

Triage of Persons With Tuberculosis Symptoms Using Artificial Intelligence–Based Chest Radiograph Interpretation: A Cost-Effectiveness Analysis

Ntwali Placide Nsengiyumva,^{1,a} Hamidah Hussain,^{2,a} Olivia Oxlade,¹ Arman Majidulla,² Ahsana Nazish,³ Aamir J. Khan,² Dick Menzies,^{1,4} Faiz Ahmad Khan^{1,4} and Kevin Schwartzman^{1,4}

¹McGill International Tuberculosis Centre, Research Institute of the McGill University Health Centre, Montreal, Canada, ²Interactive Research and Development Global, Singapore, ³Ghori Tuberculosis Clinic, Indus Hospital, Karachi, Pakistan, and ⁴Department of Medicine and Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

Background. In settings without access to rapid expert radiographic interpretation, artificial intelligence (AI)–based chest radiograph (CXR) analysis can triage persons presenting with possible tuberculosis (TB) symptoms, to identify those who require additional microbiological testing. However, there is limited evidence of the cost-effectiveness of this technology as a triage tool.

Methods. A decision analysis model was developed to evaluate the cost-effectiveness of triage strategies with AI-based CXR analysis for patients presenting with symptoms suggestive of pulmonary TB in Karachi, Pakistan. These strategies were compared to the current standard of care using microbiological testing with smear microscopy or GeneXpert, without prior triage. Positive triage CXRs were considered to improve referral success for microbiologic testing, from 91% to 100% for eligible persons. Software diagnostic accuracy was based on a prospective field study in Karachi. Other inputs were obtained from the Pakistan TB Program. The analysis was conducted from the healthcare provider perspective, and costs were expressed in 2020 US dollars.

Results. Compared to upfront smear microscopy for all persons with presumptive TB, triage strategies with AI-based CXR analysis were projected to lower costs by 19%, from \$23 233 per 1000 persons, and avert 3%–4% disability-adjusted life-years (DALYs), from 372 DALYs. Compared to upfront GeneXpert, AI-based triage strategies lowered projected costs by 37%, from \$34 346 and averted 4% additional DALYs, from 369 DALYs. Reinforced follow-up for persons with positive triage CXRs but negative microbiologic tests was particularly cost-effective.

Conclusions. In lower-resource settings, the addition of AI-based CXR triage before microbiologic testing for persons with possible TB symptoms can reduce costs, avert additional DALYs, and improve TB detection.

Keywords. artificial intelligence; chest radiography; cost-effectiveness; deep learning; tuberculosis.

Tuberculosis (TB) remains a leading cause of morbidity and mortality worldwide. With 10 million estimated cases annually [1], a large gap exists between estimated incidence and reported cases, partly due to underdiagnosis. In an effort to reduce underdiagnosis, it was recommended at the 2018 United Nations high-level meeting on tuberculosis [1] to improve diagnostic tests to promote early identification, which in turn will improve health outcomes and reduce transmission [2].

One of the most needed diagnostic tools in high-TB burden settings is a low-cost, efficient TB triage test for first contact health

centers. Chest radiographs (CXRs) have been considered potential triage tests, to focus further microbiologic evaluation of persons presenting with symptoms suggestive of pulmonary TB [1].

CXRs read by radiologists are suitably accurate, but the costs and scarcity of radiologists and other trained readers impede usage as triage tests in most low-income, high-TB-burden countries. In settings without access to experienced readers, artificial intelligence (AI)–guided CXR interpretation could potentially be used for triage, particularly given the recent advances in performance.

In December 2020, the World Health Organization (WHO) announced that forthcoming guidelines on TB screening and diagnosis will state that AI-based CXR analysis software can replace human CXR interpretation [3]. A recent market mapping study documented 27 developers of AI-guided CXR analysis systems and described 11 of these systems; 8 are commercially available for TB diagnosis [4]. Deployment, data sharing, compatibility and other elements relevant to integrating these systems into existing TB diagnostic algorithms were reviewed and the diagnostic accuracy of AI-based CXR interpretation has been evaluated in various high-TB burden settings [5–8].

Received 26 July 2021; editorial decision 3 November 2021; accepted 5 November 2021; published online 15 December 2021.

^aN. P. N. and H. H. contributed equally to this work.

Correspondence: Kevin Schwartzman, MD, MPH, McGill University Health Centre, 1001 Blvd Décarie, Room D05.2511, Montreal, QC, H4A 3J1, Canada (kevin.schwartzman@mcgill.ca).

Open Forum Infectious Diseases® 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com <https://doi.org/10.1093/ofid/ofab567>

Table 1. Diagnostic Algorithms

Strategy	Description of Diagnostic Algorithm
Smear as microbiologic test	
I. Status quo: AFB smear	Upfront sputum AFB smear × 2 for all persons with symptoms suggestive of TB. CXR performed, with human interpretation, for individuals with negative smears and persistent symptoms after 7 days of antibiotics.
1A. Triage with CXR and AI-based CXR interpretation. AFB smears only if CXR suggests TB	Triage using AI-based CXR interpretation for all persons with presumptive TB based on symptoms. If CXR suggests pulmonary TB based on score, the person then has 3 sputum AFB smears sent. If those smears are then negative, but the person has persistent symptoms despite an antibiotic trial, they are started on TB treatment based on clinical diagnosis. If the CXR is not suggestive based on score, there is no further testing.
1B. Reinforced follow-up based on triage with CXR and AI-based CXR interpretation before AFB smears	Triage using AI-based CXR interpretation for all persons with suspected TB based on symptoms, plus reinforced follow-up. If the CXR suggests pulmonary TB based on score, the person then has 3 sputum AFB smears sent. If these are negative, the person undergoes repeat sputum AFB smear within 2 weeks.
2. Triage after a negative AFB smear	Upfront sputum AFB smear × 3 for all persons with suspected TB based on symptoms. AI-based CXR interpretation done for those with negative smears; if score above threshold for suspected TB, the person is referred for Xpert testing; if the score is below the threshold, there is no further testing at that time.
Xpert as microbiologic test	
II. Status quo: Xpert	Upfront sputum Xpert for all persons with suspected TB based on symptoms. No further testing protocol if Xpert negative.
3A. Triage with AI-based CXR interpretation before Xpert	Triage using AI-based CXR interpretation for all persons with suspected TB based on symptoms. If CXR suggests pulmonary TB based on score, the person then has a sputum sample sent for Xpert. No further testing if Xpert is negative. If the CXR does not suggest pulmonary TB based on score, there is no further testing.
3B. Reinforced follow-up based on triage with AI-based CXR interpretation before Xpert	Triage using AI-based CXR interpretation for all persons with suspected TB based on symptoms, plus reinforced follow-up. If the CXR suggests pulmonary TB based on score, the person then has 1 sputum sample sent for Xpert. If this is negative, the person then undergoes repeat sputum Xpert within 2 weeks.

Abbreviations: AFB, acid-fast bacilli; AI, artificial intelligence; CXR, chest radiograph; TB, tuberculosis.

