Title: Estimating the number of incorrect tuberculosis diagnoses in low- and middle-income countries.

Authors: Ana van Lieshout Titan BSc [1,2], Prof. Peter J. Dodd PhD [3], Prof. Ted Cohen MD DPH [4], Nicolas A Menzies PhD [1,5].

Affiliations:

1. Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
2. Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands.
3. School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom.
4. Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA.
5. Center for Health Decision Science, Harvard TH Chan School of Public Health, Boston MA, USA.

Corresponding author: Nicolas A. Menzies (nmenzies@hsph.harvard.edu), Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

ABSTRACT

Tuberculosis (TB) is the greatest cause of infectious disease deaths worldwide. In highly-affected countries, effective tuberculosis control requires prompt identification and treatment of individuals with active disease. We examined the performance of tuberculosis case-finding in low- and middle-income counties, based on a comprehensive analysis of tuberculosis diagnosis data reported to WHO. Using these data we estimated the total number of individuals correctly and incorrectly diagnosed with tuberculosis, for 111 countries with a collective 6.8 million tuberculosis notifications in 2023. Here we estimate that in 2023, 2.05 (1.83-2.27) million individuals were incorrectly diagnosed with tuberculosis (false-positive), and 1.00 (0.71-1.36) million received a false-negative diagnosis, at an assumed 25% disease prevalence among individuals evaluated for TB. As many as three out of every ten tuberculosis notifications may not have tuberculosis, and many individuals with tuberculosis receive false-negative diagnoses. As compared to current diagnostic performance, scaling-up new PCR-based diagnostics would substantially reduce under-diagnosis but only produce a small reduction in false-positive diagnoses. Major improvements in TB diagnosis will likely require higher-sensitivity bacteriological tests combined with reduced reliance on clinical diagnosis.

Introduction

Routine facility-based evaluation of individuals with signs and symptoms of tuberculosis (TB) plays a central role in efforts to address TB (1). However, the test most commonly used to diagnose TB in many countries—sputum smear microscopy—has limited sensitivity (2, 3). Several cartridge-based polymerase chain reaction (PCR) tests have recently been developed, and while these rapid diagnostic tests (RDTs) provide substantial improvements in sensitivity compared to smear microscopy they are not yet universally available (4). Both smear microscopy and RDTs are bacteriological tests, which provide the strongest evidence for a TB diagnosis. However, reviews of cohort studies of individuals found to be bacteriologically negative on initial TB evaluation have reported high rates of subsequent TB diagnosis (5). For these reasons, a negative result in initial bacteriological testing does not conclusively exclude TB, and many TB diagnoses are based on clinical evaluation, following a negative bacteriological test result. For 2023, 38% of TB diagnoses were made clinically, without bacteriological confirmation (6). Clinical evaluation can include chest radiography (if available), reported symptoms, and the presence of health conditions or other patient characteristics suggestive of TB (7). Although clinical evaluation will result in additional individuals being diagnosed with TB, studies have not demonstrated high sensitivity and specificity (8). For these reasons, routine TB diagnosis will incorrectly identify some individuals as having TB, and fail to diagnose some individuals with the disease.

Both false-positive and false-negative diagnoses can harm patients. A false-positive diagnosis exposes patients to the health risks and financial burden associated with TB treatment (9), as well as the social and psychological consequences of a major disease diagnosis. False-positive diagnosis also delays treatment of the health condition that led the patient to seek care (10-12). A false-negative diagnosis will delay TB treatment, allowing ongoing lung damage, mortality risks, and transmission (13).

In this study we estimated the performance of TB diagnosis in 2023 for 111 low- and middle-income countries (98% of global incidence). To do so, we synthesized national data on notified TB diagnoses with published evidence on sensitivity and specificity at each step of TB diagnosis, identifying the combination of diagnostic parameters and outcomes consistent with available evidence. For the cohort of individuals evaluated for TB, we re-estimated results for values of initial TB prevalence ranging from 5% to 50%, using 25% as a base-case), based on the distribution of values reported by diagnostic accuracy studies (14, 15). From this analysis we estimate the number of false-positive and false-negative TB diagnoses resulting from current diagnostic approaches. Based on this analysis we explore the potential impact of alternative approaches that could be taken to improve diagnostic outcomes, including further adoption of current RDTs to replace smear microscopy, changes in clinical practices, and further improvements in RDT performance.

Results

The WHO Global TB database includes data for 134 low- and middle-income countries. We excluded 8 countries with missing data on laboratory-confirmed (i.e., bacteriologically positive) or clinically diagnosed cases, and a further 15 countries with <100 pulmonary TB notifications (Table S1). The final analysis included 111 countries, with 4,205,535 laboratory-confirmed cases (62%) and 2,562,902 clinically diagnosed cases (38%). The fraction laboratory-confirmed ranged from 13% to 97% (interquartile range (IQR): 64-84%), with this fraction lower on average for countries with higher TB rates (Fig. 1). Across countries 5% of notifications were HIV-positive (IQR: 1-12%), and 48% were tested using an RDT (IQR: 31-80%).

[Fig. 1]

Diagnostic algorithm performance

Assuming initial TB prevalence (true prevalence of TB amongst people evaluated for TB) of 25%, we estimated an overall algorithm sensitivity (fraction of individuals with TB who receive a TB diagnosis) of 82.6% (95% uncertainty interval: 78.1, 86.6) and specificity (fraction of individuals without TB who receive a TB-negative diagnosis) of 88.0% (85.6, 90.2), when performance was pooled across all countries included in the analysis. The positive predictive value (fraction diagnosed TB-positive who truly have TB) was estimated as 69.7% (66.4, 73.0), and the negative predictive value (fraction diagnosed TB-negative who do not have TB) was 93.8% (92.5, 95.1). Overall, 86.7% (85.3, 87.8) of diagnoses were estimated to be correct. The estimated positive predictive value increased with higher values of initial TB prevalence, from 39.7% (33.9, 45.8) at 5% prevalence to 79.6% (76.5, 82.7) at 50% prevalence. The negative predictive value decreased with higher TB prevalence, from 98.7% (98.5, 99.0) at 5% prevalence to 85.6% (82.1, 88.8) at 50% prevalence (Table 1). Estimates for algorithm sensitivity and specificity also changed with different assumptions for initial TB prevalence, with the inferred performance of clinical diagnosis adjusting to match reported data. Extended Data Table 1 shows global estimates for performance at each step of TB diagnosis (bacteriological testing vs. clinical diagnosis).

[Table 1]

Estimated diagnostic performance varied across world regions, with the highest algorithm sensitivity estimated for the European region (86.2% (81.8, 90.1)), and lowest for the Americas region (74.7% (69.3, 79.7)). This variation arises from differences in the use of clinical diagnosis across regions, as well as differences in RDT coverage. Extended Data Table 2 shows performance estimates by world region, country income level, and for WHO-identified high-TB burden countries.

Total numbers with each diagnostic outcome

Assuming 25% initial TB prevalence, an estimated 22.86 (20.98, 25.00) million individuals were evaluated for TB in 2023, of whom 5.72 (5.25, 6.25) million had TB disease. Of those with TB, an estimated 4.72 (4.49, 4.94) million were correctly diagnosed with TB, and 1.00 (0.71, 1.36) million did not receive a TB diagnosis (false-negative). Of the 17.15 (15.74, 18.75) million without TB, an estimated 2.05 (1.83, 2.27) million were incorrectly diagnosed with TB (false-positive). Overall yield (number diagnosed with TB divided by number evaluated for TB) was estimated as 29.7% (27.1, 32.2). Fig. 2 shows global outcomes for each stage of diagnosis, with clinical diagnosis estimated to be responsible for 22% of all true-positive diagnoses, and 75% of all false-positives). These diagnostic outcomes were sensitive to assumptions about initial TB prevalence (Extended Data Table 3). The estimated number of false-positive diagnoses declined with higher values for initial TB prevalence, ranging from 4.08 (3.67, 4.47) million at 5% TB prevalence to 1.38 (1.17, 1.59) million at 50% TB prevalence. The estimated number of false-negative diagnoses exhibited a non-monotonic relationship with initial TB prevalence, varying from 0.80 (0.57, 1.07) million at 5% TB prevalence, to 1.02 (0.74, 1.36) million at 15% TB prevalence, to 0.82 (0.53, 1.20) million at 50% TB prevalence.

[Fig. 2]

Table 2 shows the number of true-positive, true-negative, false-positive, and false-negative diagnoses by country group, with the number of false-positive and false-negative diagnoses greatest in the South-East Asia region, consistent with the share of overall TB burden in this region. The relative number of false-positive and false-negative diagnoses varied across world regions, from a low of 0.60 (0.40, 0.88) false-positive diagnoses for every false-negative

diagnosis estimated for the Americas region, up to 4.03 (2.52, 6.09) in the Western Pacific region (global average: 2.12 (1.39, 3.09)).

