

RESEARCH ARTICLE

# Impact of Molecular Diagnostics for Tuberculosis on Patient-Important Outcomes: A Systematic Review of Study Methodologies

Samuel G. Schumacher<sup>1,2</sup>, Hojoon Sohn<sup>1,2</sup>, Zhi Zhen Qin<sup>1,2</sup>, Genevieve Gore<sup>3</sup>, J. Lucian Davis<sup>4</sup>, Claudia M. Denking<sup>1,2,5</sup>, Madhukar Pai<sup>1,2\*</sup>

**1** McGill University Department of Epidemiology & Biostatistics, Montreal, Canada, **2** McGill International TB Centre, Montreal, Canada, **3** McGill University, Schulich Library of Science and Engineering, Montreal, Canada, **4** UCSF Pulmonary & Critical Care Medicine, San Francisco, United States of America, **5** Beth Israel Deaconess Medical Centre, Division of Infectious Disease, Boston, MA, United States of America

\* [madhukar.pai@mcgill.ca](mailto:madhukar.pai@mcgill.ca)



OPEN ACCESS

**Citation:** Schumacher SG, Sohn H, Qin ZZ, Gore G, Davis JL, Denking CM, et al. (2016) Impact of Molecular Diagnostics for Tuberculosis on Patient-Important Outcomes: A Systematic Review of Study Methodologies. *PLoS ONE* 11(3): e0151073. doi:10.1371/journal.pone.0151073

**Editor:** Delia Goletti, National Institute for Infectious Diseases (L. Spallanzani), ITALY

**Received:** December 22, 2015

**Accepted:** February 23, 2016

**Published:** March 8, 2016

**Copyright:** © 2016 Schumacher et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This project was funded by a grant from the Canadian Institutes of Health Research, Ottawa, ONT, Canada (grants MOP-89918). MP is supported by the Fonds de Recherche du Québec – Santé, Montreal, QC, USA. CMD was supported by a Richard Tomlinson Fellowship at McGill University and a fellowship of the Burroughs-Wellcome Fund from the American Society of Tropical Medicine and Hygiene, Deerfield, IL, USA. SGS was supported by the Max Binz Fellowship at McGill University. The

## Abstract

### Background

Several reviews on the accuracy of Tuberculosis (TB) Nucleic Acid Amplification Tests (NAATs) have been performed but the evidence on their impact on patient-important outcomes has not been systematically reviewed. Given the recent increase in research evaluating such outcomes and the growing list of TB NAATs that will reach the market over the coming years, there is a need to bring together the existing evidence on impact, rather than accuracy. We aimed to assess the approaches that have been employed to measure the impact of TB NAATs on patient-important outcomes in adults with possible pulmonary TB and/or drug-resistant TB.

### Methods

We first develop a conceptual framework to clarify through which mechanisms the improved technical performance of a novel TB test may lead to improved patient outcomes and outline which designs may be used to measure them. We then systematically review the literature on studies attempting to assess the impact of molecular TB diagnostics on such outcomes and provide a narrative synthesis of designs used, outcomes assessed and risk of bias across different study designs.

### Results

We found 25 eligible studies that assessed a wide range of outcomes and utilized a variety of experimental and observational study designs. Many potentially strong design options have never been used. We found that much of the available evidence on patient-important outcomes comes from a small number of settings with particular epidemiological and operational context and that confounding, time trends and incomplete outcome data receive insufficient attention.

funders had no role in the study design, data collection, analysis, decision to publish, or manuscript preparation.

**Competing Interests:** The authors declare that they have no financial or industry conflicts of interest. MP serves as a consultant for the Bill and Melinda Gates Foundation (Seattle, WA, USA), which had no involvement in this study. At the time of conception, conduct and write-up of this research SGS and HS were PhD students and CMD was a postdoctoral fellow at McGill University, Montreal, Canada. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

**Abbreviations:** TB, tuberculosis; NAAT, Nucleic Acid Amplification Tests; WHO, World Health Organization; LPA, Line Probe assay; Xpert, Xpert® MTB/RIF assay; MTB, *Mycobacterium tuberculosis*; RIF, Rifampicin; HIV, human immunodeficiency virus; RCT, randomized controlled trials.

## Conclusions

A broader range of designs should be considered when designing studies to assess the impact of TB diagnostics on patient outcomes and more attention needs to be paid to the analysis as concerns about confounding and selection bias become relevant in addition to those on measurement that are of greatest concern in accuracy studies.

## Introduction

Tuberculosis (TB) continues to take a massive toll on human health globally, causing 1.5 million deaths and 9 million new cases annually[1]. The World Health Organization (WHO) recently set ambitious targets in the End TB Strategy to eradicate TB globally, with accurate and rapid detection of TB and drug resistance as critical components of their strategy[2]. Although several new diagnostics are now available for TB, sputum smear microscopy continues to be widely used, despite its limited sensitivity and inability to detect drug-resistance[3]. Mycobacterial culture has high sensitivity but takes weeks or months to yield results, such that its impact on clinical decision-making and patient-important outcomes is often limited[4–7].

Newer molecular TB diagnostics (nucleic acid amplification tests, NAATs) have been shown to have good accuracy and can produce rapid results[8–10], characteristics which have led to their endorsement by the WHO[11]. However, as it has become apparent that higher accuracy does not necessarily translate into improved patient care, policy makers have begun to demand more direct evidence that improved diagnostics positively affect health or other outcomes that matter to patients[12–14]. Much uncertainty remains about how new tests should be implemented to maximize their impact, which patient-important outcomes should be measured, and which designs should be used to assess them[13,15].

Given the recent increase in research evaluating the impact of TB NAATs on patient-important outcomes and the growing list of TB NAATs expected to reach the market in coming years, there is a need to summarize the existing evidence and best practices on methodologies for evaluating the impact of TB diagnostics[16]. Various designs have been used to measure a wide range of patient-important outcomes, but no systematic review has been done of such studies and the specific methodological issues that arise in them.

In this systematic review we aim to critically assess the approaches that have been employed to measure the impact of TB NAATs on patient-important outcomes (such as time to treatment initiation or mortality) in adults with possible pulmonary TB and/or drug-resistant TB from currently available evidence. The specific objectives of the systematic review of available studies were (i) to develop a conceptual framework for relevant patient-important outcomes and to assess which outcomes have been measured; (ii) to outline which designs and methodologies might be employed to measure the impact of diagnostics on patient-important outcomes and to assess which ones have been utilized; (iii) to propose criteria for assessing risk of bias for each design based on sound epidemiological principles where such tools do not already exist and to assess risk of bias and quality of reporting.

## Methods

We drafted a protocol before commencing the review, following standard guidelines[17,18]. We then developed a conceptual framework to clarify how improved test performance may lead to improved patient health via intermediate outcomes[19–23]. We also specified a framework for classifying research designs that have been (or could be) used to assess impact, using

standard classifications from the field of epidemiology, clinical and diagnostic research[24–26], economics, and social sciences[27], and included designs only recently[28,29] or never previously suggested to our knowledge. This framework provides structure to the review and points to designs that have not been used, and provides a basis for assessing risk of bias, because threats to validity may differ between designs.

## Eligibility Criteria

We restricted our review to adult pulmonary TB, including drug-resistant TB, because diagnosis of childhood and extrapulmonary TB pose special diagnostic challenges and warrant consideration of contextual factors and methodological issues outside the scope of this review. We focused on the WHO-approved TB NAATs, i.e. two Line Probe assays (the GenoType MTBDRplus and the Inno-LiPA Rif.TB, both referred to simply as LPAs from here on) and the Xpert® MTB/RIF assay (Xpert), because these tests are currently of greatest interest globally and are being implemented in many countries around the world. We included only peer-reviewed studies that measured at least one patient-important outcome, i.e. outcomes that directly reflect some improvement in the patient's experience that may directly affect health (such as a more rapid diagnosis, more rapid initiation of treatment, reduced mortality etc.). While cost is an important factor for patients, this was outside of the scope of this review and studies focusing on cost were excluded. We did not exclude any study design and did not restrict based on region, setting, years, or language. We only considered studies that included primary data, and excluded meta-analyses and compartmental and decision-analytic modeling studies. Studies that provided diagnostic test accuracy only were excluded.

## Information Sources

We searched MEDLINE, EMBASE, Web of Knowledge and Cochrane CENTRAL through January 31, 2015. We also searched the metaRegister of Controlled Trials (mRCT) and the WHO International Clinical Trials Registry Platform to identify ongoing trials. The full electronic search strategy can be found in '[S1 Appendix](#)'. To identify additional studies, we reviewed reference lists of included articles and of systematic reviews on the diagnostic accuracy of NAATs, and contacted researchers at FIND, members of the Stop TB Partnership's New Diagnostics Working Group, and other experts on TB diagnostics.

## Study Selection

Two review authors (SGS, ZQ) independently assessed titles and abstracts identified by electronic literature searching to identify potentially eligible studies (screen 1). Any citation identified by either review author during screen 1 was selected for full-text review. Two authors (SGS, HS) independently assessed articles for inclusion using predefined inclusion and exclusion criteria (screen 2), with discrepancies resolved by discussion. We maintained a list of excluded studies by reason for exclusion.

## Data Collection Process and Data Items

We developed a standardized data extraction form using Google Forms (Google Inc., Mountain View, CA, USA) to minimize data-entry errors[30]. Two authors (SGS, HS) piloted and revised the form to improve clarity. They then independently extracted data on study design, key contextual factors, patient-important outcomes, results for these outcomes and on design-specific risks for bias on a quarter (6/25) of studies. For the remainder of the studies, one author (SGS) extracted the data and the second (HS) crosschecked all extracted items.

## Risk of Bias in Individual Studies

We assessed risk of bias of individual studies with component questions for risk of bias assessment dependent on the study design. For randomized controlled trials we used the Cochrane risk of bias tool [31]. For the other study designs there is currently no suitable validated tool for risk of bias assessment [32]. We therefore decided on methodological components separately for each study design that likely place a study at higher risk of bias. Our choices were informed by other existing tools for risk of bias assessment [33–35], approaches taken by health technology assessment units [36–38] as well as generally recognized epidemiological and statistical principles [24]. The design-specific questions and guidance for how to judge each item is available in ‘S2 Appendix’. We did not attempt any assessment of risk of bias across studies.

## Synthesis of Results

We summarized characteristics of included studies by design and index test. Because of heterogeneity at all levels of abstraction, we did not plan to provide a meta-analysis, but instead a narrative synthesis of results structured around the conceptual frameworks. We used descriptive statistics to summarize key characteristics that we abstracted, stratified by study design. We used vote counting to assess how often authors reported that use of a TB NAAT positively influenced a particular outcome. We displayed results from the bias risk assessment graphically by study design across included studies and for each individual study. We calculated binomial 95% confidence intervals if not reported. Data management and descriptive statistics were done using STATA (version 12, Stata Corp, College Station, Texas, USA).

