Challenges in Evaluating the Cost-effectiveness of New Diagnostic Tests for HIV-Associated Tuberculosis

Jason R. Andrews,1,2 Stephen D. Lawn,3,4 David W. Dowdy,5,6 and Rochelle P. Walensky1,2

1Division of Infectious Diseases, and 2Medical Practice Evaluation Center, Department of Medicine, Massachusetts General Hospital, Boston; 3Desmond Tutu HIV Centre, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, South Africa; 4Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, United Kingdom; 5Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, and 6Center for Tuberculosis Research, Johns Hopkins University, Baltimore, Maryland

With an emerging array of rapid diagnostic tests for tuberculosis, cost-effectiveness analyses are needed to inform scale-up in various populations and settings. Human immunodeficiency virus (HIV)–associated tuberculosis poses unique challenges in estimating and interpreting the cost-effectiveness of novel diagnostic tools. First, gains in sensitivity and specificity do not directly correlate with impact on clinical outcomes. Second, the cost-effectiveness of implementing tuberculosis diagnostics in HIV-infected populations is heavily influenced by downstream costs of HIV care. As a result, tuberculosis diagnostics may appear less cost-effective in this population than among HIV-uninfected individuals, raising important ethical and policy questions about the design and interpretation of cost-effectiveness analyses in this setting. Third, conventional cost-effectiveness benchmarks may be inadequate for making decisions about whether to adopt new diagnostics. If we are to appropriately deploy novel diagnostics for tuberculosis to people living with HIV in resource-constrained settings, these challenges in measuring cost-effectiveness must be more widely recognized and addressed.

Keywords. tuberculosis; HIV; diagnostics; cost-effectiveness; health policy.

The development of rapid, reliable diagnostics for tuberculosis has generated great enthusiasm among practitioners, public health agencies, and tuberculosis control programs in low- and middle-income countries. In 2010, the World Health Organization (WHO) endorsed the Xpert MTB/RIF (Mycobacterium tuberculosis/rifampicin) assay (“Xpert”), a rapid molecular diagnostic for tuberculosis and rifampicin resistance [1, 2]. In 2011, South Africa moved to purchase and deploy Xpert MTB/RIF machines in laboratories throughout the country, and many other countries have incorporated this assay into their national programs [3]. Now, within 3 years of the WHO endorsement of Xpert, a number of additional diagnostic tests are either available or expected to enter the market soon, including rapid molecular assays based on isothermal amplification, microcolony culture techniques, and a lateral flow assay for lipoarabinomannan (LAM) for use on urine samples [4]. We use this latter test, which is most accurate among individuals with advanced immunosuppression (sensitivity 67% at a CD4 count of <50 cells/mL) [5], as a representative example of a diagnostic test for HIV-associated tuberculosis below. As further diagnostic tests for tuberculosis are developed, countries will continue to face difficult choices about which diagnostic tests to scale up, when, and for whom. Cost and cost-effectiveness will be primary considerations in those decisions.

Last year, 1.1 million tuberculosis cases and 430 000 tuberculosis deaths occurred among people living with HIV [6]. Given differences in burden of disease, mortality, care costs, and test characteristics in HIV/tuberculosis coinfected individuals compared to those with...
tuberculosis alone, projections of the impact and cost-effectiveness of novel diagnostics may differ substantially when considering populations in which the majority of people with active tuberculosis are HIV-infected versus HIV-uninfected. A number of cost-effectiveness analyses of tuberculosis diagnostics for HIV-infected populations have been published, but with substantial differences in natural history and cost assumptions, leading to considerable variability in quantitative results [7–14]. Because the outcomes of cost-effectiveness analyses frequently guide and motivate program policies, it is important for decision makers and clinicians to understand how these analyses are performed, the differences between them, and the particular challenges that arise when evaluating and interpreting the cost-effectiveness of tuberculosis diagnostic tests in people living with HIV.

