What if they don't have tuberculosis? The consequences and trade-offs involved in false-positive diagnoses of tuberculosis

Houben RMGJ [1][2], Lalli M [1][2], Kranzer K [2][3], Menzies NA [4], Schumacher SG [5], Dowdy DW [6]

Affiliations
1. TB Modelling Group, TB Centre and Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, UK
2. Infectious Disease Epidemiology Department, Faculty of Epidemiology and Public Health, London School of Hygiene and Tropical Medicine, London, UK
3. Research Centre Borstel, National and Supranational Reference Laboratory, Borstel Germany
4. Dept. of Global Health and Population, Harvard TH Chan School of Public Health, Boston, MA, USA
5. FIND, Geneva, Switzerland
6. Dept. of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore MD USA

Corresponding Author
Rein Houben
London School of Hygiene and Tropical Medicine
104b, Keppel Street
WC1E 7HT
London, UK

Rein.Houben@lshtm.ac.uk

Alternate corresponding author
David W Dowdy
615 N. Wolfe St., E6531
Baltimore, MD 21205, USA
Phone: +1 410-614-5022
Fax: +1 410-614-0902
ddowdy1@jhmi.edu

Keywords
Tuberculosis, False-Positive diagnosis, Trade-off

Running Title
Impact of false-positive TB diagnoses

40 word summary
False-positive diagnoses are frequent in tuberculosis and have profound consequences, yet are usually overlooked in policy decision-making. We describe pathways and trade-offs involved, and their consequences for individuals, households and health systems. We then discuss steps to mitigate their impact.

1

Abstract

To find the millions of missed tuberculosis (TB) cases, national TB programmes are under pressure to expand TB disease screening, and target populations with lower disease prevalence. Together with imperfect performance and application of existing diagnostic tools, including empirical diagnosis, broader screening risks placing individuals without TB on prolonged treatment. These false-positive diagnoses have profound consequences for TB patients and prevention efforts, yet are usually overlooked in policy decision-making.

In this paper we describe the pathways to a false-positive TB diagnosis, including trade-offs involved in the development and application of diagnostic algorithms. We then consider the wide range of potential consequences for individuals, households, health systems, and reliability of surveillance data. Finally, we suggest practical steps that the TB community can take to reduce the frequency and potential harms of false-positive TB diagnosis and to more explicitly assess the trade-offs involved in the screening and diagnostic process.

1. Introduction

1.1 Background

To achieve ambitious targets to reduce incidence and deaths due to tuberculosis (TB), global and national policies emphasize the need to diagnose and treat a larger fraction of the 10.4 million individuals who develop TB disease each year. [1, 2] For this reason, National Tuberculosis Programmes (NTPs) are under pressure to expand access to TB screening and diagnosis. [3] Similarly, the need for more sensitive diagnostic tools to detect patients earlier in the course of their disease and care seeking process has been highlighted, but increased sensitivity may come at the cost of reduced specificity. [4, 5]

Current TB policy discussions, programme targets and indicators usually do not consider the risk of false-positive TB diagnoses [2, 6, 7], despite recognition of the issue in World Health Organization (WHO) recommendations of systematic screening programmes. [8]

In this paper we describe how false-positive diagnoses are part of TB clinical practice, and their consequences for individuals, households, health systems and surveillance data. Our aim is to enable comprehensive discussions on the trade-offs involved for expanded TB screening programmes. We then propose concrete actions that can be undertaken to mitigate the negative effects of false-positive TB diagnosis.

1.2 Definition of false-positive TB diagnosis

We define a false-positive TB diagnosis as one where *an individual, who does not have active TB disease, incorrectly receives a diagnosis of TB disease*. In this vein, we use the term false-positive (a term widely used in clinical epidemiology) to suggest the absence of active TB disease, not the absence of symptoms or the absence of other non-TB illness (e.g., bacterial pneumonia, obstructive lung disease). Importantly also, we use this term also

to apply only to the diagnostic process itself, not the individual undergoing diagnosis – as the latter may be stigmatising. [9, 10]

1.3 Path towards false-positive TB diagnosis

Figure 1 outlines the screening and diagnostic pathway for TB. A screening population is composed of individuals with and without TB. Some individuals without TB may have an underlying illness that presents with similar symptoms to those of TB. A diagnostic algorithm, with imperfect sensitivity and specificity, is then applied to the screening population. All those diagnosed with TB (correctly or not) are eligible to start treatment and should be notified as part of national and global surveillance.