[Table 2]

Approaches for improving diagnostic outcomes

We compared our main analysis with several hypothetical scenarios exploring approaches for improving TB diagnostic outcomes. When we assumed RDT coverage would increase to fully replace smear microscopy (***Full RDT adoption*** scenario), overall algorithm sensitivity increased by 5.2 (2.5, 8.1) percentage points, enabling an additional 0.30 (0.14, 0.48) million individuals to be correctly diagnosed with TB. This change reduced the number of false-negative diagnoses by one third, increased the number of individuals receiving a bacteriologically-confirmed TB diagnosis by 0.89 (0.61, 1.19) million, and produced a small, non-significant reduction in the number of false-positive diagnoses. When we allowed for reduced clinician willingness to diagnose patients clinically (***Reduced clinical diagnosis*** scenario), this increased overall algorithm specificity to 92.5% (90.9, 93.9) and produced large reductions in the number of false-positive diagnoses (0.77 (0.68, 0.86) million). However, this scenario also resulted in an additional 0.23 (0.18, 0.29) million false negative diagnoses. When we allowed for improvements in practices around clinical diagnosis (***Improved clinical algorithms*** scenario), algorithm specificity increased to 94.0% (92.4, 95.3) and false-positive diagnoses dropped by 1.02 (0.88, 1.17) million, with no loss of sensitivity. When we allowed for introduction of improved, more sensitive RDTs, with concomitant reductions in clinical diagnosis (***Improved RDTs, reduced clinical diagnosis***) this produced the greatest increases in algorithm sensitivity and accurate diagnosis, with the number of false-negative and false-positive diagnoses reduced by 0.45 (0.28, 0.64) million (a 45% reduction) and 0.73 (0.45, 1.00) million (a 47% reduction), respectively.

Table 3 reports the number of individuals receiving each diagnostic outcome under these counterfactual scenarios, as compared to the main analysis. Extended Data Tables 4 and 5 report the implications for algorithm sensitivity, specificity, and other measures of diagnostic performance.

[Table 3]

Sensitivity analyses

Fig. S1 shows partial rank correlation coefficients quantifying the sensitivity of results to parameter changes. In these analyses, total false-positive diagnoses was most strongly associated with specificity parameters, with higher specificity associated with lower numbers of false-positive diagnoses. Total false-negative diagnoses had a strong negative association with the sensitivity of clinical diagnosis, and a strong positive relationship with the fraction of culture-negative TB. Extended Data Tables 6 and 7 reports estimated diagnostic outcomes under alternative analytic specifications. When we used published reviews of Xpert MTB-RIF as the source of RDT sensitivity and specificity (vs. Xpert Ultra in the main analysis) results were largely similar, with a small increase in false-negative diagnoses (to 1.14 (0.82, 1.54) million) and a small reduction in false-positive diagnosis (to 1.85 (1.64, 2.06) million). When we re-estimated results assuming higher sensitivity and specificity for clinical diagnosis, the estimated number of false-negative and false-positive diagnoses were both reduced, to 0.72 (0.47, 1.04) and 1.78 (1.52, 2.03) million respectively. When we assumed lower sensitivity and specificity for clinical diagnosis both false-negative and false-positive diagnoses increased (1.30 (0.96, 1.70) and 2.35 (2.14, 2.55) million respectively). When we assumed 25% of individuals never receive an initial bacteriological test, false-positive diagnoses declined slightly (1.90 (1.60, 2.19) million) and false-negative diagnoses increased substantially (2.77 (1.63, 4.52) million). Extended Data Table

7 reports algorithm sensitivity, specificity, positive predictive value, and negative predictive value for these alternative specifications. Fig. 3 shows how the probabilities of false-positive diagnosis and false-negative diagnosis change with different values for initial TB prevalence and the fraction laboratory-confirmed. Extended Data Fig. 1 shows similar results for the three alternative specifications.

Discussion

This study examined the performance of routine TB diagnosis in low- and middle-income countries. Assuming 25% TB prevalence among individuals evaluated for TB, we estimated average algorithm sensitivity to be approximately 80% and specificity approximately 90%, such that individuals evaluated for TB had a one-in-eight chance of receiving an incorrect diagnosis. For those with TB, these results imply that one million could have received a false-negative diagnosis in 2023. For those without TB we estimated that as many as two million could have received an incorrect diagnosis of TB. If correct, these results imply that as many as three out of every ten individuals diagnosed with TB may not have TB. While numbers of false-positive and false-negative diagnoses varied across settings, most settings were estimated to have at least as many false-positive diagnoses as false-negative, and particularly so in the high-incidence settings representing the majority of global TB cases.

False-negative diagnoses stem from inadequate algorithm sensitivity. While novel RDTs have better sensitivity than smear microscopy (14), they will still give false-negative results for some patients, and coverage is not universal (6). For these reasons clinical diagnosis still plays a major role, and in our analysis was responsible for one-fifth of all true-positive diagnoses. The average algorithm sensitivity estimated in this analysis is generally consistent with studies of TB care cascades in high-burden settings, which have identified diagnosis as a key point at which individuals with TB are lost from the cascade (16-18). This large number of missed diagnoses,

and their negative consequences (ongoing morbidity, mortality, and transmission) have motivated major investments in TB diagnostics over the last 20 years.

False-positive TB diagnosis has received substantially less attention. Most studies that have examined the potential numbers incorrectly diagnosed with TB have focused on active case-finding interventions, for which starting TB prevalence is typically low (19-21), or the risks posed by low-specificity serological tests (22, 23). However, one study has estimated the positive predictive value of TB diagnosis in India could be as low as 62% (24), consistent with the results of our analysis.

The comparatively low attention paid to false-positive diagnosis could stem from more minor health consequences being attributed to these diagnostic errors, compared to false-negative diagnoses. However, while the harms associated with false-negative diagnosis are well understood, little is known about the health consequences of false-positive diagnosis (7). Studies that have examined the relative weight placed on false-positive and false-negative diagnoses have found clinicians to consider false-negative diagnosis approximately twice as harmful as false-positive diagnosis, and have argued this ratio should be substantially higher (25, 26). These studies considered only the treatment costs of false-positive diagnosis, and the risk of treatment side effects. Many economic evaluations of TB diagnostics have taken the same approach (27, 28), or have just considered the economic implications. However, false-positive diagnosis can also lead to harms associated with delayed treatment of the health condition causing the individual's symptoms, patient-incurred costs of unnecessary TB treatment (29), and the emotional toll and stigma attached to a TB diagnosis. Moreover, the side-effects of current TB treatment regimens are neither trivial nor rare, even for first-line regimens (30). A false-positive diagnosis may also trigger unnecessary services to identify and treat TB infection among household members. While some false-positive diagnoses will result from transient respiratory infections, others will reflect more serious infections (e.g., pneumonia), or

progressive, high-morbidity conditions such as lung cancer, heart failure, and COPD (31). In an analysis of individuals incorrectly diagnosed with TB in Brazil, estimated mortality was 2-3 times higher than for individuals with true-positive TB over the 2 years following diagnosis, with lung cancer and non-TB respiratory disorders being the most common causes (32). While lung cancer likely represents a minority of false-positive diagnoses, it is clear that at least some individuals incorrectly diagnosed with TB will face major health consequences.

As part of our study we explored the potential impact of approaches that could be taken to improve TB diagnosis. Full adoption of currently-available RDTs is likely the most immediately actionable of the counterfactual scenarios, with many countries making concerted efforts to increase RDT coverage. We found that full adoption of RDTs could reduce the number of false-negative diagnoses by one-third. The absolute increase in algorithm sensitivity in this scenario (5.2%) is smaller than the increase in sensitivity estimated for RDTs when compared to smear microscopy in diagnostic trials (14), illustrating the important role currently played by clinical diagnosis in identifying TB cases previously missed by smear microscopy (33). There are many challenges to achieving high RDT coverage, but these results highlight the benefits that would be realized with higher coverage, which could both reduce TB mortality and shorten the duration of infectiousness for individuals who would otherwise receive a false-negative diagnosis. While full RDT adoption was also projected to produce improvements in algorithm specificity, the absolute number of false-positive diagnoses only dropped by a small amount. This decrease was driven by reductions in clinical diagnosis following a negative RDT result (as compared to smear microscopy) (34), and had substantial uncertainty. As the majority of false-positive diagnoses result from clinical diagnosis, a more direct approach to addressing this issue would be to increase the level of clinical suspicion required to diagnose TB. However, while we found that this approach could potentially produce substantial reductions in false-positive diagnosis (0.8 million), it would do so at the cost of additional false-negative TB diagnoses (0.2 million). These additional missed TB diagnoses would likely represent substantial increases in TB morbidity and

death, suggesting that efforts to discourage clinical diagnosis could have harmful consequences that do not outweigh the benefits. In contrast, efforts to identify improved clinical diagnostic algorithms (potentially including greater access to non-bacteriological diagnostics such as chest radiography) could improve algorithm specificity while protecting sensitivity. In a final scenario, we estimated the potential impact of improved RDTs (to achieve the same sensitivity as culture) with concomitant changes in clinical diagnosis. Under this scenario, both false-negatives and false-positives were substantially reduced. Beyond full adoption of current RDTs, this scenario may represent the best target for future innovation in TB diagnostics, as the development of high-sensitivity diagnostics reduces the need for clinical diagnosis to catch those missed by the initial test. Such behavior change in response to higher sensitivity diagnostics has already been observed for currently-available RDTs (34). While not examined in this analysis, improvements in diagnostics for other conditions with a similar clinical presentation could also play a role in reducing false-positive TB diagnosis, by reducing the pool of individuals with unexplained TB-like symptoms.

