either confirmed TB cases only, or for both confirmed TB and non-TB cases.</p>
<i>Sample size (discussed further in section 3.6)</i>	<p>The minimum number of TB and non-TB cases required for analysis are the same for both study types, however, the total sample size required for each study differs as explained below:</p>	
	<p>In a cross-sectional study, individuals undergoing screening are prospectively sampled without knowledge of their TB status, which will be determined <i>after</i> enrolment.</p> <p>Sample size calculations will be driven by the expected level of precision around the estimation, the expected sensitivity of CAD product, as well as the underlying TB prevalence; i.e., depending on the prevalence of TB in the target population, a larger sample size will be required to identify the requisite number of TB cases required from a population that also includes those without TB.</p>	<p>In a case-control study, TB cases and non-TB cases are sampled separately and deliberately on the basis of their TB status, which is known <i>prior</i> to enrolment.</p> <p>Sample size calculations will be driven by the expected level of precision around the estimation and the expected sensitivity of CAD product.</p>
<i>Benefits</i>	<p>The calibration results will accurately represent how the CAD technology will perform in the population and setting in which it will be used, as the</p>	<p>Smaller total study sample size required due to ability to purposively sample confirmed TB and non-TB cases separately,</p>

	<p>study population will be drawn from the same setting in which CAD is intended for use.</p> <p>The study can provide useful descriptive information about the prevalence of TB and the presentation of lung abnormalities in the population.</p>	<p>requiring less time and resources to conduct the study.</p> <p>Smaller number of confirmatory tests required, due to smaller sample size and use of known-TB cases where confirmatory testing has already been carried out.</p> <p>Given lower requirements in terms of time and sample size, this design would allow for studies to be undertaken more quickly and easily in different regions/locations, or among particular subgroups etc.</p>
<i>Disadvantages</i>	<p>Greater enrolment size is required which increases time and resources needed to conduct the study.</p> <p>Requires a greater number of confirmatory tests to cover both newly detected cases, and non-TB cases.</p>	<p>Given the intentional sampling among TB and non-TB cases separately, it is very difficult to ensure that the study sample will be fully representative of the broader population in which CAD will be used and may result in selection and/or spectrum bias.</p> <p>As the prevalence of TB in the sample will be determined by the investigator, the study cannot provide any information about the prevalence of TB and the distribution of other lung conditions in the population in which CAD is to be used.</p>
<i>Requirements for conducting study design</i>	<p>Ability to conduct large number of confirmatory bacteriological tests in a timely manner.</p>	<p>Availability of digital CXR and bacteriological test results for known-TB cases or both known TB and non-TB cases.</p>

3.3 Bacteriological reference standard

Bacteriological reference standards may include smear microscopy, culture, or mWRD, such as Xpert MTB/RIF, which employ molecular or biomarker-based techniques for the diagnosis of TB(8). The use of mWRDs, such as Xpert MTB/RIF, is recommended as the reference standard. However, in situations where mWRDs have not been conducted on previously-diagnoses participants, and/or it is not feasible to conduct mWRDs for all prospectively identified individuals, other bacteriological tests (such as culture) may be considered.

The outcome of the reference standard tests (TB/not-TB) will be compared to the diagnostic performance of CAD at various thresholds (abnormal/normal) (Figure 2).

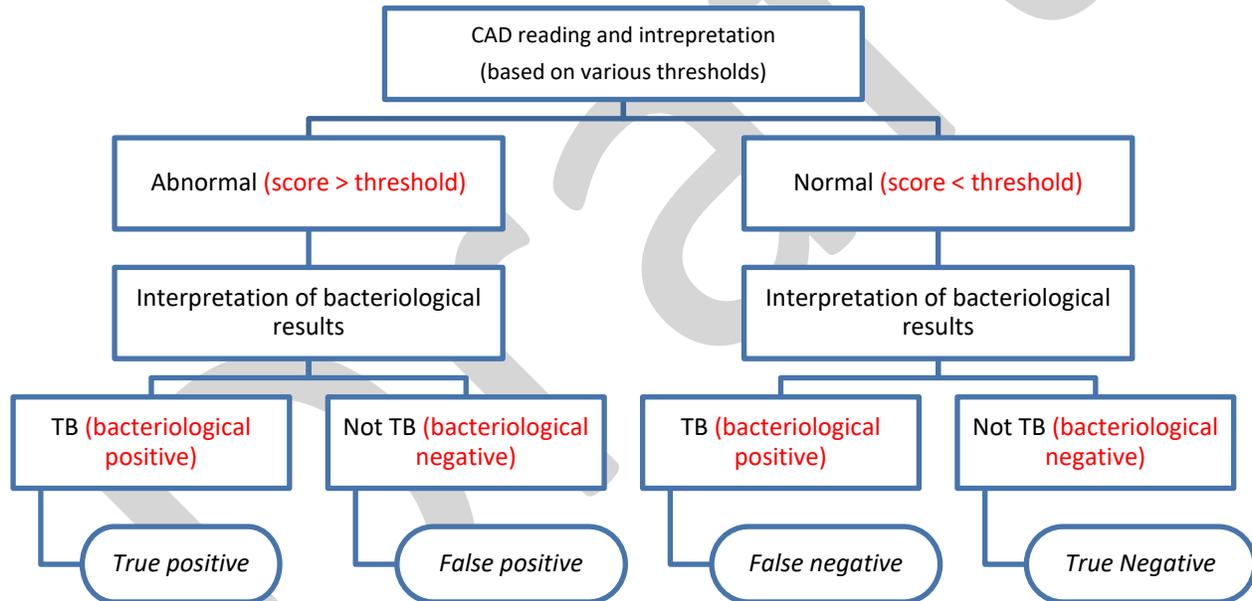


Figure 2: Comparison of CAD readings with bacteriological test results

For participants who are prospectively selected, 1 – 2 sputum samples should be collected for testing which is conducted as per regular local standard operating procedures. For retrospective case-control studies, outcomes of the chosen reference standard test conducted at the time of first consultation should be extracted from the selected data source.

3.4 Number of participants required for calibration studies

Table 5 presents the different and overall sample sizes required to estimate the CAD threshold, assuming a precision level of $\pm 5\%$, and a desired sensitivity of 50-90% for the two study designs.

Table 5: Sample size estimations by study design and level of sensitivity (based on $\pm 5\%$ precision)

Case-control design	Sensitivity				
	50%	60%	70%	80%	90%
Number of confirmed TB cases required (assuming TB prevalence of 100%)	384	369	323	246	138
Number of confirmed non-TB cases required (assuming the same precision and similar specificity as above)	384	369	323	246	138
Overall enrolment size required (assuming the same % and precision for Sp)	768	738	646	492	276
Cross-sectional design	Sensitivity				
	50%	60%	70%	80%	90%
Number of confirmed TB cases required (assuming TB prevalence of 100%)	384	369	323	246	138
Number of persons to screen for reaching the expected number of TB cases, using an example of TB prevalence of 2% or 2000 cases per 100,000 persons	384*100000/ 2000= 19,200	369*100000/ 2000= 18,450	323*100000/ 2000= 16,150	246*100000/ 2000= 12,300	138*100000/ 2000= 6,900
Overall enrolment size required (assuming the same % and precision for Sp)	19,200	18,450	16,150	12,300	6,900

Box 3: Illustrative example of sample size calculation for a combined case-control study

An NTP has decided to implement CAD to support active case finding efforts and plans to run the calibration study within a community-based screening activity using combined prospective/retrospective selection approach and aiming for a sensitivity of 70%. Based on Table 4, the total estimated sample size is 646, comprising 323 known TB cases, and an equivalent number of non-TB cases to determine the CAD threshold. The study team consecutively enrolls all eligible individuals who are identified during the screening activity, while retrospectively selecting known TB cases identified through a recent TB prevalence survey. Recruitment continues until 323 TB cases and non-TB cases respectively have been identified.

Box 4: Illustrative example of a sample size calculation for a cross-sectional study

An NTP wishes to calibrate CAD using a cross-sectional study. The estimated TB prevalence in this country is 300 cases per 100 000 persons. Aiming for a sensitivity of 50% (and $\pm 5\%$ precision) and based on table 4 they determine that 384 TB cases will need to be identified. Based on the estimated prevalence, an overall sample of around 128 000 persons will need to be screened in order to identify and get data on approximately 384 TB cases.

Table 6 presents the total number of individuals that need to be recruited at varying underlying TB prevalence rates to achieve the minimum number of TB cases at different assumptions for CAD sensitivity and a precision level of $\pm 5\%$. The number of TB cases required for each level of sensitivity is the same as in Table 5. As can be seen, TB prevalence and total enrolment size are inversely related; that is, the required sample size gets smaller as the TB prevalence increases. This is a reflection of the relative larger proportion of TB cases in settings where the underlying TB prevalence is higher. The estimates provided in Table 6 therefore reflect the overall enrolment sizes for studies relying only on prospective sampling to recruit participants.

Because the required sample size for prospective studies is dependent and inversely related to the prevalence of TB in the population being screened, this type of study design may not be appropriate for populations with lower TB prevalence (e.g. below 1%), as the number of people needed to be enrolled, and the associated costs for gathering the requisite clinical data and performing confirmatory diagnostic testing, may become prohibitive.

