

Cross-Sectional Evaluation of Universal and Stratified Computer-Aided Detection Thresholds for Tuberculosis Detection on Chest Radiographs

Omar Zayed^{1*}, Leila Ibrahim¹, Ahmed Mansour²

¹ Department of Pulmonary Medicine and TB Diagnostics, Faculty of Medicine, Cairo University, Cairo, Egypt.

² Department of Computer-Aided Diagnosis and Public Health, Faculty of Medicine, Alexandria University, Alexandria, Egypt.

*E-mail ✉ omar.zayed@cu.edu.eg

Abstract

Computer-aided detection (CAD) programs examine chest radiographs to identify signs of tuberculosis (TB) and generate a numerical abnormality score. However, assessments of CAD performance in TB screening are limited by insufficient confirmatory testing data, especially for individuals with lower scores or no symptoms. Moreover, optimal score cutoffs for triggering additional testing may vary by population traits and patient characteristics. This study sought to measure CAD accuracy across all screened people and determine if tailoring thresholds by age and sex could enhance results. This cross-sectional diagnostic accuracy study involved TB screening of individuals aged 15 years and older in Uganda using portable chest x-rays interpreted with CAD software (qXR version 3.2). Screening was offered to participants not receiving active TB treatment, irrespective of symptoms. We analyzed data from all eligible participants screened at facility- and community-based locations between June 1, 2022, and March 31, 2024. Those with scores at or above 0.1 (on a 0–1 scale) were requested to submit sputum samples for Xpert MTB/RIF Ultra testing. Diagnostic accuracy metrics (sensitivity, specificity, and area under the curve [AUC]) for detecting Xpert-positive TB were calculated using a uniform threshold, under varying assumptions about TB prevalence in those with scores below 0.1. These were then compared to stratified approaches based on age, sex, or both. Of 54,840 individuals evaluated for eligibility, 52,835 underwent CAD screening. Median age was 38 years (IQR: 26–50), with 23,586 (44.6%) males and 29,249 (55.4%) females. A total of 8,949 (16.9%) had scores of 0.1 or higher. Among 7,219 participants with valid Xpert results, 382 (5.3%) tested positive, including 81 trace results. Assuming 0.1% TB prevalence among those with scores below 0.1, qXR achieved an estimated AUC of 0.92 (95% CI: 0.90–0.94) for Xpert-positive TB. Stratification by age and sex improved performance; for instance, at 96.1% specificity (95% CI: 95.9–96.3), sensitivity reached 75.0% (69.9–79.5) with a universal threshold (≥ 0.65) versus 76.9% (71.9–81.2) with age- and sex-adjusted thresholds ($P = 0.046$). These results indicate that CAD accuracy for TB screening across all participants—including asymptomatic individuals and those with normal-appearing X-rays—is higher than previously reported. Customizing score thresholds based on factors such as age and sex may further improve performance, supporting a more targeted and efficient screening strategy.

Keywords: Tuberculosis, Computer-aided detection, Chest radiography, TB screening, Diagnostic accuracy, Threshold optimization

Introduction

Each year, over ten million people are estimated to develop tuberculosis, yet more than three million cases

go unreported to health authorities [1]. Enhanced case-detection strategies are essential to reduce the worldwide impact of TB [2]. Chest X-rays serve as an effective screening method, offering better sensitivity than symptom screening while remaining affordable and scalable [3]. Recent advances in artificial intelligence have produced computer-aided detection (CAD) systems that analyze chest radiographs, positioning them as valuable tools for expanding x-ray-based TB screening.

Research in context

Access this article online

<https://smerpub.com/>

Received: 06 September 2024; Accepted: 29 October 2024

Copyright CC BY-NC-SA 4.0

How to cite this article: Zayed O, Ibrahim L, Mansour A. Cross-Sectional Evaluation of Universal and Stratified Computer-Aided Detection Thresholds for Tuberculosis Detection on Chest Radiographs. J Med Sci Interdiscip Res. 2024;4(2):156-66. <https://doi.org/10.51847/ZRwoaKy1H7>

Evidence before this study

The World Health Organization (WHO) supports CAD for TB screening, but the score threshold prompting further diagnostic workup depends on the specific population. WHO advises setting context-appropriate thresholds and evaluating distinct cutoffs for subgroups, such as older adults, where CAD performance may be reduced. A PubMed search conducted up to September 9, 2024, using terms “tuberculosis” AND (“computer-aided detection” OR “computer aided detection” OR “CAD” OR “computer-aided reading” OR “computer aided reading” OR “artificial intelligence”) yielded 704 articles. English-language studies assessing CAD for screening, along with relevant cited works, were reviewed. Prior research typically reported AUCs between 0.76 and 0.88, but evaluations were often confined to symptomatic patients or those with visibly abnormal radiographs. Certain studies reported lower AUCs in subgroups such as older individuals or those with prior TB, and authors noted that the distribution of such traits influences the optimal population-level threshold. However, no studies have quantified potential accuracy gains from personalizing CAD thresholds based on individual characteristics.

Added value of this study

This investigation provided chest X-ray screening to consenting adolescents and adults in a high-burden setting, regardless of symptoms. A low-score cutoff (qXR version 3.2 TB score of 0.1 on a 0–1 scale) was applied to select individuals for confirmatory sputum molecular testing, which helped identify milder radiographic forms of TB and enabled evaluation of accuracy across various thresholds. To gauge CAD performance across all participants able to provide sputum, the authors assumed a 0.1% TB prevalence among those with scores below 0.1 who did not undergo testing. This symptom-agnostic approach with broad bacteriological confirmation yielded an AUC of 0.92, suggesting potentially higher accuracy than previously reported—though results hinge on the actual prevalence in low-scoring individuals. The analysis also showed that a very low threshold (< 0.1 for qXR v3.2) is required to satisfy WHO’s minimum target product profile of $\geq 90\%$ sensitivity and $\geq 60\%$ specificity for a high-sensitivity screen. Furthermore, aligning positive predictive values across demographic groups via adjusted thresholds yielded modest sensitivity gains (1–2% improvement

after age- and sex-stratification) while maintaining specificity.