The WHO's target product profile for TB diagnostics emphasizes affordability [1]. Previous reports from Pakistan and South Africa estimated costs per radiograph and per TB diagnosis with AI-based CXR interpretation software, but did not consider overall testing and treatment costs (including false-positive treatment starts), nor broader health outcomes [7, 9]. To date there has been no published assessment of the cost-effectiveness of triage using AI-based CXR interpretation. To explore the potential cost and effectiveness of this technology, a decision analysis model was developed, building on a prospective field study of diagnostic accuracy completed at Indus Health Networks (IHN) TB clinic in Karachi, Pakistan [6].

METHODS

Study Setting

In 2019, Pakistan was 1 of 8 countries that together accounted for two-thirds of the global TB burden; only 50% of notified cases were bacteriologically confirmed. The estimated incidence in Pakistan in 2019, was 263 per 100 000 population [10]; in Sindh Province in 2011, it was 454 per 100 000 [11]. We simulated a cohort reflecting the population who participated in the field study in Karachi [6]. This population was representative of persons presenting with symptoms suggestive of TB from the Indus Hospital catchment area, a low-income, high-TB-incidence area of Karachi. The prospective study evaluated the diagnostic accuracy of 2 analysis software packages: qXR version 2.0 (qXRv2, qure.ai, Mumbai, India) and CAD4TB version 6.0 (CAD4TBv6, Delft, Veenendaal, the Netherlands). It included 2198 adults (52% men) aged ≥15 years (median age,

33 years [interquartile range, 23–49 years]), who presented with TB symptoms.

Participants were enrolled from March 2017 to July 2018 and liquid culture of 2 sputum specimens was the reference standard test for pulmonary TB. Persons with at least 1 positive culture were classified as having TB, and those with 2 negative cultures were classified as not having pulmonary TB. The study informed cohort and diagnostic accuracy parameters in our model. Other inputs, such as costs and treatment outcomes, were obtained from the IHN TB clinic and the Pakistan National TB Program.

Overview

A trial-based economic evaluation of triage, using AI-based CXR interpretation for persons presenting with symptoms suggestive of pulmonary TB, used a decision analysis model (TreeAge Pro 2020; TreeAge Software, Williamstown, Massachusetts). Model outputs included TB diagnoses (both true and false positives) and costs (in 2020 US dollars), as well as deaths, numbers of microbiologic tests, and disability-associated life-years (DALYs). Where appropriate, incremental costs were estimated: per additional person correctly diagnosed with TB, per TB-related death averted, and per DALY averted. All analyses were conducted from the perspective of the healthcare payer.

Simulated Cohort

The cohort included human immunodeficiency virus (HIV)-negative individuals with TB symptoms referred for further testing at a TB clinic. In all scenarios, simulations began with persons ≥15 years with symptoms suggestive of TB who

presented at a TB clinic. In this simulated cohort, 91% were referred for microbiologic testing [11], and 12% of tests yielded culture-confirmed TB [6]. The analysis focuses on individuals without risk factors for drug-resistant TB.

AI-Based CXR Interpretation as Triage Test

The diagnostic accuracy of 2 deep learning-based AI software packages for CXR was evaluated in the prospective study. The software analyzes CXR images and outputs a TB abnormality score, with higher scores indicating greater abnormality. Thresholds are preset: Scores above the threshold suggest possible pulmonary TB and the need for further microbiologic testing; scores below the threshold suggest that the CXR is sufficient to rule out pulmonary TB without further testing. The

test properties—that is, sensitivity and specificity—reflect this categorical classification.

Several potential diagnostic algorithms were simulated, with test properties for the 2 AI-based interpretation software packages taken from the field study. Since the diagnostic performance of both software packages was similar, we present the detailed cost-effectiveness analysis for one in the main text, and the other in [Supplementary Tables 4–6](#). Parallel diagnostic algorithms where acid-fast bacilli smear microscopy or GeneXpert is used were considered. The algorithms are listed in [Table 1](#), with further detail and scenarios in [Supplementary Table 1](#): the integration of AI-based CXR interpretation parallels the approaches described in the market mapping of these technologies [4]. These algorithms were compared to the current

Table 2. Model Parameters

Parameter	Value, % (Range)	Reference
Epidemiologic parameters		
Culture-confirmed TB prevalence in individuals who present to the clinic with symptoms suggestive of TB	12 (10–14) ^a	[6]
Adherence to referrals for upfront microbiologic testing in the status quo strategies	91 (80–99)	[12]
Proportion smear-positive (of all persons with TB)	78 (64–94)	[6]
Annual rate of spontaneous recovery from active TB without treatment		
Smear-positive patients	23 (18–29)	[13, 14]
Smear-negative patients	13 (7–21)	[13, 14]
Annual rate of TB-related death for untreated		
Smear-positive patients	39 (34–45)	[15, 16]
Smear-negative patients	2.5 (1.7–3.5)	[15, 16]
Probability of completing treatment for active TB	85 (80–90)	Register TB-09 for 2017 ^b
Probability of death during TB treatment	5	Register TB-09 for 2017
Specificity of smear microscopy	98 (97–99)	[17]
Specificity of culture	96	[18]
Sensitivity of human-read CXR		
Smear negative	80 (74–85)	[19]
Smear positive	94 (88–98)	[20]
Specificity of CXR for active TB when read by		
Clinical officer ^c	46 (33–59)	[21]
Physician/radiologist	77 (73–80)	[19]
Sensitivity of clinical-radiographic diagnosis of TB after negative smears	57	[19], Register TB-09 for 2017
Specificity of clinical-radiographic diagnosis of TB after negative smears	78	[19], Register TB-09 for 2017
Sensitivity of Xpert		
Smear negative	59 (51–67)	[22]
Smear positive	100 (95–100)	[23]
Specificity of Xpert		
Smear negative	98 (96–99)	[22]
Smear positive	99 (98–100)	
Software 1: AI-based CXR diagnostic performance		
Sensitivity of AI-based CXR		
Smear negative	82 (70–91)	[6]
Smear positive	97 (93–99)	[6]
Specificity of AI-based CXR	69 (67–71)	[6]
Software 2: AI-based CXR diagnostic performance		
Sensitivity of AI-based CXR		
Smear negative	80 (68–89)	[6]
Smear positive	96 (93–98)	[6]
Specificity of AI-based CXR	75 (73–77)	[6]

Abbreviations: AI, artificial intelligence; CXR, chest radiograph; TB, tuberculosis.

^aThe 95% confidence interval for TB prevalence among symptomatic persons in the parent Karachi study ranged from 11% to 13%, but we used a wider range for purposes of sensitivity analysis.

^bReport form TB-09: Quarterly report on treatment outcomes; it is produced by extracting data from the TB register TB-03.

^cA standardized approach for radiograph reading for TB.

