Do high rates of empirical treatment undermine the potential effect of new diagnostic tests for tuberculosis in high-burden settings?

Grant Theron, Jonny Peter, David Dowdy, Ivor Langley, S Bertel Squire, Keertan Dheda

In tuberculosis-endemic settings, patients are often treated empirically, meaning that they are placed on treatment based on clinical symptoms or tests that do not provide a microbiological diagnosis (eg, chest radiography). New tests for tuberculosis, such as the Xpert MTB/RIF assay (Cepheid, Sunnyvale, CA, USA), are being implemented at substantial cost. To inform policy and rationally drive implementation, data are needed for how these tests affect morbidity, mortality, transmission, and population-level tuberculosis burden. If people diagnosed by use of new diagnostics would have received empirical treatment a few days later anyway, then the incremental benefit might be small. Will new diagnostics substantially improve outcomes and disease burden, or simply displace empirical treatment? Will the extent and accuracy of empirical treatment change with the introduction of a new test? In this Personal View, we review emerging data for how empirical treatment is frequently same-day, and might still be the predominant form of treatment in high-burden settings, even after Xpert implementation; and how Xpert might displace so-called true-positive, rather than false-positive, empirical treatment. We suggest types of studies needed to accurately assess the effect of new tuberculosis tests and the role of empirical treatment in real-world settings. Until such questions can be addressed, and empirical treatment is appropriately characterised, we postulate that the estimated population-level effect of new tests such as Xpert might be substantially overestimated.

Introduction

Although several factors, including reduction of poverty and improved access to treatment, are crucial to reduce the global burden of tuberculosis, accurate and rapid diagnostic tests are a major unmet need. Xpert MTB/RIF—an automated real-time PCR platform for diagnosis of tuberculosis and detection of rifampicin resistance—is endorsed by WHO and the USA Food and Drug Administration and is undergoing implementation in several high-burden countries. Xpert is usable at the point-of-care and can detect about two-thirds of smear-negative tuberculosis cases in less than 2 h. The widespread implementation of Xpert will need substantial investment by international donors and governments of resource-poor countries.

Modelling studies have indicated that accurate and potentially same-day tuberculosis diagnostics could reduce mortality by 20–35% by enabling earlier initiation of tuberculosis treatment. However, in HIV-endemic settings with a high tuberculosis-related mortality, clinicians compensate for the shortcomings of smear microscopy (frequently the only routinely available tuberculosis test) with the initiation of treatment on the basis of clinical symptoms, less specific tests (such as chest radiography), or absence of a response to broad-spectrum antibiotics. The initiation of treatment in the absence of a bacteriologically confirmed diagnosis is often referred to as empirical tuberculosis treatment.

In settings with high rates of empirical treatment, the effect of Xpert and other new tuberculosis tests such as the urine LAM (lipoolarabinomannan) lateral flow assay on individual-level outcomes and population-level epidemiology might be lower than predicted (table). Although the number of bacteriologically confirmed diagnoses will increase with the roll-out of Xpert, how many of these newly detected patients would have been placed on treatment in the absence of Xpert, and when this would have occurred, is unknown. A proposed benefit of Xpert is improved outcomes (eg, lower mortality) in the sickest individuals; however, doctors are most likely to treat the same patients empirically (and treat them rapidly), such that the incremental benefit of Xpert might be diminished. Thus, certain key questions remain: will Xpert actually decrease the time to treatment initiation in high-burden settings with high rates of empirical treatment to an extent that affects outcomes for patients and ongoing transmission, or will it only replace empirical tuberculosis treatment that would otherwise occur near the same time? Will Xpert change empirical tuberculosis treatment practice, reduce the proportion of false-negative diagnoses, and reduce the proportion of patients with false-positive results who are placed on treatment inappropriately? Might some patients with tuberculosis but a negative Xpert result not receive treatment because of increased confidence in Xpert?