Implications of all the available evidence

Achieving high sensitivity in CAD-based screening in high-prevalence areas may require lower score thresholds. Yet resource-limited high-burden countries may lack capacity to test everyone above such cutoffs. In these contexts, personalizing thresholds based on TB risk factors (e.g., male sex) and false-positive risks (e.g., advanced age) can boost program efficiency.

A key uncertainty remains the ideal threshold for referring individuals to further evaluation. CAD tools output a score reflecting the likelihood of pulmonary TB. WHO encourages calibration studies [4] to establish suitable thresholds for each setting. Most prior evaluations focused on symptomatic patients in triage [5–10] or restricted sputum testing for asymptomatic people to those with suggestive X-rays [11–15]. Consequently, limited information exists regarding CAD accuracy and optimal thresholds for individuals with subtle radiographic changes and no reported symptoms—data that matter because asymptomatic TB likely contributes to community transmission [16].

When establishing thresholds for community screening, demographic details such as age and sex warrant attention. Earlier studies generally applied one universal threshold [17–20], overlooking established links between TB risk and male sex [21, 22] as well as the increased chance of non-TB abnormalities on X-rays in older adults [23]. Therefore, adapting thresholds by age and sex may enhance overall screening performance. This analysis used data from an ongoing Ugandan TB case-finding project to evaluate CAD accuracy in the full screening population and examine the benefits of demographic stratification.

Materials and Methods*Study design and participants*

This cross-sectional diagnostic accuracy investigation formed part of the ongoing cluster-randomized CHASE-TB trial in Uganda (NCT05285202). It involved community-oriented tuberculosis screening using portable digital chest radiography and CAD. Eligibility extended to all adolescents and adults aged 15 years or older who were not currently receiving treatment for active tuberculosis, regardless of symptoms.

Recruitment occurred in testing tents located either adjacent to district health facilities (facility-based locations) or in busy public spaces such as transport centers and markets within neighborhoods and villages with elevated tuberculosis rates (community-based locations) in peri-urban and rural zones around Kampala. At facility-based sites, enrolment included not only patients attending the clinics but also accompanying persons, employees, and people passing through the area. In community settings, staff approached individuals walking by and, during slower periods, went door-to-door to nearby residences and businesses. The current evaluation retrospectively examined information from every participant screened at both types of sites between June 1, 2022 (when the study began) and March 31, 2024. Ethical approval was granted by the institutional review boards of Johns Hopkins University School of Medicine (Baltimore, MD, USA; IRB00300939) and Makerere University School of Public Health (Kampala, Uganda; SPH-2021-181). All participants provided oral informed consent (or assent accompanied by parental consent).

Procedures

Every consenting individual answered a standardized questionnaire covering demographic details (self-reported age, sex [male or female], and race), smoking history (introduced five months after study launch), known contacts with tuberculosis, symptoms in the preceding 30 days, prior tuberculosis treatment, and HIV status. Non-pregnant participants were invited to undergo digital chest radiography using a mobile X-ray unit. The resulting images were immediately analyzed by CAD software (qXR version 3.2) without reference to any clinical information. Individuals whose radiographs received qXR tuberculosis scores (X-ray scores; scale 0–1) exceeding the current cutoff were requested to submit expectorated sputum samples spontaneously. These were forwarded for Xpert MTB/RIF Ultra (Xpert) testing at a nearby health center, submitted without clinical details or imaging findings. Although X-ray scores help distinguish tuberculosis status, they are not calibrated as direct probability measures. The CAD cutoff for triggering Xpert testing started at 0.5, was reduced to 0.2 after the first month, and further lowered to 0.1 after another four months, guided by the observed score distribution and the goal of efficiently using available testing resources for the research.

Statistical analysis

Descriptive summaries of participant features used medians with interquartile ranges (IQR) for continuous variables and percentages for categorical ones. Group comparisons employed t-tests and χ^2 tests.

Performance of CAD (sensitivity, specificity, and area under the curve [AUC]) in identifying Xpert-positive tuberculosis was calculated among non-pregnant individuals capable of producing expectorated sputum. Because participants with X-ray scores below 0.1 were not asked to provide sputum, the main analysis assumed that 0.1% of such individuals would test Xpert-positive (a figure informed by CAD data from a prior universal Xpert Ultra screening study in Uganda). Sensitivity analyses explored alternative assumptions ranging from 0.05% (half the primary value, as a lower bound) to 0.3% (approximating Uganda's national Xpert-positive prevalence [24]). For those with scores below 0.1, the likelihood of successfully producing a sputum sample was modeled as comparable to that observed in the 0.1–0.19 score range, with sputum provision and Xpert outcomes assigned randomly. We assessed the association between X-ray scores and semi-quantitative Xpert results among Xpert-positive cases using the Spearman correlation coefficient.