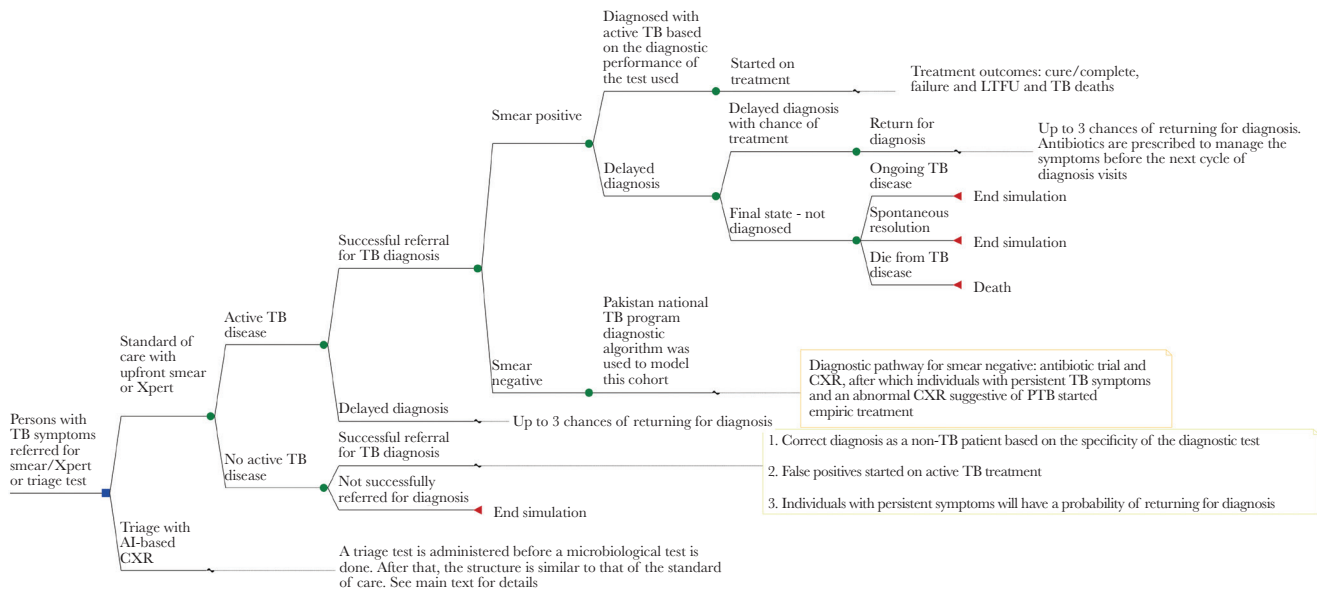


Figure 1. Simplified schematic of model structure for strategies to diagnose active TB. Probabilities related to each decision node are not shown. Abbreviations: AI, artificial intelligence; CXR, chest radiograph; LTFU, lost to follow-up; TB, tuberculosis.

standard of care (“status quo”), which involves upfront microbiological testing using (1) smear microscopy, or (2) GeneXpert, for all persons with presumptive TB based on symptoms (ie, no triage before microbiologic testing).

Model Structure and Key Assumptions

The simulation began with initial presentation to a health center with symptoms consistent with TB. Persons with active TB who were correctly diagnosed were assumed to be diagnosed, treated, and followed up during a single year. Those who remained undiagnosed were assumed to again seek care and to undergo microbiologic testing, 3 months after their previous false-negative testing episode if they had not died in the interval. Persons with TB and false-negative CXRs only underwent microbiologic testing when they returned with persistent symptoms. These persons remained undiagnosed for 3 months and during the 3-month delay, they became disabled, died, or were spontaneously cured. Survivors with persistent TB returned to the health center for reevaluation every 3 months, to a maximum of 3 visits after which they remained undiagnosed.

Persons without underlying TB continued in the simulation until they were discharged from medical evaluation and care for suspected TB—that is, all TB-related tests, as well as any TB treatment related to empiric therapy or false-positive test results, were considered. Discounting was not used given the short analytic horizon of 1 year.

Microbiologic testing required referral, sometimes to other sites, for acid-fast smears or GeneXpert, so persons initially evaluated for TB symptoms might not present for microbiologic testing. We assumed that when CXR was used as an initial triage test, microbiologic testing would be done immediately

on site, for all persons presenting with symptoms and imaging deemed consistent with TB. A model schematic is shown in [Figure 1](#).

TB Epidemiology Inputs

Model parameters related to TB pathogenesis, performance of microbiological tests, and conventional CXR were obtained from published literature ([Table 2](#)). TB diagnosis and treatment parameters were obtained from the prospective study [6] and 2018 Indus Health Network TB clinic data. The annual mortality rate from untreated TB was assumed to be 39% for smear-positive and 2.5% for smear-negative TB patients [13,15,16]. The annual rate of spontaneous resolution of untreated TB was assumed to be 23% for smear-positive and 13% for smear-negative TB patients [13,15,16].

DALYs

A weight of 0.331 DALYs was attributed to active TB [24]. Years lived with disability (YLD) and years of life lost due to premature mortality (YLL) were used to calculate DALYs (YLD + YLL). We assumed that those undergoing treatment for TB would only experience disability for the duration of their 6-month treatment regimen, and thus the yearly disability weight would be half (0.1655 DALYs) of that reported for a full year (0.331 DALYs) with untreated active TB. We defined YLD as the duration of time on TB treatment or time with TB before death multiplied by the disability weight. YLL was defined as remaining life expectancy at the age of death (the parent study median population age was 33). Life expectancy in Pakistan was 75 years in 2021 and the mean age at onset of TB disease was 33 years [25]. DALYs lost in future years were discounted at 3%.

Table 3. Cost Inputs

Cost/ Fee	Value (2021 USD)	Data Source (Year)
AI-based CXR interpretation	\$2.70	Fee schedule IHN TB Clinic, GHD-IHN, Delft imagery systems (2018)
AFB smear microscopy	\$1.26	Fee schedule Dow diagnostic (2019)
Xpert test	\$21.28	Fee schedule IHN TB clinic (2018)
Digital CXR in Karachi	\$1.70	Fee schedule IHN TB clinic (2018)
Prediagnosis antibiotics ^a	\$1.29	[30] (2019)
Radiograph reading by clinical officer	\$0.30	Personal communication with Khan RM (2019)
Radiograph reading by doctor	\$0.45	Personal communication with Dow University of Health Sciences (2018)
Standard TB treatment ^b	\$114.82	Calculated from various sources (NTP, IHN TB clinic, Aga Khan University, Dow University) (2017–2018)

Abbreviations: AFB, acid-fast bacilli; AI, artificial intelligence; CXR, chest radiograph; GHD-IHN, Global Health Directorate–Indus Health Network, IHN, Indus Health Network; NTP, Pakistan National TB Program; TB, tuberculosis; USD, United States dollars.

^aIndividuals missed with each subsequent microbiological test were assumed to take ciprofloxacin 500 mg (14 tablets) for 7 days based on NTP diagnosis guidelines.

^bMonthly and directly observed therapy visits, 6-month medication, treatment monitoring, and hospitalizations.

Cost Inputs

All component costs for diagnostic testing, treatment, and clinical care were obtained from the IHN TB clinic, published literature, and other Karachi reference laboratories [26] (Table 3). Technology costs included image analysis and equipment depreciation. Clinic overhead costs were not included as they are constant regardless of the diagnostic algorithm. Costs were converted from rupees to US dollars using the United States Treasury historical reported exchange rate, and inflated from 2018 to 2021 using Pakistan Bureau of Statistics consumer price index [27–29]. A detailed breakdown of cost estimates for TB diagnosis and treatments is provided in Supplementary Table 2.