## Results

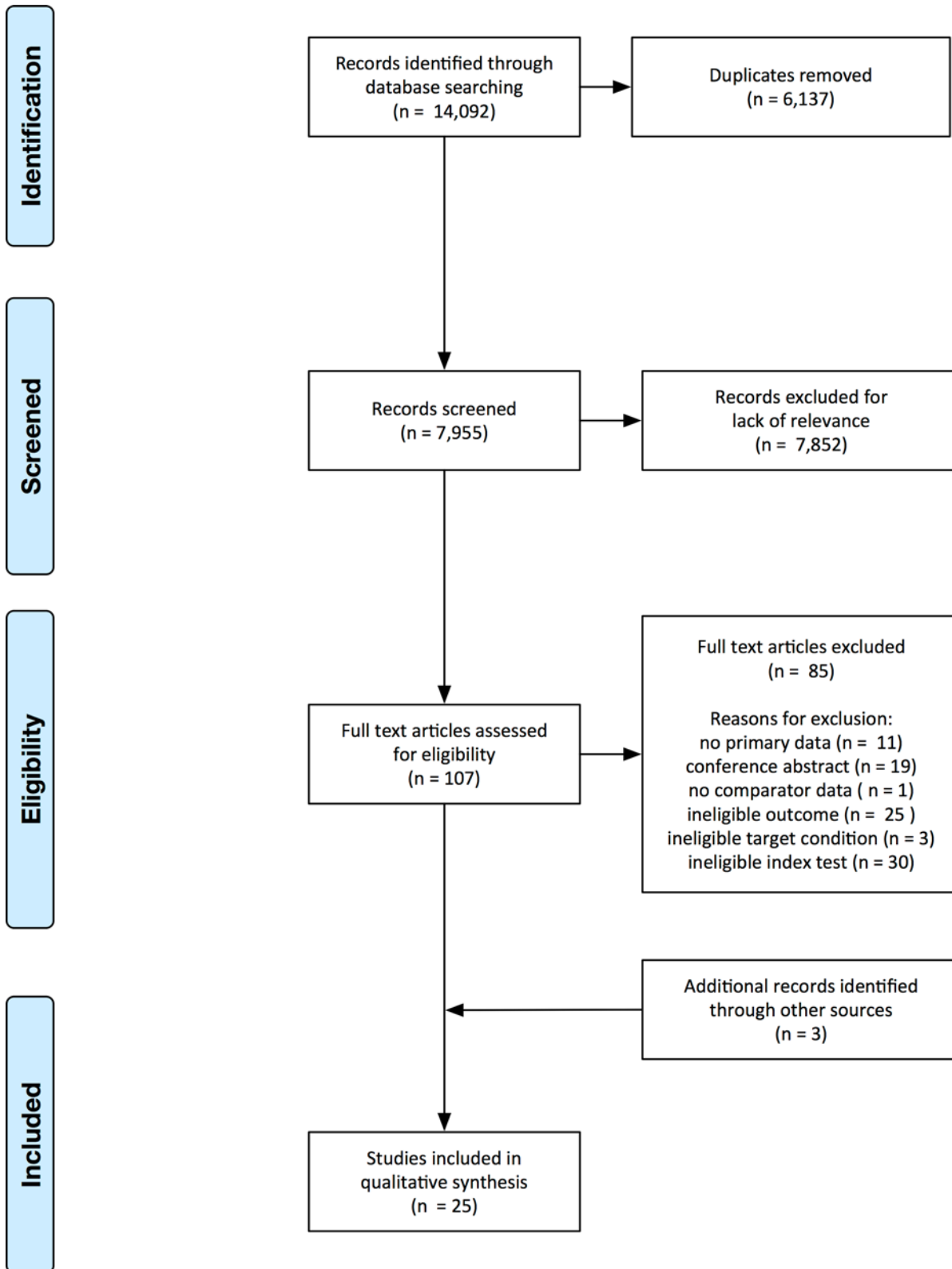
### Study Selection

We screened the 7,995 abstracts that remained after removal of duplicates and identified 107 potentially eligible articles for which we obtained the full texts (Fig 1). We excluded 85 articles for reasons listed in Fig 1. We identified three additional studies, for a total of 25 included studies.

### Study Characteristics and Contextual Factors

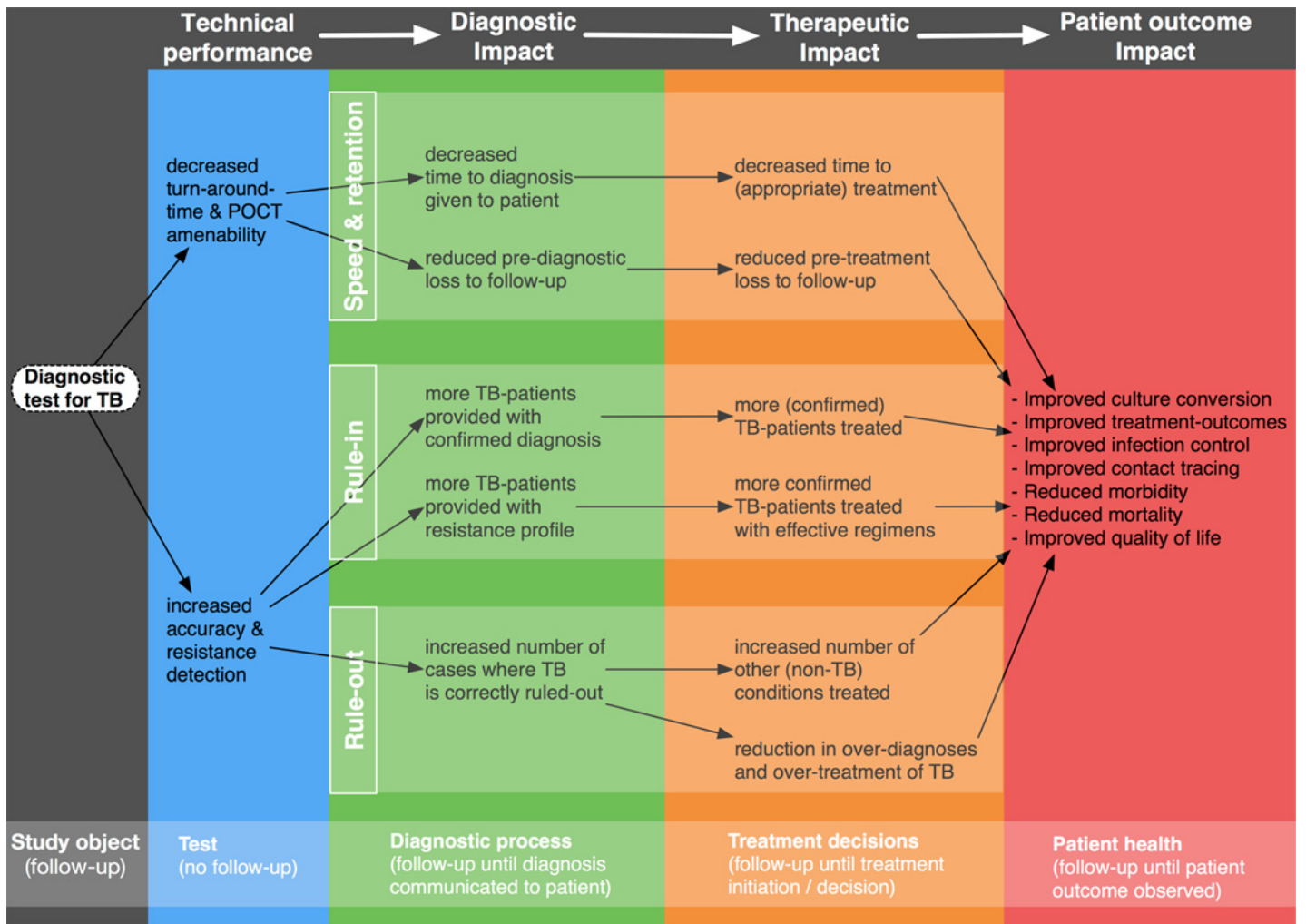
Of 25 included studies (study characteristics in ‘S3 Appendix’), eleven (44%) were done in South Africa or involved a study site in South Africa as part of a multi-center study [39–48]. Most studies were carried out in urban hospitals (56%,  $n = 14$  [29,46,49–59]) or urban primary care centers (32%,  $n = 8$  [39,40,42,43,47,48,60]) with chest radiography reported to be available for TB diagnosis in almost three-quarters of studies (72%,  $n = 18$ ) [29,40,43,44,46–55,58–60]. Of the 22 studies providing results from human immunodeficiency virus (HIV)-testing, 12 (55%) reported HIV prevalence greater than 50% [39,40,42,43,45–49,55,59]. Losses to follow-up before treatment initiation were reported in 10 studies and ranged from 7% to 39% [39,40,42,46–48,55,56,59]. The proportion of patients treated empirically (i.e. without microbiological confirmation) was reported in 12 studies and ranged from 10% to 69% [39,40,44,46–48,52,55,58–60].

Samples were transported for off-site testing in a laboratory in a different location from where patient care took place in 12 studies (48%) [29,40–43,45,46,49,51,54,55,60], and were tested on-site in the other 12 [39,44,47,48,50,52,53,56,58,59] (location was unclear in the one remaining study) [57]. Of the 18 studies evaluating Xpert—a test yielding results within 2h—only four (22%) [48,53,58] attempted to implement the test embedded within a point-of-care testing program, where the goal is same-day treatment initiation. No other co-interventions aimed at ensuring or improving the continuum of care—from patient screening, through diagnostic testing, providing the TB or alternative diagnosis, ensuring linkage to care (avoiding loss



**Fig 1. Study selection.** Flow diagram of studies in the review. Note: Some studies had more than one reason for exclusion.

doi:10.1371/journal.pone.0151073.g001



**Fig 2. Conceptual framework of outcome measures.** Conceptual framework outlining the pathways through which improved TB diagnostics may lead to improved patient outcomes.

doi:10.1371/journal.pone.0151073.g002

to follow-up), initiating TB or alternative treatment, providing support for treatment adherence and ensuring successful treatment outcomes—were reported.

### Patient-Important Outcomes

The conceptual framework of the main pathways through which new TB diagnostics may affect patient health outcomes is shown in Fig 2. For clarity, and for consistency with what has been suggested previously [61], we differentiate four categories of outcome measures based on their object of study and required follow-up as indicated at the bottom of the figure: (i) measures of technical performance, (ii) measures of diagnostic impact, (iii) measures of therapeutic impact and (iv) measures of impact on patient outcomes. Definitions and examples of these four categories of outcome measures are provided in Table 1.

Measures of technical performance (e.g. test accuracy) were often reported but not the focus of this review. The frequency of reporting of the various patient-important outcomes across included studies is shown in Fig 3. The most common measure of diagnostic impact was the ‘number of patients with confirmed diagnosis’ (28%, n = 7). We found that most studies

included at least one measure of therapeutic impact with ‘reduced time to treatment initiation’ representing by far the most commonly used outcome measure overall (84%, n = 21), followed by ‘reduced loss to follow-up’ (36%, n = 9). Only about half of studies (n = 13) reported on a measure of patient outcome impact, mostly on mortality (24%, n = 6) and outcomes related to infection control or contact tracing (20%, n = 5).

No studies assessed the potential effects on increasing the number of diagnoses of other respiratory diseases, the number of patients treated with an effective regimen or health-related quality of life measures.