Stratified CAD thresholds by age, sex, or both were developed on the basis that, under limited confirmatory testing resources, Xpert tests should be allocated to achieve comparable minimum pretest probabilities across demographic subgroups. This involved fitting shape-constrained (monotonically increasing) generalized additive models [25] within each subgroup, with Xpert outcome as the dependent variable and X-ray score as the predictor. For each age–sex stratum, we identified the score yielding an estimated Xpert-positivity prevalence nearest to 2% (and, separately, 1% in sensitivity analyses); these corresponded to resource intensities of 50 and 100 Xpert tests per positive result, respectively. The resulting scores served as subgroup-specific CAD thresholds. Overall sensitivity and specificity were then computed using age-, sex-, and age-by-sex-stratified thresholds. To contrast with a uniform approach, universal CAD thresholds were selected to match the specificity achieved in each stratified set—reflecting equivalent capacity to test individuals without tuberculosis. Sensitivity differences between universal and stratified strategies were calculated to quantify potential gains in case detection under resource constraints. Trace-positive Xpert Ultra results were classified as positive in primary analyses [26], with

additional sensitivity analyses treating them as negative. The sample size provided 80% power to detect a 3-percentage-point difference in sensitivity between strategies at a two-sided α of 0.05.

Because sputum collection was not requested for scores between 0.1 and 0.49 during the initial five months of the 22-month study, threshold derivation and accuracy assessments were restricted to participants enrolled after the cutoff reached 0.1. A sensitivity analysis incorporated earlier data via bootstrapping to simulate populations of equivalent size and score distribution but with full Xpert information. Performance estimates used the same thresholds as the primary (smaller) dataset, along with appropriate uncertainty measures. Two-sided $P < 0.05$ was considered statistically significant. All analyses were performed in Stata version 16.1 and R version 4.3.2.

Role of the funding source

The study sponsor was not involved in study design, data acquisition, analysis, interpretation, or manuscript preparation.

Results and Discussion

Out of 54,840 individuals evaluated for eligibility, 77 were already on tuberculosis treatment, 12 declined consent, 1,374 were pregnant and offered sputum testing without radiography, and 542 eligible persons lacked documented X-ray results. Consequently, 52,835 participants underwent tuberculosis screening via AI-interpreted digital chest X-rays with valid scores, of whom 45,758 were screened after the CAD threshold was reduced to 0.1. Among these 52,835 screened individuals, median age was 38 years (IQR 26–50), 23,586 (44.6%) were male, 29,249 (55.4%) were female, 3,478 (6.6%) self-reported HIV infection, 724 (1.4%) had a previous tuberculosis history, and 16,857 (31.9%) reported cough in the prior 30 days (**Table 1**).

Table 1. Characteristics of participants who received digital chest X-ray-based tuberculosis screening with computer-aided detection software

Empty cell	X-ray score < 0.1 (n = 43 886)	X-ray score 0.1–0.49 (n = 6107)	X-ray score ≥ 0.5 (n = 2842)	Total (n = 52835)
Median age, years	35 (25–48)	51 (37–65)	51 (37–67)	38 (26–50)
Sex				
Female	25 026 (57.0%)	3086 (50.5%)	1137 (40.0%)	29 249 (55.4%)
Male	18 860 (43.0%)	3021 (49.5%)	1705 (60.0%)	23 586 (44.6%)
HIV status*				
Known positive	2726 (6.2%)	457 (7.5%)	295 (10.4%)	3478 (6.6%)
Negative, with a test within the past year	21 776 (49.6%)	2704 (44.3%)	1189 (41.8%)	25 669 (48.6%)
Unknown†	19 384 (44.2%)	2946 (48.2%)	1358 (47.8%)	23 688 (44.8%)
Previous tuberculosis treatment	257 (0.6%)	123 (2.0%)	344 (12.1%)	724 (1.4%)
Previous household tuberculosis exposure	595 (1.4%)	104 (1.7%)	54 (1.9%)	753 (1.4%)
Current smoking‡	410/12 081 (3.4%)	114/1501 (7.6%)	79/753 (10.5%)	603/14 335 (4.2%)
Symptoms within 30 days of enrolment				
Any cough	12 945 (29.5%)	2299 (37.6%)	1613 (56.8%)	16 857 (31.9%)
Cough ≥ 2 weeks	6681 (15.2%)	1435 (23.5%)	1227 (43.2%)	9343 (17.7%)
Fever	8105 (18.5%)	1320 (21.6%)	728 (25.6%)	10 153 (19.2%)
Night sweats	1439 (3.3%)	294 (4.8%)	236 (8.3%)	1969 (3.7%)
Weight loss	909 (2.1%)	193 (3.2%)	181 (6.4%)	1283 (2.4%)
Any tuberculosis symptom§	17 876 (40.7%)	2963 (48.5%)	1865 (65.6%)	22 704 (43.0%)

Data are median (IQR), n (%), or n/N (%).

* Based on self-reporting.

† Includes 23 participants who declined to disclose HIV status.

‡ Only asked for participants enrolled after Nov 11, 2023.

§ Any cough, fever, night sweats, or weight loss.

Among the 52,835 screened participants, 43,886 (83.1%) recorded an x-ray score below 0.1, 6,107 (11.6%) fell in the 0.1–0.49 range, 1,616 (3.1%) had scores from 0.5 to 0.89, and 1,226 (2.3%) showed scores of 0.9 or higher (**Figure 1a**). Sputum testing was proposed to 8,038 (15.2%) individuals. Of these, 7,239 (90.1%) successfully submitted sputum specimens, and 7,219 (89.8%) yielded valid Xpert Ultra outcomes. Among the

valid results, 301 (4.2%) were positive at levels above the trace level, while an additional 81 (1.1%) were trace-positive. Factors linked to positive Xpert findings (including trace results) included younger age, male sex, presence of tuberculosis symptoms, and a history of prior tuberculosis treatment. These associations held even within the subset of participants whose X-ray scores ranged from 0.1 to 0.5.