Sensitivity Analysis

One-way sensitivity analysis for all model parameters was conducted, and tornado diagrams were generated accordingly. Distributions for the input parameters are provided in the Supplementary Table 3. Several scenarios for the underlying prevalence of TB among persons presenting with compatible symptoms were considered. Sensitivity and specificity thresholds for AI-based CXR interpretation corresponding to both the minimal and optimal target product profiles of triage tests [1] were considered. A scenario where the CXR is read by a radiologist, with increased specificity, was also considered. A scenario analysis incorporated 100% patient adherence to referrals for upfront microbiologic testing in the status quo strategies.

Probabilistic sensitivity analysis used 10 000 Monte Carlo simulations to obtain 95% uncertainty ranges (URs; 2.5th to 97.5th percentiles) around point estimates for all projected outcomes. For our willingness-to-pay threshold, and related cost-effectiveness acceptability curves, we used a value of \$195 per DALY for Pakistan, based on observed health opportunity costs in Pakistan; these may be lower for low- and middle-income country settings [31, 32]. Hence, this has been suggested as a more appropriate approach [32] than the often-used willingness below the thresholds of either per-capita

gross domestic product (GDP), or 3 times the per-capita GDP, making the analysis more conservative in this regard [33, 34].

Patient Consent Statement

For the parent study in Karachi, written informed consent was obtained from all study participants. The parent study was approved by the ethics review boards of Interactive Research and Development and the Research Institute of the McGill University Health Centre. Additional approval for the present analysis was not required, as it used only aggregate, previously reported data from the parent study without any individual patient data or identifiers.

RESULTS

The projected costs and outcomes for the various diagnostic algorithms are summarized in Table 4.

Base Case I: Acid-Fast Smear Microscopy as the Standard Microbiologic Test

Relative to a current standard of care based on upfront smear microscopy, the AI-based triage strategies without or with enhanced follow-up (1A and 1B) were projected to save 19% from the base value of \$23 233 per 1000 persons, and to increase TB detection by 0.5%–1.2% from the base value of 117.5 TB patients correctly diagnosed per 1000 persons (perfect case detection would correspond to 120 TB patients correctly diagnosed per 1000 persons evaluated).

The AI-based triage strategies reduced false-positive clinical diagnoses by 78% (from 69.5 to 15.1 per 1000 persons). False-positive treatments were a key determinant of cost, accounting for 37% of the total cost (\$8635 per 1000 persons) and this proportion was reduced to 11% (\$2061 per 1000 persons) with the AI-based triage strategies. Although diagnostic test costs doubled with AI-based triage strategies, this increase was more than offset by savings from averted false-positive treatment starts.

Table 4. Projected Cost and Effectiveness per 1000 Persons (Software 1)

Outcomes per 1000 Persons Per Strategy	Cost of Diagnosis (95% UR)	Cost of Treatment (95% UR)	True-Positive TB Diagnoses (95% UR)	No. of Microbiological Tests (95% UR)	False-Positive TB Diagnoses ^a (95% UR)	TB Deaths (95% UR)	DALYs Accrued (95% UR)
Smear as microbiologic test							
I. Status quo A: AFB smear	\$1756 (\$1663–\$1905)	\$21 477 (\$18 646–\$24 861)	1175 (99.6–138.2)	949 (935–973)	69.5 (63.1–79.8)	6.8 (5.7–8.0)	372 (298–462)
1A. Triage with AI-based CXR interpretation before AFB smears	\$747 (\$657–\$924)	\$15 290 (\$12 887–\$18 300)	118.1 (100.0–138.8)	415 (395–462)	15.1 (13.1–20.3)	6.4 (5.3–7.4)	359 (285–459)
1B. Reinforced follow-up based on triage with AI-based CXR interpretation before AFB smears	\$754 (\$664–\$934)	\$15 400 (\$12 990–\$18 429)	119.0 (100.8–139.9)	416 (395–464)	15.1 (13.1–20.3)	6.2 (5.3–7.4)	356 (285–459)
Xpert as microbiologic test							
II. Status quo B: Xpert	\$19 932 (\$19 754–\$20 312)	\$14 414 (\$12 065–\$17 364)	1175 (99.5–138.0)	936 (928–952)	8.0 (8.0–9.4)	6.7 (5.7–8.0)	369 (298–473)
3A. Triage with AI-based CXR interpretation before Xpert	\$5321 (\$4959–\$6052)	\$13 692 (\$11 239–\$16 498)	118.2 (99.9–138.5)	243 (227–275)	1.1 (1.1–1.4)	6.3 (5.4–7.6)	356 (288–462)
3B. Reinforced follow-up based on triage with AI-based CXR interpretation before Xpert	\$5342 (\$4975–\$6091)	\$13 758 (\$11 321–\$16 498)	118.8 (100.6–139.5)	244 (227–277)	1.1 (1.1–1.4)	6.2 (5.3–7.4)	354 (283–456)

Abbreviations: AFB, acid-fast bacilli; AI, artificial intelligence; CXR, chest radiograph; DALY, disability-adjusted life-year; TB, tuberculosis; UR, uncertainty range.

^aFalse positives include clinically diagnosed patients (ie, individuals who truly do not have TB, who have negative microbiologic tests but have been diagnosed with active TB on the basis of abnormal CXR and persistent symptoms). They are eventually started on empiric treatment.

Base Case II: Xpert as the Standard Microbiologic Test

Relative to a current standard of care based on upfront Xpert, the AI-based triage strategies without or with enhanced follow-up (3A and 3B) were projected to reduce costs by 37% compared to upfront Xpert (from \$34 346 per 1000 persons). The same strategies averted 4% additional DALYs (from 369 DALYs per 1000 persons) and increased patients correctly diagnosed by 1%.

AI-based triage strategies reduced the number of microbiologic tests by 74% relative to a status quo with upfront Xpert (from 936 tests per 1000 persons). Diagnosis with upfront Xpert accounted for 58% of the total cost in the standard of care strategy (\$19 932 per 1000 persons); it was reduced by 73% (to \$5321 per 1000 persons) in the AI-based triage strategies.

Incremental cost-effectiveness ratios comparing strategies are summarized in Tables 5 and 6. AI-based triage strategies (1A, 1B, 3A, and 3B) were consistently cheaper with better outcomes when compared to upfront smear microscopy and Xpert, in all simulations. The “reinforced follow-up after AI triage” strategies (1B and 3B) were projected to be the dominant strategies when either smear microscopy or Xpert was used for microbiologic diagnosis. Figures 2 and 3 show how these strategies were consistently cost-effective with respect to cost per DALY averted when compared to the next best alternative strategies: \$39 (95% UR, \$32–\$46) and \$40 (\$12–\$41) per DALY averted, for the reinforced follow-up vs AI-based triage only strategies. These estimates fell well below the willingness-to-pay threshold of \$195 per DALY averted.