Table 6: Sample size estimations by TB prevalence and level of sensitivity (based on $\pm 5\%$ precision)

TB prevalence	Sensitivity				
	50%	60%	70%	80%	90%
0.5% (i.e. 500 /100,000)	76800	73800	64600	49200	27600
2% (i.e. 2,000 /100,000)	19200	18450	16150	12300	6900
5% (i.e. 5,000 /100,000)	7680	7380	6460	4920	2760
10% (i.e. 10,000 /100,000)	3840	3690	3230	2460	1380
15% (i.e. 15,000 /100,000)	2560	2460	2154	1640	920
20% (i.e. 20,000 /100,000)	1920	1845	1615	1230	690

Box 5: Illustrative example of a sample size calculation for a cross-sectional study

An NTP plans to use CAD for TB triaging and has decided to run the calibration study within a major, urban TB clinic. Prospective, consecutive selection will be used to identify and enroll all eligible presumptive cases attending the TB clinic. The study team decides to aim for 70% sensitivity and estimates the underlying prevalence is approximately 20%, given that study population will comprise individuals who have been referred for TB testing on the basis of signs and symptoms suggestive of TB. Using Table 5, the estimated minimum sample size needed for this study design is **1615** participants.

3.5 Sub-analysis to determine unique CAD thresholds for specific populations

Certain TB risk factors, comorbidities or other characteristics may influence the reading of CXR by CAD (9-13). These include:

- Age: CAD performance can be affected by age. There is some evidence (though not consistent across all technologies and settings) that CAD-interpreted CXR is less accurate for detecting TB disease in older age groups (over age 55). Further research and product development is required to fully understand the performance of CAD among children and as such, the exclusion of individuals aged 15 and under from calibration studies is currently recommended.
- HIV status: The specificity of CAD is lower among people living with HIV, compared to HIV negative individuals.
- Previous TB: There is some evidence that CAD-interpreted CXR is less accurate for detecting TB disease in individuals with a history of TB, likely due to scarring in the lungs.

Users may wish to conduct sub-analyses to determine specific thresholds for individuals represented by these categories. Requirements for sub-analyses includes a study sample which

is sufficiently powered and includes data on the risk factor or characteristic under review. The data collection template provided in Annex includes variables related to the categories described above and can be adapted as required and used to support sub-analyses.

Sample size guidance provided in Table 6 can be used to determine the required number of individuals with and without the condition of interest for both TB and non-TB cases. Depending on the characteristics of the primary sample, study teams may be required to purposively recruit or identify additional participants to achieve the required sample size to allow for stratification by a particular characteristic of interest.

Box 6: Illustrative example of a sample size calculation for sub-analyses

Country X has a high prevalence of TB/HIV co-infection and the NTP are interested in conducting a sub-analysis to determine the ideal threshold for this specific subgroup on top of the non-HIV group. The study team decides to aim for 70% sensitivity and will purposively identify known TB/HIV coinfecting participants who have been previously diagnosed and treated at the study site. Based on Table 5, the team therefore requires to collect data for 323 TB and non-TB cases with HIV, and 323 TB and non-TB cases without HIV

The study team review the original data set to determine how many participants fit in either of the four categories and conducts additional recruitment to reach the remaining sample size to ensure the sample is sufficiently powered for sub-analysis.

3.6 Study outcomes of interest

The key outcome of interest is the diagnostic performance of CAD compared to a bacteriological reference standard. Table 7 presents the various outcomes used to measure diagnostic performance.

The predictive probabilities of correctly indicating TB by CAD at various thresholds will also be used to estimate cost and to inform programmatic implications of the number of TB cases missed or misdiagnosed and additional costs of follow up testing incurred or saved.

The online [CAD for TB detection calibration](#) tool has been developed to support data analysis for CAD calibration studies by estimating the primary outcomes described below.

Table 7: Key outcome definitions

Outcome	Definition
Sensitivity (Se)	The proportion of cases correctly identified as positive (i.e., <i>true positives</i> , TP). Sensitivity is also referred to as the <i>true positive rate</i> (TPR).
# correctly diagnosed TB cases % total number of true TB cases	For tests that are highly sensitivity, the percentage of false negative (missed cases) are reduced to zero (i.e 100% sensitivity) but at the price of an increased number of false positive results (i.e., non-TB cases will be incorrectly classified as having TB)
Specificity (Sp)	The proportion of cases correctly classified as negative (i.e., <i>true negatives</i> , TN). Specificity is also known as the <i>true negative rate</i> (TNR)
# correctly diagnosed non-TB cases % total number of true non-TB cases	For tests that are highly specific, the percentage of false positive are reduced to zero (i.e., 100% specificity) but at the price of an increased number of false negative results (i.e., true TB cases are incorrectly diagnosed as not suffering from TB and will be missed)
False positive (FP)	The number of non-TB cases incorrectly classified as a TB case
False negative (FN)	The number of TB cases incorrectly classified as a non-TB case
True positive (TP)	The number of TB cases correctly classified as a TB case
True negative (TN)	The number of non-TB cases correctly classified as a non-TB case
Positive predictive value (PPV)	PPV represents the probability that a positive test result represents a true positive, calculated as the number of correctly diagnosed cases divided by the total number of positive tests. This parameter is related to the test's specificity and varies depending on the TB prevalence in the target population; the PPV is low when the prevalence of the disease is low and the specificity of the test is not high.
Negative predictive value (NPV)	NPV is used to describe the performance of a test and represents the probability that a negative test result represents a true negative, calculated as the number of correctly diagnosed non-TB cases divided by the total number of negative tests. This parameter varies depending on the TB prevalence in the target population; the NPV is high when the prevalence of the disease is low and the sensitivity of the test is not high.
CXR abnormality rate (%)	The proportion of all tested with abnormal test results
Total costs for diagnostic testing	Estimated costs of diagnostic evaluation for the entire population to be screened at a selected threshold
Cost per TB case detected	Estimated costs of diagnostic evaluation per true TB case detected at a selected threshold

3.7 Data collection and analysis

Required data

Table 8 provides an overview of the variables that should be included in data collection. Optional variables are noted by an asterisk; users may choose to include or exclude these based on the relevance to local contexts and intention for sub-analyses. The data collection template in Annex can be adapted based on variables of interest. A template csv file has been developed based on the variables presented in Table 8 that can be used or adapted to support data management and analysis and can be access via the [CAD for TB detection calibration tool](#) (under the ‘instructions’ tab) or via the [home page for the CAD calibration toolkit](#). This file will be uploaded to the [CAD for TB detection calibration tool](#) which has been developed to support data analysis for calibration studies (described further below).

Users may choose to digitise the data collection process by adapting the questions and fields presented in the data collection template into an electronic form using appropriate software such as ODK, REDCap or Microsoft Excel. Collection of data from individuals should be recorded against a unique participant ID to ensure the privacy and protection of personal data.

Note that the online tool allows for CAD scores with a range of 0 -100. Any CAD scores produced using a different range should be covered to meet this requirement. If a CAD product produces a score in the 0 – 1 range, these scores should be multiplied by 100 in the dataset before the final file is uploaded to the online tool.

Table 8: Variables for data collection

Variable name	Description	Data type	Codes/ranges
Patient	Unique number used to identify participants	Numeric	1 >
CAD_score	Numerical output score produced by a CAD reading	Numeric	0 – 100
TB_conf	Outcome of TB confirmation by bacteriological testing (or other reference standard)	Numeric	0 = non-TB case 1 = TB case
Age_gp*	Classification of patient aged >15 years and over based on age groups	Numeric	0= Aged between 15 and 55 years 1 = Aged 55 and above
Previous_tb*	Indicates previous diagnosis of TB	Numeric	0 = no 1 = yes

			2 = unknown/not disclosed
HIV *	Indicates HIV positivity in patient	Numeric	0 = no 1 = yes 2 = unknown/not disclosed
DM *	Indicates diabetes mellitus in patient	Numeric	0 = no 1 = yes 2 = unknown/not disclosed
Smoke *	Indicates smoking status of patient at time of contact	Numeric	0 = Never smoke (i.e., patient does not smoke and has no history of smoking) 1 = Current smoker (i.e., patient currently smokes) 2 = Past smoker (i.e., patient has history of smoking but does not currently smoke)

* optional indicator

In addition to the data collected for each participant, Users must enter the following setting-specific parameters into the tool prior to data analysis:

- The size of the population in which CAD is to be used;
- The TB prevalence in the population in which CAD is to be used (this can be estimated or calculated from the calibration study);
- The cost of performing diagnostic evaluation for TB among those who are referred for such evaluation from the CAD result (e.g., the cost of an Xpert MTB/RIF cartridge).

Data analysis

The [CAD for TB detection calibration tool](#) works by analysing data uploaded by the user (in excel or csv file) to calculate the accuracy and outcomes of a screening or triage project using CAD across a range of possible threshold values. A user guide for this tool is included in Part C.

The tool will estimate the primary outcomes of yield and cost at every possible CAD threshold, including: yields of TP, TN, FP, FN; sensitivity and specificity; negative and positive predictive values; proportion of prevalent TB cases diagnosed and missed; cost implications including total costs for diagnostic evaluation, and cost per true TB case detected.

Many of the outcomes of interest cannot be directly calculated using data collected in case-control-style studies. Instead, estimates of true- and false-positive screening results, yield of

prevalent TB detected, predictive values, and diagnostic costs can be calculated based on the estimated population size, TB prevalence, and costs provided by the user, for each CAD threshold.

Finally, the TP, TN, FP and FN will be used to construct a Receiver Operator Characteristic (ROC; Figure 3) curve, illustrating how sensitivity and specificity vary across the range of possible CAD thresholds. The ROC curve also displays the 95% confidence intervals of sensitivity and specificity at various thresholds, the area under the curve (AUC) and Youden’s Index (see box below) Together, these outputs allow users to determine the ideal threshold score for a given CAD implementation, based on the desired accuracy, yield, and cost implications.