This study has several limitations. First, there is little evidence on true TB prevalence among individuals evaluated for TB. This is an important input to our analysis, and likely varies across country settings. For this reason, we estimated results for a range of prevalence values, and even with high initial prevalence the number of false-positive diagnoses was still substantial. Moreover, initial TB prevalence in many settings could be lower than the 25% used on our main analysis, and previous studies have assumed values between 10–20% (22, 35). In South Africa, RDT-positivity has averaged 9% since 2011, and has fallen progressively over this period (36). If true TB prevalence were 10%, the number of false-positive diagnoses would be higher than estimated in our main analysis. Additionally, as we calculate our results for fixed values of TB prevalence, the reported uncertainty intervals do not include this source of uncertainty. Second, while we estimated the performance of clinical diagnosis from a range of studies, most were from high HIV-prevalence settings, and we had limited ability to consider the variation in

289 performance that likely exists across settings. While we adjusted the results of these studies to
290 account for potential misclassification of culture-negative TB, the limitations of culture as a
291 reference standard adds uncertainty to our estimates. Third, there is limited empirical evidence
292 to validate one of our main findings—the potentially large number of false-positive diagnoses
293 produced by current diagnostic approaches. In part this should be expected – if most false-
294 positive individuals have self-resolving conditions, they would improve on TB treatment similar
295 to individuals with TB, with initially incorrect diagnoses unlikely to be revisited. In addition,
296 several studies have reported on diagnostic practices in routine settings that increase the risks
297 of false-positive diagnosis, with sensitivity prioritized over specificity (37, 38). Perhaps the best
298 supporting evidence is the multiple studies showing a substantial fraction of clinically-
299 diagnosed individuals to be negative when tested with culture, both from the pre-Xpert era (8)
300 and during Xpert roll-out (27, 39). For example, of 139 individuals treated clinically following a
301 negative Xpert result in the TB-NEAT trial, only 31 (22%) were culture-positive (39). While culture
302 has limitations for routine TB diagnosis it should identify the large majority of adults with
303 pulmonary TB, so the high fraction of clinically-diagnosed culture-negative TB cases in these
304 empirical studies supports our findings. Under an alternative model specification that reduced
305 the sensitivity of initial diagnostic tests (assuming a greater fraction of symptomatic TB is
306 bacteriologically-negative) and made optimistic assumptions about the ability of clinical
307 diagnosis to identify these bacteriologically-negative cases, the number of false-positive
308 diagnoses was substantially reduced, but still greater than the number of false-negative
309 diagnoses. Fourth, we conceptualized TB diagnosis as a single event, yet many individuals with
310 TB make repeated diagnostic attempts before being diagnosed correctly (40). In our analysis,
311 these multiple attempts serve to increase initial TB prevalence, by inflating the number of times
312 an individual with TB gets assessed. As noted above, our results are robust to alternative
313 assumptions about initial TB prevalence, and may be conservative on this point. Fifth, we used
314 the sensitivity and specificity reported for Xpert Ultra to represent all RDTs, even though several

RDTs are now available (41). Country-reported notifications data do not record the mix of RDTs used, and we chose Xpert Ultra to represent this class of diagnostic given the substantial evidence available on its performance and its increasing use across high-burden countries. Similarly, we did not include sputum culture in the algorithms assessed in our study, given its limited use in many settings. Sixth, we did not consider age in our analysis, as several of the required variables were not stratified by age. Diagnosis of TB in children (<10% of all notifications) shares many of the challenges of adult TB diagnosis, though these challenges are magnified, with poorer sensitivity of available bacteriological tests and difficult sample collection. Seventh, our assumptions about bacteriological test sensitivity and specificity were based on data collected under research conditions. While we adjusted test sensitivity downwards to allow for culture-negative pulmonary TB, there is evidence of lower sensitivity and specificity in routine healthcare, and reporting gaps could affect the communication of laboratory results (42-45). Finally, our estimates don't consider TB diagnoses not captured by routine reporting data. While there have been major efforts over the past decade to address under-reporting, there will still be some individuals diagnosed with TB that are not included in available data. As under-reporting of TB is most common with informal and private providers, there is little reason to believe that the performance of TB diagnosis for these individuals would be better than estimated in this analysis.

The results of this analysis have several implications. Given the potentially large number of individuals receiving false-positive diagnoses, collection of empirical evidence to corroborate this finding is urgently needed, in addition to research examining the health consequences for these individuals, which are poorly understood. While potentially less surprising, the 1 million individuals estimated to receive a false-negative diagnosis is also notable, and together with the number of false-positive diagnoses provides a strong reminder of the deficiencies of current diagnostic approaches. While much progress has been made over the past 15 years, better ways to diagnose TB are urgently needed. This includes the development of more accurate RDTs that

can be used across a wider range of samples and clinical settings, and concomitant scale up of RDTs to replace lower sensitivity smear microscopy. These changes would reduce the need for clinical diagnosis, which was responsible for most false-positive diagnoses in our analysis. Beyond patient care, these findings raise questions about how to track trends in TB incidence and mortality. For many countries, epidemiological estimates depend on country-reported notifications data, assuming these notifications represents true TB disease cases (4). In settings with more false-positive diagnoses, failing to account for this could distort epidemiological estimates (7). Moreover, programmatic initiatives to improve TB case detection will also need to avoid creating incentives for overly inclusive diagnostic approaches, which could increase false-positive diagnosis.

Diagnosis is one of several steps in the TB care cascade, and major challenges have been documented at other parts of this cascade (40, 46, 47). However, our results reinforce the critical challenges faced to diagnose TB, and highlight the importance of achieving good diagnostic outcomes for all individuals evaluated for TB, including both individuals with and without TB.

Acknowledgements

This work was supported by the National Institute of Allergy And Infectious Diseases of the National Institutes of Health (Award Number U01AI152084 to NAM). The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank Nim Arinaminpathy for feedback on an earlier version of this manuscript.

Author Contributions Statement

AvLT, PD, TC and NAM conceptualized the research. AvLT, TC, and NAM developed the methodology used, and PD, NAM, and TC validated the methodology. The research was supervised by NAM, and coordinated by NAM. AvLT curated data, performed the formal analysis, did the investigation, and administered the project. AvLT and NAM made the visualisations. NAM did the funding acquisition, and resources were provided by NAM. The original draft was written by AvLT and NAM, and reviewed, edited, and approved by all authors. AvLT and NAM have full access to all data in the study. All authors read and approved the final version of the article.

Competing Interests Statement

We declare no competing interests.

TABLES

Fraction of cohort with TB	Algorithm sensitivity (%)	Algorithm specificity (%)	Positive predictive value (%)[&]	Negative predictive value (%)[#]	Received correct diagnosis (%)
5%	77.2 (72.5, 81.4)	93.8 (92.0, 95.2)	39.7 (33.9, 45.8)	98.7 (98.5, 99.0)	92.9 (91.3, 94.3)
10%	79.0 (74.5, 83.1)	92.4 (90.6, 93.9)	53.8 (48.7, 58.7)	97.5 (97.0, 98.0)	91.1 (89.6, 92.3)
15%	80.4 (75.8, 84.4)	91.0 (89.1, 92.7)	61.3 (57.1, 65.4)	96.3 (95.6, 97.0)	89.4 (88.0, 90.6)
20%	81.6 (77.0, 85.6)	89.6 (87.5, 91.5)	66.2 (62.5, 69.8)	95.1 (94.0, 96.1)	88.0 (86.6, 89.1)
25%	82.6 (78.1, 86.6)	88.0 (85.6, 90.2)	69.7 (66.4, 73.0)	93.8 (92.5, 95.1)	86.7 (85.3, 87.8)
30%	83.6 (79.0, 87.5)	86.3 (83.6, 88.9)	72.4 (69.3, 75.6)	92.5 (90.7, 94.0)	85.5 (84.1, 86.7)
35%	84.5 (79.8, 88.4)	84.5 (81.3, 87.5)	74.6 (71.6, 77.8)	91.0 (88.9, 92.9)	84.5 (83.1, 85.8)
40%	85.3 (80.6, 89.2)	82.5 (78.8, 86.0)	76.5 (73.5, 79.6)	89.4 (86.9, 91.7)	83.6 (82.1, 85.0)
45%	86.1 (81.4, 90.0)	80.2 (75.8, 84.4)	78.1 (75.1, 81.2)	87.6 (84.6, 90.4)	82.9 (81.2, 84.4)
50%	86.9 (82.2, 90.7)	77.6 (72.4, 82.6)	79.6 (76.5, 82.7)	85.6 (82.1, 88.8)	82.3 (80.4, 83.9)

Table 1: Global average estimates for the sensitivity, specificity, positive predictive value, and negative predictive value for different values of TB prevalence among individuals evaluated for TB, based on data reported through routine notifications systems.