Scenario and Sensitivity Analyses

Sensitivity of conventional CXR read by either clinical officers or radiologists is comparable to AI-based CXR interpretation sensitivity, resulting in similar numbers of patients correctly diagnosed with TB. The lower specificity when the CXR was read by clinical officers led to more false positives. False positives with the upfront smear microscopy strategy were reduced by 44% when the subsequent CXR (done for persons with negative smears but persistent symptoms) was read by a radiologist (to 38.8 false positives per 1000 persons). Hence, the impact of CXR triage on empiric treatment and total cost was reduced (for details, see Supplementary Table 7).

When the patient referral rate for upfront microbiologic testing was increased from 91% to 100% in status quo simulations, costs increased as did diagnostic yield of upfront testing. However, the triage strategies remained cheaper, while the enhanced AI-based triage strategies (1B and 3B) were associated with slightly more true-positive diagnoses (see Supplementary Table 8 for details).

When the diagnostic performance of AI-based CXR interpretation was varied, more accurate testing predictably led to additional savings and increased TB detection. The better the diagnostic performance, the lower the TB prevalence can be for CXR-based triage to be cost-effective. Similarly, as the

prevalence of TB increased among persons evaluated for symptoms, the triage tests became more cost-effective. TB treatment cost was the most influential cost variable, as shown in the tornado diagrams in [Supplementary Figure 1](#).

In probabilistic sensitivity analyses, AI-based triage strategies with or without enhanced follow-up (1A, 3A, 1B, 3B) were projected to be cost-saving relative to the status quo in 100% of simulations, reducing false-positive treatment ([Supplementary Figures 2 and 3](#)).

Additional scenario analyses included triage scenarios where the CXR was read by a human reader. These scenarios were dominated when compared to the AI-based triage strategies ([Supplementary Table 9](#)).

We also tabulated TB treatment starts according to bacteriological and clinical diagnoses. The triage strategies reduced inappropriate treatment starts and excessive microbiological tests—generating cost savings with comparable health outcomes ([Supplementary Table 10](#)).

DISCUSSION

Our analysis builds on a trial-based evaluation of 2 packages for deep learning AI-based CXR interpretation among persons with suspected TB in Karachi, Pakistan. It suggests that triage using AI-based CXR interpretation can be cost-effective and even cost-saving relative to standard practice. The strategy with the most clinical benefit appeared to be the one where persons with TB symptoms and CXRs with AI-based interpretation compatible with possible TB underwent microbiologic testing; if microbiologic results were negative, they were retested 2 weeks later, that is, reinforced microbiologic follow-up testing.

The analysis further suggests that the use of AI-supported CXR interpretation could potentially reduce unnecessary

empiric treatment in individuals with persistent symptoms, a diagnostic limitation which has been documented elsewhere [10, 35–37]. As compared with the upfront use of Xpert for persons with presumptive TB, incorporating AI-based CXR triage is expected to generate substantial savings by reducing the number of Xpert tests, without sacrificing diagnostic yield.

These results are likely applicable to other low-income, high-TB-burden settings. Key costs as well as diagnostic performance were taken directly from a rigorous prospective cohort study in Pakistan [6]. The diagnostic algorithms we considered were reflective of current practices, or potential use cases for CXR among persons with presumptive TB. To further address the robustness and generalizability of these results, we conducted extensive scenario and sensitivity analyses that incorporated different triage thresholds. The specificity of human readers and a wide range of TB prevalence in the target population were also considered

We used local data, obtained in the type of setting where this technology will most likely be useful. The analysis reflected estimated test characteristics for 2 commercially available software packages in this setting.

One limitation is that the analysis reflects use of AI-based CXR interpretation in a setting with low HIV prevalence. The performance of these specific software packages in persons living with HIV has not been evaluated, but it is well established that HIV modifies the sensitivity and specificity of CXR for the diagnosis of pulmonary TB, and hence is likely to affect the accuracy of these software. Moreover, limited evidence suggested that accuracy was lower among people living with HIV for prior versions of one of the software programs, which were not based on deep learning [5, 8]. Similarly, this analysis evaluated a cohort with a low prevalence of drug-resistant TB. The lower specificity of the software among persons with prior tuberculosis,

Table 5. Projected Incremental Savings and Health Outcomes per 1000 Persons

Diagnostic Strategy ^a	Incremental Savings vs Status Quo (95% UR) ^b	Additional TB Patients Diagnosed vs Status Quo (95% UR)	TB Deaths Averted vs Status Quo (95% UR)	DALYs Averted vs Status Quo (95% UR)
Smear as microbiologic test				
I. Status quo A: AFB smear (comparator)				
1A. Triage with AI-based CXR interpretation before AFB smear	\$4500 (\$3593–\$5474)	0.6 (0.0–1.2)	0.4 (0.3–0.59)	12.8 (3.6–17.8)
1B. Reinforced follow-up based on triage with AI-based CXR interpretation before AFB smear	\$4383 (\$3466–\$5350)	1.5 (0.9–2.1)	0.5 (0.2–0.7)	15.7 (6.6–20.6)
Xpert as microbiologic test				
II. Status quo B: Xpert (comparator)				
3A. Triage with AI-based CXR interpretation before Xpert	\$12 637 (\$12 229–\$13 093)	0.7 (0.7–0.9)	0.4 (0.1–0.6)	12.8 (3.6–17.8)
3B. Reinforced follow-up based on triage with AI-based CXR interpretation before Xpert	\$12 550 (\$12 092–\$12 989)	1.2 (0.8–1.9)	0.5 (0.3–0.8)	15.0 (7.6–26.6)

Abbreviations: AFB, acid-fast bacilli; AI, artificial intelligence; CXR, chest radiograph; DALY, disability-adjusted life-year; TB, tuberculosis; UR, uncertainty range.

^aResults for software 1, using the status quo AFB smear and Xpert strategies as comparators. (As all strategies were associated with savings relative to the status quo, incremental cost-effectiveness ratios are not shown.)

^bNegative values indicate incremental cost.

Table 6. Projected Incremental Costs and Health Outcomes per 1000 Persons, Based on Rankings by Cost

Strategy ^a	Incremental Cost (95% UR) ^b	Additional TB Patients Diagnosed (95% UR)	Incremental Cost per TB Patient Diagnosed (95% UR)	TB Deaths Averted (95% UR)	Incremental Cost per TB Death Averted (95% UR)	DALYs Averted (95% UR)	Incremental Cost per DALY Averted (95% UR)
Smear as microbiologic test
2. Triage after a negative AFB smear
1A. Triage with AI-based CXR interpretation before AFB smears	\$474 (-\$452 to \$1203)	0.72 (0.0–1.45)	\$654 ^c (-\$2278 to \$4949)	0.36 (0.06–0.52)	\$1329 (-\$2474 to \$6437)	11.1 (1.6–16.3)	\$43 (-\$78 to \$229)
1B. Reinforced follow-up based on triage with AI-based CXR interpretation before AFB smears	\$117 (\$60–\$209)	0.96 (0.47–1.72)	\$122 ^c (\$110–\$136)	0.12 (0.06–0.21)	\$988 (\$900–\$1111)	3.0 (1.5–5.4)	\$39 (\$32–\$46)
I. Status quo: AFB smear	\$4383 (\$3460–\$5339)	Dominated ^d	Dominated ^e	Dominated	Dominated	Dominated	Dominated
Xpert as microbiologic test
3A. Triage with AI-based CXR interpretation before Xpert
3B. Reinforced follow-up based on triage with AI-based CXR interpretation before Xpert	\$87 (\$55–\$202)	0.58 (0.37–1.36)	\$152 (\$136–\$163)	0.08 (0.07–0.44)	\$1052 (\$372–\$1056)	2.2 (1.8–13.3)	\$40 (\$12–\$41)
II. Status quo: Xpert	\$12 550 (\$12 101–\$12 985)	Dominated	Dominated	Dominated	Dominated	Dominated	Dominated

Abbreviations: AFB, acid-fast bacilli; AI, artificial intelligence; CXR, chest radiograph; DALY, disability-adjusted life-year; TB, tuberculosis; UR, uncertainty range.