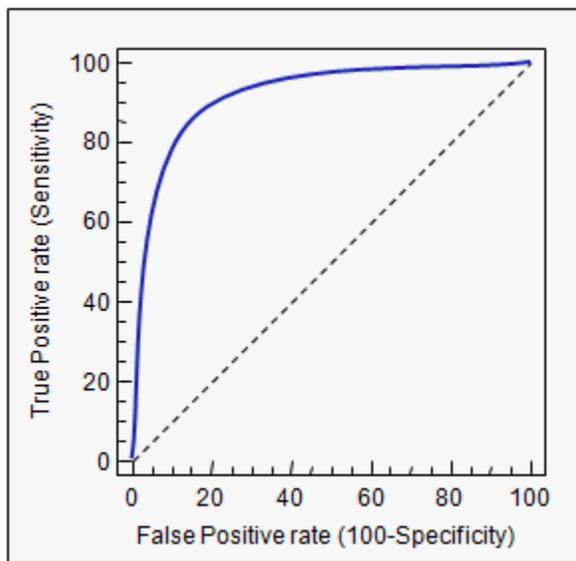


Figure 3: Illustrative receiver operator characteristic (ROC) curve

ROC curves plot the true positive rate against the false positive rate at various thresholds to show the possible trade-offs between sensitivity and specificity at various thresholds levels.

ROC curves can be used to calculate the area under the curve (AUC) which can aid in the comparison of the diagnostic performance of different CAD products under various use cases and/or settings.

Interpretation of ROC curves can be aided by Youden’s Index, a summary measure that integrates sensitivity and specificity into a single measure that ranges from 0 to 1. The highest score suggests a threshold where Se and Sp are balanced. However, this measure should be considered in light of its clinical relevance and specific programme priorities.

3.8 Study duration

For prospective studies, the length of time required to conduct this calibration study will depend chiefly on the study design, sample size target and daily throughput for bacteriological testing and/or CXR imaging. The following calculation can be used to estimate the length of time (in number of months) required per study site:

$$\frac{\left(\frac{\text{sample size}}{\text{number of study sites}} \right) / \text{daily throughput per site}}{\text{number of days of work per month}}$$

For example: a study aims to recruit 1600 participants from two study sites (which plan to enrol the same number of people per site), each of which can process approximately 10 Xpert MTB/RIF tests per day. Each site is operational for 20 days per month. Using the above calculation:

$$\frac{(1600 \div 2) / 10}{20} = 4$$

Using this example, the study would require approximately four months to complete.

For retrospective studies, study duration should be estimated based on the total sample size, number of available study personnel, the average number of individual patient files/data sets that can be reviewed per day, and the number of working days per month.

4. Countries with future plans to use CAD

Countries who are intending to use CAD but who do not yet have a functional CAD set up in place may choose to start preparing for future calibration studies by assembling an appropriate dataset that can be used for retrospective CAD analysis at a later date. This dataset should comprise digital CXR and bacteriological results for both TB and non-TB cases taken from participants who are representative of the intended population targeted for CAD use.

Box 7: Illustrative example of a country with future plans to use CAD

The NTP in Country X plans to begin using CAD in Q2 of 2022 within TB clinics as a screening tool. In preparation for the introduction of CAD, the NTP begins preparing a dataset that can be used for future calibration studies Q4 2021 and Q1 2022 (depending on the number of TB cases expected to be found in a certain period). This involves performing CXR and bacteriological testing on each new individual that attends for TB screening until the desired sample size is achieved. CAD readings will be performed on each CXR once CAD is available.

5. Calibration study costs

Table 9 presents key budget lines and potential expenses associated with a calibration study and may be used to develop a study budget. As this protocol assumes that a decision has already been made to use CAD, cost associated with the product (e.g., software, digital CXR etc.) have not been included.

Table 9: Suggested budget lines and estimated expenses for a CAD study

Item	Description
Study/data manager	Full/part time coordination role to provide oversight to study implementation and data collection and management, calculated at market monthly rate for full/part time role x number of months
Gene Xpert diagnostics	Assumed cost of USD 13 per test; costs differ based on study design selected and different sample size requirements. Total cost will depend on chosen study design and sample size. Note that costs association with samples transportation may also need to be considered depending on the study design
Ethics application	Fee for submission of protocol for ethical approval (approximately USD 1000)

Draft

PART B: GENERIC CAD CALIBRATION STUDY PROTOCOL



TB screening in Pakistan
Jacob Creswell, Stop TB Partnership

PART B: GENERIC CAD CALIBRATION STUDY PROTOCOL

Part B contains a generic protocol that can be adapted by users to support submission of calibration study proposals to local ethics committees. Users should follow the instructions (denoted by red text) included throughout to tailor the protocol as required.

Protocol summary

Study title	Determining the local calibration of computer-aided detection (CAD) software thresholds and other parameters for pulmonary TB detection in [add study location]
Investigators	[Add principal investigators and their affiliation]
Study rationale	<p>Chest radiography (CXR) plays a key role in the screening and triage of pulmonary tuberculosis (TB) and can guide the effective use of diagnostic testing to improve case detection and cost-efficiency. CAD products use artificial intelligence (AI) to analyse CXR for the presence of abnormalities suggestive of pulmonary TB and can improve the feasibility and performance of CXR for TB screening and triage. CAD technologies for TB detection have recently been recommended for use by WHO among adults aged 15 years or more, in place of human readers for interpretation of digital chest radiography in both screening and triage for TB disease.</p> <p>CAD products produce an abnormality score that can be used to signal probable TB and trigger further TB diagnostic evaluation relative to a selected threshold. Effective integration of CAD products into routine programming requires the calibration of CAD products based on the local context and intended use, as well as decision making around the goals for screening and acceptable costs. This study seeks to calibrate CAD to local settings by comparing the diagnostic accuracy of CAD compared to a reference standard.</p>
Study objective	<i>Delete as appropriate</i>

	To determine the most appropriate thresholds for a chosen CAD product calibrated to local context and use case, in order to support enhanced TB triage screening / community-based active case finding.
Study design	<i>Delete as appropriate based on selected design</i> Prospective, cross-sectional study. OR Retrospective case-control study. OR Combined retrospective/prospective case-control study.
TB diagnosis reference standard	[describe selected bacteriological reference standard]
Research outcomes of interest	Sensitivity, Specificity, false positive rate, false negative rate, predictive positive value, predictive negative value, cost at various threshold scores.
Settings	[Describe the setting(s) where the calibration study will be conducted]
Study population	[Provide a general description of the intended participant and the setting in which they will be identified from]
Eligibility criteria	Retrospective- case control study <i>(For combined retrospective/prospective case-control study, the following text should be edited to include the prospective data collection; the text under the below can be adapted for this purpose)</i> Cases will be defined as individuals aged 15 years and above for whom a CXR and a bacteriological confirmatory diagnosis of pulmonary TB exists. Controls will be defined as individuals aged 15 years and above for whom a CXR was conducted and TB diagnosis ruled out based on the negative bacteriological test results.

<p>Sample size estimates</p>	<p><i>Prospective cross-sectional study (delete text as appropriate based on intended use case for CAD)</i></p> <p><i>Community-based TB screening</i></p> <p>Individuals aged 15 and over eligible for community-based TB screening based on local TB case finding definitions/programme, who have provided consent to participate in the study, and have received or are willing to receive CXR and bacteriological testing are eligible for inclusion. Pregnant women or participants with medical conditions that are not compatible with performing a CXR may be considered ineligible, based on local guidelines.</p> <p><i>TB triage screening setting</i></p> <p>Individuals aged 15 and over presenting to selected study site/s who are eligible for TB triage screening based on local guidelines and algorithms for clinical care (e.g., patients with signs and symptoms suggestive of TB), and who have provided consent to participate in the study and have received or are willing to receive CXR and bacteriological testing will be eligible. There is no restriction based on presence of comorbidity.</p> <p>Pregnant women or participants with medical conditions that are not compatible with performing a CXR may be considered ineligible, based on local guidelines.</p> <p><i>Delete sample size estimates as appropriate based on selected study design</i></p> <p><i>Prospective cross-sectional study</i></p> <p>An estimated sample size of <i>[enter number based on Table 5 and 6]</i> individuals will be required for this study, based on a precision of 5% and</p>
------------------------------	--

	<p>a sensitivity of [enter selected Se] % and an estimated TB prevalence of [enter estimated TB prevalence] in this population.</p> <p>Retrospective case-control study</p> <p>An estimated sample size of [enter number based on Table 5] individuals/ presumptive TB will be required for this study, comprising [n] participants with TB (cases) and [n] participants without TB (controls) based on a precision of 5% and a sensitivity [enter selected Se] %</p>
Participant selection and recruitment	<p><i>Delete as appropriate based on selected approach for participant selection and study design</i></p> <p>Cross-sectional study</p> <p>Eligible participants will be prospectively and consecutive selected among those attending the study site during the study period.</p> <p>Case-control study</p> <p>Eligible participants who have been previously confirmed as TB and non-TB cases will be retrospectively identified from [describe data source that will be used]</p> <p>Or</p> <p>A combination of retrospective and prospective sampling will be used to identify cases and controls for this study. Eligible participants with previously-confirmed TB will be retrospectively identified from [describe pre-existing data source that will be used], while consecutive, prospective selection will be undertaken among eligible participants attending the study site until the required number of non-TB cases have been identified.</p>
Duration of study	[add approximate duration of study based on guidance provided in Part A]

Anticipated study dates	From [month, year] to [month, year]
-------------------------	-------------------------------------

Determining the local calibration of computer-aided detection (CAD) software thresholds and other parameters for TB detection in [add study location]: a protocol

1. Background

Chest radiography (CXR) play a key role in the detection of pulmonary tuberculosis (TB). While current World Health Organization (WHO) guidelines recommend diagnosis on the basis of bacteriological or molecular findings, the integration of CXR into screening and triaging algorithms can assist with early detection of TB through screening, triage, and as part of a comprehensive clinical evaluation. CXR is a highly-sensitive screening tool for detecting people who require diagnostic testing for TB disease, and can guide the effective allocation of molecular WHO-recommended rapid diagnostic tests (mWRDs) to improve case detection and cost-efficiency (1). CXR can also assist in the enhanced detection of low-grade and/or sub-clinical TB, thereby facilitating early detection and treatment initiation, and reducing the likelihood of onward transmission (2, 4). However, factors such as high variability in human interpretation of CXR, suboptimal specificity and limited availability of appropriately trained personnel to interpret CXR images may limit the potential utility of CXR in some settings.

