Open Forum Infectious Diseases

MAJOR ARTICLE

Diagnostic Accuracy of Computer-Aided Detection During Active Case Finding for Pulmonary Tuberculosis in Africa: A Systematic Review and Meta-analysis

Alex J. Scott,[1,2] Tahlia Perumal,[1,2] Ameer Hohlfeld,[3] Suzette Oelofse,[1,2] Louié Kühn,[1,2] Jeremi Swanepoel,[1,2] Coralie Geric,[4] Faiz Ahmad Khan,[4] Aliasgar Esmail,[1,2] Eleanor Ochodo,[5,6] Mark Engel,[7,8,a] and Keertan Dheda[1,2,9,a]

[1]Division of Pulmonology, Department of Medicine, Centre for Lung Infection and Immunity, University of Cape Town Lung Institute, Cape Town, South Africa, [2]Centre for the Study of Antimicrobial Resistance, South African Medical Research Council and University of Cape Town, Cape Town, South Africa, [3]Health Systems Research Unit, South African Medical Research Council, Cape Town, South Africa, [4]McGill International TB Centre, McGill University, Montreal, Quebec, Canada, [5]Kenya Medical Research Institute, Nairobi, Kenya, [6]Division of Epidemiology and Biostatistics, Department of Global Health, Stellenbosch University, Cape Town, South Africa, [7]Department of Medicine, University of Cape Town, Cape Town, South Africa, [8]Cochrane South Africa, South African Medical Research Council, Cape Town, South Africa, and [9]Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK

Background. Computer-aided detection (CAD) may be a useful screening tool for tuberculosis (TB). However, there are limited data about its utility in active case finding (ACF) in a community-based setting, and particularly in an HIV-endemic setting where performance may be compromised.

Methods. We performed a systematic review and evaluated articles published between January 2012 and February 2023 that included CAD as a screening tool to detect pulmonary TB against a microbiological reference standard (sputum culture and/or nucleic acid amplification test [NAAT]). We collected and summarized data on study characteristics and diagnostic accuracy measures. Two reviewers independently extracted data and assessed methodological quality against Quality Assessment of Diagnostic Accuracy Studies–2 criteria. Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines were followed.

Results. Of 1748 articles reviewed, 5 met with the eligibility criteria and were included in this review. A meta-analysis revealed pooled sensitivity of 0.87 (95% CI, 0.78–0.96) and specificity of 0.74 (95% CI, 0.55–0.93), just below the World Health Organization (WHO)–recommended target product profile (TPP) for a screening test (sensitivity ≥0.90 and specificity ≥0.70). We found a high risk of bias and applicability concerns across all studies. Subgroup analyses, including the impact of HIV and previous TB, were not possible due to the nature of the reporting within the included studies.

Conclusions. This review provides evidence, specifically in the context of ACF, for CAD as a potentially useful and cost-effective screening tool for TB in a resource-poor HIV-endemic African setting. However, given methodological concerns, caution is required with regards to applicability and generalizability.

Keywords. active case finding; Africa; computer-aided detection; diagnostic accuracy; tuberculosis.

Tuberculosis (TB) is a major cause of global mortality and morbidity. Of the estimated 10.6 million people who developed TB in 2022, 3.1 million (almost 1 in 3 people) remained undiagnosed or undetected [1]. These "missed" individuals are a potential source of TB transmission, severely undermining TB prevention and care, and potentially hindering the World Health Organization (WHO) END TB strategy goals [2, 3]. This demonstrates the need for community-based active case finding (ACF; provider-initiated screening and testing to identify people at risk for TB disease in a predetermined target group). Active case finding, if implemented adequately, has the potential to change TB epidemiology through improved access to care and a reduction in community transmission [4, 5].

Detecting TB in the community, however, has been restricted by a lack of sensitive and user-friendly point-of-care (POC) screening tools that meet the WHO-recommended target product profile (TPP) for screening tests (sensitivity ≥0.90 and specificity ≥0.70) [6]. Chest x-ray (CXR) is one of the most sensitive tests for detecting active TB [7]. However, it is often hindered by scarcity of resources and trained personnel and suffers from low specificity, especially in endemic areas where the prevalence of previous TB disease is high [8]. A potential answer to these challenges is ultraportable chest radiography

Received 29 September 2023; editorial decision 27 December 2023; accepted 15 January 2024; published online 17 January 2024

[a]Joint senior authors.

Correspondence: Keertan Dheda, PhD, Division of Pulmonology, Department of Medicine, Centre for Lung Infection and Immunity, University of Cape Town Lung Institute, George Street, Cape Town 7700, South Africa (keertan.dheda@uct.ac.za); or Alex J. Scott, MBChB, Division of Pulmonology, Department of Medicine, Centre for Lung Infection and Immunity, University of Cape Town Lung Institute, George Street, Cape Town 7700, South Africa (alex.scott@uct.ac.za).

Open Forum Infectious Diseases®

https://doi.org/10.1093/ofid/ofae020

and artificial intelligence (AI)–based computer-aided detection (CAD) software. CAD interprets abnormalities on CXRs suggestive of TB and expresses results as abnormality scores (either 0–100 or 0–1), which are deemed positive or negative if the abnormality score is above or below a precalibrated threshold. Recent data suggest that CAD performs on par with human readers to identify potential TB on CXR, helping to focus limited resources on the relevant cases [9, 10]. In 2021, the WHO conditionally endorsed the use of chest radiography and CAD for pulmonary TB screening [11].