While included studies universally (100%, 14/14) showed benefit of the TB NAATs on the assessed measures of diagnostic impact and none showed any harm, only about three quarters (76%, 29/38) of measures of therapeutic impact and only half (50%, 8/16) of measures of impact on patient outcomes were shown to improve. None (0%, 0/7) of the studies was able to show a benefit on morbidity or mortality. However, point-estimates for changes in mortality and other measures of patient outcome impact often suggested improvements in the TB NAAT arm, but since these outcomes are relatively rare, confidence intervals were wide such that a “null-effect” could not be excluded. Unfortunately this was frequently interpreted as “no difference between arms” but wide intervals that include the null should not be interpreted as evidence favoring the null hypothesis[62,63]. For example, if a mortality-reduction of 25% or more is considered important from a clinical or public health standpoint, all findings on this outcome were inconclusive, rather than proving the absence of any relevant effect.

### Designs to Assess Impact of Diagnostic Tests

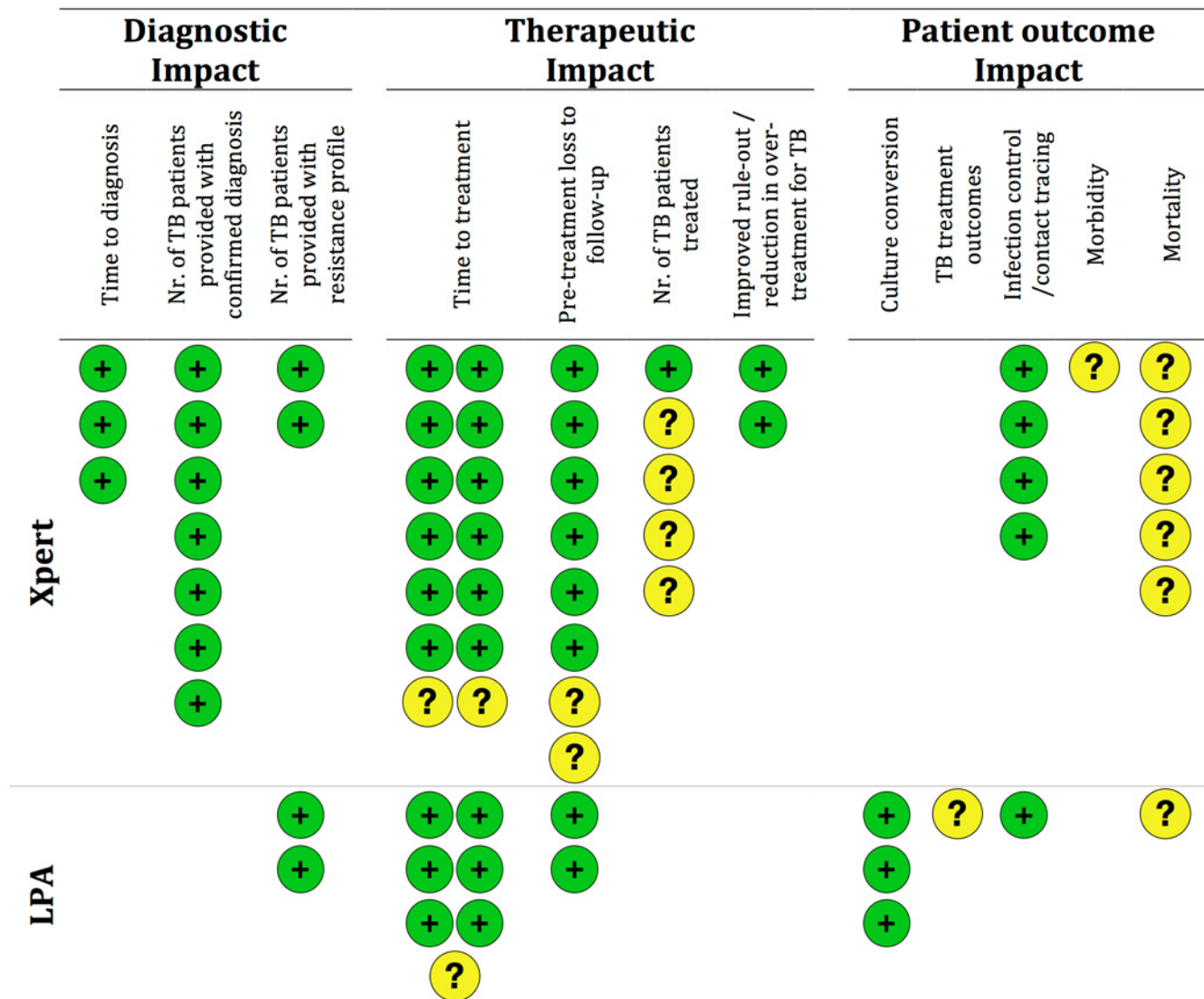
The classification of design options as used in this review is shown in Fig 4. We first differentiate two major design architectures: (i) multi-cohort designs, i.e. designs that involve two (or more) cohorts, each exposed to only a single test or testing algorithm and (ii) single-cohort designs, i.e. designs that involve a single cohort exposed to multiple tests simultaneously. We then further differentiate important sub-types of these and brief descriptions for each design with references for further details in Table 2.

As shown in Table 2, we identified two individually-randomized trials, two parallel-arm cluster-randomized trials and one stepped-wedge cluster-randomized trial. There were nine studies using a quasi-experimental design, all of which were uncontrolled pre/post implementation studies. We also found four single-cohort hypothetical studies and seven single-cohort observational studies but no diagnostic before/after studies. Authors did not provide any name for their study design in three studies and 13 different names were suggested by study authors for the remaining 17 non-randomized studies; we classified all 20 non-randomized studies into one of three study designs (Table 2). RCTs (randomized controlled trials) tended to assess outcomes across the whole range (i.e. from diagnostic impact to patient outcome impact) while pre/post and observational studies mostly assessed therapeutic impact and hypothetical studies mostly patient outcomes (Fig 5).

**Table 1. Definitions and examples of categories of outcome measures.**

Outcome measure	Definition	Examples
<b>Technical performance</b>	Outcome measures relating to the technical performance characteristics of the test.	Diagnostic accuracy, laboratory turn-around-time
<b>Diagnostic impact</b>	Outcome measures relating to how the test affects diagnosis and diagnostic processes.	Time to diagnosis, number of confirmed diagnoses provided
<b>Therapeutic impact</b>	Outcome measures relating to how the test affects treatment decisions.	Time to treatment, number of patients placed on appropriate therapy
<b>Patient outcome impact</b>	Outcome measures relating to patient health and/or quality of life.	TB treatment outcomes, mortality.

doi:10.1371/journal.pone.0151073.t001



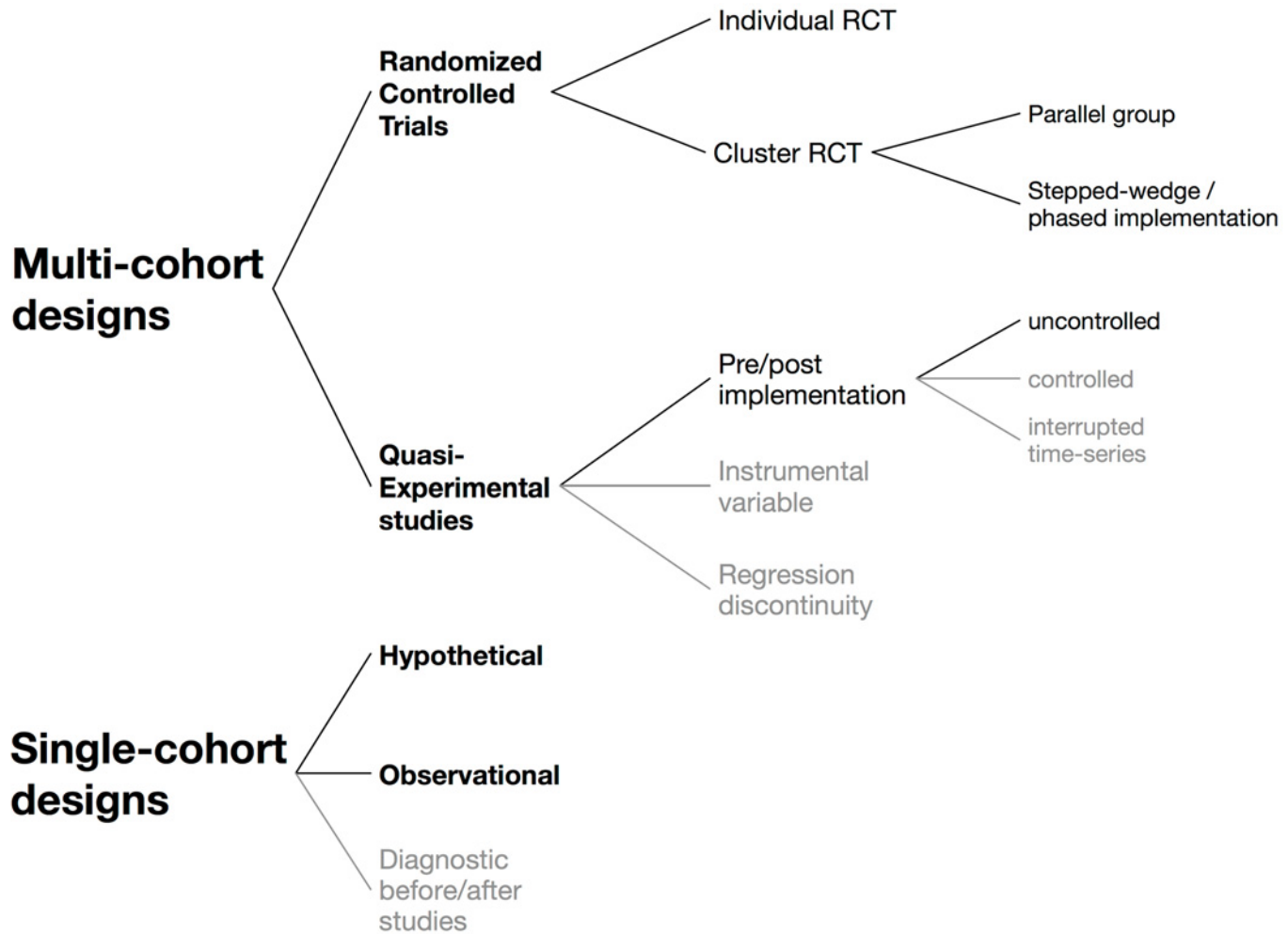
**Fig 3. Reporting and vote counting of results on the different outcome measures.** Each circle represents one study. Green circles represent a study finding that the TB NAAT improved the outcome, yellow circles represent a study with inconclusive findings where confidence intervals of the effect estimate included clinically relevant improvements or where confidence intervals were not provided and raw data for re-calculation was not accessible from the manuscript. Note: One study assessed both Xpert and LPA and is accounted for in both the upper section on Xpert and the lower section on LPA; some outcomes shown in Fig 2 were not reported in any study

doi:10.1371/journal.pone.0151073.g003

### Risk of Bias within Studies

Results from the assessment of risk of bias are shown separately for each study design across included studies in Fig 6 and for each individual included study in 'S4 Appendix'. For all designs incomplete outcome data was more common for measures of patient outcome impact than for outcomes further upstream in the causal chain. Where outcome data was missing, regression adjustment, imputation or sensitivity analyses were rarely attempted and this may have led to bias in some studies[95–97].





