UCSF

UC San Francisco Previously Published Works

Title

Impact of GeneXpert MTB/RIF on Patients and Tuberculosis Programs in a Low-Burden Setting. A Hypothetical Trial

Permalink

https://escholarship.org/uc/item/6wr7f4rs

Journal

American Journal of Respiratory and Critical Care Medicine, 189(12)

ISSN

1073-449X

Authors

Davis, J Lucian

Kawamura, L Masae

Chaisson, Lelia H

et al.

Publication Date

2014-06-15

DOI

10.1164/rccm.201311-1974oc

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

ORIGINAL ARTICLE

Impact of GeneXpert MTB/RIF on Patients and Tuberculosis Programs in a Low-Burden Setting

A Hypothetical Trial

J. Lucian Davis[1,2], L. Masae Kawamura[3], Lelia H. Chaisson[1], Jennifer Grinsdale[3], Jihane Benhammou[4], Christine Ho[5], Anna Babst[6], Houmpheng Banouvong[3], John Z. Metcalfe[1,2], Mark Pandori[6], Philip C. Hopewell[1,2], and Adithya Cattamanchi[1,2]

[1]Division of Pulmonary and Critical Care Medicine and [2]Curry International Tuberculosis Center, Department of Medicine, San Francisco General Hospital, University of California San Francisco, San Francisco, California; [3]Tuberculosis Section and [6]Public Health Laboratory, San Francisco Department of Public Health, San Francisco, California; [4]Department of Medicine, University of California, Los Angeles, Los Angeles, California; and [5]Centers for Disease Control and Prevention, Atlanta, Georgia

Abstract

Rationale: Guidelines recommend routine nucleic-acid amplification testing in patients with presumed tuberculosis (TB), but these tests have not been widely adopted. GeneXpert MTB/RIF (Xpert), a novel, semiautomated TB nucleic-acid amplification test, has renewed interest in this technology, but data from low-burden countries are limited.

Objectives: We sought to estimate Xpert's potential clinical and public health impact on empiric treatment, contact investigation, and housing in patients undergoing TB evaluation.

Methods: We performed a prospective, cross-sectional study with 2-month follow-up comparing Xpert with standard strategies for evaluating outpatients for active pulmonary TB at the San Francisco Department of Public Health TB Clinic between May 2010 and June 2011. We calculated the diagnostic accuracy of standard algorithms for initial empiric TB treatment, contact investigation, and housing in reference to three *Mycobacterium tuberculosis* sputum cultures, as compared with that of a single sputum Xpert test. We estimated the incremental diagnostic value of Xpert, and the hypothetical reductions in unnecessary treatment, contact investigation, and housing if Xpert were adopted to guide management decisions.

Measurements and Main Results: A total of 156 patients underwent Xpert testing. Fifty-nine (38%) received empiric TB treatment. Thirteen (8%) had culture-positive TB. Xpert-guided management would have hypothetically decreased overtreatment by 94%, eliminating a median of 44 overtreatment days (interquartile range, 43–47) per patient and 2,169 total overtreatment days (95% confidence interval, 1,938–2,400) annually, without reducing early detection of TB patients. We projected similar benefits for contact investigation and housing.

Conclusions: Xpert could greatly reduce the frequency and impact of unnecessary empiric treatment, contact investigation, and housing, providing substantial patient and programmatic benefits if used in management decisions.

Keywords: tuberculosis; diagnosis; health care quality assurance; operations research; public health

(Received in original form November 9, 2013; accepted in final form May 14, 2014)

Supported by American Lung Association grant CG-197164 (J.L.D.); the UCSF Center for AIDS Research (J.L.D.); National Institutes of Health grants (K23 AI080147 (J.L.D.), K23 HL094141 (A.C.), K23 AI094251 (J.Z.M.), and R01 AI076476 and U01 AI088679 (P.C.H.); and National Center for Research Resources grant KL2 RR024130 (UCSF Clinical and Translational Sciences Institute). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the U.S. CDC.

Author Contributions: Study design, J.L.D., L.M.K., C.H., J.G., J.Z.M., P.C.H., and A.C. Laboratory testing, A.B. and M.P. Data collection, J.L.D., L.H.C., J.G., J.B., and H.B. Data analysis, J.L.D., L.H.C., J.G., and J.B. Drafting of the manuscript, J.L.D. L.H.C. and J.L.D. had full access to all data in the study and had final responsibility for the decision to submit for publication.

Correspondence and requests for reprints should be addressed to J. Lucian Davis, M.D., M.A.S., San Francisco General Hospital, Room 5K1, 1001 Potrero Avenue, San Francisco, CA 94110. E-mail: lucian.davis@ucsf.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 189, Iss 12, pp 1551–1559, Jun 15, 2014
Published 2014 by the American Thoracic Society
Originally Published in Press as DOI: 10.1164/rccm.201311-1974OC on May 28, 2014
Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the Subject: Although clinical practice guidelines have recommended routine use of nucleic-acid amplification testing in evaluating patients for tuberculosis (TB) since 1996, clinicians and laboratory personnel have not implemented these recommendations widely. The recent development and US Food and Drug Administration approval of GeneXpert MTB/RIF (Xpert), a next-generation, semiautomated TB nucleic-acid amplification test, has renewed interest in this technology. Data on the influence and impact of Xpert on clinical and public health decisions and outcomes are needed to inform its uptake.

What This Study Adds to the Field: Among outpatients undergoing evaluation for active pulmonary TB, we found that conventional clinical algorithms used to guide initial management decisions frequently led to unnecessary treatment and housing of patients who did not have TB, and unnecessary contact investigation. Replacing these algorithms with Xpert testing could eliminate most unnecessary interventions, benefitting both patients and public health programs.

Nucleic-acid amplification tests (NAATs) for tuberculosis (TB) have been commercially available in the United States and Europe for almost two decades (1, 2). During that time, evidence has accumulated showing that NAATs provide excellent diagnostic accuracy (3, 4) and additional value for diagnosing TB over clinical decision-making alone (5, 6). This evidence has led the US CDC (7–9) and the British Thoracic Society (10) to recommend routine use of NAATs to guide initial management of patients with possible TB. Nevertheless, NAATs have not been widely adopted in the United States (11) or the United Kingdom (12). Most public health laboratories do not perform TB NAATs routinely (12, 13), because first-generation commercial assays are labor-intensive and have not proved cost-effective in low-burden countries (14–16). Evidence of clinical impact is mixed, with some studies suggesting that NAATs rarely change management in these settings, especially when NAAT results are negative (17, 18). Newer data, however, suggest that among the subset of individuals selected to undergo NAAT, these assays can influence a variety of management decisions, and be cost-saving in some subpopulations of patients (19).

