The Role of Clinical Suspicion in Evaluating a New Diagnostic Test for Active Tuberculosis

Results of a Multicenter Prospective Trial

Antonino Catanzaro, MD
Sharon Perry, EdM, PhD
Jill E. Clarridge, PhD
Sherry Dunbar, PhD
Sheila Goodnight-White, MD
Philip A. LoBue, MD
Christopher Peter, PhD
Gaby E. Pfyffer, PhD
Marcelino F. Sierra, PhD†
Rainer Weber, MD
Gail Woods, MD
Greg Mathews
Vivian Jonas
Katie Smith, PhD
Phyllis Della-Latta, MS, PhD

Context In laboratory trials, nucleic acid amplification tests for the diagnosis of tuberculosis (TB) are more accurate than acid-fast bacilli (AFB) smear microscopy and are faster than culture. The impact of these tests on clinical diagnosis is not known.

Objective To assess the performance of a nucleic acid amplification test, the enhanced *Mycobacterium tuberculosis* Direct (E-MTD) test, against a uniform clinical standard stratified by level of clinical suspicion.

Design Prospective multicenter trial conducted between February and December 1996, documenting the clinical suspicion of TB at enrollment and using final comprehensive diagnosis as the criterion standard.

Setting Six urban medical centers and 1 public health TB clinic.

Patients A total of 338 patients with symptoms and signs consistent with active pulmonary TB and complete clinical diagnosis were stratified by the clinical investigators to be at low (≤25%), intermediate (26%-75%), or high (>75%) relative risk of having TB.

Main Outcome Measures Sensitivity, specificity, and positive and negative predictive values of the E-MTD test in clinical suspicion of groups with low (n = 224); intermediate (n = 68); and high (n = 46) clinical suspicion of TB.

Results Based on comprehensive clinical diagnosis, sensitivity of the E-MTD test was 83%, 75%, and 87% for low, intermediate, and high clinical suspicion of TB, respectively, and corresponding specificity was 97%, 100%, and 100% (P = .25). Positive predictive value of the E-MTD test was 59% (low), 100% (intermediate), and 100% (high) compared with 36% (low), 30% (intermediate), and 94% (high) for AFB smear. Corresponding negative predictive values were 99%, 91%, and 91% (E-MTD test) vs 96%, 71%, and 37% (AFB smear).

Conclusions For complex diagnostic problems like TB, clinical risk assessments can provide important information regarding predictive values more likely to be experienced in clinical practice. For this series, a clincial suspicion of TB was helpful in targeting areas of the clinical spectrum in which nucleic acid amplification tests can make an important contribution.

JAMA. 2000;283:639-645 www.jama.com

ROBERT KOCH'S DISCOVERY IN 1882 that tuberculosis (TB) is caused by *Mycobacterium tuberculosis* was a watershed in the history of efforts to understand and control this deadly disease.[1] Since then, diagnosis of TB has focused on the detection of *M tuberculosis* using basic techniques of acid-fast bacilli (AFB) smear microscopy and culture. Detection of AFB in a smear requires more than 10 000 organisms/mL, and the test does not distinguish among mycobacteria. Even with modern radiometric detection systems, identification of *M tuberculosis* can require 2 to 5 weeks.[2-4]

Before its pathogenesis was understood, TB was recognized as a distinct clinical entity. These characterizations

Author Affiliations: University of California, San Diego Medical Center, San Diego (Drs Catanzaro, Perry, and LoBue); VA Medical Center, Houston, Tex (Drs Clarridge, Dunbar, and Goodnight-White); San Diego County Public Health Laboratory, San Diego (Dr Peter); Swiss National Center for Mycobacteria, Department of Medical Microbiology, University of Zurich, Zurich, Switzerland (Dr Pfyffer); State University of New York, Health Science Center at Brooklyn, Brooklyn, NY (Dr Sierra); Division of Infectious Diseases, University Hospital, Zurich, Switzerland (Dr Weber); Laboratory Medicine, Clinical Microbiology, University of Texas Medical Branch at Galveston (Dr Woods); Gen-Probe Inc, San Diego, Calif (Mr Mathews, Ms Jonas, and Dr Smith); and Pathology Department, Columbia-Presbyterian Medical Center, New York, NY (Dr Della-Latta). Dr Dunbar is now with Luminex, Austin, Tex; Dr LoBue is now with the Centers for Disease Control and Prevention, stationed at the San Diego County Department of Health TB Control Office.
†Deceased.
Financial Disclosure: Dr Catanzaro is a medical consultant for Gen-Probe Inc, San Diego, Calif.
Corresponding Author: Antonino Catanzaro, MD, UCSD Medical Center, 200 W Arbor Dr #8374, San Diego, CA 92103-8374 (e-mail: acatanzaro@ucsd.edu).