[&] Positive predictive value represents the fraction of individuals identified as having TB (either bacteriologically-confirmed or clinically diagnosed) who truly have TB. [#] Negative predictive value represents the fraction of individuals identified as not having TB who truly do not have TB. Values in parentheses indicate 95% uncertainty intervals.

	Cohort evaluated for TB (mil.)		Numbers of individuals with each diagnostic outcome (mil.)			
	Number evaluated for TB	Number with TB	True-positive diagnoses	False-negative diagnoses	False-positive diagnoses	True-negative diagnoses
Global	22.86 (20.98, 25.00)	5.72 (5.25, 6.25)	4.72 (4.49, 4.94)	1.00 (0.71, 1.36)	2.05 (1.83, 2.27)	15.09 (13.48, 16.90)
WHO region						
Eastern mediterranean	6.02 (5.52, 6.58)	1.50 (1.38, 1.65)	1.21 (1.15, 1.27)	0.29 (0.21, 0.40)	0.42 (0.37, 0.48)	4.09 (3.67, 4.56)
Europe	0.96 (0.88, 1.05)	0.24 (0.22, 0.26)	0.18 (0.17, 0.19)	0.06 (0.05, 0.08)	0.04 (0.03, 0.04)	0.68 (0.62, 0.76)
Africa	1.48 (1.35, 1.64)	0.37 (0.34, 0.41)	0.31 (0.30, 0.33)	0.06 (0.04, 0.08)	0.17 (0.16, 0.19)	0.94 (0.82, 1.07)
Americas	0.35 (0.32, 0.39)	0.09 (0.08, 0.10)	0.08 (0.07, 0.08)	0.01 (0.01, 0.02)	0.03 (0.03, 0.04)	0.23 (0.21, 0.26)
South-East Asia	10.36 (9.38, 11.50)	2.59 (2.34, 2.87)	2.13 (2.02, 2.25)	0.45 (0.30, 0.65)	0.92 (0.81, 1.03)	6.85 (6.01, 7.80)
Western Pacific	3.69 (3.39, 4.04)	0.92 (0.85, 1.01)	0.80 (0.76, 0.85)	0.12 (0.08, 0.17)	0.46 (0.42, 0.51)	2.31 (2.04, 2.61)
Income level[$]						
Low-income	2.94 (2.67, 3.26)	0.74 (0.67, 0.82)	0.58 (0.55, 0.61)	0.16 (0.11, 0.22)	0.22 (0.19, 0.24)	1.99 (1.76, 2.25)
Lower middle income	14.03 (12.82, 15.45)	3.51 (3.20, 3.86)	2.88 (2.74, 3.02)	0.63 (0.43, 0.87)	1.25 (1.11, 1.40)	9.27 (8.23, 10.45)
Upper middle income	5.89 (5.42, 6.44)	1.47 (1.36, 1.61)	1.26 (1.19, 1.32)	0.21 (0.15, 0.30)	0.58 (0.52, 0.64)	3.84 (3.43, 4.31)
High-TB burden*	20.04 (18.38, 21.92)	5.01 (4.60, 5.48)	4.18 (3.98, 4.38)	0.83 (0.58, 1.15)	1.91 (1.70, 2.11)	13.12 (11.69, 14.72)

Table 2: Estimated number of individuals receiving true-positive, true-negative, false-positive, and false-negative diagnoses in 2023, by world region, country income level, and high-TB burden classification.

mil. = million. [$] Income level groups based on World Bank country income classification. * High-TB burden countries represent 30 countries identified as high TB burden by WHO. Values in parentheses indicate 95% uncertainty intervals.

Scenario	Numbers of individuals with each diagnostic outcome (millions)					
	True-positive diagnoses	False-negative diagnoses	False-positive diagnoses	True-negative diagnoses	Bact.-confirmed diagnosis	Incorrect diagnosis
Main analysis	4.72 (4.49, 4.94)	1.00 (0.71, 1.36)	2.05 (1.83, 2.27)	15.09 (13.48, 16.90)	4.21 (4.19, 4.22)	3.05 (2.81, 3.34)
Counterfactual Scenario 1: Full RDT adoption	5.02 (4.71, 5.35)	0.70 (0.44, 1.03)	1.95 (1.50, 2.44)	15.19 (13.55, 17.01)	5.09 (4.81, 5.39)	2.65 (2.22, 3.12)
Counterfactual Scenario 2: Reduced clinical diagnosis	4.49 (4.28, 4.69)	1.22 (0.89, 1.64)	1.28 (1.11, 1.46)	15.87 (14.33, 17.58)	4.21 (4.19, 4.22)	2.50 (2.20, 2.88)
Counterfactual Scenario 3: Improved clinical algorithms	4.72 (4.49, 4.94)	1.00 (0.71, 1.36)	1.03 (0.85, 1.23)	16.12 (14.60, 17.82)	4.21 (4.19, 4.22)	2.03 (1.71, 2.40)
Counterfactual Scenario 4: Improved RDTs, reduced clinical diagnosis	5.16 (4.84, 5.50)	0.55 (0.32, 0.87)	1.32 (1.02, 1.65)	15.83 (14.28, 17.54)	5.54 (5.23, 5.88)	1.87 (1.51, 2.28)

Table 3: Diagnostic outcomes under hypothetical scenarios for improving TB diagnosis, compared to the main analysis.

RDT = WHO-approved rapid diagnostic test. Bact. = bacteriological. Values in parentheses indicate 95% uncertainty intervals. Scenario 1 (Full RDT adoption) represents 100% adoption of Xpert Ultra to replace smear microscopy in each modelled country. Scenario 2 (Reduced clinical diagnosis) assumes an increase in the specificity of clinical diagnosis to reduce the false-positive rate (1-specificity) of this diagnostic step by 50%, with a matching reduction in sensitivity consistent with the main analysis ROC curve shown in Fig. S2. Scenario 3 (Improved clinical algorithms) assumes improvements in practices around clinical diagnosis that allow the specificity of this diagnostic step to improve to reach the optimistic ROC curve shown on Fig. S2, with no loss of sensitivity. Scenario 4 (Improved RDTs, reduced clinical diagnosis) assumes the development and full adoption of improved RDTs with sensitivity equivalent to culture, and a change in clinical diagnosis practices matching the assumptions of Scenario 2 (increases specificity, reduced sensitivity).

FIGURE LEGENDS/CAPTIONS

Fig. 1: Number of laboratory-confirmed and clinically-diagnosed TB notifications per 100,000 for each low- and middle-income country.

'Lab-confirmed' notifications represent the sum of pulmonary TB cases bacteriologically-confirmed via smear microscopy, culture, or WHO-approved rapid diagnostic test. 'Clinically-diagnosed' notifications represent the sum of pulmonary TB cases that were not bacteriologically confirmed but diagnosed with TB by a clinician or other medical practitioner who has decided to give the patient a full course of TB treatment. Size of plotting symbols indicates total number of TB notifications for each country in 2023. Plot excludes one country that reported zero clinically-diagnosed notifications for 2023.

Fig. 2: Estimated global number of individuals receiving true-positive, true-negative, false-positive, and false-negative diagnoses, among individuals evaluated for TB disease in 2023.

Analysis assumes 25% TB prevalence among individuals evaluated for TB. Values in parentheses indicate 95% uncertainty intervals. * Values represent reported notifications data for 2023.

Fig. 3: Estimates of the probability of false-positive diagnosis (Panel A) and false-negative diagnosis (Panel B) for different values of initial TB prevalence and the percentage of notifications that are laboratory confirmed.

Probability of false-positive diagnosis defined as the probability that someone diagnosed with TB does not have TB (1 – PPV). Probability of false-negative diagnosis defined as the probability that someone diagnosed as not having TB does have TB (1 – NPV). Colors indicate different probability levels, indicated by values shown in each panel. of All inputs apart from the sensitivity and specificity of clinical diagnosis held at their global average values. Sensitivity and specificity of clinical diagnosis calculated as a function of other values, based on the ROC curve shown in Fig. S2. '+' symbol in center of each plot represents mean values from the main analysis.