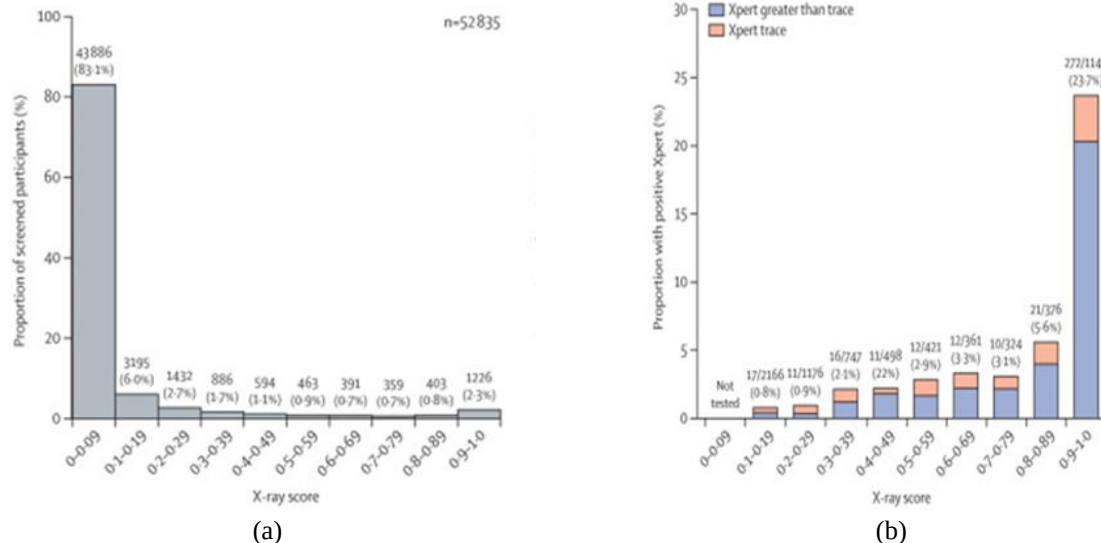


Figure 1. Chest X-ray and Xpert outcomes among community-screened adults (aged 15 years and above) in Uganda: (a) Distribution of qXR tuberculosis scores for study participants who underwent X-ray screening for tuberculosis using computer-aided detection software in Uganda; (b) Percentage of participants with positive sputum Xpert results across successive X-ray score intervals. qXR = computer-aided detection software, qXR version 3.2. Xpert = Xpert MTB/RIF Ultra.

The association between X-ray scores and Xpert outcomes appears in **Figure 1b**. Among 2,166 participants with scores ranging from 0.1 to 0.2, 17 (0.8%) showed positive Xpert results (including trace-positive). In the group with scores from 0.4 to 0.59 (919 participants), 23 (2.5%) had positive results, while 272 (23.7%) of 1,148 individuals with scores of 0.9 or above tested positive (**Table 2**). Overall, the proportion of participants with Xpert-positive sputum was higher

among men (289 [1.2%] of 23,586) than among women (93 [0.3%] of 29,249). Rates were comparable across age categories (189 [0.8%] of 24,607 individuals aged 40 years or older versus 193 [0.7%] of 28,227 younger than 40 years). Nevertheless, older adults exhibited a substantially higher rate of CAD-detected X-ray abnormalities than younger ones (X-ray score ≥ 0.1 in 6,396 [26.0%] of 24,607 versus 2,552 [9.0%] of 28,227).

Table 2. Proportion of participants with positive* sputum Xpert MTB/RIF Ultra results among those with valid results, by X-ray score

Empty cell	Male (n = 3843)	Female (n = 3376)	Aged < 40 years (n = 1999)	Aged $\geq$ 40 years (n = 5220)	All tested participants (n = 7219)
0.1–0.19	14/1042 (1.3%)	3/1124 (0.3%)	7/671 (1.0%)	10/1495 (0.7%)	17/2166 (0.8%)
0.2–0.29	8/590 (1.4%)	3/586 (0.5%)	4/299 (1.3%)	7/877 (0.8%)	11/1176 (0.9%)
0.3–0.39	9/367 (2.5%)	7/380 (1.8%)	9/173 (5.2%)	7/574 (1.2%)	16/747 (2.1%)

0.4–0.49	8/250 (3.2%)	3/248 (1.2%)	3/106 (2.8%)	8/392 (2.0%)	11/498 (2.2%)
0.5–0.59	9/240 (3.8%)	3/181 (1.7%)	3/87 (3.4%)	9/334 (2.7%)	12/421 (2.9%)
0.6–0.69	9/198 (4.5%)	3/163 (1.8%)	4/75 (5.3%)	8/286 (2.8%)	12/361 (3.3%)
0.7–0.79	9/168 (5.4%)	1/156 (0.6%)	7/89 (7.9%)	3/235 (1.3%)	10/324 (3.1%)
0.8–0.89	14/198 (7.1%)	7/178 (3.9%)	6/97 (6.2%)	15/279 (5.4%)	21/376 (5.6%)
0.9–1.0	209/789 (26.5%)	63/359 (17.5%)	150/402 (37.3%)	122/746 (16.4%)	272/1148 (23.7%)
Any x-ray score ≥ 0.1	289/3843 (7.5%)	93/3376 (2.8%)	193/1999 (9.7%)	189/5220 (3.6%)	382/7219 (5.3%)

* Including trace-positive.