Empirical tuberculosis treatment initiation

Drivers of empirical treatment

The clinical basis for empirical tuberculosis treatment varies across settings in accordance with factors that contribute to a pretest probability of a patient having tuberculosis or a poor outcome or both, which is weighted against a variable and subjective threshold for treatment initiation (figure). Such factors include baseline tuberculosis prevalence (eg, among patients with HIV with advanced immunosuppression), a clinical presentation suggestive of tuberculosis, results (if any) of adjunctive but non-confirmatory diagnostic methods
Xpert MTB/RIF assay | Empirical treatment
---|---
**Patient-level**
Time to diagnosis | Expected to be similar to smear microscopy (1–3 days); potentially offers same-day diagnosis but unlikely with centralised roll-out | Ranges from same-day to 1–2 weeks, depending on setting (highly setting and patient dependent); algorithms for smear-negative tuberculosis reduce the time to diagnosis, but not as much as do bacteriological tests
Time to treatment | Rapid (1–3 days) | Can be rapid in primary care for most patients, but will probably be 2–3 days behind Xpert (dependent on setting and patient); same-day empirical treatment can be common in primary care in Africa
Morbidity and mortality | Emerging data suggests that where high rates of rapid empirical treatment exist, the 1–3 day advantage in treatment initiation created by Xpert does not translate into improved patient-level morbidity; further research is needed | Rapid empirical treatment can reduce mortality, and clinicians empirically treat the sickest patients at the greatest risk; effect of Xpert on these endpoints might therefore be overestimated in such populations

**Population-level**
Patients treated | 90% sensitivity, susceptible to sampling error and paucibacillary disease | Sensitivity of 20–80%; can miss patients with tuberculosis who might not return after a negative test result at first visit
False-positive treatment without bacteriological confirmation of disease | Emerging data suggest a similar number of patients without microbiologically confirmed tuberculosis are placed on treatment when Xpert is available | Overtreatment for people with tuberculosis is a concern; the cost and health implications of inappropriate tuberculosis treatment needs further study
Transmission | Can reduce the infectious period, but perhaps by only a few days; whether this reduction is meaningful or cost effective is unclear; can detect rifampicin-resistant tuberculosis and thereby help to reduce its transmission, but this effect is dependent on availability of second-line drugs | Frequently started rapidly, but generally later than is treatment due to Xpert; when empirical treatment is initiated rapidly, the effect of Xpert on transmission might be overestimated because a few days difference in time to treatment is unlikely to make a meaningful difference; will not help to stop the spread of drug-resistant tuberculosis

Table: Expected effects of empirical tuberculosis treatment and Xpert MTB/RIF on key endpoints for the reduction of the tuberculosis epidemic in high-burden settings

(such as chest radiography), and concern that further delay might increase risk of severe morbidity or mortality. Inability to do microscopy or Xpert (as in sputum-scarce patients), substandard clinical training, and high likelihood of one-off encounters with patients might also drive the initiation of empirical treatment.13–21

**Timing of empirical treatment initiation**

Empirical treatment can be initiated at any stage in the diagnostic pathway; either before an available Xpert or microscopy result (eg, in very ill or immunosuppressed patients, or when a test is not available in peripheral facilities), a few days or weeks later (eg, after failure to respond to a short course of broad-spectrum antibiotics), or only once all available bacteriological tests, including culture if available, have failed to provide a positive result.24 Data for timing of empirical treatment with Xpert availability are scarce; however, in the TB-NEAT study,31% of patients with smear-negative tuberculosis who started treatment did so within 48 h of entering the health-care facility.

**Accuracy**

Because empirical tuberculosis treatment is not standardised, a global estimate of the accuracy of empirical treatment might not be measurable or useful. WHO developed an algorithm for smear-negative tuberculosis in high-burden HIV-endemic settings to standardise, and improve diagnosis and speed of initiation for, tuberculosis treatment.22 A meta-analysis23 of prospective assessments24–28 showed that empirical treatment has a pooled sensitivity of 61% (95% CI 55–67%) and specificity of 69% (66–72%) for smear-negative tuberculosis. In a post-mortem study29 in South Africa that was predominantly of individuals with HIV infection, the poor sensitivity of empirical treatment was shown by a tuberculosis prevalence of 50% for hospital inpatients not on treatment before death. In the TB-NEAT study,1 a randomised controlled trial of microscopy versus point-of-care Xpert as the initial test for 1500 patients from primary care clinics in South Africa, Tanzania, Zambia, and Zimbabwe, same-day microscopy combined with a WHO-compliant empirical treatment algorithm including chest radiography resulted in only 79% of culture-positive patients starting treatment. However, clinicians in real-world settings might do much better than standardised algorithms. For example, a review of Chinese programmatic data showed that only 3% of patients treated empirically for tuberculosis were confirmed to have an incorrect diagnosis.30