^aStrategies ranked from cheapest to most expensive, with incremental costs and effectiveness per 1000 persons and incremental cost-effectiveness ratios based on preceding strategy as comparator (Software 1).

^bNegative values indicate cost savings.

^cStrategy 1B dominates 1A by extended dominance for all outcomes. 1B also dominates 2 by extended dominance for deaths and DALYs.

^dA strategy is dominated when it is more expensive and less effective than another.

^eIn this table, each strategy is compared to the preceding strategy; in cases where the preceding strategy was dominated, the comparator becomes the last nondominated strategy.

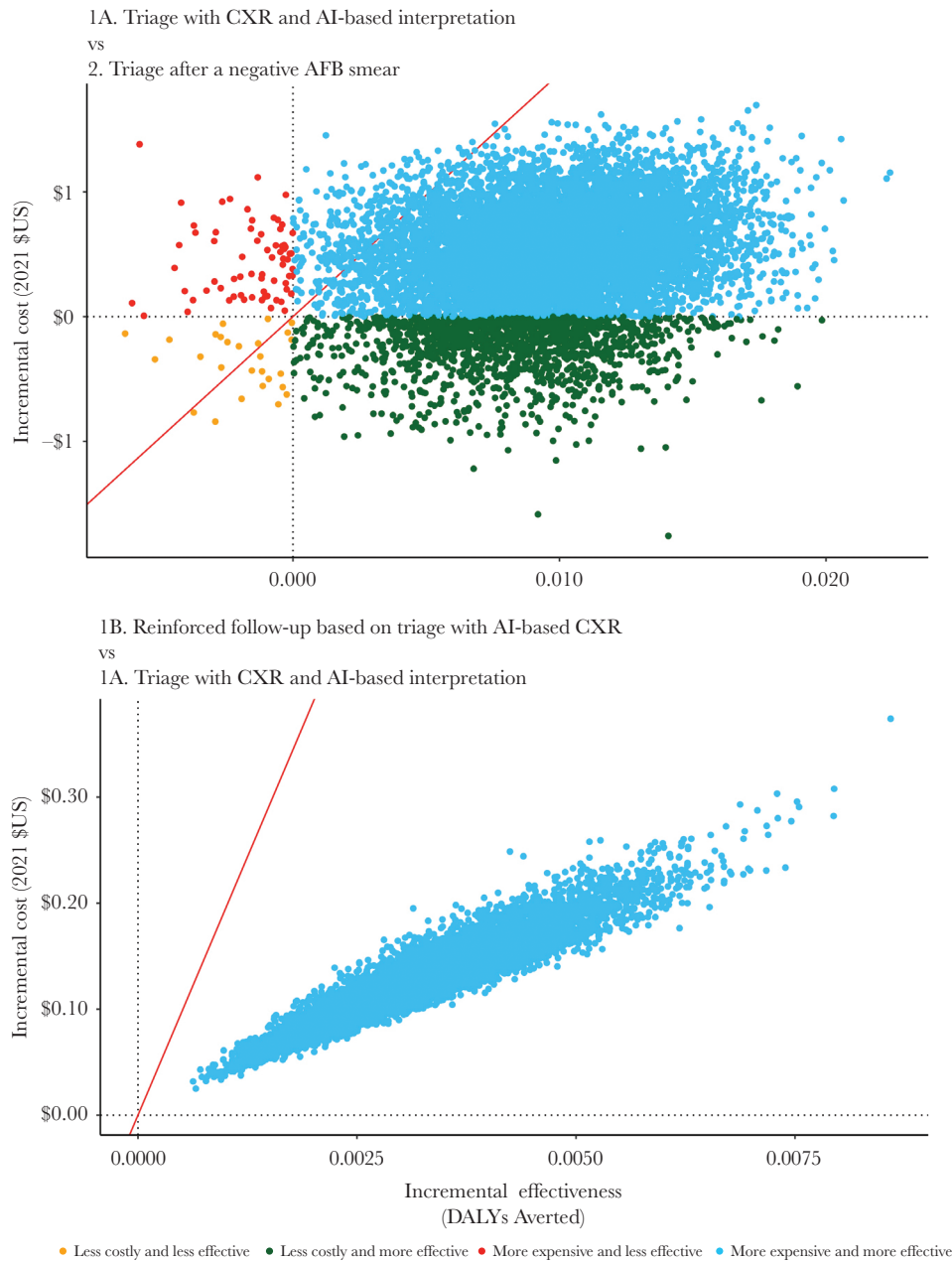


Figure 2. Probabilistic sensitivity analysis for smear-based algorithms—cost-effectiveness planes. Negative values on the y-axis indicate cost savings and the red line corresponds to the willingness-to-pay threshold per disability-adjusted life-year (DALY) averted (\$195/DALY averted). Each point reflects cost and DALY outputs from 1 of the 10 000 model runs. The status quo, upfront smear strategy is not shown as it was consistently dominated by the triage strategies. Abbreviations: AFB, acid-fast bacilli; AI, artificial intelligence; CXR, chest radiograph; DALY, disability-adjusted life-year; \$US, United States dollars.

along with the higher cost of treating drug-resistant TB, means that cost-effectiveness could change in settings where drug resistance is more prevalent.

Another limitation is that our analysis did not address the use of this technology in children. The original study was restricted to adults, and there are no published estimates of the diagnostic accuracy of these software packages in children.

We further assumed 100% adherence to microbiologic testing following positive triage CXRs, as observed in the parent

study. For the “status quo” strategy, we did not have local data for adherence to microbiologic testing, but referred to a similar Indian setting where adherence was 91%. When this parameter was increased to 100% in sensitivity analysis, key results were similar.

Perhaps most importantly, with current test characteristics, the threshold CXR score with the requisite sensitivity (95%) is associated with specificity below the optimal 80% threshold in the target product profile outlined by the WHO [1, 6].