**Fig 4. Design options to study the impact of TB diagnostics on patient health outcomes.** Designs that have not been used in any of the studies included in this review are shown in grey. Of note, quasi-experimental studies are not typically described in epidemiological textbooks but are popular among economists and other social scientists: the basic idea of these designs is to try to make causal inference by exploiting some source of exogenous variation that acts similar to randomization. The three designs listed here may appear to be quite different but have the common feature that the type of exposure/test is neither the choice of the study participants (as in traditional cohort studies) nor assigned by the investigator (as in randomized trials) but determined through some exogenous factor. Pre/post implementation studies, where ‘time’ represents this exogenous factor, were the only quasi-experimental design used in the included studies.

doi:10.1371/journal.pone.0151073.g004

For all randomized trials, blinding of physicians to what test was done was impossible since knowing which test was done is part of the intervention itself. For example, the Xpert test has higher sensitivity than smear microscopy (and also produces RIF resistance results) and physicians must be allowed to take this into account when deciding about patient management. While outcomes between patients may therefore be different due to lack of blinding this was not judged to be a source of bias but rather the mechanism through which the intervention had an effect. Outcome measurement could theoretically have been influenced by the lack of blinding but this was deemed unlikely to cause bias of important magnitude. Overall, the lack of blinding was therefore judged not to put studies at increased risk of bias. Allocation concealment was impossible for cluster RCTs since once a cluster had been assigned to one of the intervention arms, allocation of this cluster was fixed and therefore unconcealed for the remainder of the study. However, if this had led to selection bias one would expect to see

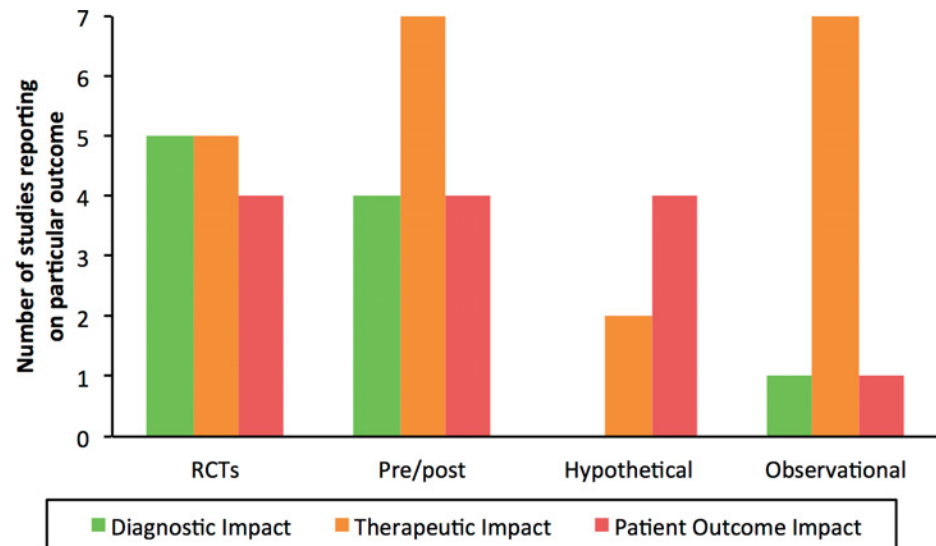
**Table 2. Descriptions of design options to study the impact of TB diagnostics on patient-important outcomes with references on methodological issues.**

Design category (and methodological references)	Design sub-type (and methodological references)	Description	References of studies included in the review
<b>Randomized Controlled Trials (RCTs)</b> [25,64]	<b>Individual RCT</b> [65]	Individuals are randomized to either receive or not receive the intervention	[55,66]
	<b>Cluster RCT</b> [25,67–70]	<u>Parallel group</u> : Clusters are randomized to either receive or not receive the intervention throughout the entire study period	[39,40]
		<u>Stepped-wedge</u> : The sequence in which clusters receive the intervention is randomized such that all clusters receive it by the end of the study (also called phased-implementation trial)	[60]
<b>Quasi-Experimental studies</b> [23,27,71–73]	<b>Pre/post implementation</b> [74–76]	<u>Uncontrolled</u> : Two cohorts are compared between two different time periods: the <i>pre</i> cohort receives the usual care during the baseline period and the <i>post</i> cohort receives the intervention during a subsequent and distinct time period.	[41–45,51,56,57,59]
		<u>Controlled</u> : Like ‘Uncontrolled pre/post implementation’ but including a contemporary control cohort that receives the usual care during both the <i>pre</i> and the <i>post</i> period, also called ‘difference in differences design’.	none
	<b>Interrupted time-series</b> [77–80]	Multiple measurements over time before and after implementation of the intervention analyzed using segmented regression or ARIMA models.	none
	<b>Instrumental variable</b> [81–85]	The effect of the intervention on the outcome is captured through another variable (the “instrument”) that affects the outcome only by affecting the intervention and does not share any causes with the outcome.	none
	<b>Regression discontinuity</b> [86–89]	The effect of the intervention on the outcome can be estimated if individuals receive the intervention based on whether they are above or below some threshold value on a continuous variable.	none
<b>Single-cohort designs</b>	<b>Hypothetical studies</b> [28]	A single cohort receives both baseline tests and the index test but results from the index tests are not used for patient management. “Hypothetical” changes in patient-important outcomes—had results been available to doctors—are estimated using a combination of study data, assumptions and potentially data from other studies.	[29,53,58,90]
	<b>Observational studies</b>	Inferences about the effect of the index test on patient-important outcomes are attempted based on a single cohort receiving both baseline tests and the index test with both being used for patient management.	[46–50,52,54]
	<b>Diagnostic before/after studies</b> [61,91–94]	Inferences about the effect of the index test on patient-important outcomes are attempted based on comparisons of the pre-test management plan (i.e. planned patient management before availability of index test results) and post-test management plan.	none

doi:10.1371/journal.pone.0151073.t002

relevant covariate-imbalance between the intervention arms—this was not observed and therefore bias of important magnitude from this source is relatively unlikely.

For pre/post implementation studies, changes of patient eligibility criteria that occurred concurrently with the change of test used in the *pre* versus the *post* period were an important threat to validity. This was usually the result of a policy broadening eligibility for drug-resistance testing from a high-risk group in the *pre* period to all patients being evaluated for TB in the *post* period, implemented alongside the introduction of the TB NAAT. The resulting selection bias may be hard or impossible to address analytically. Temporal trends (usually changes in quality of care or care delivery) pose another potential risk of bias and study authors sometimes discussed this. A figure explaining how these mechanisms lead to lack of exchangeability



**Fig 5. Frequency of studies reporting on one or several outcomes within the three outcome categories by study design.**

doi:10.1371/journal.pone.0151073.g005

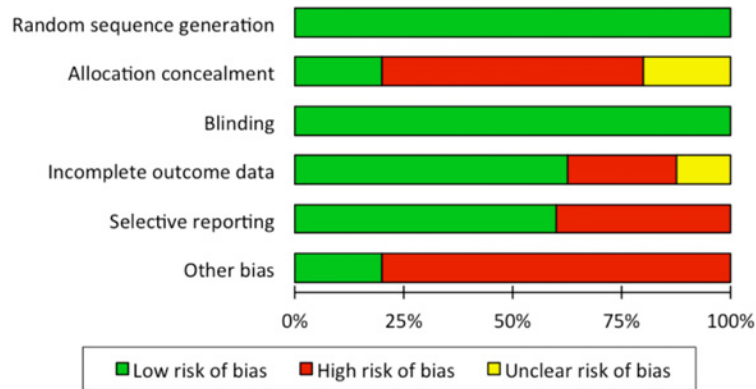
and how this could potentially be addressed analytically is shown in ‘S5 Appendix’. Only three studies addressed the potential for confounding by providing either justification for why outcomes were likely unaffected or additional analysis. A table comparing pre/post cohorts was often shown but attempts to address existing covariate imbalances, e.g. via regression, were rare. If adjustment was attempted, methods were often described in insufficient detail (e.g. no explanation how continuous covariates were modeled[98,99]) and strategies for model selection were either not described or a method known to be prone to bias was used (e.g. inclusion based on  $p < 0.05$ ) [100–103]. As mentioned earlier, controlled pre/post implementation studies (and extensions via matched cohort designs[104]) were not identified, although such designs are preferable to protect the robustness of estimates from bias related to temporal trends.

Since in single-cohort studies comparisons are made *within* individuals, confounding and selection bias are not the main concern. The challenge for these designs lies in the assumptions one needs to make to draw valid conclusions about the effect of the index test. For the single-cohort observational studies, assumptions remained implicit and were often not justified. For example, patients testing negative on smear microscopy but positive with the TB NAAT were usually assumed to have received TB therapy *due to* the TB NAAT; however, this implicitly assumes that none of these patients would have been treated empirically, which is unlikely to hold true in most settings, leading to overestimates in the value of the test in placing more patients on therapy. In contrast, in Hypothetical studies assumptions were made explicit and risk of bias was overall low. This may in part reflect the fact that authors of Hypothetical studies were explicit in their aim to estimate causal effects of TB NAATs, while authors of observational studies were often simply aiming to provide some description of the use of these tests.

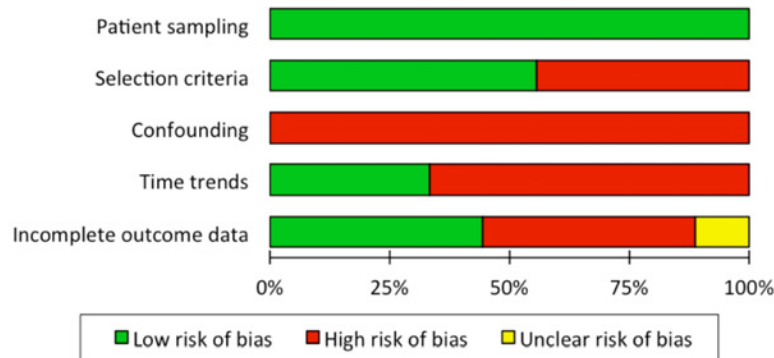
## Discussion

In this systematic review of the impact of molecular tuberculosis diagnostics on patient-important outcomes, we describe numerous challenges that may arise when choosing outcomes and designs. We describe the options that exist, the threats to validity that come with each choice, and make some suggestions about how to further raise methodological standards in design, analysis, and interpretation of results.