remain an integral part of the medical assessment. Presently, the most important criteria for establishing a presumptive diagnosis are AFB smear and a case definition, which may be based on radiographic signs, physiologic symptoms, risk factors, or a combination of these.[5] In many Western countries, the declining incidence of TB, combined with the human immunodeficiency virus (HIV) epidemic, has increased the number of mycobacteria other than tuberculosis cases, further impairing the reliability of the AFB smear to specifically predict TB.[6-8] In acute care settings, as many as 8 to 10 patients are suspected to have TB for every confirmed case.[9,10] Accurate laboratory tests that provide results in a clinically useful time frame have the potential to affect the performance of TB programs broadly, offering opportunities for more effective patient management and more efficient allocation of scarce resources.

Nucleic acid amplification (NAA) tests[11] represent a major advance in the diagnosis of TB. With the use of amplification systems, nucleic acid sequences unique to *M tuberculosis* can be detected directly in clinical specimens, offering better accuracy than AFB smear and greater speed than culture.[12-14] Early studies used in-house procedures based on polymerase chain reaction with amplification of the *IS*6110 gene sequence.[15,16] However, kits providing standard formats and reagents are now available, making the amplification technologies more practical for use in clinical laboratories.[17-24] Currently, 2 commercial kits—the *Mycobacterium tuberculosis* Direct (MTD) test (Gen-Probe Inc, San Diego, Calif) and the AMPLICOR MTB (Roche Molecular Systems, Branchburg, NJ)—have been approved by the Food and Drug Administration for use in respiratory specimens of previously untreated patients with positive AFB smear results.

The majority of studies, to date, have been based on laboratory criteria for diagnosis of disease, with clinical records used to evaluate discrepant results. Only a few studies have examined the performance of the NAA tests against clinical definitions of TB[25,26] or test performance under routine testing conditions.[27] The American Thoracic Society has stressed the need for additional studies evaluating these and other emerging diagnostics against clinical reference standards.[28] This report describes the performance of the Enhanced MTD (E-MTD) in a multicenter prospective trial that documented the physician's suspicion of TB at the time of the initial examination and a clinical and laboratory diagnosis of TB. The MTD, a 3-hour assay using transcription-mediated amplification for *M tuberculosis* complex-specific recombinant RNA,[29] was first approved by the Food and Drug Administration in 1995. The enhanced version, which accepts a larger volume of sample and has a shorter processing time than the initial kit, was approved in 1999 for use in smear negative and positive respiratory samples. The objectives of the present work were to provide a uniform clinical standard for diagnosis in the population studied, and to describe the E-MTD's performance against this standard for different levels of clinical suspicion.

METHODS

Clinical Study Design

Between February and December 1996, 425 individuals suspected of having active pulmonary TB were enrolled from 7 inpatient and specialized outpatient facilities. Two sites were located in San Diego, Calif (a university hospital and a public health TB clinic); 1 in Houston, Tex (a Veterans Affairs Hospital); 1 in Galveston, Tex (a university facility); 2 in New York, NY; and 1 center in Zurich, Switzerland. All patients were identified for enrollment by hospital or clinic site physicians based on suspicion of active pulmonary TB, including but not limited to symptoms, risk factors, tuberculin skin test reaction, and chest radiograph findings. Enrolling physicians were pulmonary or infectious disease specialists with experience in the evaluation of patients for TB. Patients were not eligible for enrollment if they received multidrug treatment for TB for more than 7 days during the 3 months prior to enrollment. Diagnosis and treatment decisions were made by site physicians according to local standards of care. During the trial, physicians were blinded to results of the NAA. All other laboratory results were available to physicians in accordance with routine laboratory procedures.