**INCREMENTAL COST-EFFECTIVENESS ANALYSIS**

While the term “cost-effective” is often used in absolute terms (eg, “Is Xpert a cost-effective test?”), cost-effectiveness is intrinsically a relative measure and cannot be evaluated without comparison to an alternative strategy (often, an existing standard of care) in a specific population [15, 16]. For example, in evaluating the cost-effectiveness of urine LAM testing among hospitalized patients, a reasonable comparison would be the collective use of sputum smear microscopy, chest radiography, and clinical judgment, against the same combination of tests with the addition of testing for urine LAM. The most widely used metric for quantifying comparisons of this “incremental” nature is the incremental cost-effectiveness ratio (ICER), defined as the additional cost, per additional unit of effectiveness, of one strategy (urine LAM testing, in this example) compared with another (eg, standard of care alternative). For novel diagnostics, this translates to:

\[
\text{ICER} = \frac{\text{Cost of Adopting New Diagnostic}}{\text{Effectiveness of New Diagnostic}} - \frac{\text{Cost of, Standard of Care}}{\text{Effectiveness of Standard of Care}}
\]

where ICER divides the overall increased cost of adopting a new diagnostic (incremental cost) by the additional effectiveness gained by making that change (incremental effectiveness, often measured as years of life gained or using “quality-adjusted” life expectancy measures such as the quality-adjusted life-year [QALY], or disability-adjusted life-year [DALY]). The cost and effectiveness of the new strategy are therefore meaningless without reference to the comparator. Importantly, both incremental costs and incremental effects should include all downstream consequences of the new strategy (eg, costs of HIV care for a person who is saved from tuberculosis death in the short term).

This method of estimating cost-effectiveness has 3 important implications for the evaluations of new tuberculosis diagnostics in people living with HIV. First, the incremental effectiveness (eg, years of life gained) of adopting a new test is particularly difficult to quantify in HIV-infected individuals. Second, the incremental cost of introducing a new tuberculosis diagnostic test is often driven by HIV-related care. Third, conventional cost-effectiveness benchmarks may be inadequate to guide policy decisions surrounding adoption of novel diagnostics. In the remainder of this article, we examine each of these implications in detail and discuss how they influence the results and interpretation of cost-effectiveness analyses.

**THE CLINICAL BENEFITS OF IMPROVED DIAGNOSIS ARE DIFFICULT TO QUANTIFY**

Novel rapid diagnostics for tuberculosis may have benefits at both an individual and population level. For individuals, earlier and more accurate diagnosis may avert morbidity and mortality. At a population level, earlier diagnosis and expedited treatment may also avert transmission [13, 17, 18]. The challenges in assessing transmission effects of better tuberculosis diagnosis are particularly great and merit separate discussion elsewhere. Most studies of novel diagnostics measure accuracy (sensitivity and specificity) as their primary outcomes. Although they are important, sensitivity and specificity are not equivalent to data on incremental effectiveness measured in terms that are relevant to patients (eg, deaths averted, years of life gained). For example, urine LAM testing may clearly diagnose more patients with tuberculosis than the standard of care, but in the absence of LAM testing, those receiving “new” (incremental) diagnoses might be treated empirically, without bacteriologic confirmation. Without direct comparative data on effectiveness that is measured in terms of health outcomes rather than sensitivity and specificity, the incremental cost-effectiveness of novel diagnostics cannot be meaningfully estimated. Furthermore, once a test has demonstrated superior accuracy, such direct comparisons (eg, randomized trials of LAM testing vs no LAM testing, with mortality outcomes) require careful ethical consideration [19].

“Pre/post” diagnostic implementation studies that capture outcomes in addition to diagnostic accuracy may provide valuable clinical data [20], but are subject to bias from secular trends. As such, there still exists an important role for randomized trials in the evaluation of diagnostic tests with uncertain population-level benefit. In some settings, individual-level randomization is possible and may be preferable to account for institutional effects, especially for patients in secondary- and tertiary-level care [21]. However, tuberculosis diagnostic tests
are most widely implemented at the primary care level, and here cluster randomized trials of diagnostics are more practical and are needed to fully capture the impact on the overall diagnostic process [19]. In settings where diagnostic rollout is planned, trial designs that capitalize on implementation, such as stepped-wedge trials, may provide valuable comparative data to inform the clinical impact of the diagnostic test. Pragmatic trials are also essential for evaluating the use of diagnostic tests in routine practice [22]. Such outcomes data will be increasingly important in evaluation of new diagnostic tests, as using accuracy alone is appropriately downgraded in systems (eg, Grading of Recommendations Assessment, Development and Evaluation [GRADE]) for assessing the strength of evidence for making policy recommendations [23–27].