The GeneXpert MTB/RIF assay (Xpert; Cepheid Diagnostics, Sunnyvale, CA) is a novel, semiautomated NAAT with similar diagnostic accuracy to first-generation commercial NAATs (20, 21). Many clinical laboratories already use the Xpert platform for other diagnostic applications, and its minimal labor requirements make it simpler, faster, and potentially cheaper than previous NAATs. The European Union and World Health Organization have endorsed Xpert for TB evaluation (22), and on July 25, 2013 the US Food and Drug Administration authorized its use for TB evaluation in the United States (23).

Despite Xpert's attractive diagnostic and operational characteristics, the poor uptake of first-generation NAATs suggests that data on diagnostic accuracy alone could be insufficient to drive adoption and that evidence on clinical and public health decision-making and outcomes may be needed (24–26). Therefore, we designed a prospective observational study to estimate the hypothetical impact of Xpert as a replacement for standard clinical and programmatic criteria used in risk stratification and triage of patents undergoing evaluation for active pulmonary TB, while awaiting results of mycobacterial culture and longitudinal clinical assessment (27).

Methods

Study Design and Population

The hypothetical trial comparing the impact of different diagnostic strategies represents a novel study design (27), and may be useful when ethical concerns, regulatory barriers, or sample size limitations make a randomized trial unfeasible (28, 29). Hypothetical trials are observational studies that make paired measures of diagnostic accuracy for different evaluation strategies, and then project how the results of novel strategies might hypothetically affect management decisions and patient outcomes relative to the actual decisions and outcomes observed for the control strategy.

In this study, we screened consecutive adults undergoing evaluation for active pulmonary TB at the San Francisco Department of Public Health TB Clinic between May 2010 and June 2011 and asked clinicians to refer individuals in whom they believed a NAAT result could inform clinical or public health decisions, a prioritized group for testing according to CDC guidelines (9). We suggested two key groups of patients for Xpert testing: those initiating empiric treatment for active TB (i.e., treatment before a confirmed mycobacterial culture result); and those coming from congregate settings (e.g., homeless shelters, behavioral treatment programs, dialysis centers), in whom an inability to rapidly assess TB transmission risk often interrupts the patient's residence or care in the congregate environment and prompts orders for housing and contact investigation. We excluded patients with incomplete microbiologic or clinical follow-up data, and patients reporting TB treatment at the time of Xpert testing.

Procedures

All patients underwent standard evaluation for TB, including a clinical interview, physical examination, and frontal chest radiography. Immediately afterward, evaluating clinicians subjectively rated the probability of TB as low, moderate, or high. Program guidelines recommend using these categories to guide initial treatment decisions pending additional test results: patients classified as moderate or high risk are usually referred for both empiric treatment and immediate contact investigation, whereas these interventions are usually withheld in patients classified as low risk (30, 31). All patients provided three daily expectorated or induced sputum specimens for acid-fast bacilli (AFB) smear microscopy and culture for *Mycobacterium tuberculosis* complex (*see* online supplement for details). The San Francisco Department of Public Health Laboratory performed all microbiologic testing according to standard protocols (20, 32). Staff set aside 0.5 ml of the remaining sputum pellet for Xpert testing, which a clinical laboratory scientist performed approximately three times weekly (33). The laboratory reported results to the TB control program with

a disclaimer that the assay was not approved by the Food and Drug Administration as a diagnostic test for TB. Therefore, we were not able to evaluate the effects of the test on actual management decisions in this study.

Measurements and Statistical Analysis

We collected clinical and demographic information from the clinic's customized electronic clinical record. We replaced missing results with the median if less than five values were unavailable, and used a multivariate normal model if five or more were unavailable. Using culture of three sputum samples collected within 7 days of initial evaluation as a reference standard (with one or more positives defining TB and two or more negatives with no positives defining non-TB status), we calculated and compared the sensitivities, specificities, and positive and negative predictive values (PPV, NPV) of Xpert and of key clinical and public health decisions. These included decisions to (*1*) initiate TB treatment, (*2*) conduct contact investigation, and (*3*) provide subsidized housing. For discordant results, we reviewed patient records, and reported final clinical diagnoses in accordance with American Thoracic Society TB diagnostic standards (34). We used the McNemar test for paired proportions to assess the statistical significance of differences in sensitivity and in specificity, and the large sample test for unpaired proportions to assess differences in predictive values.

Next, for the period before the availability of final culture results only, we measured the consequences of Xpert-guided and standard decisions on treatment, contact investigation, and subsidized housing for individuals and for the program in aggregate over the approximately 1-year period of the study. Using measures of the time to report results for all diagnostic assays, we calculated differences between standard and Xpert strategies for the following outcomes among those with and without culture-confirmed TB: days of treatment, numbers of close contacts undergoing TB contact investigation, and days of subsidized housing (*see* online supplement). We compared differences in medians using the Wilcoxon signed rank test and differences in proportions using the chi-square test.

For all analyses, we defined significance in reference to the probability of a two-tailed, type I error (P value) less than 0.05. Because the sample size arose from convenience, we estimated the precision of outcomes using 95% confidence intervals (95% CI) (35). We performed all statistical analyses using Stata version 11.0 (Stata Corporation, College Station, TX).

Ethics Approval

The University of California San Francisco Committee on Human Research approved the study protocol, and waived the requirement for informed consent on grounds of minimal risk. The CDC author provided technical support only. This role did not constitute engagement in human subjects research; therefore, CDC institutional review board review and approval was not required. Some of these results have been previously reported in the form of an abstract (36).