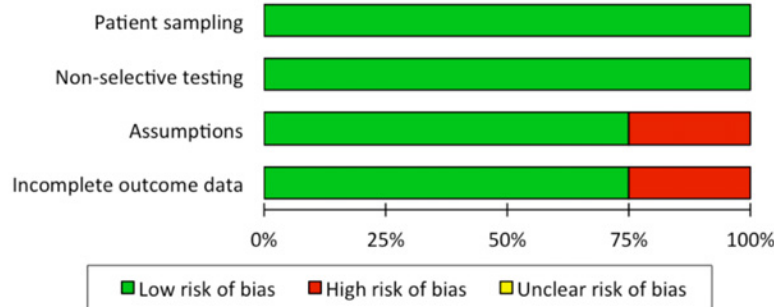
**RCTs**



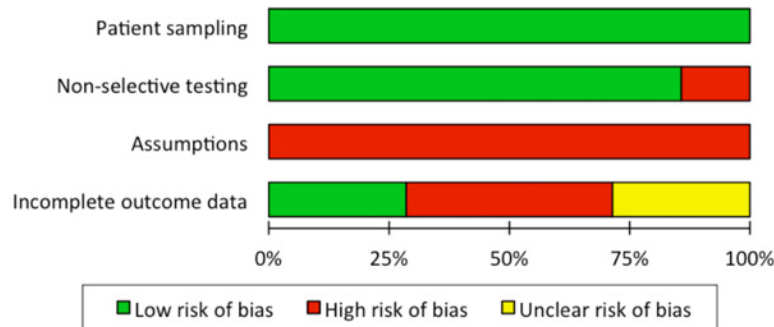
**Pre/post implementation studies**



**Hypothetical studies**



**Observational studies**



**Fig 6. Assessment of risk of bias for each study design across included studies.** Review authors' judgments about each domain presented as percentages across the 25 studies, separately by design. Judgments were based on criteria outlined in 'S2 Appendix'.

doi:10.1371/journal.pone.0151073.g006

We found that most of the evidence on patient-important outcomes comes from a small number of settings with a particular epidemiological and operational context. Therefore, general conclusions about “the impact of TB NAATs” should be made with caution. The settings are not necessarily representative of the global TB epidemiology but may reflect the settings where research is most feasible due to availability of trained personnel, expertise and interest in research methodology, and beliefs about where the impact of the evaluated tests may be greatest.

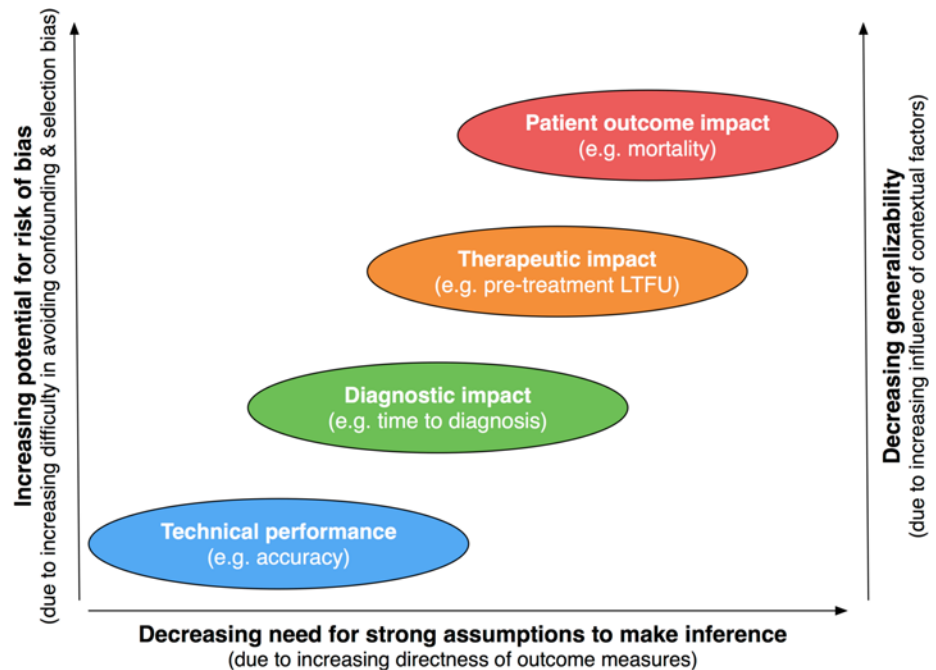
Few studies assessed new tests implemented as part of point-of-care testing programs[105] and none aimed at ensuring a continuum of TB care. The importance of the cascade of care has been described extensively in the HIV literature[106] and it is becoming apparent that similar challenges in delivering services in a timely and reliable sequence exist in TB programs[107] and in point-of-care testing in global health in general[108]. Future studies and real-world implementation plans may need to take a more patient-centered approach in order for novel tests to reach their full potential in terms of improving patient and population health.

Included studies looked at a large variety of outcomes but “time to treatment initiation” was by far the most common one. Most studies showed benefit of TB NAATs and studies that did not show benefit were usually inconclusive (rather than affirming true absence of benefit) although this was generally poorly reported. Effect estimates should always be accompanied by Confidence Intervals[109] and resampling methods (such as the non-parametric bootstrap [110]) may be used if simple procedures to obtain them are not directly available in software packages, as is the case for changes in time to diagnosis or morbidity.

There is a trade-off when choosing outcome measures: while one may try to measure effects on patient health directly, this can come at greater risk of confounding, selection bias, and difficulties with generalizing findings to other settings; on the other hand one may resort to measure effects on patient health very indirectly, which can avoid or lower these risks but any conclusions about actual effects on patient health then require us to make a number of potentially untenable assumptions that are needed to such extrapolations (Fig 7). This is in fact a problem that health technology assessment units are routinely facing when trying to decide whether a new diagnostic test or screening program should be introduced[19,32,38,111]. High-quality evidence showing positive effects directly on patient outcomes is usually lacking which often leads assessors to use decision-analytic modeling as a way to integrate different pieces of evidence[28,112,113], and this approach may also be fruitful when assessing the value of TB diagnostics.

While randomized trials provide the strongest counterfactual, non-randomized studies will continue to play an important part in providing evidence on effects on patient-important outcomes. However, stronger non-randomized designs that can be based on routine data sources remain unused so far but may be well-suited for high-quality operational research studies and their use should be explored in future research. Missing outcome data was a relatively frequent problem and greater efforts should be made to address it using established methods[95–97]. Pre/post implementation studies suffered from selection bias and confounding, but either no attempt was made to address this analytically or attempts were methodologically problematic. Hypothetical trials can be an attractive option to estimate impact on patient outcomes by extrapolating more formally from test performance data based on explicitly stated and justified assumptions[28]. Single-cohort observational studies (as defined in our review) are probably not very suitable for inference on impact on patient outcomes but may still provide valuable insights about how tests are used in various settings.

We aimed to review the methodology of studies assessing two classes of TB NAATs, using a variety of study designs to measure a large number of different outcomes in settings with very different epidemiological and operational context. The multitude of these variables are both a



**Fig 7. Directness of outcome measures, risk of bias and generalizability.** Studies evaluating outcomes that provide very direct evidence for impacting patient outcomes may be—on average—more prone to confounding and selection bias and may lead to results that are less easily generalizable. The risk of confounding is likely increased because the number of covariates that have an influence on downstream outcomes (for which balance between compared cohorts needs to be ensured) increases. The risk of selection bias increases because the required length of follow-up increases as one assesses further downstream outcomes. Generalizability of specific estimates may become increasingly questionable because the role of contextual factors that vary from setting to setting also have increasing influence on the further downstream outcomes. In contrast, studies providing only very indirect evidence may have lower risk of bias but require much stronger assumptions if we try to extrapolate from their findings to make statements about downstream patient outcomes. In general it is therefore important to take both risk of bias and applicability into account to come to an overall conclusion about the likely impact of diagnostic tests on patient outcomes.

doi:10.1371/journal.pone.0151073.g007

strength and a weakness of our review as this allowed great breadth but limited depth in discussing methodological issues and their potential effects in making overall conclusions about the impact of the tests on patient-important outcomes.

We included non-randomized designs for which no validated tool to assess the risk of bias existed. The tools we used have not been validated and it is possible that relevant criteria were omitted or that other improvements could be made to our assessment of risk of bias. However, our tool was based on extensive search and review of the literature on assessing risk of bias in intervention studies, particularly those involving diagnostics[36–38] and included simple yes/no questions with explicit guidance on how to make judgments, as has been recommended for the development of new risk-of-bias tools[36].

We did not explicitly evaluate the potential for measurement bias because we did not feel that this was a big concern in general, which would need to be assessed for each individual study. However, we emphasize that the lack of a gold standard for TB (even mycobacterial culture has imperfect sensitivity), and the resulting reliance on clinical diagnosis, complicates the interpretation of outcomes such as the ‘number of patients put on treatment’, as this is only a proxy for the ‘number of true TB patients put on treatment’. Like any measurement error, use of this proxy leads to some degree of bias towards the nul[114]. Importantly though, it also has

potential to introduce additional measurement bias towards the null because clinical diagnosis and empiric therapy is more common in the baseline arm (using smear microscopy), which likely leads to more *non-TB patients* put on TB treatment than in the TB NAAT arm. These are then counted as TB patients, thus seemingly improving the performance of the baseline arm, while actually representing in part over-treatment. It may be worthwhile to explore simple adjustments for this in sensitivity analyses in future studies[115,116], using a range of estimates of the accuracy of empiric therapy.

Our focus on adult pulmonary TB and WHO-approved NAATs was mostly a pragmatic decision. However, we believe that most of our conclusions will apply to other forms of TB as well as different infectious diseases and other types of tests. We focused on studies based on primary data but modeling studies have a key role in assessing patient (and population) impact as well. General advice on methodology for such studies exist [112,117–119] and an assessment of how they have been implemented in the case of point of care testing strategies for active tuberculosis[120] and for interferon-gamma release assays for latent TB infection[121] have been published.

## Conclusions

In conclusion, generating evidence on the impact of molecular TB diagnostics on patient-important outcomes is challenging and there is no simple or ideal choice for design or outcome. Choices will often be dictated by availability of routine data or limitations in funding to carry out primary data collection but an awareness of the trade-offs in choosing outcomes and designs will hopefully help make the best choices possible. Some designs that have the potential to yield strong evidence without requiring large-scale primary data collection have not been used to date and may have great potential for future research. Once data are collected, doing the best possible job during analysis is relatively inexpensive and should always be possible. As the analytic challenges are very different from those in accuracy research, including a methodologist during data analysis and ideally also early during study planning is advisable.

## Supporting Information

**S1 Appendix. Full electronic search strategy.**  
(DOC)

**S2 Appendix. Design-specific risk of bias tools used and guidance for judgment.**  
(DOCX)

**S3 Appendix. Study characteristics.**  
(DOCX)

**S4 Appendix. Assessment of risk of bias for each individual included study.**  
(PDF)

**S5 Appendix. Explanatory figure for Confounding and selection bias.**  
(PDF)

**S6 Appendix. PRISMA 2009 checklist.**  
(DOC)

## Acknowledgments

This project was funded by a grant from the Canadian Institutes of Health Research, Ottawa, ONT, Canada (grants MOP-89918). MP is supported by the Fonds de Recherche du Québec

Santé, Montreal, QC, USA. CMD was supported by a Richard Tomlinson Fellowship at McGill University and a fellowship of the Burroughs-Wellcome Fund from the American Society of Tropical Medicine and Hygiene, Deerfield, IL, USA. SGS was supported by the Max Binz Fellowship at McGill University. The funders had no role in the study design, data collection, analysis, decision to publish, or manuscript preparation.