The proportion and the underlying reasons for false-positive TB test results vary, and are highly setting-dependent as local guidelines and policy will dictate different diagnostic algorithms in different populations. Differences in background prevalence of TB or co-morbidity patterns (e.g. HIV, silicosis) as well as human performance and laboratory capacity can influence test performance and result interpretation. A key example is empirical (or 'clincal') diagnosis, where TB is diagnosed (and treatment is started) in the absence of a recorded positive bacteriological test. Empirical diagnosis accounted for 43% of all cases reported to WHO globally in 2017. [7] Decisions to treat empirically are based on a mix of symptomatic presentation, co-morbidities, chest X-ray and other ancillary tests, non-response to other therapeutic maneuvers (e.g., trials of broad-spectrum antibiotics), and individual clinician assessment – all of which vary from one population (and one provider) to the next. Empirical diagnosis is a critical part of the diagnostic arsenal, especially in low-resource settings where health workers may see a high number of TB patients to inform their clinical judgement. However, the limited data available suggests that the sensitivity and specificity of empirical diagnosis is both highly variable and suboptimal. A multi-country diagnostics trial found sensitivity ranged from 16% to 44.4% and the specificity ranged from

86.9% to 95.3% across study sites, and was significantly influenced by coverage of chest X-Ray [11].

In high-income settings, where greater resources are available, initial screening and diagnostic approach are strongly tilted towards increasing specificity, thus decreasing the risk of a false-positive diagnosis despite the low prior of TB in the screening population.

By contrast, TB diagnosis in a rural health centre in a resource-limited setting will often rely on smear microscopy, symptoms and clinical examinations only. Variability in the accuracy of empirical diagnosis – and the attendant risk of false-positive diagnosis – is likely to be enhanced in the highly heterogenous private sector. A study in India found that TB diagnosis was more often reliant on clinical opinion and less on bacteriological confirmation. [12]

As countries look to expand screening programmes beyond individuals self-reporting to TB clinics (i.e. passive case finding), the screening population will likely have a lower prevalence of disease, either at the start, or the prevalence could drop after repeated rounds of screening. **Table 1** illustrates how such a drop will also affect the rates of false-positive diagnoses, and why different algorithms may therefore need to be considered in these screening programmes as compared to evaluating self-reported individuals. If a standard algorithm of symptom screen, sputum microscopy and empirical diagnosis (Algorithm 1, column 2 and 3) was used, a drop from the current approximate 10% prevalence among patients submitting sputum for TB diagnosis in healthcare settings to a high-risk screening population with 1% prevalence (1000/100,000), can result in more than 70% of all TB diagnoses being made among people who do not have underlying TB disease, i.e. nearly 2.5 false-positive TB diagnoses for each true-positive.

To mitigate the high probability of a false-positive diagnosis, screening programmes could look to increase the specificity, or consider improving sensitivity to diagnose more true-

positive cases. Figures 2a-b show how the positive predictive value (i.e. the proportion of individuals with TB diagnosis that actually have TB disease) changes with sensitivity (**Figure 2a**) or specificity (**Figure 2b**). From these, it becomes clear that improvements in sensitivity will have limited impact on the PPV, and that prevalence in the screening population is the key driver. However, the PPV is much more dependent on small increases or declines across a narrow range in specificity (Figure 2b).

Ideally, diagnostic algorithms in screening programmes would improve both sensitivity and specificity, for example resembling the protocols used in prevalence surveys (**Table 1**, Algorithm 2, column 4 and 5). However, even with an algorithm of 99.0% specificity, a screening population with a 0.5% prevalence (500/100,000) of TB would lead to the number of false-positive TB diagnoses outnumbering true-positive diagnoses.

It is important to note that these scenarios use best available, but sometimes weak, current estimates for the sensitivity and specificity of all tests, including empirical diagnosis [8]. However, in the absence of wide application of improved diagnostic tools, the stronger the external push to increase the number of TB diagnoses (whether by using more sensitive tests or empirically treating more people), the more specificity will become compromised, and the higher the number of false-positive diagnoses will be.