However, there are limited published data that have evaluated CAD as a screening tool to detect TB during active case finding in the community, despite this being the key population that harbors most of the undetected patients with TB. Thus, the majority of the literature has focused on passive case finding (PCF; patients self-reporting to health care facilities) [10, 12–15], including a recent individual patient data meta-analysis [16]. Two earlier systematic reviews reported on the utility of CAD to detect TB [12, 13]. The first, published almost a decade ago (using older generation CAD software), focused on passive case finding only, while the second did not report on pooled estimates of diagnostic accuracy. Furthermore, few studies have been conducted on the African continent, which includes more than half of the 30 high–TB burden countries and accounted for almost a quarter (23%) of all people who developed TB in 2022 [1]. Special considerations in this region include being severely under-resourced and having a high HIV burden (the latter leading to atypical presentation and radiological features and a higher proportion of asymptomatic and sputum-scarce TB) [1, 17]. To address this important gap in our knowledge, we sought to systematically review the utility of CAD as a screening tool to detect TB during active case finding in an African setting. This review is timely, given the exponential rise in AI use in chest radiographical diagnosis.

METHODS

The review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration number CRD42022364968). The Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines were followed [18, 19]. Supplementary Table 1 depicts the completed PRISMA checklist.

Eligibility Criteria

Studies

The search included randomized controlled trials and observational studies (prospective and retrospective cohort studies, case–control studies, and cross-sectional studies) that were conducted in Africa. Although case–control studies, especially those with healthy controls, are known to overestimate sensitivity and specificity [20], they were included due to expected paucity of data. Case reports and case series were excluded.

Participants

The review included participants aged ≥15 years who underwent chest radiography and CAD analysis during active case finding for TB in Africa. The age cutoff was based on the WHO recommendation that CAD may be used in place of human readers during TB screening for individuals aged ≥15 years [11].

Index Test

Studies were included if the index test was CAD software analysis for the detection of pulmonary TB in participants who underwent chest radiography.

Target Condition

In this review, microbiologically confirmed pulmonary TB was the target condition.

Reference Test

Only studies that used sputum culture and/or nucleic acid amplification test (NAAT) positivity to diagnose pulmonary TB were included.

Outcome Measures

Diagnostic test accuracy estimates were considered the primary outcome measures of this review, with studies required (i) to report sensitivity and specificity, (ii) to report 2-by-2 data (true positive [TP], false positive [FP], true negative [TN], and false negative [FN]), or (iii) to have reported sufficient information to derive these numbers, from which estimates of test accuracy could be computed.

Exclusion Criteria

Studies were excluded if (i) the above criteria were not met, (ii) CAD results were reported for diagnostic modalities other than chest radiography, (iii) the study did not use a microbiological reference standard to determine CAD accuracy, or (iv) the required diagnostic accuracy data could not be acquired or calculated from the text and/or appendices.

Information Sources and Search Strategy

We systematically searched electronic databases, including PubMed/MEDLINE, Scopus, and the Cochrane Central Register of Control Trials (CENTRAL). A combination of Medical Subject Heading (MeSH) terms and free text terms was searched. The search was restricted to studies published in English, and the search period was limited to papers published from January 2012 to February 2023. Supplementary Table 2 demonstrates the detailed search strategy. To retrieve additional articles not found during the initial search, a manual review of

citation indexes and reference lists of studies identified through the electronic search was undertaken. Furthermore, a gray literature search was conducted to include conference papers and theses (WorldCat.org and The Union conferences).

Study Records

Selection Process

Five independent reviewers (A.S., T.P., S.O., J.S., and L.K.) screened all titles and abstracts identified by the search strategy against the predefined inclusion and exclusion criteria. Rayyan software [21] was used for screening of titles and abstracts. After removal of duplicate studies, 2 reviewers (A.S. and T.P.) subsequently reviewed the full texts of studies identified during initial screening. Reviewers were blinded to each other's decisions. Conflicts were resolved by discussion and/or consultation with another reviewer (M.E.) until consensus was reached.

Data Extraction and Management

Data from each selected study were collected onto a data extraction form that was developed and piloted before review. Two reviewers (A.S. and T.P.) independently extracted the following data: study identifiers (authorship, year of publication, journal, etc.), study characteristics (study design, funding sources, study country/context/setting, study population/participants, sample size, recruitment methods, eligibility criteria, participant demographic and clinical characteristics, etc.), index test details (CAD software and version, CAD threshold, etc.), description of the reference standard, and diagnostic accuracy measures (sensitivity, specificity, TP, FP, TN, FN, and area under the receiver operating characteristic curve [AUC]).

Risk of Bias

Two reviewers (A.S. and T.P.) independently assessed the risk of bias for included articles using a modified Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) (Supplementary Table 3) [22]. Disparities were resolved through consensus or discussion with a third reviewer (M.E.). Included studies were assessed across the 4 domains (patient selection, index test, reference test, and flow and timing). Risk of bias is assessed for each domain, whereas applicability concerns are assessed only in the first 3 domains. Signaling questions, answered as "yes," "no," or "unclear," were presented to assist the reviewer in judging the risk of bias and potential concerns regarding applicability. Risk of bias was reported as "low risk of bias," "high risk of bias," or "unclear risk of bias." Concerns regarding applicability were rated as "low," "high," or "unclear." Review Manager (RevMan, version 5.4) [23] was used to generate a summary and graphical representation of risk of bias and applicability concerns.

Quality of Evidence

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) is a methodological framework of assessing the quality of evidence and providing health care recommendations [24]. However, given certain methodological challenges with applying the GRADE approach for diagnostic test accuracy reviews [25, 26], we instead provided a description of the assessment of quality of evidence covering the key domains of GRADE: precision of the study estimates, heterogeneity in study findings, risk of bias, and concerns about applicability.

Data Synthesis and Statistical Analysis

A PRISMA flow diagram was used to depict the number of articles included or excluded from this review [27]. Included studies were descriptively presented in a summary table including study design, participant characteristics, details of the index test and reference standard, and outcome measures.