## Author Contributions

Conceived and designed the experiments: SGS JLD CMD MP. Performed the experiments: SGS HS ZZQ GG. Analyzed the data: SGS. Wrote the paper: SGS JLD CMD MP.

## References

1. World Health Organization. Global Tuberculosis Report 2015.
2. Uplekar M, Weil D, Lonroth K, Jaramillo E, Lienhardt C, Dias HM, et al. WHO's new End TB Strategy. *Lancet*. 2015; 385: 1799–1801. doi: [10.1016/s0140-6736\(15\)60570-0](https://doi.org/10.1016/s0140-6736(15)60570-0) PMID: [25814376](https://pubmed.ncbi.nlm.nih.gov/25814376/)
3. Kik SV, Denkinger CM, Chedore P, Pai MP. Replacing smear microscopy for the diagnosis of tuberculosis: what is the market potential? *Eur Respir J*. 2014. doi: [10.1183/09031936.00217313](https://doi.org/10.1183/09031936.00217313)
4. Stall N, Rubin T, Michael JS, Mathai D, Abraham OC, Mathews P, et al. Does solid culture for tuberculosis influence clinical decision making in India? 2011; 15: 641–646. doi: [10.5588/ijtld.10.0195](https://doi.org/10.5588/ijtld.10.0195)
5. Hepple P, Novoa-Cain J, Cheruiyot C, Richter E, Ritmeijer K. Implementation of liquid culture for tuberculosis diagnosis in a remote setting: lessons learned. 2011; 15: 405–407.
6. Reddy EA, Njau BN, Mopeth SC, Lancaster KE, Tribble AC, Maro VP, et al. A randomized controlled trial of standard versus intensified tuberc. . .—PubMed—NCBI. *BMC Infect Dis*. BioMed Central Ltd; 2014; 14: 89. doi: [10.1186/1471-2334-14-89](https://doi.org/10.1186/1471-2334-14-89) PMID: [24552306](https://pubmed.ncbi.nlm.nih.gov/24552306/)
7. Lam E, Nateniyom S, Whitehead S, Anuwatnonthakate A, Monkongdee P, Kanphukiew A, et al. Use of drug-susceptibility testing for management of drug-resistant tuberculosis, Thailand, 2004–2008. 2014; 20: 408–416. doi: [10.3201/eid2003.130951](https://doi.org/10.3201/eid2003.130951)
8. Ling DI, Zwerling AA, Pai MP. GenoType MTBDR assays for the diagnosis of multidrug-resistant tuberculosis: a meta-analysis. *Eur Respir J*. 2008; 32: 1165–1174. doi: [10.1183/09031936.00061808](https://doi.org/10.1183/09031936.00061808) PMID: [18614561](https://pubmed.ncbi.nlm.nih.gov/18614561/)
9. Denkinger CM, Schumacher SG, Boehme CC, Dendukuri N, Pai MP, Steingart KR. Xpert MTB/RIF assay for the diagnosis of extrapulmonary tuberculosis: a systematic review and meta-analysis. *Eur Respir J*. 2014; 44: 435–446. doi: [10.1183/09031936.00007814](https://doi.org/10.1183/09031936.00007814) PMID: [24696113](https://pubmed.ncbi.nlm.nih.gov/24696113/)
10. Detjen AK, DiNardo AR, Leyden J, Steingart KR. Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in children: a systematic review and meta-analysis. *The Lancet Respiratory*. 2015. doi: [10.1016/S2213-2600\(15\)00095-8](https://doi.org/10.1016/S2213-2600(15)00095-8)
11. Weyer K. WHO Policy Statement: Xpert. 2011;: 1–35.
12. Schünemann HJ, Oxman AD, Brozek JL, Glasziou P, Jaeschke R, Vist GE, et al. GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008; 336: 1106–1110. PMID: [18483053](https://pubmed.ncbi.nlm.nih.gov/18483053/)
13. Pai MP, Minion J, Steingart KR, Ramsay ARC. New and improved tuberculosis diagnostics: evidence, policy, practice, and impact. *Curr Opin Pulm Med*. 2010; 16: 271–284. doi: [10.1097/MCP.0b013e328338094f](https://doi.org/10.1097/MCP.0b013e328338094f) PMID: [20224410](https://pubmed.ncbi.nlm.nih.gov/20224410/)
14. Cobelens FG, van den Hof S, Pai MP, Squire SB, Ramsay ARC, Kimerling ME, et al. Which New Diagnostics for Tuberculosis, and When? *J Infect Dis*. 2012; 205: S191–S198. PMID: [22476716](https://pubmed.ncbi.nlm.nih.gov/22476716/)
15. Ramsay ARC, Steingart KR, Pai MP. Assessing the impact of new diagnostics on tuberculosis control. 2010; 14: 1506–1507.
16. World Health Organization. Tuberculosis Diagnostics Technology and Market Landscape, 3rd edition. 2014 pp. 1–97.
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009; 6: e1000097. doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097) PMID: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)
18. Collaboration The Cochrane. *Cochrane Handbook for Systematic Reviews of Interventions*. 2011;: 1–317.



19. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, et al. Current methods of the U.S. Preventive Services Task Force. *American Journal of Preventive Medicine*. 2001; 20: 21–35. doi: [10.1016/S0749-3797\(01\)00261-6](https://doi.org/10.1016/S0749-3797(01)00261-6)
20. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review—a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy*. SAGE Publications; 2005; 10 Suppl 1: 21–34. doi: [10.1258/1355819054308530](https://doi.org/10.1258/1355819054308530) PMID: [16053581](https://pubmed.ncbi.nlm.nih.gov/16053581/)
21. Anderson LM, Petticrew M, Rehfuess E, Armstrong R, Ueffing E, Baker P, et al. Using logic models to capture complexity in systematic reviews. *Research Synthesis Methods*. John Wiley & Sons, Ltd; 2011; 2: 33–42. doi: [10.1002/jrsm.32](https://doi.org/10.1002/jrsm.32) PMID: [26061598](https://pubmed.ncbi.nlm.nih.gov/26061598/)
22. White H. Theory-based Impact Evaluation. 3ie. 2009.
23. Gertler PJ, Martinez S, Premand P, Rawlings LB. Impact evaluation in practice. World Bank Publications. 2011.
24. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Lippincott Williams & Wilkins; 2008.
25. Friedman LM, Furberg C, DeMets DL. Fundamentals of Clinical Trials. 2010.
26. Knottnerus JA, Buntinx F. The Evidence Base of Clinical Diagnosis—Theory and methods of diagnostic research. 2009. pp. 1–314.
27. Angrist JD, Pischke J-S. Mostly Harmless Econometrics. Princeton University Press; 2008.
28. Lord SJ, Irwig LM, Bossuyt PMM. Using the principles of randomized controlled trial design to guide test evaluation. *Med Decis Making*. 2009; 29: E1–E12. doi: [10.1177/0272989X09340584](https://doi.org/10.1177/0272989X09340584) PMID: [19773580](https://pubmed.ncbi.nlm.nih.gov/19773580/)
29. Davis JL, Kawamura LM, Chaisson LH, Grinsdale J, Benhammou J, Ho C, et al. Impact of GeneXpert MTB/RIF<sup>®</sup> on Patients and Tuberculosis Programs in a Low-Burden Setting: A Hypothetical Trial. *Am J Respir Crit Care Med*. 2014;: 140528131203007. doi: [10.1164/rccm.201311-1974OC](https://doi.org/10.1164/rccm.201311-1974OC)
30. Li T, Vedula SS, Hadar N, Parkin C, Lau J, Dickersin K. Innovations in Data Collection, Management, and Archiving for Systematic Reviews. *Ann Intern Med*. American College of Physicians; 2015; 162: 287–294. doi: [10.7326/M14-1603](https://doi.org/10.7326/M14-1603) PMID: [25686168](https://pubmed.ncbi.nlm.nih.gov/25686168/)
31. Higgins JPT, Altman DG, Gotzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. BMJ Publishing Group Ltd; 2011; 343: d5928–d5928. doi: [10.1136/bmj.d5928](https://doi.org/10.1136/bmj.d5928)
32. Staub LP, Dyer S, Lord SJ, Simes RJ. Linking The Evidence: Intermediate Outcomes In Medical Test Assessments. *Int J Technol Assess Health Care*. Cambridge University Press; 2012; 28: 52–58. doi: [10.1017/S0266462311000717](https://doi.org/10.1017/S0266462311000717) PMID: [22230006](https://pubmed.ncbi.nlm.nih.gov/22230006/)
33. Whiting PF, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess*. 2004; 8: iii, 1–234.
34. Collaboration C. Cochrane Risk of Bias Tool. 2010 Apr.
35. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011; 155: 529–536. doi: [10.1059/0003-4819-155-8-201110180-00009](https://doi.org/10.1059/0003-4819-155-8-201110180-00009) PMID: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/)
36. AHRQ. Methods Guide for Comparative Effectiveness Reviews. 2012;: 1–33.
37. AHRQ. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013.
38. AHRQ. Methods Guide for Medical Test Reviews. 2012;: 1–188.
39. Cox HS, Mbhele S, Mohess N, Whitelaw A, Muller O, Zemanay W, 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. Cattamanchi A, editor. *PLoS Med*. Public Library of Science; 2014; 11: e1001760. doi: [10.1371/journal.pmed.1001760](https://doi.org/10.1371/journal.pmed.1001760) PMID: [25423041](https://pubmed.ncbi.nlm.nih.gov/25423041/)
40. Churchyard GJ, Stevens WS, Mametja LD. 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 ... *The Lancet Global Health*. 2015; 3: e450–e457. doi: [10.1016/s2214-109x\(15\)00100-x](https://doi.org/10.1016/s2214-109x(15)00100-x) PMID: [26187490](https://pubmed.ncbi.nlm.nih.gov/26187490/)
41. Jacobson KR, Theron D, Kendall EA, Franke MF, Barnard M, van Helden PD, et al. Implementation of GenoType MTBDRplus Reduces Time to Multidrug-Resistant Tuberculosis Therapy Initiation in South Africa. *Clin Infect Dis*. 2012. doi: [10.1093/cid/cis920](https://doi.org/10.1093/cid/cis920)
42. Cox HS, Daniels JF, Muller O, Nicol MP, Cox V, van Cutsem G, et al. Impact of decentralized care and the Xpert MTB/RIF test on rifampicin-resistant tuberculosis treatment initiation in Khayelitsha, South Africa. *Open Forum Infectious Diseases*. 2015. doi: [10.1093/ofid/ofv014](https://doi.org/10.1093/ofid/ofv014)