In the absence of comparative effectiveness data, one must attempt to translate gains in diagnostic accuracy into clinical benefits (eg, deaths or disability averted). Models are often used for this purpose; these models must construct a “counterfactual” scenario that describes when individuals who were diagnosed by the novel diagnostic would have been diagnosed by the standard of care, and what their corresponding risk of death or disability would have been. There are few data and little agreement among modelers in this regard. Some models, for example, have assumed that all individuals missed by screening would never be diagnosed [28–30]. This will overestimate the benefits of the novel diagnostic. More realistically, many individuals will return to clinics, and their tuberculosis will be diagnosed (either by clinical features or with newly positive tests) at subsequent visits as their disease progresses [31]. However, particularly in the case of HIV/tuberculosis coinfection, some patients will also die prior to their diagnosis. Which of these events would occur, and when, have not been well characterized. In the case of tuberculosis among HIV-uninfected individuals, historical data from the prechemotherapeutic era provide insight on the survival of individuals with untreated tuberculosis [32]. However, because tuberculosis chemotherapy was developed decades before HIV was even discovered, we have few data on survival of untreated tuberculosis among individuals with HIV.

Furthermore, progression of disease and survival of untreated tuberculosis is strongly affected by additional clinical factors such as CD4 cell count and receipt of antiretroviral therapy; again, there are no published data on these relationships. Most cost-effectiveness models also do not consider that survival differs according to diagnostic result; positive results by smear, urine LAM, or sputum Xpert have each been associated with a different risk of death among people being treated for tuberculosis [33, 34]. For example, individuals with detectable urine LAM are more likely to have disseminated tuberculosis, which entails a high risk of death. If earlier diagnosis (despite dissemination of tuberculosis) can avert deaths in these patients, urine LAM screening would confer great clinical benefits. Alternatively, if disseminated tuberculosis disease is generally too advanced for treatment to be effective, the incremental benefits of urine LAM testing may be limited. Because such data are very difficult to obtain in relation to a counterfactual of no testing, few models have accounted for this important issue.

The impact of false-positive diagnostic results, particularly those at low CD4 cell counts, also requires additional consideration. The differential diagnosis of infections and malignancies that may be present instead of—or in addition to—tuberculosis in individuals with low CD4 cell counts is broad. Therefore, false-positive test results for tuberculosis may be more harmful in HIV-infected individuals by delaying diagnosis of other life-threatening conditions. However, because nearly all evaluations of tuberculosis diagnostics focus primarily on tuberculosis-specific outcomes, our knowledge of the harms of a false-positive test result remains poor. We desperately need more comprehensive and robust clinical outcomes data, including diseases other than tuberculosis (eg, bacterial or viral pneumonia, Pneumocystis jirovecii, malignancies), among individuals being tested with novel tuberculosis diagnostics.

**HIV CARE COSTS, NOT DIAGNOSTIC-RELATED COSTS, DRIVE INCREMENTAL COSTS**

Antiretroviral therapy is recommended by the WHO for all HIV-infected patients with active tuberculosis [35]. For tuberculosis diagnostics to be fully effective in individuals with HIV, a positive test must be followed by treatment—for both tuberculosis and HIV. Diagnostics are typically one-time costs, whereas treatment is both expensive and lifelong. As a result, if programs are willing to pay the cost of antiretroviral therapy for tuberculosis survivors, they should almost always be willing to adopt the tuberculosis diagnostic ($15 one-time cost) on cost-effectiveness grounds, even if hundreds of tests are needed to benefit one person. Similar findings have been shown in other analyses of one-time tests, including genotypes for antiretroviral therapy resistance in treatment-naive patients [36]. Thus, the more relevant question is often not, “Is adopting a novel tuberculosis diagnostic cost-effective?”, but rather, “Is a program willing to provide ART to all people diagnosed (either true-positive or false-positive) with the new test?”