Data synthesis and analysis included the use of 2-by-2 contingency tables. Counts were back-calculated, where required, for studies that reported sensitivities and specificities and where 2-by-2 tables were not available. Using these tables, diagnostic accuracy measures were evaluated, including sensitivity, specificity, forest plots with 95% CIs, summary receiver operating characteristic (SROC) curves, and AUCs. Performance was evaluated using the hierarchical summary ROC (HSROC) method, providing equivalent summary estimates for both sensitivity and specificity [28].

It has been reported that the source of heterogeneity in a systematic review of DTA studies included both within- and between-study variabilities, with traditional measurements of heterogeneity (eg, I^2 statistic, etc.) not commonly performed [28]; therefore, heterogeneity was presumed in this review. In addition to visual inspection of the forest plots of individual studies' sensitivities and specificities to initially assess for heterogeneity, we followed the Cochrane Handbook for Systematic Reviews of DTA studies to graphically depict observed heterogeneity using the SROC curve [28].

All statistical analyses were performed using RevMan (version 5.4) [23] and SAS software (version 9.4) [29].

RESULTS

As shown in Figure 1, we identified 1748 articles, from which 80 duplicates were removed and 1634 were excluded after title and abstract screening. Of the 34 articles that underwent full-text review, 5 were included in this review [30–34]. Table 2 provides a summary of findings.

Characteristics of Included Articles

Characteristics of included articles are presented in Table 1. The diagnostic accuracy of CAD was evaluated retrospectively in 3 studies [30, 32, 34], among which 2 included data from national prevalence surveys [30, 32]. The remaining 2 studies utilized a prospective study design [31, 33], of which 1 reported the data of participants recruited from both community

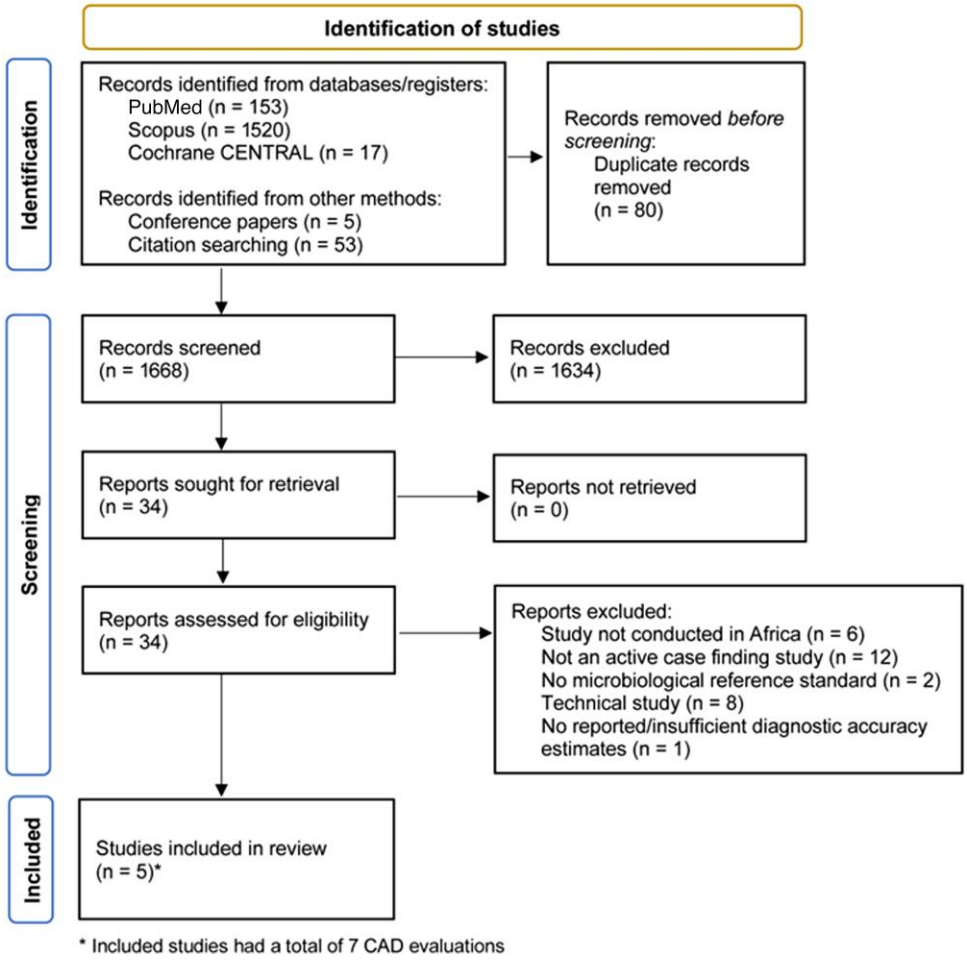

Figure 1. PRISMA flowchart of included and excluded studies. Abbreviations: CAD, computer-aided detection; PRISMA, Preferred Reporting Items for Systematic reviews and Meta-Analyses.

screening and health care facilities (ie, both active and passive case finding) [34]. CAD4TB software (Delft Imaging, 's-Hertogenbosch, the Netherlands) was used in all studies, with 1 study including an additional evaluation using qXR software (Qure.ai, Mumbai, India) [34]. Predefined thresholds based on prior evidence were used by studies using prospective study designs. Studies that utilized retrospective study designs reported using thresholds based on fixed sensitivities and specificities, or to obtain the same sensitivity as the human reader.

Diagnostic Accuracy of CAD

Four studies reported diagnostic accuracy measures of sensitivity, specificity, and AUC [30–32, 34]. Figure 2 depicts the forest plot of reported sensitivities and specificities of included studies. The remaining study reported only CAD positivity and TB yield, and accuracy estimates were back-calculated [33]. Overall, pooled estimates of sensitivity and specificity were 0.87 (95% CI, 0.78–0.96) and 0.74 (95% CI, 0.55–0.93), respectively (Figure 2). As 1 study included both active and passive case finding recruitment, sensitivity analysis was undertaken with the study removed from the overall analysis, resulting in a pooled sensitivity and specificity of 0.86 (95% CI, 0.72–1.00) and 0.81 (95% CI, 0.62–1.00), respectively (Supplementary Figure 1). Four studies reported AUC estimates, including 6 CAD evaluations, which ranged from 0.78 to 0.97 (Supplementary Table 4) [30–32, 34].