Computer-aided detection (CAD) software products are increasingly used as a tool to enhance the feasibility and accuracy of CXR interpretation. CAD use artificial intelligence (AI) to analyse CXR for abnormalities suggestive of pulmonary TB which produce an abnormality score that can be used to determine the need for follow-on diagnostic testing for TB relative to a selected threshold. Such technology can be used to replace or augment human expert interpretation of plain CXR when screening for pulmonary TB and can address inter-reader variability and reduce delays in reading radiographs when skilled personnel are scarce.

CAD technologies for TB detection were considered by the WHO in 2020 as part of an update to TB screening guidelines and recommendations (5). The performance of multiple CAD software products for automated reading of digital CXR for the detection of TB disease were evaluated and compared to human readers. Following a review of the evidence, a new recommendation was included in the updated TB screening guidelines that supports the use of CAD for adults over the age of 15 years in place of human readers for interpretation of digital chest radiography in both screening and triage for TB disease (5, 6).

Effective integration of CAD products into routine programming requires determining an appropriate CAD threshold that will be used to signal probable TB cases and trigger further TB diagnostic evaluation. Identifying the ideal threshold for each implementation of CAD requires calibration of CAD products based on the local context and intended use case, as well as decision making around the goals for screening and acceptable costs. There is an inherent trade-off in the selection of the threshold score; a lower threshold score will maximize sensitivity of the tool to detect true TB cases among the population being screened but will incur additional costs related to unnecessary follow-on diagnostic testing due to reduced specificity. On the other hand, a higher threshold score will reduce the volume, and thus costs, of follow-on diagnostic testing and will likely identify more severe cases, but its reduced sensitivity will result in missed cases

Furthermore, the performance of CAD in terms of sensitivity and specificity of a particular threshold score are likely to differ when used in different TB epidemiological settings, and/or for certain subgroups, such as people living with HIV and people in older age strata. Threshold scores also differ substantially across various CAD products, or even across updated versions of the same software. Determining appropriate threshold scores based on local realities is therefore an integral part of the set-up and ongoing use of CAD and requires TB programmes to consider the overall aim of CAD in TB screening and diagnostic algorithms.

[This section should be supplemented by an overview of the history of, or future plans for, CAD use in your setting. Include the timelines of the first use of CAD, the intended use case / settings in which it is/will be used, and the factors that led to the decision making around CAD use. This section should also highlight the anticipated ways in which CAD will be beneficial – e.g., to increase capacity of screening efforts, to overcome personnel shortages in radiographers etc.]

2. Study objectives

The primary objective of this study is to generate necessary data to calibrate CAD based on the most appropriate thresholds for use within [add use case / study context] by comparing the performance of CAD in detecting probable pulmonary TB against a bacteriological [edit this as necessary based on intended reference standard] reference standard in [add study location].

3. Study design

Delete setting and study design as appropriate

Prospective cross-sectional study

This study will adopt a prospective cross-sectional design conducted under standard programmatic settings among individuals presenting for TB screening / participating in community-based active case finding activities.

Case-control study

This study will adopt a retrospective / combined retrospective prospective case-control study conducted under standard programmatic settings among individuals presenting for TB screening / participating in community-based active case finding activities.

3.1 Study settings

Describe the settings where the study will take place. For community-based screening activities, include details on the population targeted for screening such as geographical location, risk factors, demographic profile (if known), and estimated number of people in the community. Also describe the nature of screening activity, such as frequency of screening activities, procedures for screening etc.

For TB triage settings, describe the characteristics of the selected setting/s, including location, typical case load, staffing etc.

This section should also elaborate on the current or planned integration of CAD within TB detection algorithms (refer to Section 1.3 in Part A).

3.2 Study population

Inclusion criteria should reflect the requirements for participation in either community-based TB screening, or TB triage, depending on the intended use case for CAD and reflect the participant selection methods for the chosen study type, noting that there may be specific and unique requirements for individuals who are retrospectively or prospectively identified. Note that, as

further research and product development is required to fully understand the performance of CAD among children, the exclusion of individuals aged 15 and under from calibration studies is currently recommended.

Retrospective case-control study

The following text applies to retrospective case-control studies. For combined retrospective/prospective case control studies, the following text must be adapted to reflect the prospective component of the study. The text provided for cross-sectional studies below can be added here and edited as necessary.

Cases will be defined as individuals aged 15 years and above for whom a CXR and a bacteriological confirmatory diagnosis of pulmonary TB exists.

Controls will be defined as individuals aged 15 years and above for whom a CXR was conducted and TB diagnosis ruled out based on the negative bacteriological test results.

Prospective cross-sectional study

Delete the following paragraph as appropriate based on CAD use.

Community-based TB screening

All participants aged 15 who are eligible for TB screening based on local TB case finding definitions/programme, who have provided consent to participate in the study, and have received or are willing to receive CXR and bacteriological testing are eligible for inclusion. Given the utility of CAD in detection of subclinical TB within community-based screening settings, there are no inclusion requirements around the presence of signs and/or symptoms suggestive of TB. There is no restriction based on presence of comorbidity. Pregnant women or participants with medical conditions that are not compatible with performing a CXR are excluded.

TB triage

All participants aged 15 and over presenting to selected study site/s who are eligible for TB triage screening based on local TB case finding definitions/programme priorities (namely, signs and symptoms symptom suggestive of TB), and who have provided consent to participate in the study and have received or are willing to receive CXR and bacteriological testing will be eligible. There

is no restriction based on presence of comorbidity. Pregnant women or participants with medical conditions that are not compatible with performing a CXR are excluded.

3.3 Participant selection and sample size

Delete paragraph as appropriate based on the selected study design. If study involves multiple study sites, ensure that the number of participants to be recruited at each site is specified. List any other assumptions or calculations that have been used to determine final sample size requirements.

Prospective cross-sectional study

Individuals meeting the eligibility criteria will be prospectively and consecutively selected during the study period. The estimated minimum sample size needed for this study design is [N], which is based on a precision level of 5%, a sensitivity of at least [add %] and assumes an underlying TB prevalence of [add %].

Retrospective case-control study

Cases and controls will be retrospectively identified using data from [describe pre-existing data source where cases and controls will be drawn from, for example, TB patient records or TB prevalence survey]. Based on a 5% level of precision and at least [add %] sensitivity, a total sample size of [N] will be required, comprising [n] TB cases and [n] controls (non-TB cases).

Combined prospective/retrospective case-control study

A combined prospective/retrospective approach will be used to identify eligible individuals into this study. Cases meeting the eligibility criteria will be retrospectively identified through [describe source of known TB cases, for example, TB patient records or TB prevalence survey]. In parallel, prospective sampling will be conducted until [n] non-TB cases have been identified to act as controls among eligible individuals. Any TB cases identified through prospective sampling will also be included in the sample.

In total, [N] participants will be required, comprising [n] TB cases and [n] non-TB cases, based on a 5% level of precision and at least [add %] sensitivity.

4. Outcomes of interest

The outcomes of interest are related to the diagnostic performance of CAD compared to a bacteriological reference standard. The predicative probabilities of TB detection by CAD at various thresholds will be used to estimate cost and programmatic implications, in particular the number of TB cases missed or misdiagnosed and additional costs of follow up testing incurred or saved.

Definitions of the key outcomes of interest are described in Table 10.

Table 10: Key outcome definitions

Outcome	Definition
Sensitivity (Se)	The proportion of TB cases correctly identified as positive (i.e., <i>true positives</i> , TP), calculated as the number of correctly diagnosed TB cases divided by the total number of true TB cases.
Specificity (Sp)	The proportion of non-TB cases correctly classified as negative (i.e., <i>true negatives</i> , TN), calculated as the number of correctly diagnosed non-TB cases divided by the total number of true non-TB cases.
False positive (FP)	The number of non-TB cases incorrectly classified as a TB case
False negative (FN)	The number of TB cases incorrectly classified as a non-TB case
True positive (TP)	The number of TB cases correctly classified as a TB case
True negative (TN)	The number of non-TB cases correctly classified as a non-TB case
Positive predictive value (PPV)	PPV represents the probability that a positive test result represents a true positive, calculated as the number of correctly diagnosed TB cases divided by the total number of TB positive diagnoses
Negative predictive value (NPV)	NPV is used to describe the performance of a test and represents the probability that a negative test result represents a true negative, calculated as the number of correctly diagnosed non-TB cases divided by the total number of TB negative diagnoses
CXR abnormality rate (%)	The proportion of all tested with abnormal test results
Total costs for diagnostic testing	Estimated costs of diagnostic evaluation for the population to be screened at a selected threshold
Cost per TB case detected	Estimated costs of diagnostic evaluation per true TB case detected at a selected threshold

5. Study procedures

Delete paragraph as appropriate based on the selected study design

Prospective cross-sectional study

Following confirmation of eligibility, each participant will be asked to provide a minimal set of demographic and clinical information relevant to the study. [describe the specific process for data collection; for example, through the use of data collectors using paper-based forms, self-completed electronic forms etc.]. The following variables will be collected: [list the specific variables that will be included in your data collection tool; refer to the template in Annex that can be adapted as required]. Each eligible participant will then undergo: i) clinical TB screening; ii) a digital, plain postero-anterior/ antero-posterior chest X-ray for CAD reading, and; iii) collection of [one/two] sputum samples provided at time of contact for [specify bacteriological reference standard] testing.