2. Consequences of false-positive TB diagnosis

2.1 The Individual

False-positive diagnoses for TB are usually considered to have minimal long-term health implications for patients. However, they can in fact lead to substantial negative consequences – consequences that generally will not be later corrected. Specifically, patients testing false-positive for active TB will almost invariably be recommended for a treatment course that currently lasts a minimum of six months. Not only does TB treatment

carry a non-negligible risk of adverse clinical events (for example, at least a one-in-fifty risk of clinically relevant hepatotoxicity), [13] but patients also incur substantial costs. Even when clinical services are provided to the patient for free (as is the case in many settings), patients still incur high non-clinical costs, including transportation, food, childcare, and lost wages. In a systematic review of patient costs including 14 studies in low-income countries, the mean direct patient cost (plus productivity loss) was estimated at $248 per patient [14]; a second systematic review estimated that total patient costs of TB averaged 58% of annual patient income, with half of those costs occurring after treatment [15]. In addition, undergoing TB treatment (especially when directly observed and in poor structural conditions) often results in damaging social consequences – including stigma, isolation from families, gender discrimination, loss of hope, and disrespect. [16]

Less well studied are the consequences of missing other conditions that may be falsely diagnosed as TB. One small study has suggested an increase in mortality among HIV-positive Ugandan adults with false-positive TB microscopy results [17], though this result was not statistically significant. Underlying conditions in individuals falsely diagnosed as having TB can range from bacterial pneumonia to lung cancer [18]. In the case of pneumonia – potentially the most common missed diagnosis – rifampin has some activity against the most common bacterial pathogens isolated [19], but other antimicrobial regimens are more effective. Many of these "missed" conditions are more rapidly progressive or fatal than tuberculosis itself, meaning that in many such cases, the consequences of delayed diagnosis resulting from a false-positive TB diagnosis will be equally (if not more) clinically devastating than those of false-negative TB diagnoses.

In contrast to perceptions around false-positive diagnoses, the consequences of false-negative diagnoses of TB are often portrayed as being fatal. As a result, studies of relative harms often suggest that a large number of false-positive diagnoses should be tolerated to avert a single false-negative diagnosis (with ratios as high as 30:1). [20] However, indirect

evidence suggests that most individuals with TB initially diagnosed as false-negative often later start treatment – either empirically or through other diagnostic tests. For example, results from clinical trials suggest that the advantages of using more sensitive diagnostic testing with Xpert MTB/RIF relative to sputum smear microscopy is at least partially counterbalanced by existing practices of empirical treatment (i.e., treatment of patients testing false-negative by Xpert). [21-24]

Thus, while much of the TB literature has focused on the potential consequences to patients of false-negative diagnoses, the implications of false-positive diagnoses – from life-threatening side effects to social stigmatisation to morbidity and mortality from other conditions – may be no less dire. The relative harm of false-positive versus false-negative diagnosis is an important consideration in, for example, decisions to screen for TB in lower-risk populations or to use more sensitive, less specific tests (e.g., Xpert Ultra). [25]

2.2. Household

Households of individuals who receive false-positive diagnoses are also negatively affected. Prolonged treatment for TB leads not only to additional healthcare-related expenditures but also to loss of income for the household, which can cause households to experience catastrophic costs. [15, 26] Such a descent into poverty will have a substantial and long-term impact on all household members including education, income and health. [27]

2.3 Health System

Recent modelling analyses have focused on the impact of resource constraints in the health system on the ability to implement ambitious interventions, including constraints on health care staff and beds in the multi-drug resistant TB (MDR-TB) wards [28-30]. In the context of limited drug supply and health care workers, treating patients who received false-positive TB diagnoses may substantially worsen care for patients in other segments of the healthcare system.

Programme performance will be overestimated in the presence of false-positive diagnosis, as treating false-postively diagnosed patients will have no impact on the TB epidemic. As highlighted by **Table 1,** reported notifications can easily be misleading as an indicator for the success of an expanded screening strategy – an increased number of notifications may simply represent a dramatic increase in the number of false-positive diagnoses.

In addition, those notified and initiating treatment are assumed to have been at risk of dying due to TB. If a substantial proportion of this cohort does not have TB disease (and instead has less serious conditions), observed mortality rates may be artificially low, leading to overestimates of treatment success both on country and global level.