**Advantages and disadvantages**

Aside from averted diagnostic costs, several benefits—including reduced morbidity, mortality, and transmission—could result from empirical treatment if it leads to treatment initiation before a confirmed diagnosis is available. Other important advantages might be an effect on tuberculosis incidence (through the treatment of latent tuberculosis) and other bacterial infections.6 Empirical treatment without a definitive diagnosis can also cause important harms, including unnecessary cost, toxic effects, and inconvenience to people without tuberculosis, increased morbidity and mortality from other underlying diagnoses that are not considered, inability to collect information about drug resistance, stigmatisation of patients, and economic losses due to inappropriate tuberculosis diagnosis. This balance of harms and benefits depends on factors such as the specificity of empirical diagnosis (which can be low),32 relative cost of diagnosis versus treatment, and willingness of the clinician to...
discontinue empirical treatment if another diagnosis becomes more likely.

**Xpert MTB/RIF and how its effect might be modulated by empirical treatment**

Xpert detects 40–60% more tuberculosis cases than does microscopy. Because of the high rates of empirical treatment for smear-negative patients, it is unclear which patients empirically treated for smear-negative tuberculosis overlap with those who would be detected by Xpert. A multicentre study showed that Xpert reduced the proportion of untreated culture-positive patients from 40% to 15%; however, this reduction in dropouts can be offset by empirical treatment in pragmatic settings. For example, in the TB-NEAT study, of the 68% of patients with smear-negative tuberculosis detected by Xpert, 93% were treated empirically anyway. Whether individuals who receive early empirical treatment are, in fact, the ones who benefit the most from an early diagnosis is unknown.

How will empirical tuberculosis treatment change with the roll-out of Xpert? An important potential advantage of Xpert is that it might reduce unnecessary empirical treatment by increasing clinicians’ confidence in negative results, provided that an adequate specimen can be obtained. Data to support this possibility are scarce, but increasing. In a recent cohort study of patients attending a primary care tuberculosis clinic in South Africa, 45% of patients who started treatment did so on the basis of empirical decision making (more than those diagnosed by Xpert alone); 60% were later shown to be likely false-positive treatment decisions (ie, patients were culture-negative). There is scarce data about the effect of Xpert on the accuracy of this false-negative treatment; however, these findings are supported by data from the TB-NEAT study, in which more than half of the patients who started treatment did so on the basis of empirical decision making. A similar proportion (roughly 22%) of culture-negative patients also received treatment, irrespective of Xpert availability, suggesting that Xpert might replace appropriate so-called true-positive rather than inappropriate false-positive empirical treatment. In a study in a South African hospital, although the proportion of patients who were started on treatment empirically dropped from 80% to 30% after Xpert implementation, the proportion of empirically treated patients who were culture-negative was still 40%. However, it should be noted that culture itself is not a perfect reference standard for tuberculosis, and patients who are culture-negative can still benefit from tuberculosis treatment.

**Figure: Factors affecting the empirical treatment threshold**

The threshold for empirical treatment might rise, stay the same, or lower after Xpert implementation. In the pre-Xpert scenario, six of ten patients without tuberculosis and eight of ten patients with tuberculosis are treated. Xpert implementation could change the threshold for empirical treatment according to one of three different scenarios: threshold raised (A), Xpert will reduce false-positive treatment of people without tuberculosis and increase true-positive treatment; threshold remains constant (B), Xpert will not change the rates of false-positive treatments, but will increase true-positive treatments; or threshold lowered (C), Xpert will increase false-positive treatment of people without tuberculosis and increase true-positive treatment. Tests are assumed to have high specificity.
Modelling the effect of Xpert MTB/RIF
Projections for the effect of Xpert

Since data for the population-level effect and cost-effectiveness of Xpert (eg, from continuing community-randomised trials) will probably not be available until mid-2014, projecting the effect of Xpert’s deployment has become a priority for mathematical modellers. Although several analyses have assessed the potential economic implications of Xpert roll-out,7–25 and other data-driven studies have estimated the effect of Xpert testing on individual patient-level outcomes,26–28 only one published article has formally modelled the effect of Xpert scale-up on population-level epidemiology, including the incidence and mortality of tuberculosis and multidrug-resistant-tuberculosis.29 Further clarification is needed as to exactly which patients are missed by empirical treatment but detected by Xpert, and for the effect these patients have on transmission at the population level.