We evaluated the effects of the overall pooled diagnostic accuracy estimates per 1000 individuals screened at various prevalence settings (Table 2). For example, at a 1% TB prevalence, 9 (95% CI, 8–10) out of 10 individuals screened would be correctly identified as having TB, whereas 1 (95% CI, 0–2) out of 10 persons with TB would be potentially missed as false negatives. Furthermore, 733 (95% CI, 541–925) out of 990 individuals would be correctly identified as not having TB, while 257 (95% CI, 65–449) out of 990 would be incorrectly identified as having TB and sent for microbiological diagnostic testing. Additional estimates at a 5% and 10% TB prevalence are summarized in Table 2.

Subgroup Analysis

Pooled estimates of sensitivity and specificity were not carried out due to paucity of reported data. However, 3 studies

Table 1. Descriptive Characteristics of Included Studies

Study	Study Design	Population and Setting	Sample Size	Total People With HIV (%)	Total TB-Positive (%)	Index Test: CAD Software	Index Test: Threshold	Reference Standard	Outcome(s)
Fehr [31]	Prospective cohort	South Africa: community-based screening program	9914	2954 (29.8)	99 (1.0)	CAD4TB v5 and v6	25 (prespecified) Postanalysis: 39 (v5) and 47 (v6)[a]	NAAT and/or culture	SN, SP, PPV NPV, AUC, NNT
Kagujje [34]	Cross-sectional (retrospective analysis)	Zambia: community-based and health facility case finding study	1884	702 (37.3)	298 (15.8)	CAD4TB v7 qXR v3	15[b] 6[b]	NAAT	SN, SP, PPV NPV, AUC
Melendez [30]	Cross-sectional (retrospective analysis)	Zambia: Zambian national TB prevalence survey	23 838	Not reported	106 (0.4)	CAD4TB v5	>60[a]	Culture	SN, SP, PPV NPV, AUC
Mungai [32]	Cross-sectional (retrospective analysis)	Kenya: Kenyan national TB prevalence survey	61 848	1577 (5.0)	298 (0.5)	CAD4TB v6	61[b] 47[c]	NAAT and/or culture	SN, SP, AUC
Velen [33]	Prospective cohort	South Africa: targeted screening program of correctional facilities	3576	584 (16.3)	33 (0.9)	CAD4TB v6	50 (prespecified)	NAAT and/or culture	TB yield, NNT

Abbreviations: AUC, area under the curve; CAD, computer-aided detection; NAAT, nucleic acid amplification test; NPV, negative predictive value; NNT, number needed to test; PPV, positive predictive value; SN, sensitivity; SP, specificity; TB, tuberculosis.

[a]Thresholded to obtain same sensitivity as human reader.

[b]Thresholded to obtain fixed sensitivity of 90%.

[c]Thresholded to obtain fixed specificity of 70%.

presented various diagnostic accuracy measures of subgroup analyses, which are reported narratively below.

One study performed subgroup analysis to evaluate the performance of CAD in individuals with a history of previous TB [34]. Based on fixed thresholds (15 for CAD4TB v7 and 6 for qXR v3) that achieved a sensitivity of ≥0.90, specificity decreased in those with vs without a history of previous TB (CAD4TB v7: 0.24 [95% CI, 0.20–0.29] vs 0.60 [95% CI, 0.57–0.63], respectively, and qXR v3: 0.22 [95% CI, 0.18–0.27] vs 0.62 [95% CI, 0.59–0.65], respectively). The study also substratified HIV status with a history of TB and reported no differences in AUC between people with and without HIV, regardless of previous TB history. One study stratified accuracy measures based on HIV status [31], reporting improved AUC in individuals with vs without HIV (CAD4TB v5: 0.80 [95% CI, 0.72–0.87] vs 0.75 [95% CI, 0.68–0.83], respectively, and CAD4TB v6: 0.81 [95% CI, 0.74–0.88] vs 0.76 [95% CI, 0.68–0.84], respectively). Finally, 1 study assessed the differences in CAD accuracy when grouping participant characteristics (sex, age, cough duration, and history of previous TB), with the threshold fixed to 55 (CAD4TB v6) to achieve an overall sensitivity of 0.95 [32]. The authors reported the lowest specificity in male participants who were older (>41 years of age), had a cough for >2 weeks, and had a history of previous TB (0.38 [95% CI, 0.30–0.46]).

Quality Assessment

Figure 3 summarizes the QUADAS-2 assessment of included studies. In the domain of patient selection, all studies were judged to have high risk of bias due to inappropriate exclusions. This was based on sputum only being collected on symptomatic and/or participants with abnormal chest x-ray findings. Additionally, the single study that analyzed data using both active and passive case finding recruitment methods (reported ~20% proportion recruited through active case finding) was deemed to have high risk of bias [34]. The CAD threshold was only predefined in 2 studies [31, 33], with the 3 remaining studies reporting threshold scores postanalysis [30, 32, 34], and therefore determined to have high risk of bias. All studies had low risk of bias with regards to the reference standard. The flow and timing were deemed to have low risk of bias in all but 1 study, which did not perform microbiological reference testing on all participants [30].