Clinical Methods

Clinical information, including patient demographic information, medical history, physical examination, and chest radiograph results, was collected by clinical research staff at each site either directly from the patient, or from the patient's chart. Completion of 3 standardized case reporting forms was required for inclusion in the final study population. These forms were completed at the first physical examination (initial enrollment form), at the end of specimen collection or at discharge (first follow-up form), and at 3 months follow-up (end-of-study form).

Clinical Suspicion of TB

The initial enrollment form requested the site physician to estimate the probability that the patient had TB by using a range from 0% to 100%. No clinical guidelines were provided to physicians for determining this estimate. The clinical suspicion of TB (CSTB) was based on the physician's clinical judgment. Laboratory results available to the physicians for this assessment may have included some AFB smears, and at some sites, the earlier version of the MTD may have been available. However, the results of the E-MTD were withheld from the chart and from the clinician.

Clinical Diagnosis of TB

To provide a uniform definition of TB across sites, criteria intended to represent a conservative consensus standard for ruling in or out pulmonary TB were established by an independent panel consisting of 3 experts in TB diagnosis and treatment. Under these standards, the combination of high clinical suspicion (>80%) and at least 2 positive cultures for *M tuberculosis* from separate specimens was considered definitive evidence of active TB. In the absence of these con-

ditions, the cases were reviewed by the independent expert panel and at least 2 of the 3 panel members had to consider the patient to have TB for a positive ruling. The combination of low clinical suspicion (<10%) and respiratory specimens consistently negative for *M tuberculosis* was considered to constitute definitive evidence for absence of active TB. In the absence of these conditions, at least 2 of the 3 panel members had to consider the patient to be free of TB for a negative ruling.

Independent Expert Panel Review

Without knowing the reason for referral, the panel reviewed all cases not meeting criteria defined above. Possible actions of the panel were to exclude a case from analysis due to insufficient information or to provide a consensus clinical diagnosis by at least 2 of 3 panel members. Information reviewed by the panel consisted of all case report forms, an assessment by the laboratory director as to whether contamination occurred in any single positive culture, and copies of initial and (when available) follow-up chest radiographs. Copies of case report forms excluded the identity of the patient, the site of enrollment, the CSTB, and the results of the NAA. As a quality-control measure to assess interobserver agreement, 6 control cases were drawn randomly from cases considered to meet the conservative case definitions. Panel determinations in all 6 cases were unanimous and consistent with the site physician's diagnosis as reported on the end-of-study form.

Laboratory Methods

Briefly, 1 to 6 sequential respiratory samples, primarily expectorated or induced sputa, were collected over 7 days following patient enrollment. For this study, only the first sample of each specimen type was accepted per day as a study sample. Specimens were processed within 3 days of collection, stored at 2°C to –8°C, and an aliquot of resuspended sediment was frozen for inhibition testing. Culture (Lowenstein-Jensen and BACTEC 460, BD Biosciences Division, Sparks, Md; Middlebrook 7H10/7H11), AFB smear (auramine O stain), and NAA tests were performed on each specimen. Experienced clinical laboratory technologists performed the NAA assays according to manufacturer's instructions.

Analysis

Data were pooled across all sites for analysis. Final comprehensive diagnosis as determined by the panel review procedure was defined as the criterion standard for computing sensitivity, specificity, and predictive values of culture, E-MTD, and AFB smear. Laboratory results were modeled at the patient level by defining a positive test result as the occurrence of at least 1 positive test in a series of up to 6 specimens per patient.

Clinical Suspicion. Physicians' clinical suspicion estimates were grouped into 3 relative risk categories: low (≤25%); intermediate (26%-75%); and high (>75%) probability of TB. For each category, we report the number and proportion of patients, the prevalence of prognostic symptoms and signs, and the observed prevalence of disease based on the panel's clinical diagnosis at the end of the study. The association of patient characteristics with risk group classification was assessed in polychotomous logistic regression using high suspicion as the reference category. Results of the univariate analysis are reported.