The effect of empirical treatment on model outcomes
Xpert is often assumed to reduce population-level transmission more than does empirical treatment through a decrease in the time to diagnosis and increase in the likelihood of diagnostic success. However, models might overestimate the incremental population-level effect of Xpert relative to empirical treatment alone to the same extent that they overestimate the number of people with Xpert-confirmed tuberculosis who would not otherwise receive empirical treatment. For example, many published models assume that, in the absence of Xpert, up to 80% of people with true, smear-negative, tuberculosis would not be empirically diagnosed and treated in any given attempt.30–31 However, in high-burden settings without a large private sector, the WHO DOTS (directly observed treatment, short-course) country statistics estimate that 30–50% of patients with notified pulmonary tuberculosis are started on treatment in the absence of a positive bacteriological result. Similarly, results of studies from Uganda32 and Kenya33 have shown that empirically treated patients in primary care can outnumber those treated with a bacteriological diagnosis.34 In some settings, reductions from 80% to 50% in the proportion of people with Xpert-positive tuberculosis who would remain undiagnosed after empirical attempts would reduce the cost-effectiveness of Xpert by half.35 Thus, a better understanding of the role of clinical and empirical diagnosis for individuals who, if tested, would be smear-negative, but Xpert-positive, is crucial to generate more accurate model-based assumptions of Xpert’s potential effect on tuberculosis epidemiology at the population level. This understanding is also crucial to project effects on costs to patients and health system. For example, more rapid diagnosis (from being diagnosed upfront by Xpert rather than delayed by negative smear and then empirical diagnosis) could lead to reduced costs incurred by patients, lower diagnostic default, and maybe even more patients seeking diagnosis. Also, effects on empirical diagnosis will change the number of patients given treatment and, therefore, treatment costs, which could affect affordability and sustainability of any new diagnostic intervention.

How can empirical treatment data be incorporated into models?
The incorporation of clinical diagnosis into models of Xpert’s effect poses a unique challenge. Specifically, most models of effect or cost-effectiveness conceptualise the diagnosis of tuberculosis as a series of diagnostic attempts that could either fail or succeed, depending on the accuracy of the test used. However, clinical diagnosis does not fit easily into this framework because decisions are often made on the basis of complex ancillary information, including the failure of the initial diagnostic test. As such, clinical diagnoses will generally happen between diagnostic attempts (a series of microbiological tests); furthermore, the availability of a new diagnostic test itself might affect the process of clinical diagnosis. During any diagnostic delay, patients might have additional morbidity, incur additional costs, and have an increased risk of mortality; however, the duration of delay between a false-negative Xpert result and a clinical diagnosis resulting in empirical treatment is unlikely to be the same as that between a false-negative Xpert result and a completely new diagnostic attempt (eg, patients presenting to another health-care facility). Combining of all these effects— which are highly heterogeneous between clinical settings—at the population level is arguably an impossible task. Thus, model-based projections must make simplifying assumptions (eg, assigning of a single sensitivity of clinical diagnosis); the alternative option is to make decisions without any population-level projections whatsoever. However, for clinical diagnosis, practitioners generally have more information than is credited to them by models, and likely make more appropriate decisions than models project. As such, models, which generally presume poor performance of clinical diagnosis, might substantially overestimate the true effect of Xpert as implemented at the population level.

Information about empirical treatment is needed to improve models
Detailed operational models36 that simulate individual patients and sample pathways can include empirical treatment. They can also include estimates for the false-positive rate of treatment for different types of patient (eg, those with and without HIV) and at different stages in the diagnostic pathway (eg, after smear microscopy or Xpert). Evidence from these models indicates that in countries such as Tanzania that have a high rate of empirical treatment (smear-negative tuberculosis represents about 45% of cases37), Xpert might lead to a substantial positive effect on numbers of individuals who are microbiologically cured.38 The increased cures result from treatment of Xpert-positive patients who might have been missed with empirical treatment decision making. An analogous
scenario is noted in antigen-detection testing when LAM identifies a subset of cases missed by empirical treatment in a hospital setting. Understanding of these opposing effects is crucial in cost-effectiveness studies and in prediction and evaluation of the effect of new diagnostics like Xpert. Operational modelling can help, field data are needed for actual numbers of patients wrongly started on treatment empirically, along with the reasons why this occurred (eg, reliance on results of chest radiography in smear-negative cases).