43. Naidoo P, Toit Du E, Dunbar R, Lombard C, Caldwell J, Detjen A, et al. A Comparison of Multidrug-Resistant Tuberculosis Treatment Commencement Times in MDRTBPlus Line Probe Assay and Xpert® MTB/RIF-Based Algorithms in a Routine Operational Setting in Cape Town. Nicol MP, editor. *Public Library of Science*; 2014; 9: e103328. doi: [10.1371/journal.pone.0103328](https://doi.org/10.1371/journal.pone.0103328)
44. Boehme CC, Nicol MP, Nabeta P, Michael JS, Gotuzzo E, Tahirli R, et al. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. *Lancet*. 2011; 377: 1495–1505. doi: [10.1016/S0140-6736\(11\)60438-8](https://doi.org/10.1016/S0140-6736(11)60438-8) PMID: [21507477](https://pubmed.ncbi.nlm.nih.gov/21507477/)
45. Hanrahan CF, Dorman SE, Erasmus L, Koornhof H, Coetzee G, Golub JE. The impact of expanded testing for multidrug resistant tuberculosis using genotype [correction of geontype] MTBDRplus in South Africa: an observational cohort study. 2012; 7: e49898. doi: [10.1371/journal.pone.0049898](https://doi.org/10.1371/journal.pone.0049898)
46. Cohen GM, Drain PK, Noubary F, Cloete C, Bassett IV. Diagnostic delays and clinical decision-making with centralized Xpert MTB/RIF testing in Durban, South Africa. *J Acquir Immune Defic Syndr*. 2014;: 1.
47. Van Rie A, Page-Shipp L, Hanrahan CF, Schnippel K, Dansey H, Bassett J, et al. Point-of-care Xpert® MTB/RIF for smear-negative tuberculosis suspects at a primary care clinic in South Africa. 2013; 17: 368–372. doi: [10.5588/ijtld.12.0392](https://doi.org/10.5588/ijtld.12.0392)
48. Hanrahan CF, Selibas K, Deery CB, Dansey H, Clouse K, Bassett J, et al. Time to Treatment and Patient Outcomes among TB Suspects Screened by a Single Point-of-Care Xpert MTB/RIF at a Primary Care Clinic in Johannesburg, South Africa. Dheda K, editor. 2013; 8: e65421. doi: [10.1371/journal.pone.0065421.t003](https://doi.org/10.1371/journal.pone.0065421.t003)
49. Balcells ME, García P, Chanqueo L, Bahamondes L, Lasso M, Gallardo AM, et al. Rapid molecular detection of pulmonary tuberculosis in HIV-infected patients in Santiago, Chile. *int j tuberc lung dis*. 2012; 16: 1349–1353. doi: [10.5588/ijtld.12.0156](https://doi.org/10.5588/ijtld.12.0156) PMID: [22863872](https://pubmed.ncbi.nlm.nih.gov/22863872/)
50. Buchelli Ramirez HL, García-Clemente MM, Álvarez-Álvarez C, Palacio-Gutierrez JJ, Pando-Sandoval A, Gagatek S, et al. Impact of the Xpert® MTB/RIF molecular test on the late diagnosis of pulmonary tuberculosis [Short Communication]. 2014; 18: 435–437. doi: [10.5588/ijtld.13.0747](https://doi.org/10.5588/ijtld.13.0747)
51. Kipiani M, Mirtskhulava V, Tukvadze N, Magee M, Blumberg HM, Kempker RR. Significant Clinical Impact of a Rapid Molecular Diagnostic Test (Genotype MTBDRplus Assay) to detect Multidrug-Resistant Tuberculosis. *Clin Infect Dis*. Oxford University Press; 2014;: ciu631. doi: [10.1093/cid/ciu631](https://doi.org/10.1093/cid/ciu631)
52. Kwak N, Choi SM, Lee J, Park YS, Lee C-H, Lee S-M, et al. Diagnostic Accuracy and Turnaround Time of the Xpert MTB/RIF Assay in Routine Clinical Practice. Wilkinson RJ, editor. 2013; 8: e77456. doi: [10.1371/journal.pone.0077456.t006](https://doi.org/10.1371/journal.pone.0077456.t006)
53. Lippincott CK, Miller MB, Popowitch EB, Hanrahan CF, van Rie A. Xpert® MTB/RIF Shortens Airborne Isolation for Hospitalized Patients with Presumptive Tuberculosis in the United States. *Clin Infect Dis*. Oxford University Press; 2014;: ciu212. doi: [10.1093/cid/ciu212](https://doi.org/10.1093/cid/ciu212)
54. Lyu J, Kim M-N, Song JW, Choi C-M, Oh Y-M, Lee SD, et al. GenoType® MTBDRplus assay detection of drug-resistant tuberculosis in routine practice in Korea. 2013; 17: 120–124. doi: [10.5588/ijtld.12.0197](https://doi.org/10.5588/ijtld.12.0197)
55. Mupfumi L, Makamure B, Chirehwa M, Sagonda T, Zinyowera S, Mason P, et al. Impact of Xpert MTB/RIF on Antiretroviral Therapy-Associated Tuberculosis and Mortality: A Pragmatic Randomized Controlled Trial. *Open Forum Infectious Diseases*. 2014; 1: ofu038–ofu038. doi: [10.1093/ofid/ofu038](https://doi.org/10.1093/ofid/ofu038) PMID: [25734106](https://pubmed.ncbi.nlm.nih.gov/25734106/)
56. Singla N, Satyanarayana S, Sachdeva KS, Van den Bergh R, Reid T, Tayler-Smith K, et al. Impact of Introducing the Line Probe Assay on Time to Treatment Initiation of MDR-TB in Delhi, India. Sola C, editor. *Public Library of Science*; 2014; 9: e102989. doi: [10.1371/journal.pone.0102989](https://doi.org/10.1371/journal.pone.0102989)
57. Skenders G, Holtz TH, Riekstina V, Leimane V. Implementation of the INNO-LiPA Rif.TB line-probe assay in rapid detection of multidrug-resistant tuberculosis in Latvia. 2011; 15: 1546–1552. doi: [10.5588/ijtld.11.0067](https://doi.org/10.5588/ijtld.11.0067)
58. Sohn H, Aero AD, Menzies D, Behr M, Schwartzman K, Alvarez GG, et al. Xpert MTB/RIF testing in a low tuberculosis incidence, high-resource setting: limitations in accuracy and clinical impact. *Clin Infect Dis*. 2014; 58: 970–976. doi: [10.1093/cid/ciu022](https://doi.org/10.1093/cid/ciu022) PMID: [24429440](https://pubmed.ncbi.nlm.nih.gov/24429440/)
59. Yoon C, Cattamanchi A, Davis JL, Worodria W, Boon den S, Kalema N, et al. Impact of Xpert MTB/RIF Testing on Tuberculosis Management and Outcomes in Hospitalized Patients in Uganda. Dheda K, editor. 2012; 7: e48599. doi: [10.1371/journal.pone.0048599.t002](https://doi.org/10.1371/journal.pone.0048599.t002)
60. Durovni B, Saraceni V, van den Hof S, Trajman A, Cordeiro-Santos M, Cavalcante S, et al. Impact of Replacing Smear Microscopy with Xpert MTB/RIF for Diagnosing Tuberculosis in Brazil: A Stepped-Wedge Cluster-Randomized Trial. Murray M, editor. *PLoS Med*. 2014; 11: e1001766.

61. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *CMAJ*. 1986; 134: 587–594. PMID: [3512062](#)
62. Altman DG, Bland JM. Statistics notes: Absence of evidence is not evidence of absence. *BMJ*. BMJ Publishing Group Ltd; 1995; 311: 485–485. doi: [10.1136/bmj.311.7003.485](#) PMID: [7647644](#)
63. Greenland S. Interval estimation by simulation as an alternative to and extension of confidence intervals. *International Journal of Epidemiology*. Oxford University Press; 2004; 33: 1389–1397. doi: [10.1093/ije/dyh276](#) PMID: [15319402](#)
64. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer*. 1976; 34: 585–612. PMID: [795448](#)
65. Schulz KF, Altman DG, Moher D, Grp C. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. *PLoS Med*. 2010; 7. doi: [10.1371/journal.pmed.1000251](#)
66. 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: 424–435. doi: [10.1016/S0140-6736\(13\)62073-5](#) PMID: [24176144](#)
67. Eldridge S, Kerry S. *A Practical Guide to Cluster Randomised Trials in Health Services Research*. John Wiley & Sons; 2012. doi: [10.1002/9781119966241.ch3](#)
68. Hayes RJ, Moulton LH. *Cluster Randomised Trials*. Chapman and Hall/CRC; 2009.
69. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *BMJ: British Medical Journal*. BMJ Publishing Group Ltd; 2012; 345: e5661–e5661. doi: [10.1136/bmj.e5661](#) PMID: [22951546](#)
70. Squire SB, Ramsay ARC, van den Hof S, Millington KA, Langley I, Bello G, et al. Making innovations accessible to the poor through implementation research. 2011; 15: 862–870. doi: [10.5588/ijtld.11.0161](#)
71. Khandker SR, Koolwal GB, Samad HA. *Handbook on impact evaluation: quantitative methods and practices*. World Bank Publications. 2010.
72. Cousens S, Hargreaves JR, Bonell C, Armstrong B, Thomas J, Kirkwood BR, et al. Alternatives to randomisation in the evaluation of public-health interventions: statistical analysis and causal inference. *Journal of Epidemiology & Community Health*. 2011; 65: 576–581. doi: [10.1136/jech.2008.082610](#)
73. Rockers PC, Røttingen J-A, Shemilt I, Tugwell P, Bärnighausen T. Inclusion of quasi-experimental studies in systematic reviews of health systems research. *Health Policy*. 2015; 119: 511–521. doi: [10.1016/j.healthpol.2014.10.006](#) PMID: [25776033](#)
74. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA*. 2014; 312: 2401–2402. doi: [10.1001/jama.2014.16153](#) PMID: [25490331](#)
75. Ryan AM, Burgess JF, Dimick JB. Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences. *Health Services Research*. 2014;: n/a–n/a. doi: [10.1111/1475-6773.12270](#)
76. Meyer BD. *Natural and quasi-experiments in economics*. Journal of business & economic statistics. 1995.
77. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. *International Journal of Epidemiology*. 2013; 42: 1187–1195. doi: [10.1093/ije/dyt092](#) PMID: [23760528](#)
78. Taljaard M, McKenzie JE, Ramsay CR, Grimshaw JM. The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care. *Implementation Sci*. 2014; 9: 77. doi: [10.1186/1748-5908-9-77](#)
79. Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. *Acad Pediatr*. 2013; 13: S38–44. doi: [10.1016/j.acap.2013.08.002](#) PMID: [24268083](#)
80. Lagarde M. *How to do (or not to do) . . . Assessing the impact of a policy change with routine longitudinal data*. Health Policy Plan. Oxford University Press; 2011; 27: czr004–83. doi: [10.1093/heapol/czr004](#)
81. Swanson SA, Hernán MA. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology*. 2013; 24: 370–374. doi: [10.1097/EDE.0b013e31828d0590](#) PMID: [23549180](#)
82. Angrist J, Krueger A. *Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments*. Cambridge, MA: National Bureau of Economic Research; 2001 Sep. doi: [10.3386/w8456](#)
83. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*. 2000; 29: 722–729. PMID: [10922351](#)