A total of 45,758 (86.6%) of the 52,835 participants underwent screening after the CAD threshold was reduced to 0.1. Within this group, the X-ray scores estimated to correspond to a 1 percent or 2 percent probability of Xpert positivity varied by subgroup as follows: 0.11 or 0.47 for males aged 15–39 years, 0.25 or 0.53 for males aged 40 years and older, 0.28 or 0.52 for females aged 15–39 years, and 0.44 or 0.89 for females aged 40 years and older. Among all individuals with positive Xpert results, semi-quantitative Xpert grades showed only a weak association with X-ray scores (Spearman's correlation coefficient $r=0.28$).

Under the assumption that 0.1% of individuals with X-ray scores below 0.1 would test positive on sputum Xpert, the estimated area under the curve (AUC) for

CAD in detecting Xpert-positive tuberculosis during community screening was 0.92 (95% CI 0.90–0.94) (Figure 2). With this assumption, the manufacturer-suggested universal threshold of 0.5 yielded an estimated sensitivity of 79.1% (95% CI 74.3–83.2) and specificity of 94.8% (94.6–95.0). Reducing the threshold to 0.1 raised sensitivity to 89.9% (86.1–92.7) while lowering specificity to 83.6% (83.2–83.9). When a higher prevalence of 0.3% for Xpert-positive tuberculosis was assumed among those with scores below 0.1, the AUC fell to 0.83 (0.81–0.86). Sensitivity at the 0.5 threshold dropped to 65.6% (60.7–70.2) and at the 0.1 threshold to 74.5% (69.9–78.7), whereas specificity values stayed nearly unchanged (94.8% [94.6–95.0] at 0.5 and 83.5% [83.2–83.9] at 0.1).

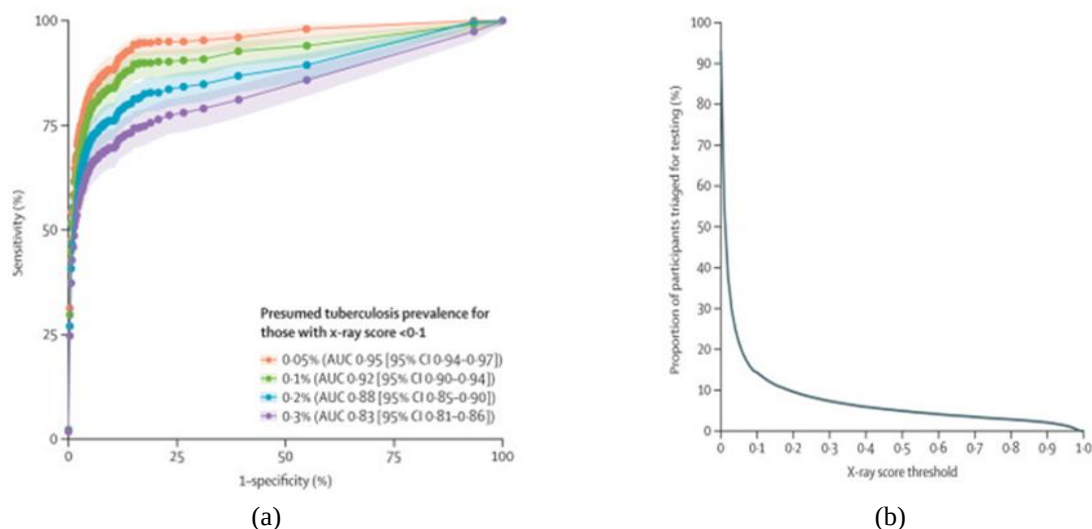


Figure 2. Diagnostic accuracy of CAD-supported tuberculosis screening in Ugandan community settings: (a) Receiver operating characteristic (ROC) curves illustrating the performance of CAD software (qXR version 3.2) in identifying sputum Xpert-positive tuberculosis. Shaded areas represent 95% confidence intervals around the ROC curves and AUC values; (b) Percentage of participants with X-ray scores at or above each possible threshold who would therefore qualify for tuberculosis confirmatory testing. AUC = area under the curve. CAD = computer-aided detection. ROC = receiver operating characteristic. Xpert = Xpert MTB/RIF Ultra.

In the sensitivity analysis including data from participants recruited during the initial study period, the AUC remained comparable at 0.93 (0.91–0.94) under the 0.1% prevalence assumption for individuals with X-ray scores below 0.1.

When compared with a single universal threshold set to achieve equivalent specificity, thresholds tailored by both age and sex (using the scores estimated to reflect 2% Xpert positivity per subgroup) delivered modestly superior sensitivity (76.9% [95% CI: 71.9–81.2] versus 75.0% [69.9–79.5]; $P = 0.046$); (**Table 3**). Sensitivity gains from age-and-sex stratification remained consistent under alternative scenarios: when thresholds were calibrated to a 1% Xpert positivity level (85.1% [80.8–88.6] versus 83.5% [79.1–87.2]; $P = 0.18$), or when different prevalence assumptions were applied to those with scores below 0.1 (81.0% [76.2–85.0] versus 79.0% [74.0–83.2]; $P = 0.046$ assuming 0.05% prevalence, and

63.8% [58.8–68.4] versus 62.2% [57.2–66.9]; $P = 0.046$ assuming 0.3% prevalence). Stratified thresholds by age and sex also produced higher sensitivities when trace-positive Xpert results were classified as negative (79.1% [73.8–83.6] versus 77.6% [72.1–82.2]; $P = 0.22$) and when the analysis encompassed all eligible participants, including those from the early screening phase (73.4% [69.4–77.5] versus 72.4% [68.5–76.5]; $P = 0.21$). Stratification based solely on sex produced smaller sensitivity improvements at higher thresholds (75.9% [70.9–80.3] versus 75.0% [69.9–79.5]; $P = 0.55$) or no difference at lower thresholds (87.0% [82.9–90.3] versus 87.0% [82.9–90.3]; $P = 1.00$). In contrast, age-only stratification did not yield consistent benefits (76.6% [71.6–80.9] versus 77.2% [72.3–81.5]; $P = 0.48$ at higher thresholds, and 82.9% [78.4–86.7] versus 82.3% [77.7–86.1]; $P = 0.68$ at lower thresholds).