Additional information should be added to this section to reflect the specific procedures that will be undertaken – for example, this section should mention the study team personnel that will be involved in data collection, responsible for the sputum collection and testing, CAD readings etc.

Retrospective case-control study

Individuals with confirmed TB status (both TB and non-TB cases) will be retrospectively identified from [describe selected pre-existing data source]. Among individuals meeting the eligibility criteria, the following demographic and clinical variables will be extracted: [list the specific variables that will be included in your data collection tool; refer to the template in Annex that can be adapted as required]. Additionally, CXR films will be extracted for CAD reading and the results of [specify bacteriological reference standard] testing used to first determine TB status. Where multiple CXR and/or [specify bacteriological reference standard] test results exist, the earliest-dated ones will be used.

Additional information should be added to this section to reflect the specific procedures that will be undertaken – for example, this section should mention the study team personnel that will be involved in data extraction, CAD readings etc.

Combined prospective/retrospective case-control study

Individuals with confirmed, previously-diagnosed TB will be retrospectively identified from [describe selected pre-existing data source]. Among individuals meeting the eligibility criteria, the following demographic and clinical variables will be extracted: [list the specific variables that will be included in your data collection tool; refer to the template in Annex that can be adapted as required]. Additionally, CXR films will be extracted for CAD reading and the results of [specify bacteriological reference standard] testing used to first determine TB status. Where multiple CXR and/or [specify bacteriological reference standard] test results exist, the earliest-dated ones will be used.

In parallel, prospective recruitment will be undertaken among eligible individuals until the required number of confirmed non-TB cases have been identified. After confirmation of eligibility, each participant will be asked to provide a minimal set of demographic and clinical information relevant to the study. [describe the specific process for data collection; for example, through the use of data collectors using paper-based forms, self-completed electronic forms etc]. The following variables will be collected: [list the specific variables that will be included in your data collection tool; refer to the template in Annex that can be adapted as required]. Each eligible participant will then undergo: i) clinical TB screening; ii) a digital, plain postero-anterior/ antero-posterior chest X-ray for CAD reading, and; iii) collection of [one/two] sputum samples provided at time of contact for [specify bacteriological reference standard] testing. Any and all individuals diagnosed with TB during this process will be retained in the final sample as cases.

Additional information should be added to this section to reflect the specific procedures that will be undertaken – for example, this section should mention the study team personnel that will be involved in data extraction, CAD readings etc.

5.1 Bacteriological testing

If use of mWRDs as reference standard is not feasible, the following section should be edited to reflect choice of substitute bacteriological reference standard. See Section 3.3 in Part A for further information.

CAD readings will be compared against [specify the specific bacteriological reference standard] as the selected reference standard for estimating the diagnostic performance of CAD.

Describe any additional information relevant to the conduct of testing, such as local standard operating procedures for collection and testing.

6. Data collection and management

This section should describe the data collection and management systems that will be put in place to support the study. The following text is based on the use of a paper-based data collection form and CSV spreadsheet for data management. Users who wish to use a different system, such as an online data collection form, should edit the following text accordingly and describe the system in detail here. Table 10 should also be edited to reflect final variables that will be collected from participants.

This section should also describe quality control and assurance mechanisms, including training of staff involved in the study.

A paper-based data collection form has been developed to facilitate collection of participant demographic and clinical data. This data will be entered into a standardized CSV spreadsheet and uploaded to the online [CAD for TB detection calibration tool](#) developed by WHO to facilitate the analysis of data needed for CAD calibration studies. Table 10 describes the variables captured in the spreadsheet that will be used for data analysis.

Table 11: Data collection variables for CAD calibration study

Variable name	Description	Data type	Codes/ranges
Patient ID	Unique number used to identify participants	Numeric	1 >
CAD_score	Numerical output score produced by a CAD reading	Numeric	0 – 100
TB_conf	Outcome of TB confirmation by bacteriological testing	Numeric	0 = non-TB case 1 = TB case
Age_gp*	Classification of patient aged >15 years and over based on age groups	Numeric	0= Aged between 15 and 55 years 1 = Aged 55 and above
Previous_tb*	Indicates previous diagnosis of TB	Numeric	0 = no 1 = yes 2 = unknown/not disclosed
HIV *	Indicates HIV positivity in patient	Numeric	0 = no 1 = yes 2 = unknown/not disclosed
DM *	Indicates diabetes mellitus in patient	Numeric	0 = no 1 = yes 2 = unknown/not disclosed

Smoke *	Indicates smoking status of patient at time of contact	Numeric	0 = Never smoke (i.e., patient does not smoke and has no history of smoking) 1 = Current smoker (i.e., patient currently smokes) 2 = Past smoker (i.e., patient has history of smoking but does not currently smoke)
---------	--	---------	--

* optional indicator

7. Data analysis

Note that any plans for sub-analyses should also be described in this section.

Data analysis will be conducted using the [online CAD for TB detection calibration tool](#). The analysis will be based on a dichotomous classification of each participant as either a TB, or non-TB case, based on the outcomes of bacteriological testing, against a binary interpretation of CAD abnormality score (abnormal/normal) at various CAD thresholds (Figure 4) in order to assess the diagnostic performance of CAD.

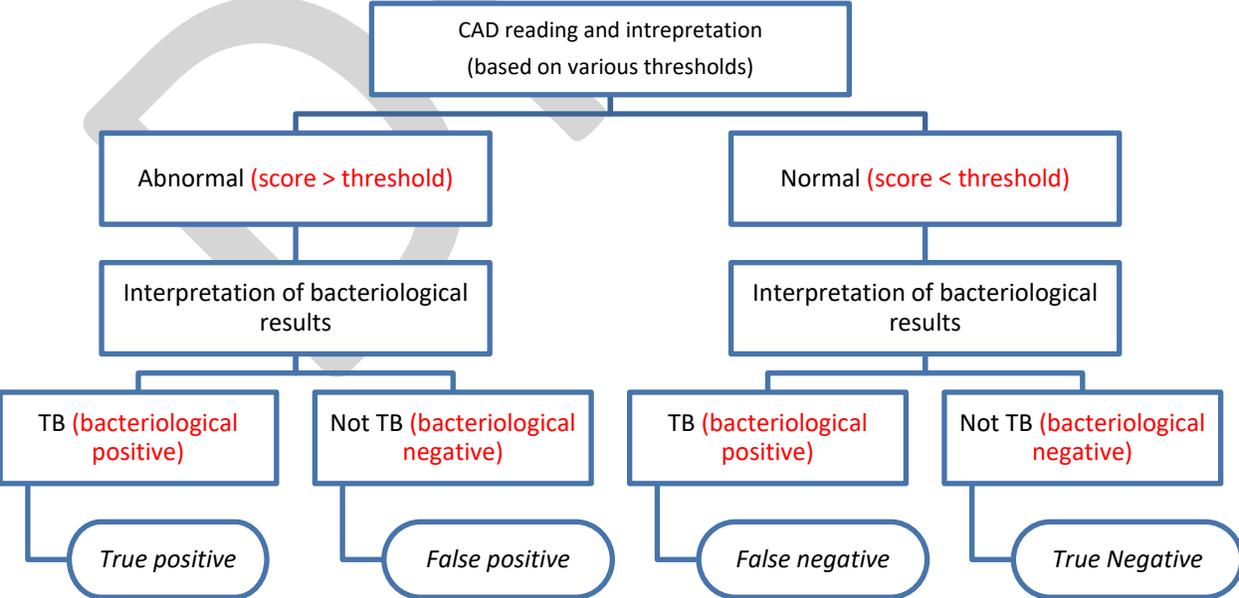


Figure 4: Comparison of CAD reading with bacteriological test results

Diagnostic performance will be estimated using the sensitivity (Se), specificity (Sp), and the number of false positives, false negatives, true positives and true negatives (FP, FN, TP and TN, respectively) of CAD performance at various thresholds compared to [specify bacteriological reference standard].

The following paragraph should be deleted for cross-sectional study protocols

Given the use of retrospectively-collected data, the positive predictive value (PPV) and negative predictive value (NPV) will be calculated based on the TB prevalence found in the study or a posteriori by applying the Se and Sp estimates for the CAD thresholds selected to the TB prevalence in the target population.

TP, TN, FP and FN will be used to construct a Receiver Operator Characteristic (ROC) curve, illustrating how sensitivity and specificity vary across the range of possible CAD thresholds. Based on the proportion of prevalent TB cases diagnosed and missed, the cost implications of various threshold levels in terms of total costs for diagnostic evaluation, and cost per true TB case detected, will be calculated, based on the average cost of confirmatory test.

Together, these outputs will be used to determine the ideal threshold score for a given CAD implementation, based on the desired accuracy, yield, and cost implications.

8. Ethical considerations

8.1 Ethics approval

This section should describe the process for local ethics approval and timelines for seeking clearance.

Ethics approval will be sought from [add name and details of relevant local ethical committee/ institutional review board] prior to commencing the study.

8.2 Protection of patient and confidentiality

All necessary steps will be taken to ensure the protection of participants and the confidentiality of their personal information. In addition to the requirement for informed patient consent (described further below), all necessary steps will be taken to protect and secure patient data. No personally-identifying information will be collected from participants during the study.

Instead, de-identified patient data will be recorded against unique patient identifiers, which will be linked back to a separate, master list. [provide further details here]. All data collection records (paper-based and/or electronic) will be securely protected by [describe security measures here, such as a password-protected file on a secure computer or in a locked cabinet on site] and will be accessible only to authorised people involved in the study.