84. Angrist JD, Imbens GW, RUBIN DB. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*. 1996; 91: 444. doi: [10.2307/2291629](https://doi.org/10.2307/2291629)
85. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology*. 2006; 17: 260–267. doi: [10.1097/01.ede.0000215160.88317.cb](https://doi.org/10.1097/01.ede.0000215160.88317.cb) PMID: [16617274](https://pubmed.ncbi.nlm.nih.gov/16617274/)
86. Bor J, Moscoe E, Bärnighausen T. Three Approaches to Causal Inference in Regression Discontinuity Designs. *Epidemiology*. 2015; 26: e28–e30. doi: [10.1097/EDE.0000000000000256](https://doi.org/10.1097/EDE.0000000000000256)
87. Bor J, Moscoe E, Mutevedzi P, Newell M-L, Bärnighausen T. Regression Discontinuity Designs in Epidemiology: Causal Inference Without Randomized Trials. *Epidemiology*. 2014; 25: 729–737. doi: [10.1097/EDE.0000000000000138](https://doi.org/10.1097/EDE.0000000000000138) PMID: [25061922](https://pubmed.ncbi.nlm.nih.gov/25061922/)
88. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *Journal of Clinical Epidemiology*. Elsevier; 2015; 68: 122–133. doi: [10.1016/j.jclinepi.2014.06.021](https://doi.org/10.1016/j.jclinepi.2014.06.021)
89. Lee DS, Lemieux T. Regression discontinuity designs in economics. *National Bureau of Economic Research*; 2009.
90. Chaisson LH, Roemer M, Cantu D, Haller B, Millman AJ, Cattamanchi A, et al. Impact of GeneXpert MTB/RIF assay on triage of respiratory isolation rooms for inpatients with presumed tuberculosis: a hypothetical trial. *Clin Infect Dis*. 2014; 59: 1353–1360. doi: [10.1093/cid/ciu620](https://doi.org/10.1093/cid/ciu620) PMID: [25091300](https://pubmed.ncbi.nlm.nih.gov/25091300/)
91. Guyatt GH, Tugwell PX, Feeny DH, Drummond MF, Haynes RB. The role of before-after studies of therapeutic impact in the evaluation of diagnostic technologies. *J Chronic Dis*. Elsevier; 1986; 39: 295–304.
92. Knottnerus JA, Dinant GJ, van Schayck OP. The Diagnostic Before–After Study to Assess Clinical Impact. *The Evidence Base of Clinical Diagnosis*. Oxford, UK: Wiley-Blackwell; 2008. pp. 83–95. doi: [10.1002/9781444300574.ch5](https://doi.org/10.1002/9781444300574.ch5)
93. Meads CA, Davenport CF. Quality assessment of diagnostic before-after studies: development of methodology in the context of a systematic review. *BMC Med Res Methodol*. 2009; 9: 3. doi: [10.1186/1471-2288-9-3](https://doi.org/10.1186/1471-2288-9-3) PMID: [19152699](https://pubmed.ncbi.nlm.nih.gov/19152699/)
94. Staub LP, Lord SJ, Simes R, Dyer S, Houssami N, Chen RY, et al. Using patient management as a surrogate for patient health outcomes in diagnostic test evaluation. *BMC Med Res Methodol*. 2012; 12: 12. doi: [10.1016/S0140-6736\(09\)62070-5](https://doi.org/10.1016/S0140-6736(09)62070-5) PMID: [22333319](https://pubmed.ncbi.nlm.nih.gov/22333319/)
95. Groenwold RHH, Moons KGM, Vandenbroucke JP. Randomized trials with missing outcome data: how to analyze and what to report. *Canadian Medical Association Journal*. Canadian Medical Association; 2014;: cmaj.131353. doi: [10.1503/cmaj.131353](https://doi.org/10.1503/cmaj.131353)
96. Groenwold RHH, Donders ART, Roes KCB, HARRELL FE, Moons KGM. Dealing with missing outcome data in randomized trials and observational studies. *Am J Epidemiol*. 2012; 175: 210–217. doi: [10.1093/aje/kwr302](https://doi.org/10.1093/aje/kwr302) PMID: [22262640](https://pubmed.ncbi.nlm.nih.gov/22262640/)
97. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. BMJ; 2009; 338: b2393–b2393. doi: [10.1136/bmj.b2393](https://doi.org/10.1136/bmj.b2393) PMID: [19564179](https://pubmed.ncbi.nlm.nih.gov/19564179/)
98. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statist Med*. 2006; 25: 127–141. doi: [10.1002/sim.2331](https://doi.org/10.1002/sim.2331)
99. Groenwold RHH, Klungel OH, Altman DG, van der Graaf Y, Hoes AW, Moons KGM, et al. Adjustment for continuous confounders: an example of how to prevent residual confounding. *Canadian Medical Association Journal*. Canadian Medical Association; 2013; 185: 401–406. doi: [10.1503/cmaj.120592](https://doi.org/10.1503/cmaj.120592) PMID: [23401401](https://pubmed.ncbi.nlm.nih.gov/23401401/)
100. Greenland S. Modeling and variable selection in epidemiologic analysis. *Am J Public Health*. 1989; 79: 340–349. PMID: [2916724](https://pubmed.ncbi.nlm.nih.gov/2916724/)
101. Greenland S, Pearce NNP. Statistical foundations for model-based adjustments. *Annu Rev Public Health*. 2015; 36: 89–108. doi: [10.1146/annurev-publhealth-031914-122559](https://doi.org/10.1146/annurev-publhealth-031914-122559) PMID: [25785886](https://pubmed.ncbi.nlm.nih.gov/25785886/)
102. Maldonado G, Greenland S. Simulation Study of Confounder-Selection Strategies. *Am J Epidemiol*. 1993; 138: 923–936. PMID: [8256780](https://pubmed.ncbi.nlm.nih.gov/8256780/)
103. Sauer BC, Brookhart MA, Roy J, VanderWeele T. A review of covariate selection for non-experimental comparative effectiveness research. *Pharmacoepidemiol Drug Saf*. 2013; 22: 1139–1145. doi: [10.1002/pds.3506](https://doi.org/10.1002/pds.3506) PMID: [24006330](https://pubmed.ncbi.nlm.nih.gov/24006330/)
104. Arnold BF, Khush RS, Ramaswamy P, London AG, Rajkumar P, Ramaprabha P, et al. Causal inference methods to study nonrandomized, preexisting development interventions. 2010; 107: 22605–22610. doi: [10.1073/pnas.1008944107/-DCSupplemental](https://doi.org/10.1073/pnas.1008944107/-DCSupplemental)

105. Pant-Pai N, Vadnais C, Denkinger CM, Engel N, Pai MP. Point-of-Care Testing for Infectious Diseases: Diversity, Complexity, and Barriers in Low- And Middle-Income Countries. *PLoS Med.* 2012; 9: e1001306. doi: [10.1371/journal.pmed.1001306.t002](https://doi.org/10.1371/journal.pmed.1001306.t002) PMID: [22973183](https://pubmed.ncbi.nlm.nih.gov/22973183/)
106. Mugavero MJ, Amico KR, Horn T, Thompson MA. The state of engagement in HIV care in the United States: from cascade to continuum to control. *Clin Infect Dis.* Oxford University Press; 2013; 57: 1164–1171. doi: [10.1093/cid/cit420](https://doi.org/10.1093/cid/cit420) PMID: [23797289](https://pubmed.ncbi.nlm.nih.gov/23797289/)
107. Sun AY, Denkinger CM, Dowdy DW. The impact of novel tests for tuberculosis depends on the diagnostic cascade. *Eur Respir J.* European Respiratory Society; 2014;: erj01110–2014. doi: [10.1183/09031936.00111014](https://doi.org/10.1183/09031936.00111014)
108. Jani IV, Peter TF. How point-of-care testing could drive innovation in global health. *N Engl J Med.* 2013; 368: 2319–2324. doi: [10.1056/NEJMs1214197](https://doi.org/10.1056/NEJMs1214197) PMID: [23758238](https://pubmed.ncbi.nlm.nih.gov/23758238/)
109. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *British Medical Journal (Clinical research ed).* BMJ Group; 1986; 292: 746–750.
110. Bradley E, Tibshirani RJ. An introduction to the bootstrap. *Monographs on statistics and applied probability*; 1993.
111. Moons KG. *Evaluation Of New Technology In Health Care.* 2014;: 1–157.
112. Trikalinos TA, Siebert U, Lau J. Decision-analytic modeling to evaluate benefits and harms of medical tests: uses and limitations. *Med Decis Making.* Rockville (MD): Agency for Healthcare Research and Quality (US); 2009; 29: E22–9. doi: [10.1177/0272989X09345022](https://doi.org/10.1177/0272989X09345022) PMID: [19734441](https://pubmed.ncbi.nlm.nih.gov/19734441/)
113. Bossuyt PMM, Reitsma JB, Linnet K, Moons KGM. Beyond Diagnostic Accuracy: The Clinical Utility of Diagnostic Tests. *Clin Chem.* 2012; 58: 1636–1643. doi: [10.1373/clinchem.2012.182576](https://doi.org/10.1373/clinchem.2012.182576) PMID: [22730450](https://pubmed.ncbi.nlm.nih.gov/22730450/)
114. Hutcheon JA, Chiolerio A, Hanley JA. Random measurement error and regression dilution bias. *BMJ.* 2010; 340: c2289–c2289. doi: [10.1136/bmj.c2289](https://doi.org/10.1136/bmj.c2289) PMID: [20573762](https://pubmed.ncbi.nlm.nih.gov/20573762/)
115. Lash TL, Fox MP, Maclehorse RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *International Journal of Epidemiology.* Oxford University Press; 2014;: dyu149. doi: [10.1093/ije/dyu149](https://doi.org/10.1093/ije/dyu149)
116. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data (Statistics for Biology and Health)* [Internet]. 1st ed. Springer; 2009. Available: <https://sites.google.com/site/biasanalysis/>.
117. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess.* 2004; 8: iii–iv, ix–xi, 1–158. Available: <http://www.hta.ac.uk/fullmono/mon836.pdf>.
118. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Eur J Health Econ.* 2013; 14: 367–372. doi: [10.1007/s10198-013-0471-6](https://doi.org/10.1007/s10198-013-0471-6) PMID: [23526140](https://pubmed.ncbi.nlm.nih.gov/23526140/)
119. Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *BMJ.* 2011; 342: d1766. doi: [10.1136/bmj.d1766](https://doi.org/10.1136/bmj.d1766) PMID: [21482590](https://pubmed.ncbi.nlm.nih.gov/21482590/)
120. Zwerling AA, Dowdy D. Economic evaluations of point of care testing strategies for active tuberculosis. *Expert Rev Pharmacoeconomics Outcomes Res.* 2013; 13: 313–325. doi: [10.1586/erp.13.27](https://doi.org/10.1586/erp.13.27)
121. Oxlade O, Pinto M, Trajman A, Menzies D. How Methodologic Differences Affect Results of Economic Analyses: A Systematic Review of Interferon Gamma Release Assays for the Diagnosis of LTBI. *Gagnier JJ, editor.* 2013; 8: e56044. doi: [10.1371/journal.pone.0056044.s008](https://doi.org/10.1371/journal.pone.0056044.s008)