Table 3. Sensitivity and specificity of qXR thresholds stratified by both age and sex for detecting Xpert MTB/RIF Ultra-positive tuberculosis, compared with universal thresholds of matching specificity

Empty cell	Fixed-specificity comparison, higher sensitivities		Fixed-specificity comparison, lower sensitivities	
	Thresholds stratified by both age and sex [‡]	Universal threshold [†]	Thresholds stratified by both age and sex [‡]	Universal threshold [§]
Sensitivity estimates				
Assumed prevalence 0.05%	89.7% (85.7–92.6)	88.0% (83.8–91.2)	81.0% (76.2–85.0)	79.0% (74.0–83.2)
Assumed prevalence 0.1%	85.1% (80.8–88.6)	83.5% (79.1–87.2)	76.9% (71.9–81.2)	75.0% (69.9–79.5)
Assumed prevalence 0.2%	77.3% (72.6–81.4)	75.9% (71.1–80.1)	69.8% (64.8–74.4)	68.1% (63.0–72.8)
Assumed prevalence 0.3%	70.6% (65.8–75.0)	69.3% (64.5–73.7)	63.8% (58.8–68.4)	62.2% (57.2–66.9)
Specificity estimates				
Assumed prevalence 0.05–0.3%[¶]	91.0% (90.7–91.3)	91.0% (90.7–91.3)	96.1% (95.9–96.3)	96.1% (95.9–96.3)

Data presented as % (95% CI). Assumed prevalences refer to the estimated tuberculosis prevalence among individuals with X-ray scores below 0.1. qXR = computer-aided detection software, qXR version 3.2.

* These values represent the estimated scores at which roughly 1% of participants within each demographic subgroup would test Xpert-positive. Thresholds were ≥ 0.11 for males aged 15–39 years; ≥ 0.25 for males aged 40 years and older; ≥ 0.28 for females aged 15–39 years; and ≥ 0.44 for females aged 40 years and older.

† Universal threshold selected to achieve the same overall specificity of 91.0% as the age- and sex-stratified thresholds. Threshold ≥ 0.26 for the entire population.

‡ These values represent the estimated scores at which roughly 2% of participants within each demographic subgroup would test Xpert-positive. Thresholds were ≥ 0.47 for males aged 15–39 years; ≥ 0.53 for males aged 40 years and older; ≥ 0.52 for females aged 15–39 years; and ≥ 0.89 for females aged 40 years and older.

§ Universal threshold selected to achieve the same overall specificity of 96.1% as the age- and sex-stratified thresholds. Threshold ≥ 0.65 for the entire population.

¶ Specificity values did not change when the assumed tuberculosis prevalence among individuals with X-ray scores below 0.1 was varied to 0.05%, 0.1%, 0.2%, or 0.3%.

This assessment of a community-based active case-finding initiative in Uganda revealed that CAD demonstrated greater diagnostic accuracy—particularly higher specificity—than earlier investigations [11, 13–15] that evaluated performance only among participants with tuberculosis symptoms or abnormal radiographs.

Assuming that 0.1% of individuals with normal or nearly normal X-rays (scores below 0.1) harbored Xpert-positive tuberculosis, the AUC for detecting Xpert-positive disease across all screening participants capable of producing sputum (irrespective of symptoms or radiographic appearance) reached 0.92 (95% CI: 0.90–

0.94). A threshold close to 0.1 could satisfy the WHO's minimum target product profile criteria [27] of at least 90% sensitivity and at least 60% specificity for a high-sensitivity screening tool. This performance may be further enhanced by stratifying CAD thresholds by participant age and sex. These observations highlight CAD's value in population-level screening and support additional research on more inclusive thresholds adjusted by straightforward demographic factors.

In the studied population, 5.4% of participants had X-ray scores at or above the manufacturer-recommended cutoff of 0.5 and exhibited the greatest likelihood of Xpert positivity. Nonetheless, the additional 11.6% of participants with scores between 0.1 and 0.49 still had a more than 1% chance of Xpert-positive tuberculosis; omitting confirmatory testing for this group reduced sensitivity to 79.1%. Thus, community screening programs may require lower CAD thresholds to attain adequate sensitivity. This observation of a broad spectrum of radiographic manifestations of tuberculosis in community settings is consistent with findings from a recent CAD evaluation of participants in a South African prevalence survey [15], in which the threshold had to be lowered to 0.18 to achieve 90% sensitivity. Given that lowering thresholds can place additional demands on scarce resources in high-burden regions, programs should carefully balance the advantages and drawbacks of more versus less sensitive strategies when establishing cutoffs. Formal decision modeling, supported by cost-effectiveness evaluations, could help determine suitable thresholds for diverse populations.