Results

Study Enrollment

Of 538 consecutive patients undergoing evaluation for possible pulmonary TB during the 13-month study, 227 met eligibility criteria, including 132 patients coming from congregate settings and 95 patients receiving empiric treatment (Figure 1). Of these 227 patients, clinicians ordered Xpert in 156 (69%), including 97 coming from congregate settings but not receiving empiric treatment, and 59 receiving empiric treatment. Nine (15%) of these 59 also came from congregate settings but we analyzed them in the empiric treatment group. Patients in whom clinicians ordered Xpert were similar to those in whom they did not, with a few exceptions. Patients from congregate settings were more likely to undergo Xpert testing if female (risk ratio, 2.2; 95% CI, 1.08–3.4; $P = 0.035$) or foreign-born (risk ratio, 2.41; 95% CI, 1.41–3.2; $P = 0.005$) (*see* Table E1 in online supplement). Patients receiving empiric treatment were more likely to undergo Xpert if they had abnormal chest radiography (risk ratio, 1.43; 95% CI, 1.14–1.80; $P < 0.0001$) (*see* Table E2).

Patient Characteristics

Median age was 52 years (interquartile range [IQR], 39–60), and 54 (35%) were women (Table 1). A total of 117 (75%) were foreign-born, of whom 46 (39%) had immigrated to the United States within the previous 5 years. Twenty (13%) patients were homeless. Thirty-three (21%) reported

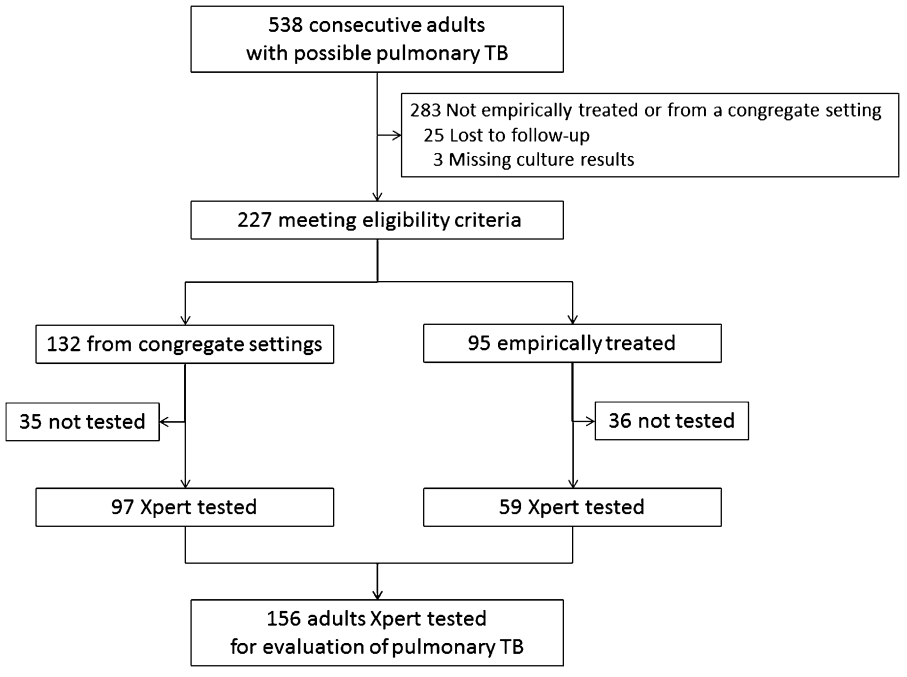

Figure 1. Patient enrollment flow diagram. TB = tuberculosis; Xpert = GeneXpert MTB/RIF.

Table 1. Demographic and Clinical Characteristics of Patients Being Evaluated for Pulmonary TB

Characteristic	From Congregate Settings (*n = 97*)	Empirically Treated* (*n = 59*)
Median (IQR) age	54 (39–59)	49 (38–63)
Women	36 (37)	18 (31)
Foreign-born	72 (74)	45 (76)
Homeless†	11 (11)	9 (15)
Clinician-estimated probability of TB		
Low	74 (76)	5 (8)
Moderate	21 (22)	23 (39)
High	2 (2)	31 (53)
Taking ⩾1 drug with a potential TB-drug interaction‡	40 (41)	24 (41)
⩾1 risk factor for TB drug–related hepatotoxicity§	17 (18)	16 (27)
Risk factors for rapid TB progression	8 (8)	9 (15)
HIV-seropositive	6 (6)	7 (12)
Immunosuppression, not caused by HIV	1 (1)	1 (2)
AFB smear-positive	6 (6)	16 (27)
Culture-confirmed TB	0 (0)	11 (19)
Culture-confirmed TB	1 (1)	12 (20)
AFB smear-positive	0 (0)	11 (19)

Definition of abbreviations: AFB = acid-fast bacilli; IQR = interquartile range; TB = tuberculosis.
Values are given as n (%) unless otherwise specified.
*A total of 9 of 59 (15%) empirically treated patients were homeless but analyzed with the empirically treated group rather than with the congregate settings group.
†Six missing observations, five from the congregate settings group and one from the empirically treated group.
‡Including antiretroviral therapy, oral contraceptives, immunosuppressive therapy, and methadone.
§Including ethanol use, chronic liver disease, and viral hepatitis.

a history of viral hepatitis, chronic liver disease, or regular ethanol use. Thirteen (8%) were HIV-infected. Although exact medication lists were not available for all patients, 64 (41%) reported taking one or more medications from drug classes commonly associated with TB drug interactions (i.e., antiretroviral therapy, oral contraceptives, immunosuppressive medications, or methadone). Fifteen (10%) were immunosuppressed. Among those tested with Xpert, the clinician-estimated probability of TB was low in 79 (51%) patients, moderate in 44 (28%), and high in 33 (21%). Twenty-two (14%) had positive sputum AFB smear microscopy results, but only 11 (50%) of these had positive results on *M. tuberculosis* complex cultures. Two patients with negative microscopy results had positive culture results. In total, 13 (8.3%) patients had culture-confirmed TB, including 1 of 97 (1.0%) from congregate settings, and 12 of 59 (20%) receiving empiric treatment. The public health laboratory completed Xpert testing in a median of 2 days (IQR, 1–3), and tested 95% of specimens within 5 days.