DISCUSSION

This systematic review provides evidence for the utility of CAD in screening for TB during active case finding in Africa. Our pooled analysis indicated that CAD accuracy was just below the WHO-recommended target product profile (sensitivity ≥0.90 and specificity ≥0.70). Our results are complemented by a systematic review conducted by Harris and colleagues [13], who reported on articles published between 2005 and 2019, including a few in the TB screening context (ie, active case finding). In the screening studies reviewed, sensitivity ranged from 0.53 to 0.89 and specificity from 0.56 to 0.98. However, importantly, no meta-analyses or pooled estimates

Table 2. Summary of Findings

Question	What Is the Diagnostic Accuracy of CAD to Detect TB During Community-Based ACF in Africa
Population	Participants aged 15 y or older who underwent chest radiography during active case finding for TB in Africa
Index test	CAD
Comparator test	None
Target condition	Pulmonary tuberculosis
Reference test	Sputum NAAT (eg, Xpert) and/or sputum culture
Role	If accurate, CAD may be screening tool to detect TB during active case finding for TB
Limitations	
Risk of bias	High risk of bias
	Patient selection: All 5 studies reported inappropriate exclusions (eg, sputum collected on symptomatic participants and/or participants with abnormal chest x-ray findings). One study included participants recruited via active and passive case finding Index test: Three studies conducted retrospective analyses with no prespecified threshold (ie, derived from each study-specific analysis). Flow and timing: One study did not perform microbiological testing on all participants
Applicability	Concern about applicability
	Patient selection: Due to the nature of active case finding strategies (eg, prevalence surveys, etc.), where cost and resource demand are high, sputum-sampling all participants is often difficult. However, participants with possible TB who did not undergo reference testing (ie, asymptomatic and/or participants with normal-appearing chest x-rays) may have been missed, ultimately affecting the true diagnostic accuracy of the index test Index test: Retrospective analysis of the index test may not truly reflect diagnostic accuracy
Findings	

Quantity of evidence	5 studies (7 CAD evaluations)	Total participants	101 060	Total participants with valid CXR and sputum results	43 270	Total participants with target condition	834

				Effect per 1000 individuals tested at different prevalence settings		
Accuracy		Test consequences		1%	5%	10%
Pooled sensitivity	0.87 (0.78–0.96)	True positives	Correctly identified as having TB	9 (8–10)	44 (39–48)	87 (78–96)
		False negatives	Incorrectly identified as not having TB, therefore TB potentially missed	1 (0–2)	6 (2–11)	13 (4–22)
Pooled specificity	0.74 (0.55–0.93)	True negatives	Correctly identified as not having TB	733 (541–925)	703 (519–887)	666 (492–841)
		False positives	Incorrectly identified as having TB, therefore incorrectly sent for diagnostic testing	257 (65–449)	247 (63–431)	234 (59–408)
Consistency			Heterogeneity present for estimates of sensitivity and specificity			

Abbreviations: ACF, active case finding; CAD, computer-aided detection; CXR, chest x-ray; NAAT, nucleic acid amplification test; TB, tuberculosis.

of accuracy were performed. Furthermore, observational studies outside of Africa evaluating CAD during active case finding have found the tool to be reasonably accurate (AUC >0.80), equaling or surpassing human reader performance [35, 36].

We found study accuracy measures to be inconsistent, with heterogeneity present for sensitivity and specificity estimates. Additionally, we identified high risk of bias as well as applicability concerns in all studies included in the review. Similar to this systematic review, high levels of heterogeneity for sensitivity and specificity estimates have been previously described [13], with high risk of bias in the domains of patient selection, index test, and applicability concerns being reported. These findings suggest that despite an increasing body of research evaluating CAD accuracy, there are ongoing methodological limitations causing uncertainty in true CAD accuracy and generalizability, especially in active case finding strategies.

Assessing the accuracy of a putative screening test requires the population and setting in which the test is used to be considered. Despite community-based active case finding being effective in reducing TB prevalence if delivered with sufficient intensity and coverage [4], operationalizing active case finding strategies is difficult due to substantial resource requirements. In resource-limited settings, as in many areas in Africa, this becomes even more challenging. Cost and resource demand are high, especially during large-scale prevalence surveys, and performing microbiological testing on all participants is often infeasible. Therefore, tools such as digital chest radiography and CAD offer a potential solution by being an accurate screening/triage test, while also being cost-effective [37, 38]. However, a proportion of participants who do not undergo microbiological diagnostic testing, for example, those who are asymptomatic or had a normal-appearing chest x-ray, may in fact have TB and be potentially missed. From a public health perspective, these missed participants further hamper efforts to curb TB transmission. From a TB screening or diagnostics perspective, and reflected in the findings of our review, this adds high risk of bias in studies using CXR and CAD, thereby requiring accuracy