8.3 Informed consent

Users should delete/edit the following text based on their local situation and selected study design

Prospective, cross-sectional study

Informed consent will be sought from all individuals willing to participate to the study and meeting the eligibility criteria during the study period prior to any investigation specific to the study is done. A plain language summary of the study (i.e., an easy-to-understand overview written in the local language) will be developed and provided to individuals during the first contact, to explain the procedures of the study and the proposed collection of use of their personal data. The patient information sheet will also cover use of anonymised CXR in any future CAD calibration or validation studies, in the case of CAD product updates or developments requiring future calibration studies. Individuals will be free to ask additional questions about the study and reassured that their care or ability to access services will not be impacted by a refusal to participate. Individuals will be informed of their right to revoke consent at any time, included after enrolment in the study. Consent will be recorded by writing, with consenting individuals providing a signature or thumb print (in the case of illiterate participants) on the provided consent form. For children aged between 15 and 18, consent must be granted from the parent or legal guardian but an assent from will be signed by them. Any child refusing to participate in the study cannot be forced to do so, even when parent or legal guardian consent has been granted.

Retrospective case-control study

We will seek a waiver of the need for consent from known, previously-diagnosed TB participants and TB suspects not diagnosed with TB who are retrospectively identified through [describe study site/data source], on account of the following factors: i) the disproportional efforts needed to obtain informed consent from participants who may no longer be under treatment at the study site (for TB triage settings) / the anonymous nature of data collected during TB prevalence

surveys (edit as required); ii) the limited risk to patient confidentiality due to the de-identified nature of data to be collected.

Combined prospective/retrospective case-control study

We will seek a waiver of the need for consent from known, previously-diagnosed TB participants who are retrospectively identified through [describe study site/data source], on account of the following factors: i) *the disproportional efforts needed to obtain informed consent from participants who may no longer be under treatment at the study site (for TB triage settings) / the anonymous nature of data collected during TB prevalence surveys (edit as required)*; ii) the limited risk to patient confidentiality due to the de-identified nature of data to be collected.

For prospective selection, informed consent will be sought from all individuals willing to participate to the study and meeting the eligibility criteria during the study period prior to any investigation specific to the study is done. A plain language summary of the study (i.e., an easy-to-understand overview written in the local language) will be developed and provided to individuals during the first contact, to explain the procedures of the study and the proposed collection of use of their personal data. The patient information sheet will also cover use of anonymised CXR in any future CAD calibration or validation studies, in the case of CAD product updates or developments requiring future calibration studies. Individuals will be free to ask additional questions about the study and reassured that their care or ability to access services will not be impacted by a refusal to participate. Individuals will be informed of their right to revoke consent at any time, included after enrolment in the study. Consent will be recorded by writing, with consenting individuals providing a signature or thumb print (in the case of illiterate participants) on the provided consent form. For children aged between 15 and 18, consent must be granted from the parent or legal guardian but an assent from will be signed by them. Any child refusing to participate in the study cannot be forced to do so, even when parent or legal guardian consent has been granted.

9. Study timelines

The expected duration of this study depends on study design, sample size and capacity to conduct CXR and bacteriological testing (for prospective studies only). For retrospective studies, study duration should be estimated on account of total sample size, number of available study personnel and average number of individual patient files/data sets that can be reviewed per day. See Section 3.8 in Part A for further guidance.

The estimated duration of this study is [enter estimated number of months], based on the following calculations:

$$\frac{\left(\frac{\text{sample size}}{\text{number of study sites}} \right) / \text{daily throughput per site}}{\text{number of days of work per month}}$$

Draft

PART C: CAD FOR TB DETECTION CALIBRATION TOOL USER GUIDE



TB screening in Pakistan
Jacob Creswell, Stop TB Partnership

Version April 2021

PART C: CAD FOR TB DETECTION CALIBRATION TOOL USER GUIDE

1. Introduction

An online tool has been developed to support the analysis and interpretation of data for CAD calibration. This tool has been designed to use data collected during the calibration study described in Part A and B to estimate sensitivity and specificity across a range of possible CAD thresholds, along with a ROC curve for the product. The tool also aims to support decision making around appropriate threshold levels by illustrating the potential implications (costs, potential missed cases, necessary/unnecessary confirmatory testing etc.) associated with various threshold choices.

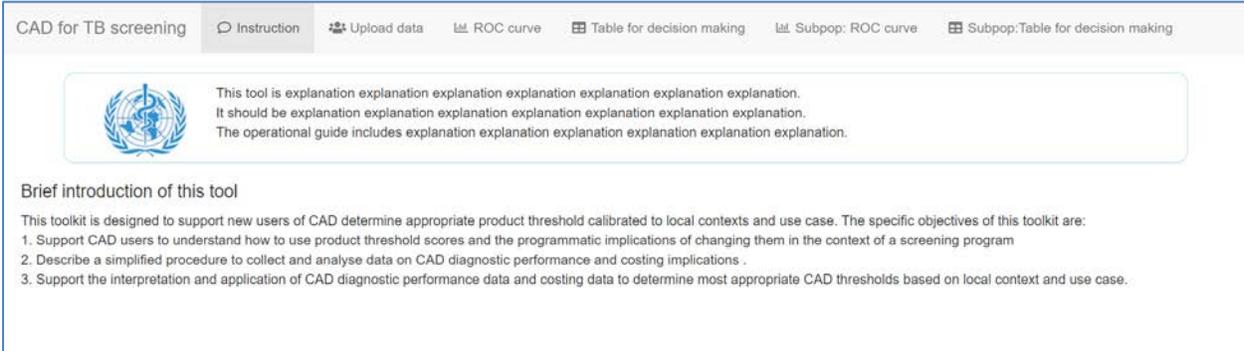
Part C of this toolkit aims to provide guidance on:

- How to use the online CAD for TB detection calibration tool, and;
- How to interpret the results in the context of programmatic decision making.

2. Overview of the CAD for TB detection calibration tool

The tool consists of the following tabs, as indicated by the horizontal tabs (Figure 5):

- Instructions
- Upload data
- ROC curve
- Table for decision making
- Subpop ROC curve
- Subpop: table for decision making



The screenshot shows the user interface of the 'CAD for TB screening' tool. At the top, there is a horizontal tab menu with the following tabs: 'Instruction', 'Upload data', 'ROC curve', 'Table for decision making', 'Subpop: ROC curve', and 'Subpop: Table for decision making'. The 'Instruction' tab is currently selected. Below the tabs, there is a content area with a blue header bar containing the WHO logo and the text: 'This tool is explanation explanation explanation explanation explanation explanation explanation. It should be explanation explanation explanation explanation explanation explanation explanation. The operational guide includes explanation explanation explanation explanation explanation explanation explanation.' Below this, there is a section titled 'Brief introduction of this tool' with the following text: 'This toolkit is designed to support new users of CAD determine appropriate product threshold calibrated to local contexts and use case. The specific objectives of this toolkit are: 1. Support CAD users to understand how to use product threshold scores and the programmatic implications of changing them in the context of a screening program 2. Describe a simplified procedure to collect and analyse data on CAD diagnostic performance and costing implications . 3. Support the interpretation and application of CAD diagnostic performance data and costing data to determine most appropriate CAD thresholds based on local context and use case.'

Figure 5: Landing page of online tool

3. Preparing the data set

Before using to tool, users must have collected and compiled the necessary data by carrying out the steps described in Part A.

Data must be prepared in a excel/CSV file that can be uploaded to the site. An excel-based template has been developed and is available for use by study teams. Table 7 in Part A outlines the recommended type and format of variables to be included in the data set.

The online tool allows for CAD scores with a range of 0 -100. Any CAD scores produced using a different range should be covered to meet this requirement. If a CAD product produces a score in the 0 – 1 range, these scores should be multiplied by 100 in the dataset before uploading the final file.

4. Uploading the data set

Once the dataset has been created, the relevant excel or CSV file can be uploaded by clicking *Browse* under the *choose CSV file* heading on the **Upload data** tab. This will open a local window where the user should navigate to and select the appropriate file (Figure 6).