Such unintentional misrepresentations can, when revealed through for example prevalence surveys, undermine faith from domestic and international funders of the national TB response.

These issues also affect burden estimation and reporting at the global level. Current TB estimation methods by WHO and Institute for Health Metrics and Evaluation do not consider false-positive diagnoses, and instead assume all notified cases represent patients with a true-positive diagnosis. [31, 32] As a consequence the inclusion of false-positive diagnoses artificially increases official incidence estimates, and leads to an underestimation of the gap between incident cases and those notified for treatment. [7, 32]

3. What can be done

3.1 Acknowledge the problem

The first step will be to explicitly include the challenge of false-positive TB diagnosis in TB policy discourse. In these discussions, policy bodies should explicitly consider false-positive TB diagnosis as a challenge that already inflates reported TB notifications, and acknowledge

that these distortions may increase further as new ambitious programmes are rolled out. In addition, efforts should be made to estimate the proportion of false-positive TB diagnoses in surveillance data as part of country-level reporting to WHO-GTB and the Global Fund against AIDS, Tuberculosis and Malaria. As multiple high-level indicators for policy evaluation and advocacy may be strongly affected by false-positive diagnosis, an explicit evaluation of the potential bias introduced in different settings would be sensible.

To inform such and future evaluations, a concrete step could be to generate a quantitative estimate for the sensitivity and specificity of the diagnostic algorithms as applied in each country, and then to use those estimates to inform estimates of false-positive diagnoses (as in **Table 1**). While heterogeneity in diagnostic test performance and implementation is likely, strengthening the current data through for example programmatic reviews and national strategic planning is of high value. Outcomes could include identified opportunities for improvement (e.g. in the protocol for empirical diagnosis) and data strengthening, better estimates of the proportion false-positive TB diagnoses, and more realistic estimates of country programme performance.

3.2 Evaluate trade-offs in screening programmes

As outlined in the previous section, false-positive TB diagnoses follow from the trade-off between potential positive and negative effects of real-life screening and diagnostic processes. We propose that rather than focusing solely on the positive potential of new screening strategies, new strategies should be evaluated through a comprehensive conceptual framework that also acknowledges the potential negative consequences of false-positive diagnosis.

Here, any change in screening strategies should be evaluated in terms of delivering the correct diagnosis for all individuals evaluated for TB, including those without TB. In other

words, in a setting where a more sensitive screening test is being considered, programmes could explicitly consider the aim of increasing the number of individuals who receive a true positive diagnosis instead of a false negative diagnosis and balance this against the number of individuals who may receive a false positive rather than a true negative diagnosis. Where more specific confirmatory testing is being considered, it is important to consider the number of false positive diagnoses that could be averted, balancing this against the number of true positive diagnoses that may be missed.

The practical consequences of these choices are shown **Figure 2** and the final three columns in **Table 1** where the balance between additional true positive and true negative diagnoses shifts notably as the prevalence of disease changes, an observation masked by simply counting the number of additional TB diagnoses.

With each policy decision around e.g. diagnostic approach or population screened, the relative health and financial cost for each diagnosis should be considered. Together, these will provide a comprehensive overview of the cost and benefit of a specific change, for example, informing the choice to try and replace empirical diagnosis with microbiological testing.

In this framework, the trade-offs between sensitivity and specificity of the diagnostic algorithm, as well as the disease prevalence in the screening population, become more explicit. Moving beyond a simple comparison of sensitivity and specificity, the potential performance of new tests or case finding approaches could be simply summarized in the change in the ratio false-positive/true-positive across different prevalences of disease in the target population (see Table 1). Alternatively, more sophisticated decision curve analyses can incorporate additional heterogeneity and complexity in given settings. [25, 33]

While quantifying the consequences of each trade-off more comprehensively is critical, decisions regarding acceptable levels of false-positive diagnosis (relative to the number of true-positive diagnoses) will still need to be made. Such choices are also multidimensional and highly setting-specific, but explicitly considering these trade-offs represents an important first step.