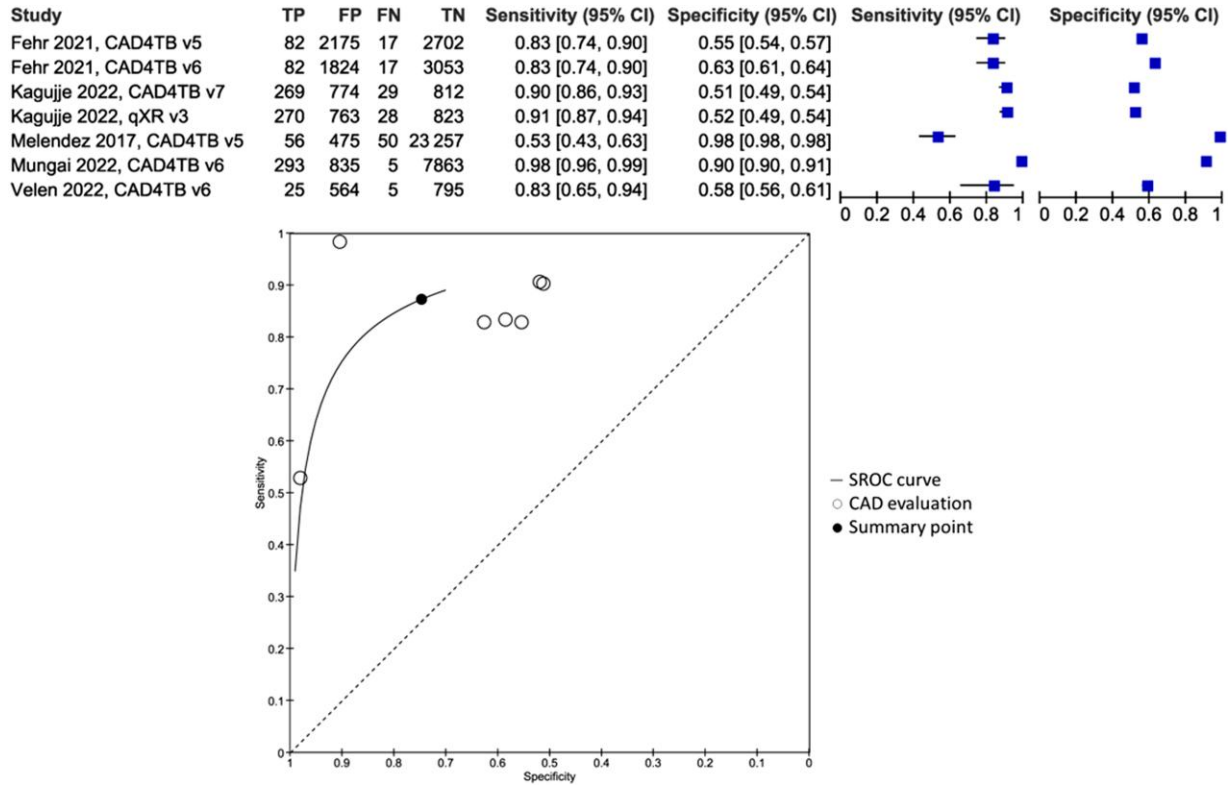

Figure 2. Accuracy measures for included studies (n = 7 CAD evaluations): Forest plot of reported sensitivities and specificities (top) and SROC curve (bottom) depicting pooled estimates of sensitivity (0.87 [0.78–0.96]) and specificity (0.74 [0.55–0.93]). Abbreviations: CAD, computer-aided detection; FN, false negative; FP, false positive; SROC, summary receiver operating characteristic; TN, true negative; TP, true positive; v, version.

estimates to be interpreted with caution. Another vital consideration when undertaking an evaluation of CAD is threshold determination, as a more sensitive threshold will result in increased cost due to a higher number of required confirmatory microbiological tests, whereas a less sensitive threshold will potentially miss individuals with TB. We advocate for careful consideration when determining the most appropriate threshold for various settings and populations.

To improve clinical applicability, organizations including the WHO, the Foundation for Innovative New Diagnostics (FIND), and the Stop TB Partnership have developed toolkits and practical guides to assist researchers in planning and implementing CAD in various settings and populations [39–41]. Researchers are encouraged to use these guides in their projects to mitigate methodological shortfalls and be able to report accurate and reliable data. In addition, we suggest that researchers and relevant stakeholders consider the technical, economic, public policy, and usability aspects of CAD before implementation, all of which are available and have been described in published reviews and reports [42–45].

Our review has several strengths. First, we only included studies that compared CAD with a microbiological reference standard, as opposed to human readers, for example, allowing more robust interpretation of accuracy estimates. Second, our review included a large number of participants from high-quality studies, including national prevalence surveys. Third, we reported pooled accuracy estimates, improving the accuracy of results and overcoming some of the shortcomings of previous systematic reviews. Finally, this review adds to the important body of research evaluating CAD diagnostic performance during active case finding for TB, especially on the African continent where there is a large TB/HIV burden. Additionally, it has been purported that more than half of people diagnosed with TB are subclinical, that is, asymptomatic persons with microbiologically proven TB [46]. These individuals may never have attended a health care facility and may be transmitting disease. We have reported CAD to be an accurate screening tool that may be used during large-scale screening with the goal to detect all individuals with TB.

There are some limitations of our review. We only included published, peer-reviewed studies in English, which may have introduced publication bias. Furthermore, we only included studies that recruited participants aged ≥15 years, and we can therefore not comment on the diagnostic accuracy of CAD in children and young adolescents, an important population that contributes ~5%–10% of the global TB burden. However, the WHO only recommends CAD as a screening or triage tool in individuals aged ≥15 years. Finally, we were unable to provide