References

1. World Health Organization. The End TB strategy. Geneva, Switzerland; 2015.
2. Steingart KR, Henry M, Ng V, Hopewell PC, Ramsay A, Cunningham J, et al. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. The Lancet infectious diseases. 2006;6(9):570-81.
3. Cattamanchi A, Dowdy DW, Davis JL, Worodria W, Yoo S, Joloba M, et al. Sensitivity of direct versus concentrated sputum smear microscopy in HIV-infected patients suspected of having pulmonary tuberculosis. BMC infectious diseases. 2009;9:1-9.
4. World Health Organization. Global TB Report 2024. Geneva2024.
5. Sossen B, Richards AS, Heinsohn T, Frascella B, Balzarini F, Oradini-Alacreu A, et al. The natural history of untreated pulmonary tuberculosis in adults: a systematic review and meta-analysis. Lancet Respir Med. 2023;11(4):367-79.
6. WHO Global TB Database [http://www.who.int/tb/country/data/download/en/] [Internet]. WHO Global TB Programme. 2023.
7. Houben RMGJ, Lalli M, Kranzer K, Menzies NA, Schumacher SG, Dowdy DW. What if They Don't Have Tuberculosis? The Consequences and Trade-offs Involved in False-positive Diagnoses of Tuberculosis. Clinical Infectious Diseases. 2018;68(1):150-6.
8. Walusimbi S, Bwanga F, De Costa A, Haile M, Joloba M, Hoffner S. Meta-analysis to compare the accuracy of GeneXpert, MODS and the WHO 2007 algorithm for diagnosis of smear-negative pulmonary tuberculosis. BMC Infect Dis. 2013;13:507.
9. World Health Organization. National surveys of costs faced by tuberculosis patients and their households 2015-2021: World Health Organization; 2022.
10. Laurence YV, Griffiths UK, Vassall A. Costs to Health Services and the Patient of Treating Tuberculosis: A Systematic Literature Review. PharmacoEconomics. 2015;33(9):939-55.
11. Tostmann A, Boeree MJ, Aarnoutse RE, De Lange WCM, Van Der Ven AJAM, Dekhuijzen R. Antituberculosis drug-induced hepatotoxicity: Concise up-to-date review. Journal of Gastroenterology and Hepatology. 2008;23(2):192-202.
12. Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A, Volmink J. Patient Adherence to Tuberculosis Treatment: A Systematic Review of Qualitative Research. PLOS Medicine. 2007;4(7):e238.

13. Moreira J, Bisig B, Muwawenimana P, Basinga P, Bisoffi Z, Haegeman F, et al. Weighing Harm in Therapeutic Decisions of Smear-Negative Pulmonary Tuberculosis. Medical Decision Making. 2009;29.

14. Horne DJ, Kohli M, Zifodya JS, Schiller I, Dendukuri N, Tollefson D, et al. Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Database of Systematic Reviews. 2019(6).

15. Zifodya JS, Kreniske JS, Schiller I, Kohli M, Dendukuri N, Schumacher SG, et al. Xpert Ultra versus Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults with presumptive pulmonary tuberculosis. Cochrane Database of Systematic Reviews. 2021(2).

16. Subbaraman R, Nathavitharana RR, Satyanarayana S, Pai M, Thomas BE, Chadha VK, et al. The Tuberculosis Cascade of Care in India's Public Sector: A Systematic Review and Meta-analysis. PLoS Med. 2016;13(10):e1002149.

17. Naidoo P, Theron G, Rangaka MX, Chihota VN, Vaughan L, Brey ZO, et al. The South African Tuberculosis Care Cascade: Estimated Losses and Methodological Challenges. J Infect Dis. 2017;216(suppl_7):S702-s13.

18. Emani S, Alves K, Alves LC, da Silva DA, Oliveira PB, Castro MC, et al. Quantifying gaps in the tuberculosis care cascade in Brazil: A mathematical model study using national program data. PLoS Med. 2024;21(3):e1004361.

19. Lalli M, Hamilton M, Pretorius C, Pedrazzoli D, White RG, Houben R. Investigating the impact of TB case-detection strategies and the consequences of false positive diagnosis through mathematical modelling. BMC Infect Dis. 2018;18(1):340.

20. Cilloni L, Kranzer K, Stagg HR, Arinaminpathy N. Trade-offs between cost and accuracy in active case finding for tuberculosis: A dynamic modelling analysis. PLoS Med. 2020;17(12):e1003456.

21. Chadha VK, Praseeja P. Active tuberculosis case finding in India - The way forward. Indian J Tuberc. 2019;66(1):170-7.

22. Dowdy DW, Steingart KR, Pai M. Serological testing versus other strategies for diagnosis of active tuberculosis in India: a cost-effectiveness analysis. PLoS Med. 2011;8(8):e1001074.

23. Steingart KR, Flores LL, Dendukuri N, Schiller I, Laal S, Ramsay A, et al. Commercial serological tests for the diagnosis of active pulmonary and extrapulmonary tuberculosis: an updated systematic review and meta-analysis. PLoS medicine. 2011;8(8):e1001062.

24. Shewade HD, Satyanarayana S, Kumar AM. Does active case finding for tuberculosis generate more false-positives compared to passive case finding in India? Indian J Tuberc. 2021;68(3):396-9.

25. Moreira J, Bisig B, Muwawenimana P, Basinga P, Bisoffi Z, Haegeman F, et al. Weighing harm in therapeutic decisions of smear-negative pulmonary tuberculosis. Med Decis Making. 2009;29(3):380-90.

26. Basinga P, Moreira J, Bisoffi Z, Bisig B, Van den Ende J. Why are clinicians reluctant to treat smear-negative tuberculosis? An inquiry about treatment thresholds in Rwanda. Med Decis Making. 2007;27(1):53-60.

27. Vassall A, van Kampen S, Sohn H, Michael JS, John KR, den Boon S, et al. Rapid diagnosis of tuberculosis with the Xpert MTB/RIF assay in high burden countries: a cost-effectiveness analysis. PLOS Med. 2011;8(11):e1001120-e.

28. Menzies NA, Cohen T, Lin HH, Murray M, Salomon JA. Population health impact and cost-effectiveness of tuberculosis diagnosis with Xpert MTB/RIF: a dynamic simulation and economic evaluation. PLoS Medicine. 2012;9(11):e1001347-e.

29. Portnoy A, Yamanaka T, Nguhiu P, Nishikiori N, Garcia Baena I, Floyd K, et al. Costs incurred by people receiving tuberculosis treatment in low-income and middle-income countries: a meta-regression analysis. Lancet Global Health. 2023;11(10):e1640-e7.

30. Burman W, Rucsineanu O, Horsburgh CR, Johnston J, Dorman SE, Menzies D. Research on the treatment of rifampin-susceptible tuberculosis-Time for a new approach. PLoS Med. 2024;21(7):e1004438.

31. Jayasooriya S, Dimambro-Denson F, Beecroft C, Balen J, Awokola B, Mitchell C, et al. Patients with presumed tuberculosis in sub-Saharan Africa that are not diagnosed with tuberculosis: a systematic review and meta-analysis. Thorax. 2023;78(1):50-60.

32. Thompson RR, Kim S, Pelissari DM, Harada LO, Júnior J, Oliveira PB, et al. High mortality rates among individuals misdiagnosed with tuberculosis: a matched retrospective cohort study of individuals diagnosed with tuberculosis in Brazil. J Infect Dis. 2024.

33. Theron G, Peter J, Dowdy D, Langley I, Squire SB, Dheda K. Do high rates of empirical treatment undermine the potential effect of new diagnostic tests for tuberculosis in high-burden settings? Lancet Infect Dis. 2014.

34. Kim S, Can MH, Agizew TB, Auld AF, Balcells ME, Bjerrum S, et al. Factors associated with tuberculosis treatment initiation among bacteriologically negative individuals evaluated for tuberculosis: An individual patient data meta-analysis. PLoS Med. 2025;22(1):e1004502.

35. Keeler E, Perkins MD, Small P, Hanson C, Reed S, Cunningham J, et al. Reducing the global burden of tuberculosis: the contribution of improved diagnostics. Nature. 2006;444 Suppl 1:49-57.

36. da Silva MP, Cassim N, Ndlovu S, Marokane PS, Radebe M, Shapiro A, et al. More Than a Decade of GeneXpert(®)Mycobacterium tuberculosis/Rifampicin (Ultra) Testing in South Africa: Laboratory Insights from Twenty-Three Million Tests. Diagnostics (Basel). 2023;13(20).

37. McDowell A, Pai M. Treatment as diagnosis and diagnosis as treatment: empirical management of presumptive tuberculosis in India. Int J Tuberc Lung Dis. 2016;20(4):536-43.

38. Mesfin MM, Tasew TW, Richard MJ. The quality of tuberculosis diagnosis in districts of Tigray region of northern Ethiopia. Ethiopian Journal of Health Development. 2005;19(I):13.

39. Theron G, Zijenah L, Chanda D, Clowes P, Rachow A, Lesosky M, et al. Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. Lancet. 2014;383(9915):424-35.

40. Hanson C, Osberg M, Brown J, Durham G, Chin DP. Finding the Missing Patients With Tuberculosis: Lessons Learned From Patient-Pathway Analyses in 5 Countries. J Infect Dis. 2017;216(suppl_7):S686-s95.

41. World Health Organization. WHO consolidated guidelines on tuberculosis. Module 3: diagnosis–rapid diagnostics for tuberculosis detection: World Health Organization; 2024.