Performance of AFB Smear and E-MTD by Level of Clinical Suspicion. Sensitivity and specificity of E-MTD and AFB smear are reported for low, intermediate, and high categories of CSTB. Variations in these test properties with respect to clinical suspicion level were assessed in dichotomous logistic regression predicting the probability of a positive test result given final clinical diagnosis and clinical risk classification. Statistical evidence of variation by CSTB is reported as the χ^2 statistic for a linear model (low, intermediate, high) of clinical suspicion. This analysis was conducted separately for AFB smear and E-MTD, and each model was validated by the Hosmer-Lemeshow test.

Estimating Clinical Utility. Positive predictive values (PPVs) and negative predictive values (NPVs) of E-MTD and AFB smear are reported for low, intermediate, and high categories using the observed prevalence of TB in each group as a proxy for the prior (pretest) risk of disease. Because this was a partially blinded observational study, the objective of this analysis is descriptive, not prescriptive.

RESULTS

Initial Study Population

A total of 425 patients were enrolled at 6 sites in the United States and 1 European site. Of these, 341 (80%) had complete clinical and laboratory data. Eighty-four patients (20%) were excluded for the following reasons: 46 did not have at least 1 valid specimen, 8 received more than 7 days of treatment for TB within the previous 3 months, and 30 had an incomplete set of clinical data forms. TABLE 1 summarizes characteristics of the 84 excluded patients and the 341 patients eligible for analysis. Among the 7 trial sites, the number of evaluable patients ranged from 13 (4%) in Europe to 122 (36%) at a university hospital in San Diego. Site exclusions averaged 23%, ranging from 11% (7 patients) to 43% (10 patients) of a site's enrollments.

Final Study Population

Of the 341 patients with complete study forms and valid laboratory results, 303 (89%) had an end-of-study report that satisfied the case definitions. Thirty-eight cases (11%) were referred to the panel for further review, including 15 patients with a 10% or greater probability of TB and less than 2 positive cultures, and 23 patients with an 80% or less probability of TB but at least 1 positive smear and/or culture. Three of these cases were excluded by the panel because the patient record provided insufficient information on which to base a final diagnosis. Of the remaining 35 cases, 11 were classified as TB and 24 were classified as not TB. These determinations were unanimous in 22 cases (58%; 7 TB, 15 not TB) and rendered by two-thirds consensus in 13 cases (4 TB, 9 not TB). Following panel review, there were a total of 338 patients for analysis, including 72 (21%) considered to have active TB and 266 considered to be free of TB.

Diagnosis by Culture

Among the 72 patients diagnosed as having active TB, 45 (63%) had 2 or more cultures positive for *M tuberculosis* in a series of 6, 20 (28%) had 1 positive culture, and all cultures were negative for *M tuberculosis* in 7 cases (10%). One to 6 specimens collected on different consecutive days were accepted for the study. The average number of specimens cultured per patient was 2.5 (2.6 with TB and 2.5 without TB). When the culture result was compared with final comprehensive diagnosis, the sensitivity of a series yielding only 1 positive culture was 90%, but was 63% when 2 or more positive cultures were required to meet the case definition. The number of specimens with *M tuberculosis* isolated was established by the site physician end-of-study report and panel consensus. One patient determined to be free of TB had a positive culture, resulting in a specificity of 99.6%.

Table 1. Characteristics of Patients Eligible and Ineligible for Analysis*

Risk Factor	Final Study Population (n = 341)†	Excluded (n = 84)‡
Male	77	76
Homeless	13	13
HIV/AIDS	31	31
Foreign-born§	32	40
Hispanic	19	17
Black	29	35
Asian	15	19
Corrections institution	23	21
Nursing home resident	2	2
Halfway house	9	7
Alcohol dependence	31	32
Drug dependence	13	17
Contact of tuberculosis case	19	18
Positive skin test	35	32
Unemployed	70	76
Average age, mean (SD), y	46 (16)‡	46 (15)‡

*Values are expressed as percentages unless otherwise indicated. HIV indicates human immunodeficiency virus; AIDS, acquired immunodeficiency syndrome.
†Only 338 had a complete diagnosis.
‡Reasons for exclusion: no valid specimens (46 patients); more than 7 days of treatment within previous 3 mo (8 patients); incomplete case report forms (30 patients).
§Includes 15 European enrollees described as foreign-born.