We illustrate this point with a very simple model of HIV and tuberculosis cost of care in South Africa. We assume that the diagnostic is 20% more sensitive in absolute terms than the standard of care; that 25% of persons with suspected tuberculosis have tuberculosis; and that early diagnosis conservatively averts just 1 death among every 25 correctly diagnosed patients (ie, the “number needed to test” to save one life is 500). If the diagnostic test costs $15 more than the standard of care, tuberculosis treatment costs $150/month and HIV care costs $2500/
Figure 1. Cumulative incremental costs over time after the introduction of a novel tuberculosis diagnostic test among patients with human immunodeficiency virus (HIV). We assume the following: tuberculosis prevalence is 25%; the novel diagnostic is 20% more sensitive than the standard of care in absolute terms; early diagnosis results in treatment in 20% of patients who otherwise would not receive it; and, conservatively, additional treatment averts death in just 20% of cases, such that just 1 in 500 deaths is averted through use of the diagnostic test. Additionally, we assume that life expectancy among survivors averages 10 years; diagnostic cost is $15, tuberculosis treatment costs are $150/month; HIV care costs are $2500/year, and discounting is 3% per year. Cumulative incremental costs are given on the vertical axis and include incremental diagnostic costs (blue), tuberculosis treatment costs (yellow), and HIV care costs (red) over time (horizontal axis). The downstream costs of tuberculosis and HIV care are estimated as $\sum (1 - mortality)^t \cdot Cost/(1+\text{discount})^t$, using a 0.1-year time-step. Tuberculosis treatment costs only apply for the first 6 months. Abbreviations: HIV, human immunodeficiency virus; TB, tuberculosis; USD, US dollars.

The fact that most downstream costs associated with novel tuberculosis diagnostics among HIV-infected populations reflect HIV care and treatment (rather than tuberculosis diagnosis and treatment) has important implications. Lifetime costs of care for HIV-infected individuals who survive as a result of using the diagnostic test are higher than for HIV-uninfected people. Therefore, tuberculosis diagnostics may appear to be less cost-effective when implemented in HIV-infected populations. This discrepancy raises ethical concerns about the fairness of comparing cost-effectiveness across HIV-infected and -uninfected subpopulations. Methodologically, the same cost-effectiveness analysis might show dramatically different results if conducted from the perspective of the tuberculosis program alone (ignoring HIV care costs but with more stringently constrained resources and willingness to pay) versus the HIV care system (including HIV care costs but with more resources available). Recent cost-effectiveness analyses for the same diagnostic test have differed on this critical point of whether to include HIV care costs, leading to estimates of incremental cost-effectiveness that differ by nearly a factor of 10 [10, 13].

In many low- and middle-income countries, HIV programs operate separately and maintain independent budgets from tuberculosis control programs. Over the past decade, HIV programs have received far more funding than tuberculosis (even relative to the burden of disease) [42], and their implicit cost-effectiveness thresholds for adopting new interventions are generally much higher. As many countries face funding shortfalls in their HIV budgets as well, they must make decisions about which HIV treatment recommendations (expanding access to ART, use of newer regimens, CD4 thresholds, second- and third-line therapies) to prioritize [43]. Although tuberculosis diagnostic tests generally do not enter into these discussions, tuberculosis remains the largest cause of mortality among individuals with HIV infection, such that implementation of tuberculosis diagnostic tests might save more lives of people living with HIV, at lower cost, than many of the alternatives being considered. Decisions regarding tuberculosis diagnosis strategies among HIV-infected individuals will therefore need to involve joint planning—despite separate budgeting—by HIV and tuberculosis programs.

Finally, estimates of the annual costs of care for HIV, including medical visits, antiretroviral therapy, ancillary medications, hospitalizations, etc, vary substantially. In South Africa alone, for example, studies have reported annual costs that ranged from $580 to $5220 (adjusted to 2011 US dollars), nearly a 10-fold difference [40, 41]. Given the importance of HIV care costs in driving ICERs, the uncertainty expressed in these costs is among the greatest sources of uncertainty in cost-effectiveness analyses of tuberculosis diagnostics among people living with HIV—far more important, for example, than uncertainty in the sensitivity of the tuberculosis diagnostic test, or its cost. A better understanding of the downstream costs of care will lead to more accurate projections of the true costs and cost-effectiveness of implementing novel tuberculosis diagnostics.

CONVENTIONAL COST-EFFECTIVENESS BENCHMARKS MAY BE INAPPROPRIATE

Despite our current limitations in estimation of costs and effectiveness, it is qualitatively clear that adoption of novel tuberculosis diagnostics will be deemed cost-effective by conventional thresholds for most settings—as long as ART costs are not disproportionately large. Although there is disagreement about the threshold at which a health intervention becomes "cost-effective," the most widely used metric is that of the WHO Commission on Macroeconomics and Health, which denoted interventions “highly cost-effective” if the cost, per DALY saved,
is less than the per capita Gross Domestic Product (GDP) and “cost-effective” if the cost is <3 the per capita GDP [44].