3.3 Research

Work is needed to provide quantitative estimates for the weight a false-positive diagnosis should carry compared to a false-negative, or example, by quantifiying the costs and quality-adjusted life years lost through incorrect diagnoses for individuals with vs without TB. Such work could examine the acceptable number of false-positive TB diagnoses for one additional true positive diagnosis, across settings. A number of papers have already highlighted strong discrepancies between perceived and data-based thresholds, as well as variation between the preferences of clinicians, public health officials and patients. [20, 34, 35].

Another key area of research is to improve the empirical data underlying our current estimates of test performance, particularly the accuracy of empirical diagnosis and the specificity of "definitive' diagnostic tests. While reviews [8] and examples of research exist [36], substantial uncertainty remains on the baseline performance of key tests, particularly as implemented in actual field settings. One example is the specificity of sputum smear, which has shown great variability in recent prevalence surveys, with up to 43% of smear-positive cases not confirmed by culture. [37] Again, some of these data will be setting-specific, in particular for empirical diagnosis. But a better understanding of how the performance depends on specific factors (e.g. background TB and HIV prevalence, use/availability of Xray, etc..) will help inform estimates. Recent studies have also highlighted the value of considering urine LAM and C-reactive Protein as part of TB screening in high-risk HIV positive populations. [38, 39]

3.4 Test application

False positive TB diagnoses are not exclusively caused by suboptimal diagnostic accuracy, but are also driven by pre- and post-analytic errors, adminstrative errors as well as laboratory errors in handling of reagents and/or maintance of instruments. By strengthening domestic and international laboratory networks, TB diagnostics can be embedded in well-functioning systems of training and quality control to help reduce false-positive TB diagnoses.

4. Conclusion

As the TB community looks to close the diagnostic gap through expanded screening and improved diagnostic tools, it is important to acknowledge false-positive TB diagnoses as part of current reality and consequence of expanded screening efforts. By recognizing and addressing the size as well as consequences of false-positive diagnoses, we can improve clinical outcomes for individuals and focus the limited resources available to 'End TB' to achieve stronger population impact.

leads evaluations, takes positions, and accelerates access to tools identified as serving its mission. It provides indirect support to industry (e.g., access to open specimen banks, a clinical trial platform, technical support, expertise, laboratory capacity strengthening in LMICs, etc.) to facilitate the development and use of products in these areas. FIND also supports the evaluation of publicly-prioritized TB assays and the implementation of WHO-approved (guidance & PQ) assays using donor grants. In order to carry out test evaluations, has product evaluation agreements with several private sector companies for TB and other diseases, which strictly define its independence and neutrality vis-a-vis the companies whose products get evaluated and describes roles and responsibilities. The other authors declare they have no potential conflicts of interest to disclose