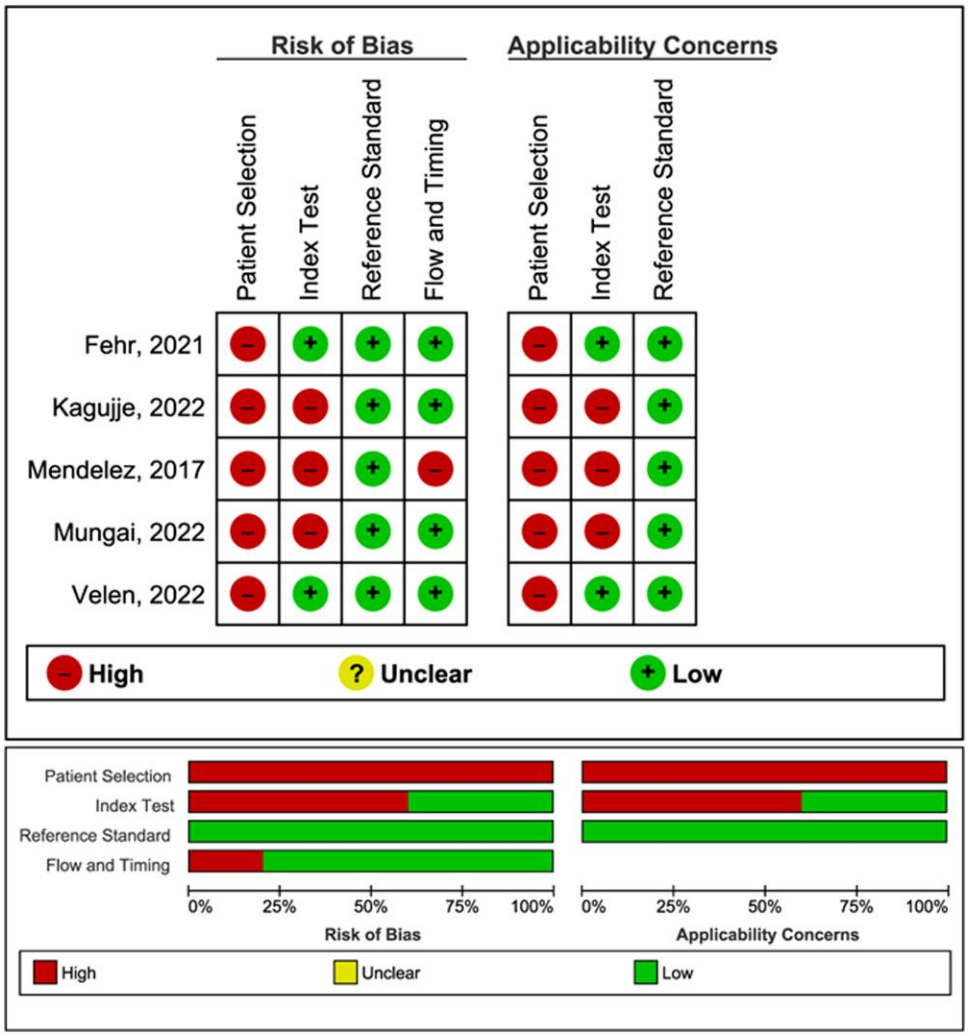

Figure 3. QUADAS-2 summary (top) and graph (bottom) of included studies. Abbreviation: QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies–2.

pooled accuracy estimates for subgroups, particularly people with HIV, due to paucity of reported data. It is imperative that future studies provide detailed analysis on subgroups, including people with HIV, asymptomatic persons, older participants, and individuals with a history of TB, to elucidate variations of CAD accuracy between populations and clinical characteristics.

CONCLUSIONS

In the context of active case finding, CAD has the potential to be a useful and cost-effective screening tool for TB in a resource-poor HIV-endemic African setting, assisting active case finding strategies to break the TB transmission cycle. However, given methodological limitations resulting in high risk of bias, caution is required with regards to applicability and generalizability. Further evidence should focus on an adaptive approach to using CAD, with prospective studies predetermining thresholds with prior calibration and analyzing results by stratifying according to population and individual clinical characteristics.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

A.J.S. is a Harry Crossley Senior Clinical Fellow and acknowledges the support of the Harry Crossley Foundation.

Author contributions. Conceptualization and study design: all authors. Database searching: A.J.S., A.H., M.E. Abstract screening: A.J.S., T.P., S.O., L.K., J.S. Full-text review and data extraction: A.J.S., T.P., M.E. Data synthesis and analysis: A.J.S., E.O., M.E. Drafting the manuscript: A.J.S., A.E., M.E., K.D. Manuscript review and editing: all authors.

Patient consent. This study does not include factors necessitating patient consent.

Financial support. The work reported herein was made possible through funding by the South African Medical Research Council through its Division of Research Capacity Development under the SAMRC Internship Scholarship Programme. The content hereof is the sole

responsibility of the authors and does not necessarily represent the official views of the SAMRC.

Potential conflicts of interest. All authors: no reported conflicts.