42. Nsawotebba A, Ibanda I, Mujuni D, Nabadda S, Nadunga D, Kabugo J, et al. Impact of randomized blinded rechecking program on the performance of the AFB Microscopy Laboratory Network in Uganda: a decadal retrospective study. BMC Infect Dis. 2023;23(1):494.

43. Otero L, Van Deun A, Agapito J, Ugaz R, Prellwitz G, Gotuzzo E, et al. Quality assessment of smear microscopy by stratified lot sampling of treatment follow-up slides. Int J Tuberc Lung Dis. 2011;15(2):211-6, i.

44. Mekonen A, Ayele Y, Berhan Y, Woldeyohannes D, Erku W, Sisay S. Factors which contributed for low quality sputum smears for the detection of acid fast bacilli (AFB) at

selected health centers in Ethiopia: A quality control perspective. PLoS One. 2018;13(6):e0198947.

45. Desalegn DM, Kitila KT, Balcha HM, Gebeyehu CS, Kidan YW, Amare K, et al. Misdiagnosis of pulmonary tuberculosis and associated factors in peripheral laboratories: a retrospective study, Addis Ababa, Ethiopia. BMC Res Notes. 2018;11(1):291.

46. Murongazvombo AS, Dlodlo RA, Shewade HD, Robertson V, Hirao S, Pikira E, et al. Where, when, and how many tuberculosis patients are lost from presumption until treatment initiation? A step by step assessment in a rural district in Zimbabwe. Int J Infect Dis. 2019;78:113-20.

47. Chihota VN, Ginindza S, McCarthy K, Grant AD, Churchyard G, Fielding K. Missed Opportunities for TB Investigation in Primary Care Clinics in South Africa: Experience from the XTEND Trial. PLoS One. 2015;10(9):e0138149.

Methods

Population and data

The target population included individuals evaluated for pulmonary TB disease through routine healthcare in low- and middle-income countries. We extracted data from the WHO's Global TB Database (1) on TB notifications for 2023 (Table S2). Using these data we categorized total pulmonary notifications into the number bacteriologically-confirmed and the number clinically-diagnosed. We also extracted data on the number HIV-positive and the number evaluated with an RDT. We excluded countries with <100 pulmonary TB notifications and countries with missing data on laboratory-confirmed or clinically diagnosed cases. If no values were reported for the number HIV-positive or receiving an RDT, we assumed these values were zero. This study used publicly-accessible aggregate data, and did not represent human subjects research.

TB diagnosis model

We analyzed notifications data using a mathematical model of TB diagnosis (schematic shown in Extended Data Fig. 2). This model assumed patients would receive an initial bacteriological test, by smear microscopy or WHO-approved RDT. Positive results would be recorded as bacteriologically-confirmed TB cases. Bacteriologically-negative patients would be evaluated clinically, and if diagnosed positive would be recorded as clinically-diagnosed TB cases. Patients not determined to have TB would not receive a TB diagnosis. We did not consider under-reporting (TB diagnoses missing from national notifications data). Model equations are shown below.

Calculation of diagnostic outcomes

Equations 1-6 were used to calculate the probability of the starting cohort (total individuals evaluated for TB) experiencing each diagnostic outcome by country, stratified by HIV status and whether initial testing was with smear or RDT.

$$f_{h,b}^{TP-bact} = p^{tb} * p_h^{hiv} * p_b^{test} * se_{h,b}^{t1} * \left(1 - p^{cult-neg}\right) * (1 - \kappa) \quad [1]$$

$$f_{h,b}^{FP-bact} = \left(1 - p^{tb}\right) * p_h^{hiv} * p_b^{test} * \left(1 - sp_{h,b}^{t1}\right) * (1 - \kappa) \quad [2]$$

$$f_{h,b}^{TP-clin} = p^{tb} * p_h^{hiv} * p_b^{test} * \left(\left(1 - se_{h,b}^{t1} * \left(1 - p^{cult-neg}\right)\right) * (1 - \kappa) + \kappa\right) * se_{h,b}^{t2} \quad [3]$$

$$f_{h,b}^{FP-clin} = \left(1 - p^{tb}\right) * p_h^{hiv} * p_b^{test} * \left(sp_{h,b}^{t1} * (1 - \kappa) + \kappa\right) * \left(1 - sp_{h,b}^{t2}\right) \quad [4]$$

$$f_{h,b}^{FN} = p^{tb} * p_h^{hiv} * p_b^{test} * \left(\left(1 - se_{h,b}^{t1} * \left(1 - p^{cult-neg}\right)\right) * (1 - \kappa) + \kappa\right) * \left(1 - se_{h,b}^{t2}\right) \quad [5]$$

$$f_{h,b}^{TN} = \left(1 - p^{tb}\right) * p_h^{hiv} * p_b^{test} * \left(sp_{h,b}^{t1} * (1 - \kappa) + \kappa\right) * sp_{h,b}^{t2} \quad [6]$$

In these equations, the following country-specific parameters were defined for individuals in the starting cohort, stratified by HIV stratum h (1 = HIV positive, 2 = HIV-negative) and bacteriological test received b (1 = smear, 2 = RDT): $f_{h,b}^{TP-bact}$, the probability of receiving a bacteriologically-confirmed true-positive TB diagnosis, $f_{h,b}^{FP-bact}$, the probability of receiving a false-positive TB diagnosis; $f_{h,b}^{TP-clin}$, the probability of receiving a clinically-diagnosed true-positive TB diagnosis; $f_{h,b}^{FP-clin}$, the probability of receiving a clinically-diagnosed false-positive TB diagnosis; $f_{h,b}^{FN}$, the probability of receiving a false-negative TB diagnosis; and $f_{h,b}^{TN}$, the probability of receiving a true-negative TB diagnosis. p^{tb} was defined as the probability of having TB for individuals in the starting cohort (varied from 0.05 to 0.50, with 0.25 used for the base-case analysis). p_h^{hiv} was defined as $1 - p_{hiv}$ for h = 1 and p_{hiv} for h = 2. p_{hiv} (HIV prevalence in the starting cohort) was computed from country-reported data defined in Table S2 (newrel_hivpos / c_newinc). p_b^{test} was defined as the probability of receiving bacteriological test type b ($1 - p_{rdt}$ for b = 1, p_{rdt} for b = 2). p_{rdt} (fraction tested with an RDT) was computed from country-reported data defined in Table S2 (newinc_rdx / c_newinc, or (newinc_pulm_labconf_rdx + newinc_pulm_clindx_rdx + newinc_ep_rdx) / c_newinc for countries reporting disaggregated data). Sensitivity and specificity values for bacteriological test and clinical diagnosis were stratified by HIV stratum h and bacteriological test received b: $se_{h,b}^{t1}$, the sensitivity of the initial bacteriological test for culture-positive TB; $sp_{h,b}^{t1}$, the specificity of the initial bacteriological test; $se_{h,b}^{t2}$, the sensitivity of clinical diagnosis; and $sp_{h,b}^{t2}$, the specificity of clinical diagnosis. $p^{cult-neg}$ represents the

probability of culture-negative TB, for individuals with TB in the initial cohort. This parameter adjusts bacteriological test sensitivity downwards to reflect sensitivity for all pulmonary TB (culture-positive and culture-negative), under the assumption that culture-negative TB will also be negative on smear and RDT. κ represents the probability that an individual does not receive a bacteriological test as part of TB evaluation, and is only evaluated clinically. This parameter was set to zero in the main analysis, with values >0 examined in sensitivity analyses.

Likelihood function for reported data

The diagnostic outcomes defined in equations 1-6 were used to parameterize a binomial likelihood function for the number of TB notifications that were bacteriologically confirmed out of the total number of notifications in each country.

$$f^{bact} = \left(\sum_{h=1}^{2}\sum_{b=1}^{2}\left(f_{h,b}^{TP-bact} + f_{h,b}^{FP-bact}\right)\right)/\left(\sum_{h=1}^{2}\sum_{b=1}^{2}\left(f_{h,b}^{TP-bact} + f_{h,b}^{FP-bact} + f_{h,b}^{TP-clin} + f_{h,b}^{FP-clin}\right)\right) \quad [7]$$

$$N^{bact} \sim \text{Binomial}\left(n = N^{notif},\ p = f^{bact}\right) \quad [8]$$

In equations 7-8, f^{bact} represents the probability that diagnosis is bacteriologically-confirmed, among individuals diagnosed with TB. N^{notif} represents the total number of TB diagnoses for a given country, computed from country-reported data defined in Table S2 (sum of new_clindx, ret_rel_clindx, new_labconf, and ret_rel_labconf variables). N^{bact} represents the total number of bacteriologically-confirmed TB diagnoses for a given country, computed from country-reported data (sum of new_labconf and ret_rel_labconf variables).