References

1. Uplekar, M., et al., WHO's new End TB Strategy. Lancet, **2015**; 385.
2. StopTB Partnership, *Global Plan To End TB 2016-2020 - The Paradigm Shift*. 2015: Geneva.
3. Global Fund against AIDS, T.a.M. *Stop TB and Global Fund Deepen Cooperation to Find Missing Cases of TB*. 2017; Available from: https://www.theglobalfund.org/en/news/2017-12-18-stop-tb-and-global-fund-deepen-cooperation-to-find-missing-cases-of-tb/.
4. Dorman, S.E., et al., Xpert MTB/RIF Ultra for detection of Mycobacterium tuberculosis and rifampicin resistance: a prospective multicentre diagnostic accuracy study. Lancet Infect Dis, **2018**; 18(1): p. 76-84.
5. Arend, S.M. and D. van Soolingen, Performance of Xpert MTB/RIF Ultra: a matter of dead or alive. Lancet Infect Dis, **2018**; 18(1): p. 8-10.
6. Global Fund against AIDS, T.a.M., *Modular Framework Handbook*. 2017.
7. World Health Organisation, *Global Tuberculosis Report 2017* 2017: Geneva.
8. WHO, *Systematic screening for active tuberculosis: principles and recommendations*, WHO, Editor. 2013, World Health Organisation: Geneva.
9. Frick, M., D. von Delft, and B. Kumar, End stigmatizing language in tuberculosis research and practice. BMJ, **2015**; 350: p. h1479.
10. stopTB Partnership, *United to End TB - Every Word Counts*. 2015.
11. Vassall, A., 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): p. e1001120.
12. McDowell, A. and M. Pai, Treatment as diagnosis and diagnosis as treatment: empirical management of presumptive tuberculosis in India. Int J Tuberc Lung Dis, **2016**; 20(4): p. 536-43.
13. Tostmann, A., et al., Antituberculosis drug-induced hepatotoxicity: concise up-to-date review. J Gastroenterol Hepatol, **2008**; 23(2): p. 192-202.
14. Laurence, Y.V., U.K. Griffiths, and A. Vassall, Costs to Health Services and the Patient of Treating Tuberculosis: A Systematic Literature Review. Pharmacoeconomics, **2015**; 33(9): p. 939-55.
15. Tanimura, T., et al., Financial burden for tuberculosis patients in low- and middle-income countries: a systematic review. Eur Respir J, **2014**; 43(6): p. 1763-75.
16. Munro, S.A., et al., Patient adherence to tuberculosis treatment: a systematic review of qualitative research. PLoS Med, **2007**; 4(7): p. e238.
17. Blount, R., et al., The Effect Of False Positive And False Negative Microscopy Results On Mortality Among HIV-infected Ugandans Undergoing Evaluation For TB. **2010**. A2257-A2257.
18. Singh, V.K., et al., A common medical error: lung cancer misdiagnosed as sputum negative tuberculosis. Asian Pac J Cancer Prev, **2009**; 10(3): p. 335-8.
19. Thornsberry, C., et al., Rifampin: spectrum of antibacterial activity. Rev Infect Dis, **1983**; 5 Suppl 3: p. S412-7.
20. Moreira, J., et al., Weighing harm in therapeutic decisions of smear-negative pulmonary tuberculosis. Med Decis Making, **2009**; 29(3): p. 380-90.
21. Theron, G., 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): p. 424-35.
22. Churchyard, G.J., et al., Xpert MTB/RIF versus sputum microscopy as the initial diagnostic test for tuberculosis: a cluster-randomised trial embedded in South African roll-out of Xpert MTB/RIF. Lancet Glob Health, **2015**; 3(8): p. e450-7.
23. Durovni, B., et al., Impact of replacing smear microscopy with Xpert MTB/RIF for diagnosing tuberculosis in Brazil: a stepped-wedge cluster-randomized trial. PLoS Med, **2014**; 11(12): p. e1001766.

24. Cox, H.S., et al., Impact of Xpert MTB/RIF for TB Diagnosis in a Primary Care Clinic with High TB and HIV Prevalence in South Africa: A Pragmatic Randomised Trial. PLoS Med, **2014**; 11(11): p. e1001760.

25. Kendall, E.A., et al., Estimated clinical impact of the Xpert MTB/RIF Ultra cartridge for diagnosis of pulmonary tuberculosis: A modeling study. PLoS Med, **2017**; 14(12): p. e1002472.

26. Ukwaja, K.N., et al., The high cost of free tuberculosis services: patient and household costs associated with tuberculosis care in Ebonyi State, Nigeria. PLoS One, **2013**; 8(8): p. e73134.

27. McIntyre, D., et al., What are the economic consequences for households of illness and of paying for health care in low- and middle-income country contexts? Soc Sci Med, **2006**; 62(4): p. 858-65.

28. Trauer, J.M., et al., Modelling the effect of short-course multidrug-resistant tuberculosis treatment in Karakalpakstan, Uzbekistan. BMC Med, **2016**; 14(1): p. 187.

29. Menzies, N.A., et al., Cost-effectiveness and resource implications of aggressive action on tuberculosis in China, India, and South Africa: a combined analysis of nine models. Lancet Glob Health, **2016**.

30. Sumner, T., et al., *Incorporating resource constraints into mathematical models of TB: an example of case finding in South Africa*, in *World Conference on Tuberculosis and Lung Health*. 2017: Guadalajara, Mexico.

31. World Health Organisation, *Global Tuberculosis Report 2017 - Technical Appendix - Methods to estimate the global burden of disease*. 2017: Geneva.

32. Kyu, H.H., et al., The global burden of tuberculosis: results from the Global Burden of Disease Study 2015. The Lancet Infectious Diseases, **2018**; 18(3): p. 261-284.

33. Vickers, A.J., B. Van Calster, and E.W. Steyerberg, Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ, **2016**; 352: p. i6.

34. Van den Ende, J., et al., Effect of applying a treatment threshold in a population. An example of pulmonary tuberculosis in Rwanda. J Eval Clin Pract, **2010**; 16(3): p. 499-508.