Analysis of CSTB

The average CSTB for patients with a final clinical diagnosis of TB was 69% (median, 80%). For patients with a final clinical diagnosis other than TB, the average CSTB was 20% (median, 15%). Two hundred twenty-four patients (66%) had low CSTB, 46 patients (14%) had high CSTB, and 68 patients (20%) had intermediate CSTB. Based on final clinical diagnosis, prevalence of TB was 5% (low CSTB), 29% (intermediate CSTB), and 87% (high CSTB) ($\chi^2_2 = 83$; $P<.001$). A total of 103 patients (30%) were prescribed on anti-TB medication presumptively, including 11% of those in the low CSTB group, 49% of those in the intermediate CSTB group, and 98% of those in the high CSTB group ($\chi^2_2 = 64$; $P<.001$).

Patient characteristics most significantly associated with CSTB group are summarized in TABLE 2. The most consistent predictor of CSTB level was a chest radiograph suggestive of current disease. The proportion of patients with a suggestive chest radiograph increased steadily from 25% in the low CSTB group, to 57% in the intermediate CSTB group, to 87% in the high CSTB group. The presence of at least 2 major symptoms predicted classification at an intermediate or high CSTB, with the most important symptoms being cough lasting more than 2 weeks and recent weight loss. Thirty-four percent of patients had a positive tuberculin skin test, and 19% were known contacts of a TB case. Having at least 1 of these latter characteristics was moderately associated with higher CSTB.

Performance of AFB Smear and E-MTD

Overall sensitivity and specificity of the E-MTD test were 83% (95% confidence interval [CI], 71%-93%) and 97% (95% CI, 95%-99%), respectively. By level of clinical suspicion, sensitivity was 83% (low CSTB), 75% (intermediate CSTB), and 87% (high CSTB). Specificity ranged from 97% at low CSTB to 100% at intermediate and high CSTB. This variation was not statistically significant ($\chi^2_2 = 2.77$; $P = .25$; FIGURE 1). By reference, performance of AFB smear varied significantly by CSTB group ($\chi^2_2 = 18.2$; $P<.001$). Sensitivity of the AFB smear was not statistically different for low and intermediate CSTB (42% and 25%, respectively, $P = .33$), but was significantly lower than sensitivity at high CSTB (83%; high vs low, $P = .008$). Specificity of the AFB varied inversely to sensitivity and also significantly by level of clinical suspicion, ranging from 96% at low CSTB, 77% at intermediate CSTB, to 67% at high CSTB (intermediate vs low, $P<.001$; high vs low, $P = .009$). The numbers for calculating sensitivity and specificity are provided in TABLE 3.

Table 2. Patient Symptoms and Medical History Associated With Physicians' Clinical Suspicion of Tuberculosis (CSTB)*

Symptoms and Medical History	Proportion of All Patients	CSTB Group		
		Low (n = 224)	Intermediate (n = 68)	High (n = 46)
Suggestive chest radiograph†	40	25‡	57‡	87‡
Weight loss	44	39	46	67§
Cough	75	69‡	87	87
Chest pain	33	38‡	21	22
Positive skin test	34	34	28	41
Tuberculosis exposure or positive skin test‖	44	44	34§	57
Immune suppression¶	34	36	43	11‡
Foreign-born	31	33	18§	46
Homeless	13	14	13	9
Institutional domicile#	28	32	17	22

*Values are expressed as percentages.
†Intake chest radiograph was suggestive of classic tuberculosis (upper-lobe fibrocavitary) disease.
‡Significant positive or negative predictor ($P<.05$) or CSTB vs 1 or more other categories in polychotomous logistic regression.
§Borderline significance of $P<.10$.
‖Recent contact with tuberculosis case (19%) or positive skin test (34%).
¶Includes patients positive for the human immunodeficiency virus (91%) and those patients with a history of steroid use or cancer chemotherapy (9%).
#Prison inmate or resident of nursing home or halfway house.