Diagnostic Accuracy

Fifty-nine (38%) patients referred for Xpert received empiric TB treatment, including 5 of 79 (6.3%) rated as low probability for TB, 23 of 44 (52%) rated as moderate probability, and 31 of 33 (94%) rated as high probability. A decision to treat empirically for TB had high sensitivity (12 of 13; 92%; 95% CI, 64–100) and high NPV (96 of 97; 99%; 95% CI, 94–100) for culture-positive TB. However, the specificity of empiric treatment decisions was poor (96 of 143; 67%; 95% CI, 59–75), and only 12 of 59 patients starting empiric treatment actually had TB (PPV, 20%; 95% CI, 11–33) (Figure 2).

Xpert had identical sensitivity to clinician-guided treatment decisions (sensitivity difference, 0%; 95% CI, −29 to +29; $P = 1.0$), detecting all 11 patients with positive AFB smear microscopy results (sensitivity, 100%; 95% CI, 72–100), and one of two patients with negative microscopy results (sensitivity, 50%; 95% CI, 1–100) (*see* Figure E1). Xpert also had a high NPV (140 of 141; 99%; 95% CI, 96–100), which did not vary significantly by smear status, indication for Xpert testing, or level of clinician-estimated probability of TB (*see* Figures E1–E3).

The specificity of Xpert (140 of 143; 98%; 95% CI, 94–100) was considerably higher than that of clinician-guided treatment decisions (difference, +31%; 95% CI, +22 to +39; $P < 0.0001$), and correctly excluded TB in three AFB smear-positive patients with *Mycobacterium avium* complex infection. Twelve of 15 patients with positive Xpert results also had positive cultures (PPV, 80%; 95% CI, 52–96). Among 15 patients testing Xpert-positive, three had positive tests for rifampin resistance; all were confirmed by phenotypic drug-susceptibility testing.

Discordant Xpert and Culture Results

One patient had a negative Xpert result but a positive sputum culture: an HIV-infected patient with negative AFB smear microscopy results, low CD4 count, and minimally abnormal chest radiography, in whom the managing clinician estimated a high clinical probability of active TB and initiated empiric treatment. Culture confirmed the TB diagnosis 14 days later. Three patients had positive Xpert results but negative sputum mycobacterial culture results: all three had positive microscopy results. Of these three, one received initial empiric treatment with clinical and radiographic improvement consistent with culture-negative TB. Another received initial empiric treatment, but failed to improve clinically or radiographically. Given this patient's prior history of TB, the program believed the positive microscopy and Xpert results likely reflected dead bacilli and discontinued treatment. The third was not initially treated, but later acknowledged having taken 2 months of active TB treatment immediately before evaluation. He restarted treatment and received a final diagnosis of culture-negative TB based on clinical and radiographic improvement with therapy.

Clinical and Public Health Impact

Among 13 patients with TB, 12 received empiric TB treatment. Xpert results were discordant with empiric treatment decisions for two patients: one untreated with negative microscopy but positive Xpert results, and one empirically treated with negative microscopy but positive Xpert results. Thus, had Xpert been used to guide initial treatment, there would have been no net

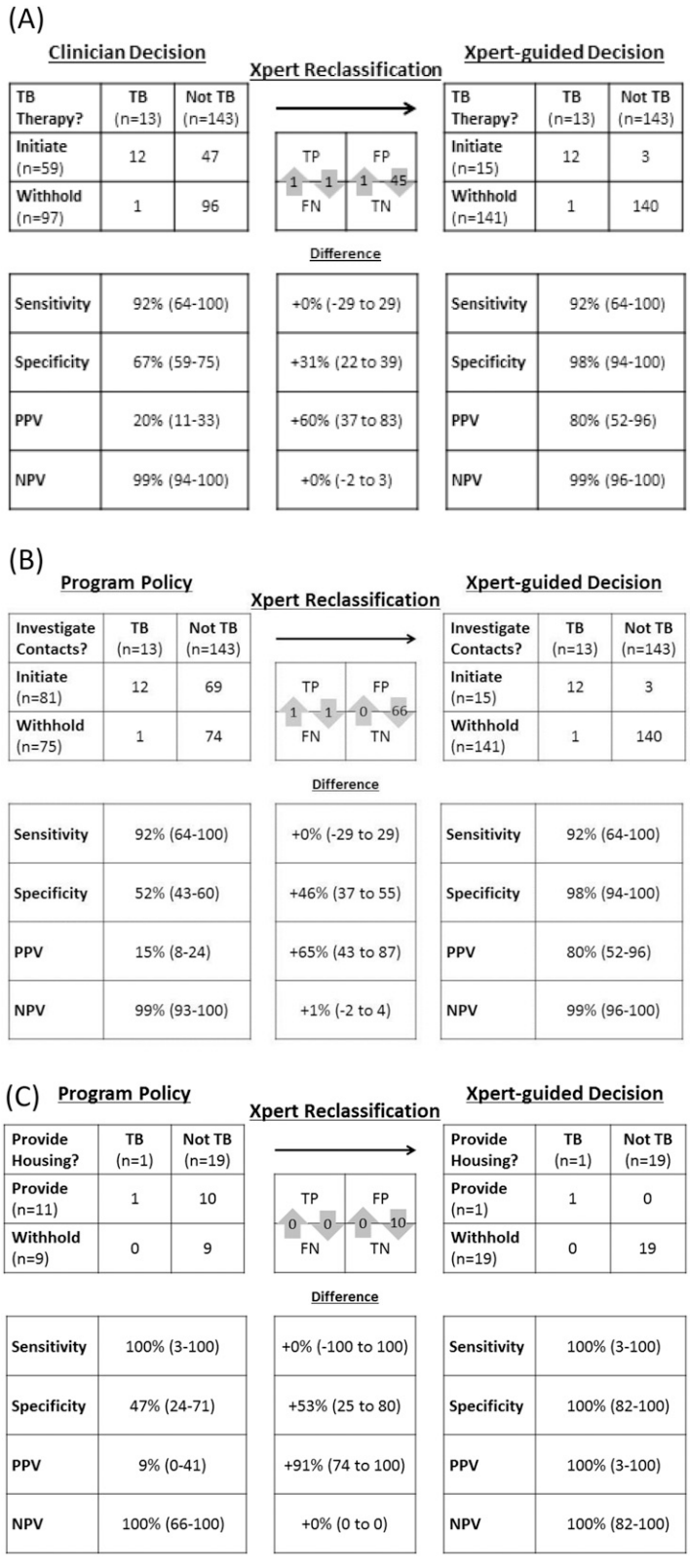

Figure 2. Comparison of diagnostic accuracy and impact of clinician- versus GeneXpert MTB/RIF (Xpert)–guided decisions on (*A*) empiric treatment, (*B*) contact investigation, and (*C*) subsidized housing. FN = false-negative; FP = false-positive; NPV = negative predictive value; PPV = positive predictive value; TB = tuberculosis; TN = true-negative; TP = true-positive.

change in the number of TB patients who received early treatment (Figure 2), and little change in median and total days of TB treatment prescribed or missed before the first positive culture (Table 2).