Implications of high rates of empirical treatment for Xpert MTB/RIF placement

The extent to which Xpert will improve the number of true-positive patients with tuberculosis on treatment will depend on the setting-specific appropriateness and timing of empirical tuberculosis treatment. For example, in settings in which empirical treatment is prevalent and rapid (within 2–3 days), Xpert is unlikely to substantially improve outcomes for patients. However, in settings in which empirical treatment is delayed, the reduced time to diagnosis (from weeks to days) offered by Xpert might result in substantial improvements in patient health and dropout rates. These data imply that Xpert might have the greatest effect in settings or groups of patients in which empirical treatment is late or occurs at a low rate, such as in low-burden countries, routine (non-tuberculosis) services within antiretroviral clinics, or used for active tuberculosis case finding. Xpert, in view of its ability to detect rifampicin drug resistance, might also have greatest effect (in terms of placing of patients on effective therapy) in settings or populations with a high suspicion of drug resistance. This placement is important, not only to individual patients, but also to populations, because appropriate treatment can substantially reduce the transmission of drug-resistant tuberculosis. The effectiveness of empirical first-line treatment in such settings will, of course, be diminished; alternatively, the benefit of Xpert might be abrogated in settings in which consistent supplies of high-quality treatment for drug-resistant tuberculosis are not routinely available. In addition, placement of Xpert in central laboratories rather than at the point-of-care will delay the reporting of results and affect dropout rates, meaning that the main benefit of Xpert is available as opposed to scenarios when it is not. Thus, whether roll-outs of new diagnostic technologies reduce false-negative treatment decisions, and hence, affect morbidity and mortality, needs clarification. Cohort studies done before and after the introduction of Xpert into a health-care system and pragmatic randomised controlled trials (including large-scale cluster randomised trials) can be informative. Furthermore, to get a better sense of the true nature of false-negative patients, more robust sampling methods (eg, sputum induction, bronchoscopy, and biopsy for extrapulmonary tuberculosis) should be incorporated into diagnostic algorithms however, these studies might have limited utility. Additional data could include observational studies of patients’ and providers’ behaviour (eg, patterns of care-seeking and empirical treatment) before and after implementation of Xpert, ascertainment of preferences for true-positive diagnoses versus false-positive and false-negative diagnoses (eg, how many false-positive treatments would be tolerated to treat one true-positive or false-negative patient?), the interviewing of clinicians about their treatment decisions before and after provision of an Xpert result, and implementation research to assess the real-world use of Xpert rather than use in tightly controlled, strictly algorithmic research studies. Such real-world studies will need to cover a wide range of clinical settings (eg, inpatient vs outpatient, high HIV-infection prevalence vs low, high prevalence of multidrug-resistant tuberculosis vs low, and public vs private settings) to fully characterise the range of empirical treatment practices and the ways in which Xpert could modify those practises. Finally, new technologies might need to be used to actively seek out cases, rather than detect them passively, if tuberculosis infection rates are to be reduced.

Conclusion

Xpert is a substantial advance over smear microscopy, but the population-level effect of Xpert scale-up remains unclear because of uncertainty about the extent and accuracy of empirical treatment patterns in high-burden settings. Compared with the existing standard of care in such settings, in which smear microscopy is the only available tuberculosis test, Xpert might increase true-positive treatment decisions, but, depending on how much empirical treatment is displaced, the major effects of Xpert might be to reduce false-negative treatment, reduce the time to treatment either marginally or substantially (the effect of which on transmission, morbidity, and mortality itself needs clarification), or even cause some individuals who would be treated empirically without Xpert to not be given treatment after a false-negative Xpert result. Future studies of real-world Xpert use, including empirical treatment practices, are essential. Until we better understand the complex interplay between Xpert-based diagnosis and empirical treatment for tuberculosis, we should continue to use Xpert as the most appropriate test for microbiologically confirmed diagnoses in many settings, but we should not simply assume that Xpert scale-up, with its associated immense costs, will have large population-level effects on tuberculosis incidence or mortality.
Contributors
GT conceptualised the Personal View. GT, JP, DD, and KD designed the Personal View. All authors wrote and critically revised the article.

Conflicts of interest
We declare that we have no conflicts of interest.

References