The study also demonstrated that customizing CAD thresholds based on simple, readily available individual characteristics, such as age and sex, can enhance screening accuracy. Tuberculosis prevalence was approximately four times higher among male than female participants. Consequently, males with X-ray scores of 0.52 or above (the sex-specific threshold linked to 2% Xpert positivity) faced a comparable tuberculosis risk to females with scores of 0.81 or higher. Although overall tuberculosis prevalence was similar between those younger than 40 years and those aged 40 years or older, older participants were nearly three times more likely to show CAD-detected radiographic abnormalities (score $\geq$ 0.1). This difference is probably attributable to a greater burden of other pulmonary conditions, including unrecognized prior tuberculosis, in older adults. As a result, among those with abnormal X-rays who underwent Xpert testing, younger individuals were three

times more likely to test positive than older ones. Younger participants with scores of 0.47 or above (the age-specific threshold for 2% Xpert positivity) exhibited a tuberculosis risk similar to that of older participants with scores of 0.62 or higher. In resource-constrained environments, therefore, age- and sex-stratified thresholds can increase tuberculosis case detection without expanding confirmatory testing capacity.

These results open the door to more individualized approaches within active case-finding efforts. In clinical practice, decisions to test for tuberculosis routinely integrate multiple elements including age, sex, HIV status, exposure history, symptoms, and imaging findings. In large-scale screening, however, such comprehensive assessment is rarely practical, leading to reliance on basic criteria such as symptoms or visibly abnormal X-rays. With CAD now supplying quantitative scores that correlate with the probability of Xpert positivity, it becomes feasible to incorporate additional risk factors into screening algorithms. While the sensitivity improvement from age- and sex-stratification was relatively small (1–2%), it remains meaningful in settings where testing resources cannot be extended to every person who exceeds a low threshold. Future economic analyses should compare the savings from reduced unnecessary testing with the modest costs of implementing stratified thresholds; for basic variables such as age and sex, these additional costs are expected to be negligible. If confirmed in prospective studies and diverse populations, adopting stratified thresholds—including lowered cutoffs for higher-risk groups—merits consideration.

Specificity estimates in this systematic screening context are higher than those reported in some prior studies due to methodological differences. Even at the lowest threshold (0.1) for which bacteriological confirmation was pursued, specificity remained 83.6%—well above the WHO minimum target of $\geq$ 60%. Three earlier studies of qXR version 3 [13, 15, 18] reported substantially lower specificities (ranging from below 50% to 62%) at thresholds achieving 90% sensitivity. Those figures arose from restricting analyses to participants who received Xpert testing—typically those with symptoms or abnormal radiographs. By contrast, applying a low threshold and estimating prevalence below it permitted evaluation across the full screening population, including the large proportion with negative symptom screens and normal X-rays who are not usually tested. Incorporating all eligible individuals increases the pool of true

negatives (no tuberculosis and low X-ray scores), yielding higher specificity estimates, as observed in other studies [28, 29] and in the present analysis, which included participants with normal radiographs.

Several limitations should be acknowledged. First, although we used a low CAD threshold, we did not perform Xpert testing for individuals with scores below 0.1. Since roughly 80% of participants fell into this category, sensitivity estimates are sensitive to the assumed tuberculosis prevalence in this group. However, AUC values and the benefits of stratification proved reasonably stable across sensitivity analyses exploring different prevalence assumptions. A single prevalence value was applied, yet the true rate likely scales with local tuberculosis burden; additional research should examine setting-specific CAD performance. Second, accuracy was measured against a practical reference standard of expectorated sputum Xpert testing (without induction or culture confirmation). While performance against a more exhaustive standard remains unknown, sputum Xpert is standard in systematic screening, so the estimates accurately reflect CAD's ability to identify cases detectable by this method. Pregnant individuals and those who failed to produce sputum (9.9% of those with scores ≥ 0.1) were excluded. Stratification was examined only for age, sex, or both, given their easy availability; future studies should investigate other factors such as HIV status (unknown for about half the participants) and tuberculosis history. Third, only one CAD product was assessed, and no comparison with human radiologists was possible, limiting insights into specific radiographic patterns linked to tuberculosis at low scores. Finally, stratified threshold performance relied on Xpert results from a modest number of individuals with scores in the bimodal distribution valley and could be affected by variations in CAD versions. Validation in other settings with additional CAD systems and exploration of implementation barriers are therefore essential. Where feasible, direct measurement of tuberculosis prevalence among those with low X-ray scores would further clarify the role of chest radiography and CAD in screening, given the strong dependence on prevalence in the large low-score group.

Conclusion

In summary, screening of more than 50,000 individuals for tuberculosis in Ugandan communities with portable chest X-ray and CAD showed that, while overall

accuracy for Xpert-positive tuberculosis is high in this context, participants with scores between 0.1 and 0.5 remained at increased risk. Stratifying CAD thresholds by age and sex can modestly improve screening performance. Although confirmation in additional studies is required, lower thresholds should be contemplated when resources allow. In high-burden areas with constrained confirmatory testing capacity, age- and sex-stratified thresholds may offer a more efficient, personalized approach to tuberculosis screening.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. World Health Organization. Global tuberculosis report 2023. Geneva: WHO; 2023.
2. Burke RM, Nliwasa M, Feasey HRA, Khan P, Henrion MYR, Mallewa J, et al. Community-based active case-finding interventions for tuberculosis: a systematic review. *Lancet Public Health*. 2021;6(5):e283-99.
3. World Health Organization. WHO consolidated guidelines on tuberculosis: module 2: screening – systematic screening for tuberculosis disease. Geneva: WHO; 2021.
4. 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. Geneva: WHO; 2021.
5. Qin ZZ, Sander MS, Rai B, Titahong CN, Sudrungrot S, Laah SN, et al. Using artificial intelligence to read chest radiographs for tuberculosis detection: a multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Sci Rep*. 2019;9:15000.
6. Breuninger M, van Ginneken B, Philipsen RH, Mhimbira F, Hella JP, Lwilla F, et al. Diagnostic accuracy of computer-aided detection of pulmonary tuberculosis in chest radiographs: a validation study