Among 143 patients without TB, 47 were treated empirically pending culture results. Xpert results were discordant with the empiric treatment decision for 46 patients, 45 of whom were treated but had negative Xpert results, and one of whom was not treated but had positive results. Of note, 39 (89%) of the 45 empirically treated patients with negative Xpert results received directly observed therapy. Only eight (18%) of these 45 received final clinical diagnoses of culture-negative TB. Thus, 37 (82%) of the 45 patients correctly reclassified by Xpert were overtreated for active TB. Five (11%) with latent TB infection and 16 (36%) with latent TB infection and abnormal chest radiographs (American Thoracic Society TB Category 4) received four-drug therapy for active TB when fewer drugs would have been adequate, and 16 (36%) received entirely unnecessary TB treatment (34). Using Xpert to guide initial treatment decisions would have also decreased the median duration of overtreatment from 46 days (IQR, 45–49) to 1 day (IQR, 1–3), a median difference of 44 days (IQR, 43–47). Had Xpert been used to guide treatment decisions in all patients, 44 fewer individuals would have started TB treatment, and the total number of overtreatment days during the 13-month study period would have decreased by 95%, from 2,280 (95% CI, 2,081–2,479) to 111 (95% CI, 0–256) days, eliminating 2,169 days (95% CI, 1,938–2,400) of unnecessary treatment for the TB program.

Contact Investigation

The standard indications for contact investigation, having positive AFB smear microscopy results or chest radiographic findings consistent with active TB and receiving empiric TB therapy, were present in 81 of 156 (52%) patients. The sensitivity of these criteria for culture-positive TB was 92% (95% CI, 64–100); specificity was 52% (95% CI, 43–60) (Figure 2). Although the NPV of standard criteria for contact investigation was high (99%; 95% CI, 93–100), only 12 of 81 (PPV 15%; 95% CI, 8–24) meeting these criteria actually had TB. Xpert identified one patient with AFB smear-negative, culture-positive TB with a normal chest radiograph that

Table 2. Impact of Xpert-guided Decisions on Individual and Total Annual Outcomes

Outcomes	Median (IQR) Individual Impact			Total Annual Impact (95% CI)		
	Standard Criteria	Xpert	Difference	Standard Criteria	Xpert	Difference
Treatment						
Mtb culture-positive						
Days of prediagnosis treatment	13 (10 to 15)	12 (9 to 15)	1 (1 to 3)	187 (86 to 288)	174 (68 to 280)	13 (−16 to 42)
Days of undertreatment	24 (—)	5 (—)	19 (—)	24 (—)	5 (—)	19 (—)
Mtb culture-negative						
Days of overtreatment	46 (45 to 49)	1 (1 to 3)	44 (43 to 47)	2,280 (2,081 to 2,479)	111 (0 to 56)	2,169 (1,938 to 2,400)
Contact investigation*						
Index case *Mtb* culture-positive						
Number of TB contacts investigated	2 (1 to 4)	1 (1 to 3)	—	30 (14 to 46)	23 (11 to 34)	—
Number of TB contacts not investigated	12 (—)	5 (—)	—	12 (—)	5 (—)	—
Index case *Mtb* culture-negative						
Number of non-TB contacts investigated	1 (1 to 1)	1 (1 to 7)	—	99 (79 to 119)	9 (0 to 35)	—
Subsidized housing						
Mtb culture-positive						
Nights of early housing	6 (—)	6 (—)	0 (—)	6 (—)	6 (—)	0 (—)
Nights of housing missed	0	0	0	0	0	0
Mtb culture-negative						
Nights of unnecessary housing	47 (46 to 49)	1 (1 to 4)	46 (38 to 47)	495 (387 to 603)	30 (6 to 54)	465 (348 to 582)

Definition of abbreviations: CI = confidence interval; IQR = interquartile range; *Mtb* = *Mycobacterium tuberculosis*; TB = tuberculosis; Xpert = GeneXpert MTB/RIF.

*Differences not calculated because of multiple imputation of missing values.

programmatic criteria did not select for contact investigation, but missed another patient with AFB smear-negative, culture-positive TB and an abnormal radiograph that programmatic criteria identified, thereby yielding no net change in the number of contact investigations (sensitivity difference, 0%; 95% CI, −29 to +29; $P = 1.0$). The specificity of Xpert, however, was 46% higher (95% CI, 37–55; $P < 0.0001$) than standard criteria for contact investigation. Therefore, Xpert would have hypothetically eliminated 66 unnecessary contact investigations, and reduced the overall number of contacts of non-TB patients screened as case contacts from 99 to 9 (Table 2).

Housing

Among the 20 homeless patients, 11 had indications for subsidized housing, including 10 receiving empiric TB treatment for whom housing was part of a broader package of social support, and one individual in whom the clinical probability for TB was moderate and housing was indicated to address infection control concerns. Both programmatic criteria for providing housing and Xpert correctly identified the single homeless patient with culture-positive TB (sensitivity, 100%; 95% CI, 3–100), and had perfect NPV for excluding TB (100% for programmatic criteria, 95% CI, 66–100; 100% for Xpert, 95% CI, 82–100) (Figure 2). However, specificity of the standard algorithm was poor (9 of 19; 47%; 95% CI, 24–71), and only 1 of 11 patients with a programmatic indication for housing actually had TB (PPV, 9%; 95% CI, 0–41). The specificity of Xpert (19 of 19; 100%; 95% CI, 82–100) was much higher (specificity difference, +53%; 95% CI, +25 to +80; $P = 0.002$) and PPV was 100% (95% CI, 3–100). Using Xpert instead of standard programmatic criteria would have decreased the median number of nights of unnecessary housing from 47 (IQR, 46–49) to 1 (IQR, 1–4), and the total number of nights of unnecessary housing from 495 (95% CI, 387–603) to 30 (95% CI, 6–54) (Table 2).