35. Basinga, P., et al., Why are clinicians reluctant to treat smear-negative tuberculosis? An inquiry about treatment thresholds in Rwanda. Med Decis Making, **2007**; 27(1): p. 53-60.

36. Schumacher, S.G., et al., Diagnostic Test Accuracy in Childhood Pulmonary Tuberculosis: A Bayesian Latent Class Analysis. Am J Epidemiol, **2016**; 184(9): p. 690-700.

37. Onozaki, I., et al., National tuberculosis prevalence surveys in Asia, 1990-2012: an overview of results and lessons learned. Trop Med Int Health, **2015**; 20(9): p. 1128-1145.

38. Yoon, C., et al., Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. Lancet Infect Dis, **2017**; 17(12): p. 1285-1292.

39. Lawn, S.D., et al., Diagnostic accuracy, incremental yield and prognostic value of Determine TB-LAM for routine diagnostic testing for tuberculosis in HIV-infected patients requiring acute hospital admission in South Africa: a prospective cohort. BMC Med, **2017**; 15(1): p. 67.

Tables

Table 1: False-positive TB diagnoses in hypothetical screening programmes

Prevalence in screening population	Algorithm 1: Any symptom --> smear --> empirical diagnosis		Algorithm 2: Any symptom or X-Ray --> GeneXpert		Impact of new algorithm in Population = 100,000		
	PPV	Ratio TP / FP	PPV	Ratio TP / FP	Change in # TB diagnoses	From FN to TP	From FP to TN
10%	73%	1:0.37	95%	1:0.05	+2261	3767	1506
5%	61%	1:0.64	90%	1:0.11	+1369	2271	902
1%	29%	1:2.50	64%	1:0.56	+157	526	369
0.5%	17%	1:4.78	67%	1:1.13	-30	268	298
0.15%	6%	1:15.29	21%	1:3.78	-166	82	248

PPV = Posititive Predictive Value, i.e. % of individuals with TB diagnosis that actually have TB disease, FP = False-positive, TP = True-positive. Cell highlighting shows scenarios where the majority of diagnosed cases are expected to be true positive cases (green) or false-positive TB diagnosis (red).

Algorithm 1: *Individuals who are positive on screen with any symptom are tested for TB using sputum smear microscopy. A proportion of individuals with a negative smear continue to empirical diagnosis. The probability of being referred towards empirical diagnosis following a negative smear is negatively correlated with the prevalence of disease in the screening population between 80% and 5%, as the negative predictive value following a negative smear decreases.*

Algorithm 2: *Individuals who screen positive either with any symptom of show any abnormality on Chest X-Ray are tested with GeneXpert. Here we assume no empirical diagnosis so that individuals with no symptoms and no CXR abnormality or a negative GeneXpert result are ruled out from having TB disease.*

In both scenarios, we assume that the prevalence of smear positive TB disease is negatively correlated with the prevalence of disease in the screening population between 60% and 40% to reflect expanding case-detection towards the general population, therefore detecting earlier disease. The screening and specificity values were obtained from [8] and [11]