Traditional cost-effectiveness measures and benchmarks, while endorsed by international public health bodies [15, 45], nevertheless cannot be the only considerations for programs and governments in deciding whether to adopt a new diagnostic. Budgetary and political decision-making horizons are typically much shorter than the lifetime cost perspective commonly recommended for cost-effectiveness comparisons. Suggested thresholds for “highly cost-effective” or “cost-effective” interventions may not be relevant to actual budgetary situations of various countries and programs, which may be able to fund only the most cost-effective interventions in some sectors (eg, tuberculosis) but adopt a more liberal constraint in others (eg, HIV). Furthermore, competing demands or priorities, existing commitments, and inelastic budgets may render even programs that are highly attractive by cost-effectiveness criteria infeasible for implementation. The fact that a diagnostic test may be a “good value” does not imply that it will be affordable. A country may be able to pay for an intervention that costs $1 million and provides 1000 life-years but not an intervention that costs $1 billion (1000 times more costly) and provides 2 000 000 life-years (2000 times as effective), even though the latter provides twice as many life-years per dollar spent (ie, is more cost-effective).

The time horizon of the costs also bears heavily on the question of affordability. In the example above, the costs of implementing a new tuberculosis diagnostic test over the first 5 years involve mostly additional tuberculosis diagnoses and treatments, whereas life-years continue to accrue in future years, with no additional tuberculosis-related costs. The upfront costs may be borne by the tuberculosis program and deemed unaffordable in the short term, despite the long-term survival benefits that patients will experience. Because diagnostics are likely to appear cost-effective from the point of view of traditional cost-effectiveness analyses (which consider full life expectancy) and yet still may not be affordable, budgetary impact analyses that consider several time horizons (eg, 5 vs 20 years), different perspectives (eg, tuberculosis program vs HIV program), and full costs (rather than just cost-effectiveness) may provide additional useful information.

CONCLUSIONS

New diagnostic tests for tuberculosis have great potential to reduce the burden of morbidity and mortality from HIV-associated tuberculosis, but this effectiveness comes with substantial cost. Given the many competing healthcare needs and constrained budgets in most high-burden settings, appropriate estimation of the cost-effectiveness of these diagnostics is of paramount importance. Unfortunately, cost-effectiveness analyses of tuberculosis diagnostics in HIV-endemic settings face unique challenges in the estimation of incremental effectiveness (which requires comparative data of scenarios with and without the new test), incremental costs (which are largely driven by HIV care), and cost-effectiveness thresholds (which may be inadequate to inform decisions on adoption of diagnostics). To confront these challenges, we need (1) studies to provide data on diagnostic effectiveness, rather than just accuracy; (2) locally relevant, high-quality data on the costs of HIV care; and (3) joint budgetary planning among HIV and tuberculosis programs to anticipate the downstream costs associated with adopting novel diagnostics.

Additionally, a better understanding of the time course of tuberculosis transmission events with respect to diagnostic testing may enable better projections of the population-level benefits of new diagnostic tests [46].

As countries and programs in HIV-endemic settings make decisions about which diagnostics to prioritize in their tuberculosis control strategies, cost-effectiveness analyses can provide valuable information. Understanding the unique limitations and mechanics of cost-effectiveness analysis when evaluating diagnostics for HIV-associated tuberculosis will enable better decision making in these settings, and ultimately better health for people living with HIV.

Notes

Author contributions. J. R. A., S. D. L., D. W. D., and R. W. P. conceived of the content; J. R. A. performed the analysis and wrote the first draft. All authors contributed to the writing of the manuscript.

Financial support. This work was supported by the National Institute of Allergy and Infectious Diseases (grant numbers R01 AI099326, R01 AI058736, and T32AI007433) and the Harvard Global Health Institute. S. D. L. is funded by the Wellcome Trust, London, UK.

Disclaimer. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, which played no role in the study design, methods, interpretation of results, the content of this manuscript, or the decision to submit it for publication.

Potential conflicts of interest. All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References