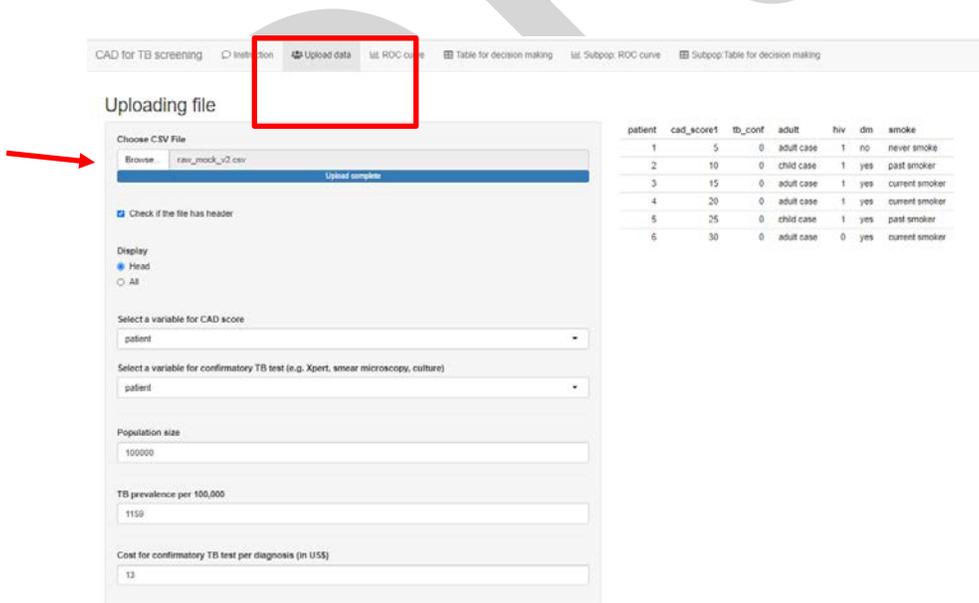


Figure 6: Uploading the dataset

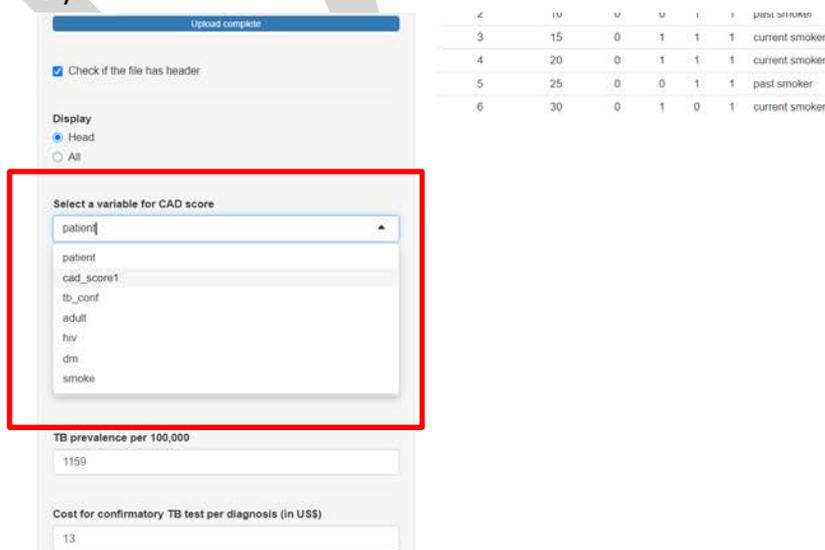
4.1 Setting the parameters

There are various options for customising the parameters of your dataset, which are listed in the grey box under *Uploading file*:

- **'Check if the file has header'**: This box should be selected if the dataset contains a heading row. This will import the variable headings present in the dataset.
- **Display**: Users can select how much of the dataset to be displayed: 'Head' will show just the first six lines, while 'all' displays the dataset in full
- **Population size**: Should be adjusted to reflect the population size of the country which the study is being conducted in.
- **TB prevalence per 100,000**: Should be adjusted to reflect the most recent overall national rate of prevalence per 100,000 people and reflect the estimated prevalence in the population for whom CAD is planned for use.
- **Cost for confirmatory TB test per diagnosis in USD**: Should be adjusted to reflect the average cost per test used as the reference standard (e.g., culture, Xpert etc.).

5. Selecting variables for ROC curve and outcomes table

Producing the ROC curve and decision-making tables requires a user to select the variable for CAD score and the variable for TB confirmation. If the excel spreadsheet template developed as part of this toolkit is being used, this will correspond to the CAD_score and TB_conf variables respectively.



The screenshot shows a software interface with several settings and a data table. The settings include:

- Upload complete** (status)
- Check if the file has header
- Display**
 - Head
 - All
- Select a variable for CAD score** (dropdown menu with options: patient, cad_score1, tb_conf, adult, hiv, dm, smoke)
- TB prevalence per 100,000** (input field with value 1159)
- Cost for confirmatory TB test per diagnosis (in US\$)** (input field with value 13)

To the right of the settings is a data table with the following content:

id	tu	u	u	1	1	prev smoker
3	15	0	1	1	1	current smoker
4	20	0	1	1	1	current smoker
5	25	0	0	1	1	past smoker
6	30	0	1	0	1	current smoker

Upload complete

Check if the file has header

Display

Head

All

Select a variable for CAD score

cad_score1

Select a variable for confirmatory TB test (e.g. Xpert, smear microscopy, culture)

patient

patient

cad_score1

tb_conf

adult

hiv

dms

smoke

Cost for confirmatory TB test per diagnosis (in US\$)

13

2	10	0	0	1	1	past smoker
3	15	0	1	1	1	current smoker
4	20	0	1	1	1	current smoker
5	25	0	0	1	1	past smoker
6	30	0	1	0	1	current smoker

Figure 7: Selecting variables for ROC curve and outcomes table

After these variables have been selected, the ROC curve and decision tables will be generated and displayed on the respective tabs.

The ROC curve displays the 95% confidence intervals of the specificity and sensitivity at specific thresholds, while the box under the curve presents the specific level and sensitivity and the value for Youden’s Index for each point selected along the curve (Figure 8)



Figure 8: ROC curve display

6. Selecting variables for sub - analyses

Users can generate specific outcomes for selected sub-groups (as described in Section 3.5 in Part A) by selecting the variable that represents the exposure of interest for the sub-group from the first drop-down box (Figure 9).

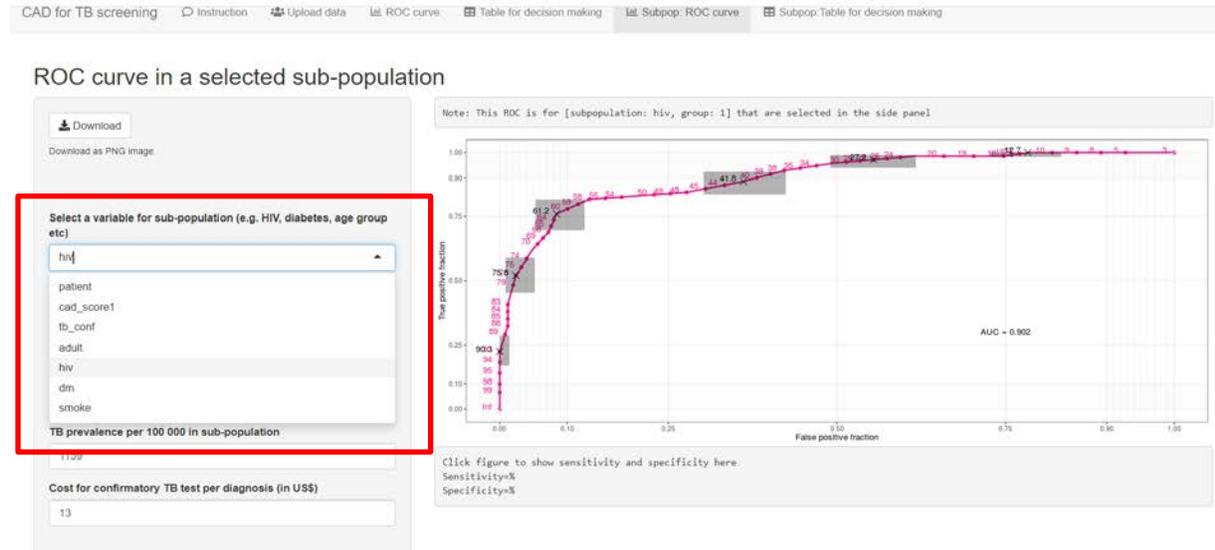


Figure 9: Selecting variables for sub-analysis

Users must then select the appropriate reference category that represents the exposure for interest. In the below example, 1 denotes individuals with HIV (Figure 10). The ROC curve for the sub analysis will be displayed on the current tab, while the tabular form of the key study outcomes is displayed on the next tab (“subpop: tables for decision making”).

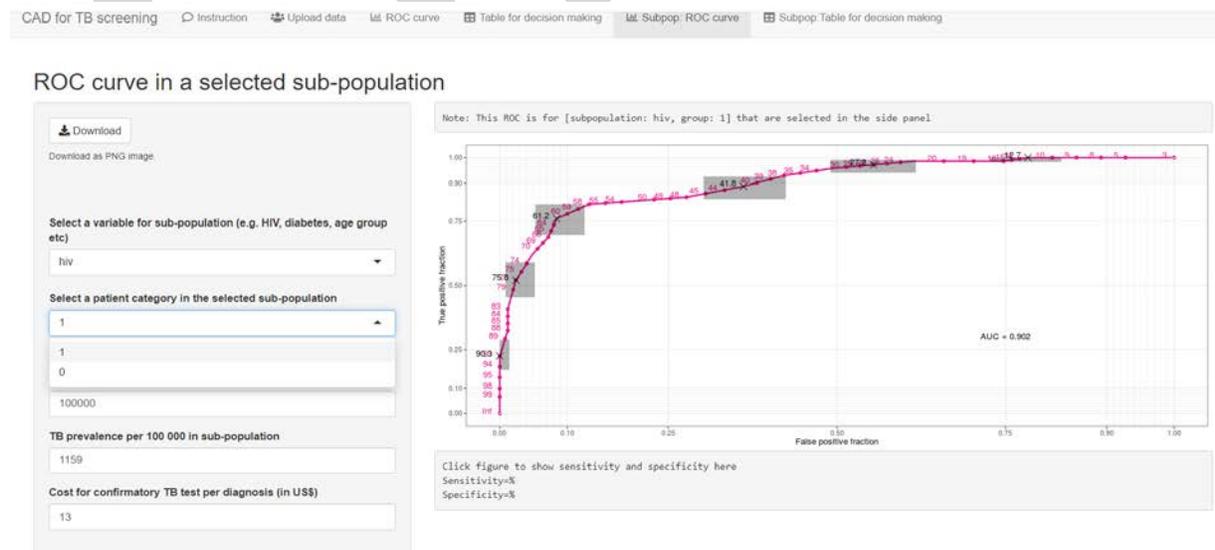


Figure 10: Selecting variables for sub-analysis

7. Interpreting calibration analysis

Once the results of a CAD calibration analysis are available, they should be carefully considered and interpreted in the context of the intended TB screening or triage programme. These considerations, along with other factors specific to the programme, will influence what CAD threshold is the most appropriate:

- In what population is CAD going to be used? What is the prevalence of TB and of other relevant comorbidities (including other lung conditions) in the population?
- Where will CAD be placed in the screening or triage algorithm?
- How will CAD be combined (or not) with human reading of CXR?
- What is the goal of CAD use? Is it to maximize the case detection of a screening programme, or to improve the efficiency of a triage or referral system?

If CAD is to be used as an early screening tool, followed by further screening or subsequent CXR reading by trained personnel, more emphasis may be placed on sensitivity. If the CAD reading is following earlier screening steps (e.g. screening for symptoms of TB) and is to be followed directly by diagnostic evaluation, efficiency of screening may take precedence.