References

1. World Health Organization. Global Tuberculosis Report 2023. World Health Organization; **2023**.
2. World Health Organization. The END TB Strategy. World Health Organization; **2015**.
3. Dheda K, Barry CE III, Maartens G. Tuberculosis. Lancet **2016**; 387:1211–26.
4. Burke RM, Nliwasa M, Feasey HRA, et al. Community-based active case-finding interventions for tuberculosis: a systematic review. Lancet Public Health **2021**; 6: e283–99.
5. Ortiz-Brizuela E, Menzies D. Tuberculosis active case-finding: looking for cases in all the right places? Lancet Public Health **2021**; 6:e261–2.
6. World Health Organization. High-Priority Target Product Profiles for New Tuberculosis Diagnostics: Report of a Consensus Meeting. World Health Organization; **2014**.
7. World Health Organization. WHO Standard: Universal Access to Rapid Tuberculosis Diagnostics. World Health Organization; **2023**.
8. World Health Organization. Chest Radiography in Tuberculosis Detection: Summary of Current WHO Recommendations and Guidance on Programmatic Approaches. World Health Organization; **2016**.
9. Melendez J, Hogeweg L, Sanchez CI, et al. Accuracy of an automated system for tuberculosis detection on chest radiographs in high-risk screening. Int J Tuberc Lung Dis **2018**; 22:567–71.
10. Qin ZZ, Ahmed S, Sarker MS, et al. Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. Lancet Digit Health **2021**; 3:e543–54.
11. World Health Organization. WHO Consolidated Guidelines on Tuberculosis. Module 2: Screening—Systematic Screening for Tuberculosis Disease. World Health Organization; **2021**.
12. Pande T, Cohen C, Pai M, Ahmad Khan F. Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. Int J Tuberc Lung Dis **2016**; 20:1226–30.
13. Harris M, Qi A, Jeagal L, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. PLoS One **2019**; 14:e0221339.
14. Khan FA, Majidulla A, Tavaziva G, et al. Chest x-ray analysis with deep learning-based software as a triage test for pulmonary tuberculosis: a prospective study of diagnostic accuracy for culture-confirmed disease. Lancet Digit Health **2020**; 2: e573–81.
15. Murphy K, Habib SS, Zaidi SMA, et al. Computer aided detection of tuberculosis on chest radiographs: an evaluation of the CAD4TB v6 system. Sci Rep **2020**; 10: 5492.
16. Tavaziva G, Harris M, Abidi SK, et al. Chest x-ray analysis with deep learning-based software as a triage test for pulmonary tuberculosis: an individual patient data meta-analysis of diagnostic accuracy. Clin Infect Dis **2022**; 74:1390–400.
17. Esmail A, Randall P, Oelofse S, et al. Comparison of two diagnostic intervention packages for community-based active case finding for tuberculosis: an open-label randomized controlled trial. Nat Med **2023**; 29:1009–16.
18. McInnes MDF, Moher D, Thombs BD. Preferred Reporting Items for a Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies: the PRISMA-DTA statement. JAMA **2018**; 319:388–96.
19. Salameh JP, Bossuyt PM, McGrath TA, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis of Diagnostic Test Accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. BMJ **2020**; 370:m2632.
20. Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem **2005**; 51: 1335–41.
21. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev **2016**; 5:210.
22. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med **2011**; 155: 529–36.
23. Review Manager (RevMan) [computer software]. The Cochrane Collaboration; **2014**.
24. Schünemann H, Brożek J, Guyatt G, Oxman A, eds. Handbook for Grading the Quality of Evidence and the Strength of Recommendations Using the GRADE Approach. GRADE Working Group; **2013**.
25. Gopalakrishna G, Mustafa RA, Davenport C, et al. Applying Grading of Recommendations Assessment, Development and Evaluation (GRADE) to diagnostic tests was challenging but doable. J Clin Epidemiol **2014**; 67:760–8.
26. Gopalakrishna G, Leeflang MM, Davenport C, et al. Barriers to making recommendations about medical tests: a qualitative study of European guideline developers. BMJ Open **2016**; 6:e010549.
27. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ **2021**; 372:n71.
28. Deeks J, Bossuyt P, Takwoingi Y, Leeflang M, Flemyng E, Mellor L, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Cochrane Screening and Diagnostic Test Methods Group; **2022**.
29. SAS Institute Inc. SAS® OnDemand for Academics: User's Guide. SAS Institute Inc.; **2020**.
30. Melendez J, Philipsen RHHM, Chanda-Kapata P, Sunkutu V, Kapata N, van Ginneken B. Automatic versus human reading of chest x-rays in the Zambia National Tuberculosis Prevalence Survey. Int J Tuberc Lung Dis **2017**; 21:880–6.
31. Fehr J, Konigorski S, Olivier S, et al. Computer-aided interpretation of chest radiography reveals the spectrum of tuberculosis in rural South Africa. NPJ Digit Med **2021**; 4:106.
32. Mungai B, Ong'ango J, Ku CC, et al. Accuracy of computer-aided chest x-ray in community-based tuberculosis screening: lessons from the 2016 Kenya National Tuberculosis Prevalence Survey. PLOS Glob Public Health **2022**; 2:e0001272.
33. Velen K, Sathar F, Hoffmann CJ, et al. Digital chest x-ray with computer-aided detection for tuberculosis screening within correctional facilities. Ann Am Thorac **2022**; 19:1313–9.
34. Kagujje M, Kerkhoff AD, Nteeni M, Dunn I, Mateyo K, Muyoyeta M. The performance of computer-aided detection digital chest x-ray reading technologies for triage of active tuberculosis among persons with a history of previous tuberculosis. Clin Infect Dis **2023**; 76:e894–901.
35. Codlin AJ, Dao TP, Vo LNQ, et al. Independent evaluation of 12 artificial intelligence solutions for the detection of tuberculosis. Sci Rep **2021**; 11:23895.
36. Qin ZZ, Barrett R, Ahmed S, et al. Comparing different versions of computer-aided detection products when reading chest x-rays for tuberculosis. PLOS Digit Health **2022**; 1:e0000067.
37. Jo Y, Kagujje M, Johnson K, et al. Costs and cost-effectiveness of a comprehensive tuberculosis case finding strategy in Zambia. PLoS One **2021**; 16:e0256531.
38. Nsengiyumva NP, Hussain H, Oxlade O, et al. Triage of persons with tuberculosis symptoms using artificial intelligence-based chest radiograph interpretation: a cost-effectiveness analysis. Open Forum Infect Dis **2021**; 8:XXX–XX.
39. World Health Organization. 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. World Health Organization; **2021**.
40. FIND. Tuberculosis Diagnostics Digital Chest Radiography and Computer-Aided Detection (CAD) Solutions for Technology Landscape Analysis. FIND; **2021**.
41. Stop TB Partnership. Practical Guide on the Screening and Triage for TB Using Computer-Aided Detection (CAD) Technology and Ultra-Portable X-Ray Systems. Stop TB Partnership; **2021**.
42. Ahmad Khan F, Pande T, Tessema B, et al. Computer-aided reading of tuberculosis chest radiography: moving the research agenda forward to inform policy. Eur Respir J **2017**; 50:1700953.
43. David PM, Onno J, Keshavjee S, Ahmad Khan F. Conditions required for the artificial-intelligence-based computer-aided detection of tuberculosis to attain its global health potential. Lancet Digit Health **2022**; 4:e702–4.
44. Geric C, Qin ZZ, Denkinger CM, et al. The rise of artificial intelligence reading of chest x-rays for enhanced TB diagnosis and elimination. Int J Tuberc Lung Dis **2023**; 27:367–72.
45. Qin ZZ, Barrett R, Del Mar Castro M, et al. Early user experience and lessons learned using ultra-portable digital x-ray with computer-aided detection (DXR-CAD) products: a qualitative study from the perspective of healthcare providers. PLoS One **2023**; 18:e0277843.
46. Frascella B, Richards AS, Sossen B, et al. Subclinical tuberculosis disease—a review and analysis of prevalence surveys to inform definitions, burden, associations, and screening methodology. Clin Infect Dis **2021**; 73:e830–41.