Additional study outcomes

Equations 9-17 were used to calculate additional study outcomes.

$$f^{notif} = \sum_{h=1}^{2}\sum_{b=1}^{2}\left(f_{h,b}^{TP-bact} + f_{h,b}^{FP-bact} + f_{h,b}^{TP-clin} + f_{h,b}^{FP-clin}\right) \quad [9]$$

$$N^{eval} = N^{notif}/f^{notif} \quad [10]$$

$$N^{TP} = N^{eval} * \sum_{h=1}^{2}\sum_{b=1}^{2}\left(f_{h,b}^{TP-bact} + f_{h,b}^{TP-clin}\right) \quad [11]$$

$$N^{FP} = N^{eval} * \sum_{h=1}^{2}\sum_{b=1}^{2}\left(f_{h,b}^{FP-bact} + f_{h,b}^{FP-clin}\right) \quad [12]$$

$$N^{FN} = N^{eval} * \sum_{h=1}^{2}\sum_{b=1}^{2} f_{h,b}^{FN} \quad [13]$$

$$N^{TN} = N^{eval} * \sum_{h=1}^{2}\sum_{b=1}^{2} f_{h,b}^{TN} \quad [14]$$

$$PPV = N^{TP}/(N^{TP} + N^{FP}) \quad [15]$$

$$NPV = N^{TN}/(N^{TN} + N^{FN}) \quad [16]$$

$$f^{correct} = (N^{TP} + N^{TN})/N^{eval} \quad [17]$$

In these equations, f^{notif} is the estimated probability of being diagnosed with TB (equivalent to 'yield' of TB diagnosis) among individuals in the starting cohort, N^{eval} is the estimated number of individuals evaluated for TB, N^{TP} is the estimated number of true-positive diagnoses, N^{FP} is the estimated number of false-positive diagnoses, N^{FN} is the estimated number of false-negative diagnoses, and N^{TN} is the estimated number of true-negative diagnoses. PPV is the estimated positive predictive value of TB diagnosis (probability that individuals diagnosed with TB truly have TB), NPV is the estimated negative predictive value of TB diagnosis (probability that individuals not receiving a TB diagnosis truly do not have TB), and $f^{correct}$ is the estimated probability of receiving a correct diagnosis, among individuals evaluated for TB.

Model parameters

Estimates of test sensitivity and specificity ($se_{h,b}^{t1}$ and $sp_{h,b}^{t1}$, respectively) were drawn from diagnostic accuracy studies of smear microscopy and Xpert Ultra (2-4). We stratified test sensitivity by HIV status, accounting for lower sensitivity among individuals with HIV. As reported sensitivity estimates are based on comparison to culture, they may overestimate true sensitivity due to the presence of culture-negative pulmonary TB. In our analysis we adjusted test sensitivity downwards to account for this possibility (via $p^{cult-neg}$ in equations 1, 3, and 5). For each country, we calculated HIV prevalence among individuals tested for TB and the fraction of

initial bacteriological tests performed with an RDT from country-reported notifications data (p_{hiv} and p_{rdt}, respectively). We assumed the fraction receiving an RDT did not vary by HIV status. Extended Data Table 8 summarizes input values and sources.

A wide range of estimates for the sensitivity and specificity of clinical diagnosis have been reported (5). We used a parametric binormal model (6) to synthesis the data from these studies (Table S3, (7-16)) and define the combinations of sensitivity and specificity consistent with published evidence (Fig. S2). This approach assumes that, while countries could achieve high sensitivity or specificity of clinical diagnosis (depending on local practices), available evidence doesn't support the assumption that clinical diagnosis can be simultaneously highly sensitive and specific. We allowed clinical diagnosis sensitivity and specificity ($se_{h,b}^{t2}$ and $sp_{h,b}^{t2}$, repectively) to vary between countries within the plausible values defined by the binormal model. We also allowed the sensitivity and specificity of clinical diagnosis to vary within countries by HIV status and by whether initial diagnostic testing was via RDT, based on a systematic review finding higher rates of clinical diagnosis for HIV-positive individuals (vs. HIV-negative) and for individuals initially tested with smear microscopy (vs. RDT) (17). Estimates for overall sensitivity and specificity (for the diagnostic algorithm overall as well as individual steps of the algorithm) are reported in the Results section.

There is limited evidence on true TB prevalence among individuals evaluated for TB in routine settings (p^{tb}). We extracted data on TB prevalence among samples of individuals with presumptive TB included in recent diagnostic accuracy studies (4, 18). These data demonstrate a wide range of study-level TB prevalence values, with a median value of 26% and an interquartile range 14-37%. In our analysis we estimated results for values from 5% to 50%, and used 25% for our main analysis.

Statistical analysis

We implemented the analysis using a Bayesian approach. Under this approach, we created prior distributions representing published evidence on each model parameter (Extended Data Table 8) and used a Hamiltonian Monte Carlo algorithm to generate 5000 fitted values for each outcome of interest. Outcomes included algorithm sensitivity, specificity, positive predictive value, and negative predictive value, as well as the number of true-positive, true-negative, false-positive, and false-negative diagnoses generated by TB diagnosis. Point estimates were calculated as the mean of the distribution of results. We used a non-parametric approach to calculate measures of uncertainty around study outcomes, with 95% uncertainty intervals calculated as the 2.5^{th} and 97.5^{th} percentiles of the Monte Carlo simulation results for each outcome. We estimated outcomes for each country independently, and pooled results across countries to report regional and global results. Uncertainty in pooled results assumed a rank correlation of 0.5 across country-specific values. Analyses were conducted using R (v4.4.2) (19), and the RStan package (v2.32.6) (20).

Counterfactual scenarios

Using the fitted models for each country, we explored counterfactual scenarios representing hypothetical alternatives for improving diagnostic performance. Under the first counterfactual scenario (***Full RDT adoption***) we recalculated diagnostic outcomes assuming 100% adoption of Xpert Ultra to replace smear microscopy in each modelled country. The second counterfactual scenario (***Reduced clinical diagnosis***) assumed there would be efforts to reduce clinician willingness to diagnose patients clinically, such that the false-positive rate (1-specificity) of clinical diagnosis is reduced by 50%. We assumed that this would not change the ROC curve for clinical diagnosis, such that improvements in specificity would come at the cost of reduced sensitivity, consistent with the main analysis ROC curve shown in Fig. S2. The third counterfactual scenario (***Improved clinical algorithms***) assumed there could improvements in practices around clinical diagnosis (such as greater use of chest radiography, or improved diagnostic criteria for bacteriologically-negative TB), allowing the specificity of this diagnostic

step to increase to reach the optimistic ROC curve shown on Fig. S2, with no loss of sensitivity. The fourth counterfactual scenario (***Improved RDTs, reduced clinical diagnosis***) assumed the development and full adoption of improved RDTs with sensitivity equivalent to culture, and contemporaneous change in clinical diagnosis matching the assumptions of the second counterfactual scenario (i.e., increased specificity and reduced sensitivity). For each counterfactual scenario we recalculated diagnostic outcomes and compared these results with the main analysis, to estimate the improvements in diagnostic outcomes that could be achieved.

Sensitivity analyses

We calculated partial rank correlation coefficients quantifying the relationship between individual parameters and study outcomes. In addition, we re-estimated results under several alternative analytic specifications. These included [1] using published evidence on Xpert MTB/RIF (instead of Ultra) as the basis for RDT sensitivity and specificity (18), adopting [2] more optimistic and [3] pessimistic assumptions for the sensitivity and specificity of clinical diagnosis (ROC curves shown in Fig. S2), and [4] assuming 25% of individuals don't receive a bacteriological test and are only evaluated clinically.

Finally, we created plots showing how the probability of false-positive diagnosis (the false discovery rate (FDR), equal to 1- PPV) and false-negative diagnosis (the false omission rate (FOR), equal to 1- NPV) change as a function of initial TB prevalence and the fraction lab-confirmed, holding other inputs at their global average. To do so we considered a simplified version of equations 1-6. These simplified equations do not differentiate bacteriological test type or HIV-infection status (subscripts 'h' and 'b' removed), and restate initial test sensitivity for all pulmonary TB (given by $se_{h,b}^{t1} * \left(1 - p^{cult-neg}\right)$ in equations 1-6) as se^{t1}. With these simplifications, the probabilities of being true-positive on bacteriological testing ($f^{TP-bact}$), false-positive on bacteriological testing ($f^{FP-bact}$), true-positive after clinical evaluation

($f^{TP-clin}$), false-positive after clinical evaluation ($f^{FP-clin}$), false-negative after clinical evaluation ($f^{FN-clin}$), and true-negative after clinical evaluation ($f^{TN-clin}$) are given by equations 18-23.

$$f^{FP-bact} = \left(1 - p^{tb}\right) * [(1 - \kappa) * (1 - sp^{t1})] \quad [18]$$

$$f^{TP-bact} = p^{tb} * [(1 - \kappa) * se^{t1}] \quad [19]$$

$$f^{FP-clin} = \left(1 - p^{tb}\right) * [(1 - \kappa) * sp^{t1} + \kappa] * (1 - sp^{t2}) \quad [20]$$

$$f^{TP-clin} = p^{tb} * [(1 - \kappa) * (1 - se^{t1}) + \kappa] * se^{t2} \quad [21]$$

$$f^{FN-clin} = p^{tb} * [(1 - \kappa) * (1 - se^{t1}) + \kappa] * (1 - se^{t2}) \quad [22]$$

$$f^{TN-clin} = \left(1 - p^{tb}\right) * [(1 - \kappa) * sp^{t1} + \kappa] * sp^{t2} \quad [23]$$