Figure legends

Figure 1: Screening and diagnostic pathway for tuberculosis (TB)
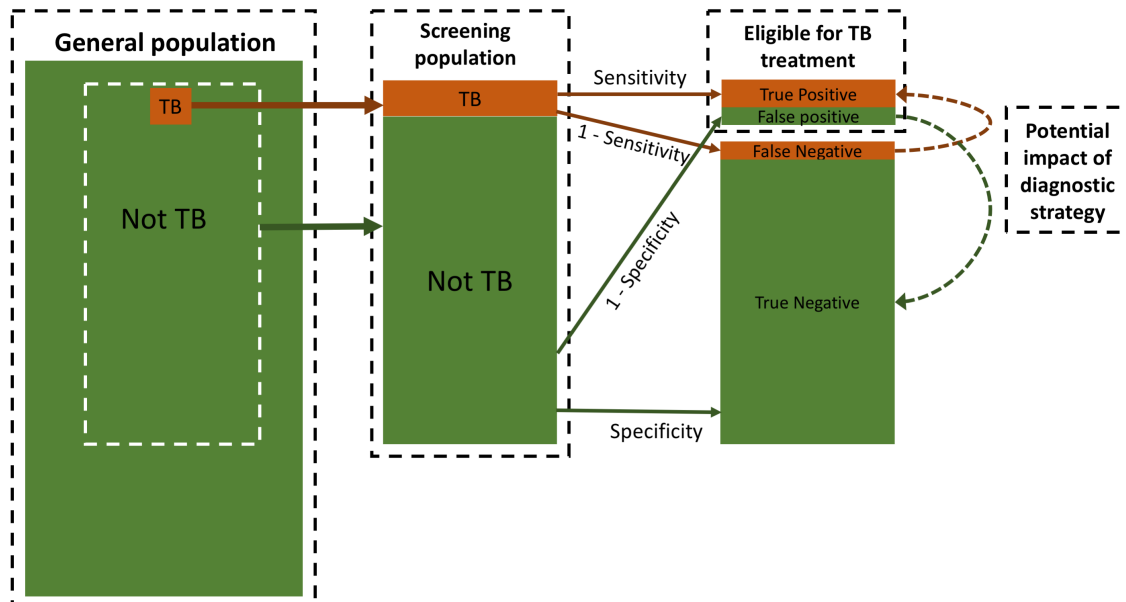
*From a general population, a screening population is formed from individuals with (orange) and without TB (green). The diagnostic algorithm is applied to the screening population, categorising individuals into those recommended for TB treatment (following a true-positive or false-positive diagnosis) or not. The contribution of false-positive TB diagnoses is mostly driven by the prevalence of TB in the screening population and the specificity of the diagnostic algorithm (see **Table 1**). The dashed arrows on the right highlight the two processes that new screening or diagnostic strategies aim to achieve: orange = convert false-negative diagnoses into true-positive, green = convert false-positive diagnoses into true-negative diagnoses.*

Figure 2a-b: Change in positive predictive value by varying sensitivity (2a) or specificity (2b)
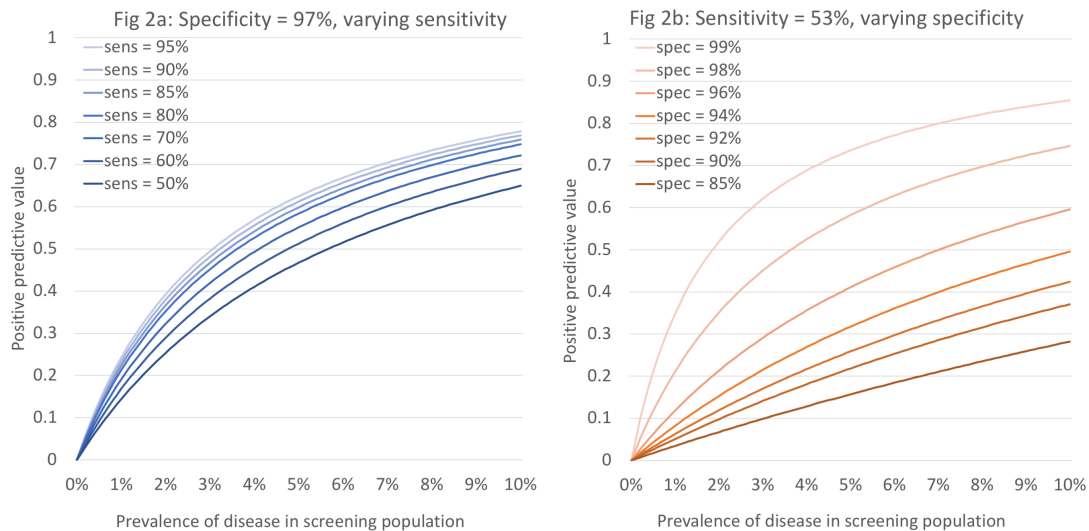
Figures show relationship between Positive Predictive Value (% of individuals with TB diagnosis that actually have TB disease) and prevalence of disease in screening population for combinations of sensitivity and specificity. Lines in Figure 2a (left) show how relationship changes if specificity for algorithm 1 (Table 1) remains constant at 97%, but sensitivity increases. Lines in Figure 2b (right) show how relationship changes if sensitivity remains constant at 53%, but specificity increases or decreases. See Table 1 for details of algorithm 1.