Keeping these considerations in mind, careful attention should be paid to the estimates of accuracy, yield, and costs at different CAD threshold scores. Total costs of diagnostic evaluation will vary greatly with CAD threshold, and budgetary requirements or amount of available consumables for diagnostic evaluation may influence the selection of a CAD threshold.

Box 8: Case study (based on a fictional example)

Country X has decided to use CAD and wish to implement it into prison settings to enhance TB screening. The team responsible for calibrating CAD plan to conduct a prospective cross-sectional calibration study, as there is not pre-existing, relevant data source that can be used for this purpose.

The team assume a TB prevalence in prison settings of approximately 3000 cases per 100,000. Based on Table 4, they estimate that a sample of that approximately 10,800 ($323 * 100,000 / 3000$) will be required, assuming a sensitivity of 70% and 5% precision.

After collecting the required data from the recruited participants, the final data set is uploaded to the online tool and considers the results.

For the highest level of sensitivity (and lowest level of specificity), only 14, or 0.05% of true TB cases will be missed. This level of sensitivity corresponds to a CAD threshold cut off of 10

(everyone with an abnormality score >10 will be referred for confirmatory testing). Assuming an average cost of USD \$20 per confirmatory test, this threshold would cost \$1,833,333, or \$614 per case. The cost of unnecessary over diagnoses will be \$1,773,605 (equivalent to 88,680 false positives).

CAD triage for TB screening Instruction Upload data ROC curve **Table for decision making** Sub-population: ROC curve Sub-population:Table for decision making

Validity of CAD screening and estimated operational costs using CAD screening + Confirmatory TB test

[Download](#)
Download as a csv file (can be read by Excel).

Note: The estimation for TP/FN/FP/TN/misssed cases/% of missed/Total cost for confirmatory test/cost per case/Extra cost due to over dx is based on population size: 100000, TB prevalence: 3000, and cost for TB confirmatory test: US\$20
*If you want to change the assumptions, please go back to a panel [Upload data], and change the parameters.

Show 10 entries Search:

cut-off	Population	Prevalence	Prevalent cases	Sens	Spec	CAD pos	PPV	TP	FN	FP	TN	Misssed cases	% of missed	Total cost for confirmatory test	Cost per case	Extra cost due to over dx
10	100,000	3,000	3,000	99.5%	15.9%	91.7%	53.2%	2,986	14	88,680	477	14	0.5%	\$1,833,333	\$614	\$1,773,605
13	100,000	3,000	3,000	99.3%	19%	90%	54.1%	2,980	20	87,020	569	20	0.7%	\$1,800,000	\$604	\$1,740,408
14	100,000	3,000	3,000	98.6%	21.6%	88.3%	54.7%	2,959	41	85,374	647	41	1.4%	\$1,766,667	\$597	\$1,707,483
15	100,000	3,000	3,000	98%	24.2%	86.7%	55.4%	2,939	61	83,728	725	61	2%	\$1,733,333	\$590	\$1,674,558
18	100,000	3,000	3,000	97.3%	26.8%	85%	56.1%	2,918	82	82,082	804	82	2.7%	\$1,700,000	\$583	\$1,641,633
19	100,000	3,000	3,000	96.6%	29.4%	83.3%	56.8%	2,898	102	80,435	882	102	3.4%	\$1,666,667	\$575	\$1,608,707
20	100,000	3,000	3,000	95.9%	32%	81.7%	57.6%	2,878	122	78,789	961	122	4.1%	\$1,633,333	\$568	\$1,575,782

Conversely, a CAD threshold of 99 represents the lowest sensitivity score (and highest specificity score), which will cost \$62,585 (at a cost of \$327 per case), but this will miss 93% of true cases.

CAD triage for TB screening Instruction Upload data ROC curve **Table for decision making** Sub-population: ROC curve Sub-population:Table for decision making

Validity of CAD screening and estimated operational costs using CAD screening + Confirmatory TB test

[Download](#)
Download as a csv file (can be read by Excel).

Note: The estimation for TP/FN/FP/TN/misssed cases/% of missed/Total cost for confirmatory test/cost per case/Extra cost due to over dx is based on population size: 100000, TB prevalence: 3000, and cost for TB confirmatory test: US\$20
*If you want to change the assumptions, please go back to a panel [Upload data], and change the parameters.

Show 100 entries Search:

cut-off	Population	Prevalence	Prevalent cases	Sens	Spec	CAD pos	PPV	TP	FN	FP	TN	Misssed cases	% of missed	Total cost for confirmatory test	Cost per case	Extra cost due to over dx
100	100,000	3,000	3,000	3.4%	99.8%	1.7%	93.7%	102	2,898	1,565	2,993	2,898	96.6%	\$33,333	\$327	\$31,293
99	100,000	3,000	3,000	6.8%	99.8%	3.3%	96.8%	204	2,796	3,129	2,993	2,796	93.2%	\$66,667	\$327	\$62,585
98	100,000	3,000	3,000	10.2%	99.8%	5%	97.8%	306	2,694	4,694	2,993	2,694	89.8%	\$1e+05	\$327	\$93,878
95	100,000	3,000	3,000	13.6%	99.8%	6.7%	98.4%	408	2,592	6,259	2,993	2,592	86.4%	\$133,333	\$327	\$125,170

Based on these two extremes, the team must decide on what is acceptable in terms of number of missed cases, number of misdiagnoses and the cost implications of both, while acknowledging that i) the level of sensitivity and the associated costs do not move

proportionally to one another and; b) while the cost implications of over-diagnoses are clearly presented, the cost-implications of missed diagnoses are less concrete and this must be taken into consideration when determining an appropriate threshold.

Draft

ANNEX

DATA COLLECTION FORM FOR CAD CALIBRATION STUDY

CAD Pilot study data collection template	
Centre/location:	
Patient ID:	[__ __ __] year/study site/patient number

1. Patient information

Date of birth (DD/MM/YYYY; or age in years at last year if unknown): _____

Man	
Woman	

Has patient previously had TB?

Yes	
No	
Unknown	

Smoker?

Yes	
No	
Unknown	

HIV status:

HIV positive	
HIV negative	
Unknown	
Pending	

2. CAD reading

Product used: _____

Score: _____

Date of digital chest radiograph (DD/MM/YYYY):__ / __/ _____

3. Status of the patient regarding TB:

Suffering from TB	
Not suffering from TB	

4. Please specify the test done that confirm the current status of the patient regarding TB, the date and its result

Test	Test performed? (Y/N)	Date (dd/mm/yyyy)	Test outcome (positive, negative, indeterminate)
Xpert MTB/Rif			
Xpert Ultra			
Truenat			
TB-LAMP			
Culture			
Clinical diagnosis			
Other			

If "other" please specify: _____

REFERENCES

1. World Health Organization. Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches. World Health Organization; 2016. Report No.: 9241511508.
2. Frascella B, Richards AS, Sossen B, Emery JC, Odone A, Law I, et al. Subclinical tuberculosis disease—a review and analysis of prevalence surveys to inform definitions, burden, associations and screening methodology. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2020.
3. Law I, Floyd K, Group ATPS, Abukaraig EAB, Addo KK, Adetifa I, et al. National tuberculosis prevalence surveys in Africa, 2008–2016: an overview of results and lessons learned. *Tropical Medicine & International Health*. 2020;25(11):1308-27.
4. Onozaki I, Law I, Sismanidis C, Zignol M, Glaziou P, Floyd K. National tuberculosis prevalence surveys in Asia, 1990–2012: an overview of results and lessons learned. *Tropical Medicine & International Health*. 2015;20(9):1128-45.
5. World Health Organization. Consolidated guidelines on tuberculosis. Module 2: screening – systematic screening for tuberculosis disease. Geneva, Switzerland: World Health Organization; 2021.
6. World Health Organization. Operational handbook on tuberculosis. Module 2: screening – systematic screening for tuberculosis disease. Geneva, Switzerland; 2021.
7. Benito N, Moreno A, Miro J, Torres A. Pulmonary infections in HIV-infected patients: an update in the 21st century. *European Respiratory Journal*. 2012;39(3):730-45.
8. World Health Organization. Definitions and reporting framework for tuberculosis – 2013 revision (updated December 2014). Geneva.
9. Kik SV GS, Ruhwald M, et al. . Diagnostic accuracy of chest-X-ray reading with three artificial intelligence-based software when used as a triage test for pulmonary tuberculosis: an individual patient meta-analysis of a global chest-x-ray library. Unpublished.
10. Gelaw SM KS, Ruhwald M et al. Diagnostic accuracy of three computer-aided detection solutions for detecting pulmonary tuberculosis on chest radiography: Study on IOM Health Assessment TB Screening of Migrants. Unpublished.
11. Kik SV GS, Ruhwald M, et al. Diagnostic accuracy of chest-X-ray reading with three artificial intelligence-based software when used as a screening test for pulmonary tuberculosis: an individual patient meta-analysis of a global chest-x-ray library. Unpublished.
12. Zhi Zhen Qin SA, Mohammad Shahnewaz Sarker, Kishor Paul, Ahammad Shafiq Sikder Adel, Tasneem Naheyan, et al. . Can artificial intelligence (AI) be used to accurately detect tuberculosis (TB) from chest X-rays? An evaluation of five AI products for TB screening and triaging in a high TB burden setting. (Unpublished).
13. Zhi Zhen Qin TM, Rachael Barrett, Sizulu Moyo, Khangelani Zuma, Samuel Manda, et al. . Head-To-Head Evaluation of Computer Automated Reading Software to Detect Tuberculosis-Related Abnormalities from Chest X-Ray Images using South Africa Prevalence Survey Results. (Unpublished).