Discussion

Recent advances in evidence synthesis for policy-making emphasize the need for data on outcomes important to patients and public health programs (24). In this study, we have addressed this important TB research priority (37) by projecting the impact of a novel, automated TB NAAT on key clinical and programmatic management decisions before the availability of final mycobacterial culture results in patients undergoing evaluation for active pulmonary TB. In our TB program clinic in San Francisco, we found that more than 80% of the treatment, contact investigation, and housing interventions provided during the initial 8-week evaluation period went to individuals who ultimately proved not to have active pulmonary TB by either the culture reference standard or final clinician determination. Furthermore, we found that replacing standard evaluation algorithms with a single sputum Xpert test could eliminate almost all unnecessary interventions in patients without TB, without adversely impacting the timely and appropriate use of these interventions in patients with TB.

There are several reasons why overuse of empiric treatment and other early TB

interventions before the availability of mycobacterial culture results is common in San Francisco and in other public health settings (17, 18). First, although standard algorithms for treating, investigating, and housing outpatients with possible TB are inefficient, they are nonetheless highly effective at detecting and excluding culture-positive TB, as shown by their high sensitivity (≥90%) and NPV (≥99%) in this study. Second, although researchers, professional societies, and TB programs have consistently highlighted the public health consequences of missing TB patients in epidemiologic studies and practice guidelines (7–9, 31, 38, 39), few investigators have examined the impact of an incorrect TB diagnosis on patients (5, 30). A recent survey found that only 3 of 96 published systematic reviews and metaanalyses of TB diagnostic strategies had addressed questions related to clinical or public health impact (40), and similar gaps exist in the primary literature (41). Because of a lack of high-quality comparative data on the costs of undermanagement and overmanagement of TB, clinicians and public health programs consistently underestimate the adverse consequences of false-positive results (and the resulting unnecessary treatments, contact investigations, and housing orders) thereby obscuring the negative effects that standard strategies have on patients and TB programs (42). As we have shown, incorrect initial management decisions adversely impact a large proportion of individuals being evaluated and have sustained adverse consequences for patients and families, including anxiety, stigma, drug toxicities and interactions, and missed primary diagnoses. In addition, misclassification errors take a heavy toll on public health departments, leading to poor allocation of medications, laboratory tests, and staff time, and undermining patient confidence in the competence of TB programs, in an era when public health funding is increasingly tight (43–45).

We have shown that introducing Xpert could reduce the need to rely on nonspecific clinical and programmatic algorithms, and accelerate and improve most initial decisions so that they better serve patients and public health programs. In our study, both Xpert and the standard algorithm correctly excluded culture-positive TB in 99% of individuals, but because of its superior specificity, Xpert would have averted many unnecessary courses of empiric four-drug treatment for active TB, many unnecessary contact investigations, and a few nights of housing. If this strategy were implemented, a few patients with culture-positive or culture-negative TB might be initially missed and have early interventions incorrectly withheld, but such misclassifications also happen with current clinical algorithms. Moreover, we would hypothesize that because Xpert-negative patients have paucibacillary disease (46), most are unlikely to experience adverse outcomes or transmit TB in the short 2–4 weeks required for liquid cultures to turn positive (47). For the minority of patients who have features of extrapulmonary TB or risk factors for rapid disease progression, who are residing or receiving care amid a vulnerable population, or for whom the quality of sputum is uncertain, clinicians may choose to conservatively provide initial empiric treatment and/or housing regardless of the Xpert result, whereas contact investigation may await the final clinical determination. In addition, for patients who later prove to be sputum culture-negative but for whom concerning clinical features of TB persist, clinicians should consider empiric treatment of presumptive culture-negative TB at that time.

Our findings complement previous studies of the impact of TB NAATs by providing hypothetical patient- and clinic-level outcome data on the new Xpert assay in a highly relevant population. A study from a North Carolina public health program projected that an idealized NAAT with characteristics similar to Xpert could reduce treatment costs by 54%, respiratory isolation costs by 75%, and contact investigation costs by 13% (45). Operational data from Georgia, Hawaii, Maryland, and Massachusetts using the older *Mycobacterium tuberculosis* Direct (MTD; GenProbe, San Diego, CA) NAAT to evaluate patients for TB showed reductions in unnecessary empiric treatment, respiratory isolation, and contact investigation, but greater overall costs (19). Finally, a mathematical model comparing Xpert, MTD, and standard microbiologic testing for patients undergoing TB evaluation in the United States found Xpert to be cost-saving relative to the other approaches. Moreover, universal Xpert testing modestly improved quality of life at a modest incremental cost relative to testing only patients with positive sputum smears (47). Future studies should also account for the economic and social costs borne by individuals to better understand the impact of TB diagnostic strategies on patient-important outcomes.

Our study has limitations. First, patients were not selected consecutively but based on the presence of diagnostic uncertainty. Although this kind of sampling may upwardly bias measures of diagnostic accuracy, our pragmatic, "intention-to-test" enrollment strategy is consistent with CDC guidelines to refer only those for whom a result would alter clinical or public health decisions (9, 48). Second, because of the small number of TB cases, our sensitivity measures for Xpert had limited precision, although the point estimates were identical to those reported in multiple studies from other low-burden settings (49), and the NPVs were high and precise. Third, we used a concentrated pellet rather than a clinical specimen for the Xpert assays, which has been associated with higher sensitivity in some studies and lower sensitivity in others, but shown to have no effect on specificity (49). Nevertheless, in our low-burden population, the reported sensitivity differences between direct and pelleted specimens would not meaningfully alter Xpert's NPV. Fourth, we did not collect data on the effects of sputum volume, type of collection, or quality on Xpert accuracy. However, limited published data suggest that reductions in sensitivity observed with low-volume or induced sputa have only modest effects on NPV, especially in low-burden settings, and that patients missed by Xpert have paucibacillary disease often detected by sputum culture within only a few weeks (46, 50).