- from sub-Saharan Africa. *PLoS One*. 2014;9(9):e106381.
7. Khan FA, Majidulla A, Tavaziva G, Nazish A, Abidi SH, Benedetti A, 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(11):e573-81.
 8. Melendez J, Sánchez CI, Philipsen RH, Maduskar P, Dawson R, Theron G, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Sci Rep*. 2016;6:25265.
 9. Muyoyeta M, Maduskar P, Moyo M, Kasese N, Mwanza W, Kapata N, et al. The sensitivity and specificity of using a computer aided diagnosis program for automatically scoring chest X-rays of presumptive TB patients compared with Xpert MTB/RIF in Lusaka Zambia. *PLoS One*. 2014;9(4):e93757.
 10. Qin ZZ, Ahmed S, Sarker MS, Paul K, Adel ASS, Naheyan T, 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(9):e543-54.
 11. Fehr J, Gunda R, Siedner MJ, Muyoyeta M, Kasese N, Mwanza W, et al. CAD4TB software updates: different triaging thresholds require caution by users and regulation by authorities. *Int J Tuberc Lung Dis*. 2023;27(2):157-60.
 12. Kik SV, Gelaw SM, Ruhwald M, Ndlovu Z, Muyoyeta M, Kasese N, et al. Diagnostic accuracy of chest x-ray interpretation for tuberculosis by three artificial intelligence-based software in a screening use-case: an individual patient meta-analysis of global data [preprint]. *medRxiv*. 2022. doi:10.1101/2022.01.24.22269730
 13. Codlin AJ, Dao TP, Vo LNQ, Forse RJ, Van Truong V, Dang HM, et al. Independent evaluation of 12 artificial intelligence solutions for the detection of tuberculosis. *Sci Rep*. 2021;11:23895.
 14. Scott AJ, Perumal T, Hohlfeld A, Osei L, Mutyambizi C, Lall M, et al. Diagnostic accuracy of computer-aided detection during active case finding for pulmonary tuberculosis in Africa: a systematic review and meta-analysis. *Open Forum Infect Dis*. 2024;11(2):ofae020.
 15. Qin ZZ, Van der Walt M, Moyo S, Sizmur S, Tembo J, Tembo T, et al. Computer-aided detection of tuberculosis from chest radiographs in a tuberculosis prevalence survey in South Africa: external validation and modelled impacts of commercially available artificial intelligence software. *Lancet Digit Health*. 2024;6(9):e605-13.
 16. Nguyen HV, Tiemersma E, Nguyen NV, Nguyen HB, Cobelens F. Disease transmission by patients with subclinical tuberculosis. *Clin Infect Dis*. 2023;76(11):2000-6.
 17. MacPherson P, Webb EL, Kamchedzera W, Khundi M, Choko AT, Msefula C, et al. Computer-aided x-ray screening for tuberculosis and HIV testing among adults with cough in Malawi (the PROSPECT study): a randomised trial and cost-effectiveness analysis. *PLoS Med*. 2021;18(9):e1003752.
 18. 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(3):e894-901.
 19. Fehr J, Konigorski S, Olivier S, Gunda R, Surujdeen A, Gareta D, et al. (Vukuzazi Team). Computer-aided interpretation of chest radiography reveals the spectrum of tuberculosis in rural South Africa. *NPJ Digit Med*. 2021;4:106.
 20. Moodley N, Velen K, Saimen A, Nakhura Z, Churchyard G, Charalambous S. Digital chest radiography enhances screening efficiency for pulmonary tuberculosis in primary health clinics in South Africa. *Clin Infect Dis*. 2022;74(9):1650-8.
 21. Horton KC, MacPherson P, Houben RM, White RG, Corbett EL. Sex differences in tuberculosis burden and notifications in low- and middle-income countries: a systematic review and meta-analysis. *PLoS Med*. 2016;13(9):e1002119.
 22. Ledesma JR, Ma J, Vongpradith A, Maddison ER, Novotney A, Biehl MH, et al. Global, regional, and national sex differences in the global burden of tuberculosis by HIV status, 1990–2019: results from the Global Burden of Disease Study 2019. *Lancet Infect Dis*. 2022;22(2):222-41.
 23. Mungai BN, Joeke E, Masini E, Onyango M, Mwangi M, Mugi B, et al. 'If not TB, what could it be?' Chest X-ray findings from the 2016 Kenya Tuberculosis Prevalence Survey. *Thorax*. 2021;76(6):607-14.

24. The Republic of Uganda Ministry of Health. The Uganda national tuberculosis prevalence survey, 2014–2015 survey report. Kampala: Ministry of Health; 2016.
25. Pya N, Wood SN. Shape constrained additive models. *Stat Comput.* 2015;25(3):543-59.
26. Sung J, Nantale M, Nalutaaya A, Nakaweesi J, Kato E, Namugosa M, et al. Long-term risk of tuberculosis among individuals with Xpert Ultra trace screening results in Uganda: a longitudinal follow-up study. *Lancet Infect Dis.* 2025 [published online Oct 7]. doi:10.1016/S1473-3099(25)00536-5
27. World Health Organization. Target product profiles for tuberculosis screening tests. Geneva: WHO; 2025.
28. Keter AK, Vanobberghen F, Lynen L, Muvunyi CM, Muvunyi R, Muvunyi E, et al. Simultaneous alleviation of verification and reference standard biases in a community-based tuberculosis screening study using Bayesian latent class analysis. *PLoS One.* 2024;19(8):e0305126.
29. Mungai B, Ong'ang'o J, Ku CC, Wambugu M, Mwangi M, Oduor C, 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(11):e0001272.