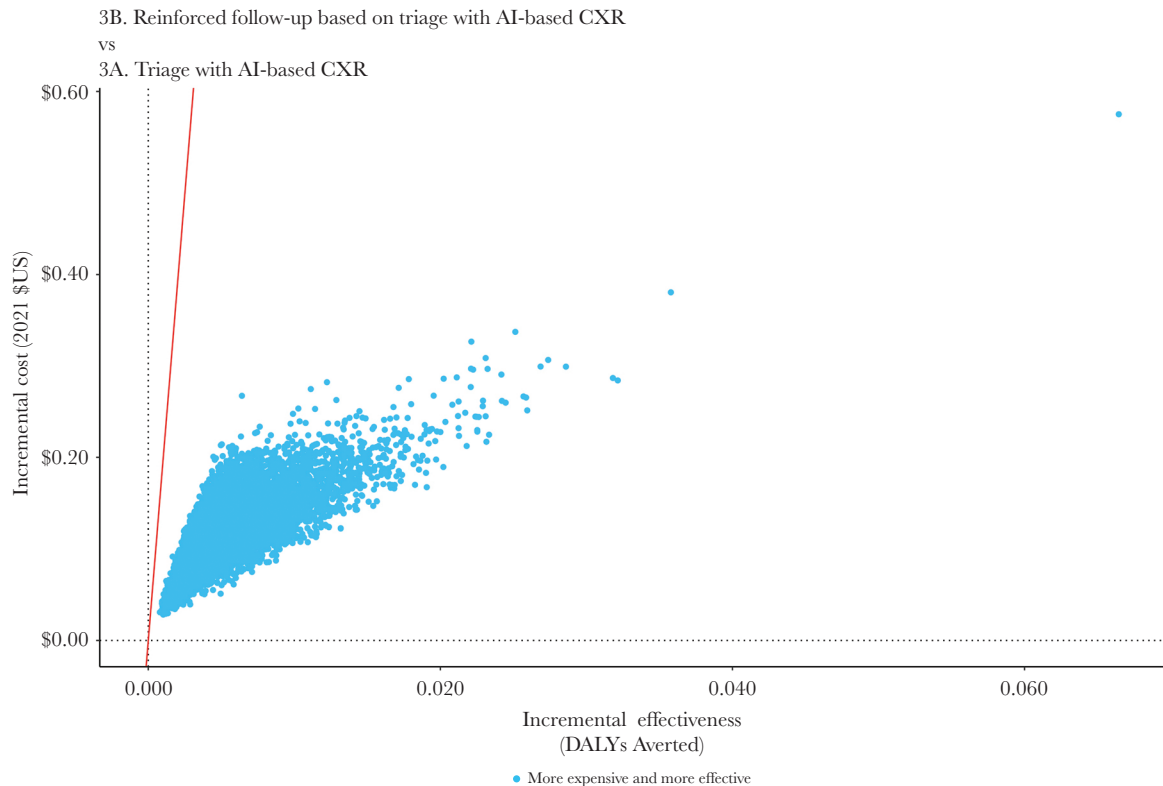


Figure 3. Probabilistic sensitivity analysis for Xpert-based algorithms—cost-effectiveness planes. The red line corresponds to the willingness-to-pay threshold per disability-adjusted life-year (DALY) averted (\$195/DALY averted). Each point reflects cost and DALY outputs from one of the 10 000 model runs. The status quo, upfront smear strategy is not shown as it was consistently dominated by the triage strategies. Abbreviations: AI, artificial intelligence; CXR, chest radiograph; DALY, disability-adjusted life-year; \$US, United States dollars.

This is the first full economic evaluation of AI-based CXR interpretation as triage tests; previous studies focused on diagnostic performance and direct testing costs [7, 9, 38–42]. A retrospective study in Nepal and Cameroon used the same software evaluated here; it suggested a substantial reduction in the need for Xpert testing [43]. In our analysis, upfront Xpert tests represent a major determinant of cost, and hence account for significant savings in the triage strategies.

The 2 AI software packages have also been evaluated as potential TB screening tools in persons who have not sought medical care (ie, active case-finding). To date, there is no economic analysis addressing their use in the screening context. To be cost-effective, these would require better diagnostic performance than for the triage use case, since the prevalence of active TB is substantially lower among unselected individuals [44–47].

Further real-world studies and economic evaluations of AI-based CXR interpretations will be warranted as future releases of these technologies eventually meet the WHO recommendations for optimal diagnostic performance in the triage setting [1]. It is likely that future versions will be even more cost-effective in low- and middle-income countries, if the cost of licensing and reading remains similar or even lower. Currently, cost is linked to volumes; the cost per read decreases with increasing volumes.

In conclusion, the addition of AI-based CXR interpretation to focus microbiologic testing for TB, in settings of low HIV prevalence, can reduce costs and empiric treatment while averting deaths and DALYs. Our study suggests that this technology is highly cost-effective, and supports its use for triage of HIV-uninfected persons presenting with symptoms consistent with TB, in settings with limited availability of radiologists or highly skilled readers.

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.

Notes

Acknowledgments. The authors would like to acknowledge the team at Indus Health Network who provided costing and treatment related parameters used in the model. We also acknowledge the assistance of Dr Jonathon Campbell with analysis insights and guidance.

Financial support. This work was supported by the Canadian Institutes of Health Research (PJT-148743).

Potential conflicts of interest. A. J. K., since October 2019, has financial interests in the company Alcela, of which qure.ai is a client, and he has been in discussions with qure.ai for additional business development. Before October 2019, he was involved with design and initiation of the parent cohort study. However, he was not involved in the analysis, reporting, writing,

editing, or decision to submit the work reported in the present manuscript. All other authors report no potential conflicts of interest.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