Finally, our estimates of the clinical and public health impact of Xpert are hypothetical. However, a clinical trial comparing clinician- and Xpert-guided decisions in low-burden settings is unlikely to be performed anytime soon, and would arguably be unnecessary and unethical given the overwhelming potential benefit in reducing unnecessary management (51). Instead, implementation studies are needed to inform the safety and acceptability of Xpert-guided management decisions in various low-burden settings, with close monitoring of populations in whom NPV is uncertain and of populations who are at high risk of transmitting TB or of progressing rapidly to more severe disease.

We did not directly examine the economic costs of the proposed strategy, although others project substantial savings (45, 52). Future analyses using these detailed data may further define the individual and public health costs and benefits, and inform use of Xpert within new TB evaluation algorithms where resources are constrained.

In summary, routine use of Xpert could potentially have substantial clinical and public health impact by reducing empiric treatment, contact investigation, and housing for many patients who do not have TB in low-burden settings. Our data provide a strong argument for using Xpert and other similar tests in most presumed pulmonary TB patients in programmatic settings, and for practice-based research using closely monitored and well-controlled research designs to evaluate their safety and benefits in patients with uncertain sputum quality, or with risk factors for rapid progression or transmission to vulnerable populations. Xpert is already revolutionizing TB diagnosis in high-burden countries, and could improve patient well-being and resource allocation in low-burden countries, enabling programs to focus on identifying and treating patients who actually have TB. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank the patients and staff of the San Francisco Department of Public Health TB Program.

References

1. Nightingale SL. From the Food and Drug Administration. *JAMA* 1996; 275:585.
2. Drobniewski FA, Caws M, Gibson A, Young D. Modern laboratory diagnosis of tuberculosis. *Lancet Infect Dis* 2003;3:141–147.
3. Greco S, Girardi E, Navarra A, Saltini C. Current evidence on diagnostic accuracy of commercially based nucleic acid amplification tests for the diagnosis of pulmonary tuberculosis. *Thorax* 2006;61:783–790.
4. Sarmiento OL, Weigle KA, Alexander J, Weber DJ, Miller WC. Assessment by meta-analysis of PCR for diagnosis of smear-negative pulmonary tuberculosis. *J Clin Microbiol* 2003;41:3233–3240.
5. Catanzaro A, Perry S, Clarridge JE, Dunbar S, Goodnight-White S, LoBue PA, Peter C, Pfyffer GE, Sierra MF, Weber R, *et al*. The role of clinical suspicion in evaluating a new diagnostic test for active tuberculosis: results of a multicenter prospective trial. *JAMA* 2000;283:639–645.
6. Taegtmeyer M, Beeching NJ, Scott J, Seddon K, Jamieson S, Squire SB, Mwandumba HC, Miller ARO, Davies PDO, Parry CM. The clinical impact of nucleic acid amplification tests on the diagnosis and management of tuberculosis in a British hospital. *Thorax* 2008;63:317–321.
7. Centers for Disease Control and Prevention (CDC). Nucleic acid amplification tests for tuberculosis. *MMWR Morb Mortal Wkly Rep* 1996;45:950–952.
8. Update: nucleic acid tests for tuberculosis. *JAMA* 2000;284:826.
9. Centers for Disease Control and Prevention (CDC). Updated guidelines for the use of nucleic acid amplification tests in the diagnosis of tuberculosis. *MMWR Morb Mortal Wkly Rep* 2009;58:7–10.
10. Joint Tuberculosis Committee of the British Thoracic Society. Control and prevention of tuberculosis in the United Kingdom: code of practice 2000. *Thorax* 2000;55:887–901.
11. Dorman SE. Editorial commentary: coming of age of nucleic acid amplification tests for the diagnosis of tuberculosis. *Clin Infect Dis* 2009;49:55–57.
12. Drobniewski FA, Watt B, Smith EG, Magee JG, Williams R, Holder J, Ostrowski J. A national audit of the laboratory diagnosis of tuberculosis and other mycobacterial diseases within the United Kingdom. *J Clin Pathol* 1999;52:334–337.
13. Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 2011. Atlanta, GA: US Department of Health and Human Services, CDC; 2012.
14. Dowdy DW, Maters A, Parrish N, Beyrer C, Dorman SE. Cost-effectiveness analysis of the gen-probe amplified *Mycobacterium tuberculosis* direct test as used routinely on smear-positive respiratory specimens. *J Clin Microbiol* 2003;41:948–953.
15. Dylewski J. Nucleic acid amplification testing for the diagnosis of tuberculosis: not for all. *Clin Infect Dis* 2009;49:1456–1457, author reply 1457.
16. Hughes R, Wonderling D, Li B, Higgins B. The cost effectiveness of nucleic acid amplification techniques for the diagnosis of tuberculosis. *Respir Med* 2012;106:300–307.
17. Guerra RL, Hooper NM, Baker JF, Alborz R, Armstrong DT, Maltas G, Kiehlbauch JA, Dorman SE. Use of the amplified *Mycobacterium tuberculosis* direct test in a public health laboratory: test performance and impact on clinical care. *Chest* 2007;132:946–951.
18. Conaty SJ, Claxton AP, Enoch DA, Hayward AC, Lipman MC, Gillespie SH. The interpretation of nucleic acid amplification tests for tuberculosis: do rapid tests change treatment decisions? *J Infect* 2005;50:187–192.
19. Marks SM, Cronin W, Venkatappa T, Maltas G, Chon S, Sharnprapai S, Gaeddert M, Tapia J, Dorman SE, Etkind S, *et al*. The health-system benefits and cost-effectiveness of using *Mycobacterium tuberculosis* direct nucleic acid amplification testing to diagnose tuberculosis disease in the United States. *Clin Infect Dis* 2013;57:532–542.
20. Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F, Allen J, Tahirli R, Blakemore R, Rustomjee R, *et al*. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med* 2010;363:1005–1015.
21. Boehme CC, Nicol MP, Nabeta P, Michael JS, Gotuzzo E, Tahirli R, Gler MT, Blakemore R, Worodria W, Gray C, *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.
22. World Health Organization. Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF system: policy statement. Geneva: World Health Organization; 2011.
23. US Food and Drug Administration. FDA permits marketing of first US test labeled for simultaneous detection of tuberculosis bacteria and resistance to the antibiotic rifampin. 2013 Jul 25 [accessed 2014 May 28]. Available from: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm362602.htm
24. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams JW Jr, Kunz R, Craig J, Montori VM, *et al*.; GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008; 336:1106–1110.
25. Small PM, Pai M. Tuberculosis diagnosis—time for a game change. *N Engl J Med* 2010;363:1070–1071.
26. Bowles EC, Freyée B, van Ingen J, Mulder B, Boeree MJ, van Soolingen D. Xpert MTB/RIF®, a novel automated polymerase chain reaction-based tool for the diagnosis of tuberculosis. *Int J Tuberc Lung Dis* 2011;15:988–989.
27. Lord SJ, Irwig L, Bossuyt PM. Using the principles of randomized controlled trial design to guide test evaluation. *Med Decis Making* 2009;29:E1–E12.
28. Bossuyt PM, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet* 2000;356: 1844–1847.
29. Dowdy DW, Gounder CR, Corbett EL, Ngwira LG, Chaisson RE, Merritt MW. The ethics of testing a test: randomized trials of the health impact of diagnostic tests for infectious diseases. *Clin Infect Dis* 2012;55:1522–1526.
30. Gordin FM, Slutkin G, Schecter G, Goodman PC, Hopewell PC. Presumptive diagnosis and treatment of pulmonary tuberculosis based on radiographic findings. *Am Rev Respir Dis* 1989;139:1090–1093.
31. National Tuberculosis Controllers Association; Centers for Disease Control and Prevention (CDC). Guidelines for the investigation of contacts of persons with infectious tuberculosis. Recommendations