In these equations, se^{t1} and sp^{t1} are the sensitivity and specificity of the initial bacteriological test, respectively, and se^{t2} and sp^{t2} are the sensitivity and specificity of clinical diagnosis among individuals testing negative on the initial bacteriological test, respectively. The FDR (equal to $\left[f^{FP-bact} + f^{FP-clin}\right] / \left[f^{FP-bact} + f^{FP-clin} + f^{TP-bact} + f^{TP-clin}\right]$) and FOR (equal to $f^{FN-clin} / \left[f^{FN-clin} + f^{TN-clin}\right]$) can then be written as

$$FDR = \frac{f^{bact}}{1+\omega*\sigma^{t1}(0)} + \frac{1-f^{bact}}{1+\omega*\sigma^{t1}(0)*\sigma^{t2}(\lambda+1)} \quad [24]$$

$$FOR = \frac{\omega*\sigma^{t1}(\lambda+1)*\sigma^{t2}(1)}{1+\omega*\sigma^{t1}(\lambda+1)*\sigma^{t2}(1)} \quad [25]$$

where $\omega = odds\left(p^{tb}\right)$, $\lambda = odds(\kappa)$, we introduce the function $\sigma^{i}(x) = \frac{se^{i}-x}{1-sp^{i}-x}$, and f^{bact} is the fraction of diagnoses that are bacteriologically confirmed ($\left[f^{TP-bact} + f^{FP-bact}\right] / \left[f^{TP-bact} + f^{FP-bact} + f^{TP-clin} + f^{FP-clin}\right]$). The expression for f^{bact} can be rearranged as

$$\frac{1}{odds\left(f^{bact}\right)} = \frac{\left(1-sp^{t2}\right)*\left(\frac{sp^{t1}+\lambda}{1-sp^{t1}}\right) + se^{t2}*\omega*\left(\frac{1-se^{t1}+\lambda}{1-sp^{t1}}\right)}{1+\omega*\sigma^{t1}(0)} \quad [26]$$

which represents a negatively sloped straight line in the terms $\left(\mathbf{1}-\boldsymbol{sp}^{\boldsymbol{t2}}\right)$ and $\boldsymbol{se}^{\boldsymbol{t2}}$. The ROC curve $\Phi(se^{t2}) = a + b * \Phi(1 - sp^{t2})$ (where Φ is the standard normal cumulative distribution, and $a, b > 0$) intersects this line at a unique point, implying that giving f^{bact}, ω, λ, se^{t1}, and sp^{t1}, together with a ROC curve, uniquely determines the accuracy of clinical evaluation, se^{t2} and sp^{t2}. We use these relationships to explore how the FDR and FOR change as we vary TB prevalence (p_{tb}) and the fraction of TB diagnoses that is bacteriologically confirmed (f^{bact}). Here we have used the base case ROC curve, taken $\kappa = 0$, and used $se^{t1} = 0.62$ and $sp^{t1} = 0.98$ from the global average model results. We repeated this analysis for the three alternative analytic specifications (using optimistic and pessimistic ROC curves, and assuming $\kappa = 0.25$).

Ethics and inclusion

This study exclusively utilized publicly available aggregate data, and collected no primary data on humans or animals. Data on TB case notifications were obtained from the WHO TB Database, reflecting the work of a large number of national TB programs to collect, confirm, and report data. We gratefully acknowledge the efforts made to make these data available. Data on the performance of clinical diagnosis were extracted from the published literature. Local and regional researchers responsible for collecting and publishing these data were appropriately cited and acknowledged throughout. All co-authors actively participated from the early stages of project design through to data interpretation and manuscript preparation.

Data availability statement

All data used in this study were drawn from publicly available datasets (downloadable from https://www.who.int/teams/global-tuberculosis-programme/data, "Case notifications" and "WHO TB burden estimates" files), as well as published studies listed in Supplement Table S3.

Code availability statement

Analytic code used to implement the analysis is available from https://zenodo.org/records/16414104.

Methods-only References

1. WHO Global TB programme. WHO Global TB Database. In: WHO Global TB programme, editor. Geneva, Switzerland2023.
2. Steingart KR, Henry M, Ng V, Hopewell PC, Ramsay A, Cunningham J, et al. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. The Lancet infectious diseases. 2006;6(9):570-81.
3. Cattamanchi A, Dowdy DW, Davis JL, Worodria W, Yoo S, Joloba M, et al. Sensitivity of direct versus concentrated sputum smear microscopy in HIV-infected patients suspected of having pulmonary tuberculosis. BMC infectious diseases. 2009;9:1-9.
4. Zifodya JS, Kreniske JS, Schiller I, Kohli M, Dendukuri N, Schumacher SG, et al. Xpert Ultra versus Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults with presumptive pulmonary tuberculosis. Cochrane Database of Systematic Reviews. 2021(2).
5. Walusimbi S, Bwanga F, De Costa A, Haile M, Joloba M, Hoffner S. Meta-analysis to compare the accuracy of GeneXpert, MODS and the WHO 2007 algorithm for diagnosis of smear-negative pulmonary tuberculosis. BMC Infect Dis. 2013;13:507.
6. Dorfman DD, Alf Jr E. Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating-method data. Journal of mathematical psychology. 1969;6(3):487-96.
7. Abebe G, Deribew A, Apers L, Abdissa A, Kiflie Y, Koole O, et al. Evaluation of the 2007 WHO guideline to diagnose smear negative tuberculosis in an urban hospital in Ethiopia. BMC Infect Dis. 2013;13:427.
8. Alamo ST, Kunutsor S, Walley J, Thoulass J, Evans M, Muchuro S, et al. Performance of the new WHO diagnostic algorithm for smear-negative pulmonary tuberculosis in HIV prevalent settings: a multisite study in Uganda. Trop Med Int Health. 2012;17(7):884-95. doi: 10.1111/j.365-3156.2012.03003.x. Epub 2012 May 11.
9. Huerga H, Varaine F, Okwaro E, Bastard M, Ardizzoni E, Sitienei J, et al. Performance of the 2007 WHO algorithm to diagnose smear-negative pulmonary tuberculosis in a HIV prevalent setting. PLoS One. 2012;7(12):e51336. doi: 10.1371/journal.pone.0051336. Epub 2012 Dec 19.
10. Koole O, Thai S, Khun KE, Pe R, van Griensven J, Apers L, et al. Evaluation of the 2007 WHO guideline to improve the diagnosis of tuberculosis in ambulatory HIV-positive adults. PLoS One. 2011;6(4):e18502. doi: 10.1371/journal.pone.0018502.

11. Siddiqi K, Walley J, Khan MA, Shah K, Safdar N. Clinical guidelines to diagnose smear-negative pulmonary tuberculosis in Pakistan, a country with low-HIV prevalence. Trop Med Int Health. 2006;11(3):323-31.

12. Soto A, Solari L, Díaz J, Mantilla A, Matthys F, van der Stuyft P. Validation of a clinical-radiographic score to assess the probability of pulmonary tuberculosis in suspect patients with negative sputum smears. PLoS One. 2011;6(4):e18486.

13. Swai HF, Mugusi FM, Mbwambo JK. Sputum smear negative pulmonary tuberculosis: sensitivity and specificity of diagnostic algorithm. BMC Res Notes. 2011;4:475.(doi):10.1186/756-0500-4-475.

14. Theron G, Zijenah L, Chanda D, Clowes P, Rachow A, Lesosky M, et al. Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. Lancet. 2014;383(9915):424-35.

15. Vassall A, van Kampen S, Sohn H, Michael JS, John KR, den Boon S, et al. Rapid diagnosis of tuberculosis with the Xpert MTB/RIF assay in high burden countries: a cost-effectiveness analysis. PLOS Med. 2011;8(11):e1001120-e.

16. Wilson D, Mbhele L, Badri M, Morroni C, Nachega J, Chaisson RE, et al. Evaluation of the World Health Organization algorithm for the diagnosis of HIV-associated sputum smear-negative tuberculosis. Int J Tuberc Lung Dis. 2011;15(7):919-24. doi: 10.5588/ijtld.10.0440.

17. Kim S, Can MH, Agizew TB, Auld AF, Balcells ME, Bjerrum S, et al. Factors associated with tuberculosis treatment initiation among bacteriologically negative individuals evaluated for tuberculosis: An individual patient data meta-analysis. PLoS Med. 2025;22(1):e1004502.

18. Horne DJ, Kohli M, Zifodya JS, Schiller I, Dendukuri N, Tollefson D, et al. Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Database of Systematic Reviews. 2019(6).

19. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2010.

20. Stan Development Team. Rstan: the R interface to Stan. R package 2.21.8 ed2023.