REFERENCES

- World Health Organization. High Priority Target Product Profiles for New Tuberculosis Diagnostics: Report of a Consensus Meeting, 28-29 April 2014. Geneva, Switzerland: WHO; 2014. <https://apps.who.int/iris/handle/10665/135617>. Accessed 19 November 2021.
- Sun AY, Pai M, Salje H, et al. Modeling the impact of alternative strategies for rapid molecular diagnosis of tuberculosis in Southeast Asia. *Am J Epidemiol* 2013; 178:1740-9.
- World Health Organization. Rapid communication on systematic screening for tuberculosis. Report No. 978-92-4-001655-2. 2020. <https://www.who.int/publications/i/item/rapid-communication-on-the-systematic-screening-for-tuberculosis>. Accessed 17 September 2021.
- Qin ZZ, Naheyan T, Ruhwald M, et al. A new resource on artificial intelligence powered computer automated detection software products for tuberculosis programmes and implementers. *Tuberculosis (Edinb)* 2021; 127:102049.
- 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.
- 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.
- 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.
- 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.
- Philipsen RH, Sánchez CI, Maduskar P, et al. Automated chest-radiography as a triage for Xpert testing in resource-constrained settings: a prospective study of diagnostic accuracy and costs. *Sci Rep* 2015; 5:12215.
- World Health Organization. Global Tuberculosis Report. Geneva, Switzerland; 2020. https://www.who.int/tb/publications/global_report/en/. Accessed 17 September 2021.
- Qadeer E, Fatima R, Yaqoob A, et al. Population based national tuberculosis prevalence survey among adults (>15 years) in Pakistan, 2010-2011. *PLoS One* 2016; 11:e0148293.
- Grzybowski S. Drugs are not enough. Failure of short-course chemotherapy in a district in India. *Tuber Lung Dis* 1993; 74:145-6.
- Ragonnet R, Flegg JA, Brilleman SL, et al. Revisiting the natural history of pulmonary tuberculosis: a Bayesian estimation of natural recovery and mortality rates. *Clin Infect Dis* 2021; 73:e88-96.
- Tiemersma EW, van der Werf MJ, Borgdorff MW, et al. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PLoS One* 2011; 6:e17601.
- Ismail IM, Kibballi Madhukeshwar A, Naik PR, et al. Magnitude and reasons for gaps in tuberculosis diagnostic testing and treatment initiation: an operational research study from Dakshina Kannada, South India. *J Epidemiol Glob Health* 2020; 10:326-36.
- Grzybowski S, Enarson DA. The fate of cases of pulmonary tuberculosis under various treatment programmes. *Bull IUAT* 1978; 53:70-5.
- Minion J, Brunet L, Pai M. Fluorescent light emitting diode (LED) microscopy for the detection of *Mycobacterium tuberculosis*: a systematic review and meta-analysis. 2009. http://www.stoptb.org/wg/gli/assets/documents/egm%20who%20led%20report%20submitted_sept_01_2009.pdf. Accessed 17 September 2021.
- Harries A. What is the additional yield from repeated sputum examinations by smear microscopy and culture? 2004. <https://apps.who.int/iris/handle/10665/42701>. Accessed 17 September 2021.
- van Cleeff MR, Kivihya-Ndugga LE, Meme H, et al. The role and performance of chest X-ray for the diagnosis of tuberculosis: a cost-effectiveness analysis in Nairobi, Kenya. *BMC Infect Dis* 2005; 5:111.
- van't Hoog AH, Meme HK, Laserson KF, et al. Screening strategies for tuberculosis prevalence surveys: the value of chest radiography and symptoms. *PLoS One* 2012; 7:e38691.
- Wisnivesky JP, Kaplan J, Henschke C, et al. Evaluation of clinical parameters to predict *Mycobacterium tuberculosis* in inpatients. *Arch Intern Med* 2000; 160:2471-6.
- Yan L, Xiao H, Zhang Q. Systematic review: comparison of Xpert MTB/RIF, LAMP and SAT methods for the diagnosis of pulmonary tuberculosis. *Tuberculosis (Edinb)* 2016; 96:75-86.
- Boehme CC, Nabeta P, Hillemann D, et al. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med* 2010; 363:1005-15.
- Salomon JA, Haagsma JA, Davis A, et al. Disability weights for the Global Burden of Disease 2013 study. *Lancet Glob Health* 2015; 3:e712-23.
- World Health Organization. Life Tables by Country—Pakistan. Geneva, Switzerland: WHO; 2019.
- Hussain H, Mori AT, Khan AJ, et al. Correction to: The cost-effectiveness of incentive-based active case finding for tuberculosis (TB) control in the private sector Karachi, Pakistan. *BMC Health Serv Res* 2019; 19:794.
- US Bureau of Labor Statistics. CPI inflation calculator. 2020. https://www.bls.gov/data/inflation_calculator.htm. Accessed 17 September 2021.
- US Department of the Treasury. Fiscal data: treasury reporting rates of exchange. 2020. <https://fiscaldata.treasury.gov/datasets/treasury-reporting-rates-exchange/treasury-reporting-rates-of-exchange>. Accessed 17 September 2021.
- Statistics PBO. CPI Inflation Calculator. Pakistan Bureau of Statistics. 2021. https://www.pbs.gov.pk/sites/default/files/press_releases/2021/cpi_press_release_sep_2021.pdf. Accessed 19 November 2021.
- National TB Control Program Pakistan. National guidelines national TB control program Pakistan, training module doctors. 2019. <https://ntp.gov.pk/resource-center/>. Accessed 17 September 2021.
- Revill P, Ochalek J, Lomas J, et al. Cost-Effectiveness Thresholds: Guiding Health Care Spending for Population Health Improvement. York, UK: Center for Health Economics, University of York; 2015. <https://www.idshealth.org/wp-content/uploads/2015/01/CE-Thresholds-iDSI-Working-Group-Final-Report.pdf>. Accessed 19 November 2021.
- Ochalek J, Lomas J, Claxton K. Estimating health opportunity costs in low-income and middle-income countries: a novel approach and evidence from cross-country data. *BMJ Glob Health* 2018; 3:e000964.
- World Health Organization. *Macroeconomics and Health: Investing in Health for Economic Development: Report of the Commission on Macroeconomics and Health*. Geneva, Switzerland: WHO; 2001:202.
- World Health Organization. The World Health Report 2002: Reducing Risks, Promoting Healthy Life. Geneva, Switzerland: WHO; 2002.
- Houben RMGJ, Lalli M, Kranzer K, et al. What if they don't have tuberculosis? The consequences and trade-offs involved in false-positive diagnoses of tuberculosis. *Clin Infect Dis* 2019; 68:150-6.
- Kendall EA, Kamoga C, Kitonsa PJ, et al. Empiric treatment of pulmonary TB in the Xpert era: correspondence of sputum culture, Xpert MTB/RIF, and clinical diagnoses. *PLoS One* 2019; 14:e0220251.
- McCarthy K, Fielding K, Churchyard GJ, Grant AD. Empiric tuberculosis treatment in South African primary health care facilities—for whom, where, when and why: implications for the development of tuberculosis diagnostic tests. *PLoS One* 2018; 13:e0191608.
- Breuninger M, van Ginneken B, Philipsen RH, et al. Diagnostic accuracy of computer-aided detection of pulmonary tuberculosis in chest radiographs: a validation study from sub-Saharan Africa. *PLoS One* 2014; 9:e106381.
- Melendez J, Sánchez CI, Philipsen RH, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Sci Rep* 2016; 6:25265.
- Muyoyeta M, Maduskar P, Moyo M, 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:e93757.
- Rahman MT, Codlin AJ, Rahman MM, et al. An evaluation of automated chest radiography reading software for tuberculosis screening among public-and private-sector patients. *Eur Respir J* 2017; 49:1602159.
- Zaidi SMA, Habib SS, Van Ginneken B, et al. Evaluation of the diagnostic accuracy of computer-aided detection of tuberculosis on chest radiography among private sector patients in Pakistan. *Sci Rep* 2018; 8:12339.
- Qin ZZ, Sander MS, Rai B, 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.
- Koesoemadinata RC, Kranzer K, Livia R, et al. Computer-assisted chest radiography reading for tuberculosis screening in people living with diabetes mellitus. *Int J Tuberc Lung Dis* 2018; 22:1088-94.
- Melendez J, Hogeweg L, Sánchez 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.
- Melendez J, Philipsen RHHM, Chanda-Kapata P, et al. Automatic versus human reading of chest X-rays in the Zambia National Tuberculosis Prevalence Survey. *Int J Tuberc Lung Dis* 2017; 21:880-6.
- Muyoyeta M, Kasese NC, Milimo D, et al. Digital CXR with computer aided diagnosis versus symptom screen to define presumptive tuberculosis among household contacts and impact on tuberculosis diagnosis. *BMC Infect Dis* 2017; 17:301.