from the National Tuberculosis Controllers Association and CDC. *MMWR Recomm Rep* 2005;54:1–47.
32. Kent PT, Kubica GP. Public health mycobacteriology: a guide for the level III laboratory. Atlanta, GA: Centers for Disease Control; 1985.
33. Marlowe EM, Novak-Weekley SM, Cumpio J, Sharp SE, Momeny MA, Babst A, Carlson JS, Kawamura M, Pandori M. Evaluation of the Cepheid Xpert MTB/RIF assay for direct detection of *Mycobacterium tuberculosis* complex in respiratory specimens. *J Clin Microbiol* 2011;49:1621–1623.
34. American Thoracic Society, Centers for Disease Control and Prevention, Infectious Diseases Society of America. Diagnostic standards and classification of tuberculosis in adults and children. *Am J Respir Crit Care Med* 2000;161:1376–1395.
35. Cummings P, Rivara FP. Reporting statistical information in medical journal articles. *Arch Pediatr Adolesc Med* 2003;157:321–324.
36. Davis JL, Ho C, Cattamanchi A, Grinsdale J, Metcalfe JZ, Pandori M, Banouvong H, Babst A, Carlson JS, Hopewell PC. The clinical and public health impact of automated nucleic acid testing for TB evaluation in San Francisco [abstract]. *Am J Respir Crit Care Med* 2011;183:A5314.
37. World Health Organization, Stop TB Partnership, Global Fund to Fight AIDS Tuberculosis and Malaria. Priorities in operational research to improve tuberculosis care and control. Geneva: World Health Organization; 2011.
38. Jensen PA, Lambert LA, Iademarco MF, Ridzon R; CDC. Guidelines for preventing the transmission of *Mycobacterium tuberculosis* in health-care settings, 2005. *MMWR Recomm Rep* 2005;54:1–141.
39. American Thoracic Society; CDC; Infectious Diseases Society of America. Treatment of tuberculosis. *MMWR Recomm Rep* 2003;52: 1–77.
40. Nicolau I, Ling D, Tian L, Lienhardt C, Pai M. Research questions and priorities for tuberculosis: a survey of published systematic reviews and meta-analyses. *PLoS ONE* 2012;7:e42479.
41. Brunet L, Minion J, Lienhardt C, Pai M. Mapping the landscape of tuberculosis diagnostic research. *Am J Respir Crit Care Med* 2010; 181:A2255.
42. Dowdy DW, Cattamanchi A, Steingart KR, Pai M. Is scale-up worth it? Challenges in economic analysis of diagnostic tests for tuberculosis. *PLoS Med* 2011;8:e1001063.
43. Gawande A. The cost conundrum. The New Yorker. New York; 2009.
44. Cassel CK, Guest JA. Choosing wisely: helping physicians and patients make smart decisions about their care. *JAMA* 2012;307:1801–1802.
45. Park PH, Holland DP, Wade A, Goswami ND, Bissette D, Stout JE. Public health costs for tuberculosis suspects in Wake County, North Carolina, United States. *Int J Tuberc Lung Dis* 2013;17:759–763.
46. Sohn H, Aero AD, Menzies D, Behr M, Schwartzman K, Alvarez GG, Dan A, McIntosh F, Pai M, Denkinger CM. 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.
47. Lawn SD, Kerkhoff AD, Vogt M, Ghebrekristos Y, Whitelaw A, Wood R. Characteristics and early outcomes of patients with Xpert MTB/RIF-negative pulmonary tuberculosis diagnosed during screening before antiretroviral therapy. *Clin Infect Dis* 2012;54:1071–1079.
48. Centers for Disease Control and Prevention (CDC). Availability of an assay for detecting *Mycobacterium tuberculosis*, including rifampin-resistant strains, and considerations for its use—United States, 2013. *MMWR Morb Mortal Wkly Rep* 2013;62:821–827.
49. Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev* 2014;1: CD009593.
50. Theron G, Peter J, van Zyl-Smit R, Mishra H, Streicher E, Murray S, Dawson R, Whitelaw A, Hoelscher M, Sharma S, *et al*. Evaluation of the Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in a high HIV prevalence setting. *Am J Respir Crit Care Med* 2011; 184:132–140.
51. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med* 2006;144:850–855.
52. Choi HW, Miele K, Dowdy D, Shah M. Cost-effectiveness of Xpert® MTB/RIF for diagnosing pulmonary tuberculosis in the United States. *Int J Tuberc Lung Dis* 2013;17:1328–1335.