Estimating Clinical Utility

In evaluating a diagnostic test, documentation of a prior or pretest risk is important for modeling PPVs and NPVs likely to operate in the clinical setting. If the prevalence of TB in each CSTB group represented a pretest risk, the predictive values shown in FIGURE 2 would be estimated. Based on final diagnosis, overall PPVs and NPVs for the AFB smear were 67% (95% CI, 56%-78%) and 90% (95% CI, 86%-94%), respectively. Corresponding values for the E-MTD were 88% (95% CI, 80%-96%) and 95% (95% CI, 93%-98%). For the low CSTB group (5% prior risk), both tests appeared useful for ruling out disease with NPVs of 96% (AFB smear) to 99% (E-MTD). While neither test provided convincing evidence for ruling in disease at this risk level, the E-MTD was potentially more useful with a PPV of 59% compared with 36% for the AFB smear. Conversely, both tests appeared to be useful for ruling in disease for the high CSTB group (87% prior risk), with PPVs of 94% for the AFB smear and 100% for the E-MTD. However, the expected NPV of the AFB smear was only 37%, compared with 91% for the E-MTD. The numbers for calculating PPV and NPV are provided in Table 3.

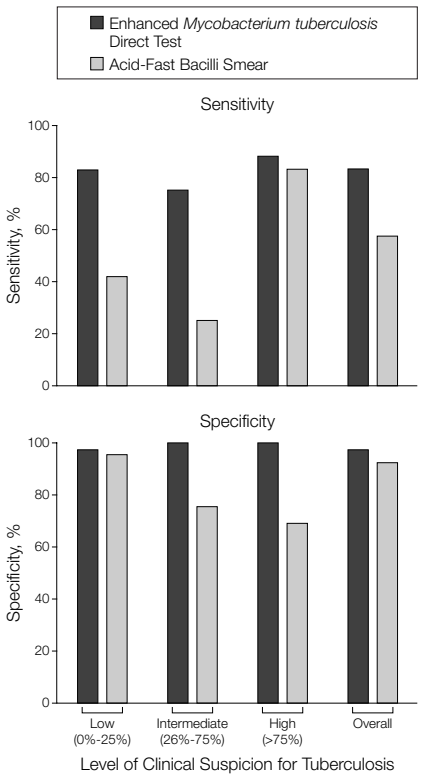
The E-MTD appeared to offer greatest utility overall in the intermediate CSTB (29% prior risk), demonstrating a PPV of 100% (vs 30% for AFB smear) and an NPV of 91% (vs 71% for AFB smear). This clinically complex group included 20 TB cases (29%) and 40 cases (18%) free of active TB (TABLE 4). Mycobacteria other than *M tuberculosis* were cultured more commonly from patients in the intermediate CSTB group than the low or high CSTB groups (22/48 [46%] vs 18% for low, 9% for high). Cases of HIV infection were somewhat more frequent (40% vs 32% for low, 11% for high). The TB and non-TB cases with CSTB estimates in the intermediate range were equally likely to have a positive AFB smear (25% TB, 23% not TB). The prevalence of the constellation of suggestive chest radiograph, cough, and weight loss was also similar (20% vs 23%). Patients with TB in this CSTB group were relatively more likely to have risk factors, such as contact exposure or positive tuberculin skin test, than their counterparts at high or low CSTB.

COMMENT

The majority of studies assessing performance of the NAA tests have used laboratory performance criteria. While the tests have performed well under these conditions, appropriate clinical uses have been more difficult to establish because clinical and laboratory definitions of disease may differ. Whereas laboratory performance is based on culture growth and is usually presented at the specimen level of analysis, clinical diagnosis is based on multiple indications including clinical signs and symptoms and response to therapy and laboratory results. Laboratory and clinical case definitions can measure clinical utility differently as shown by Bradley et al,[25] who compared the MTD against the classification system of the American Thoracic Society, and Chin et al,[26] who used empiric case definitions with the Roche AMPLICOR. Because the clinical risk assessment is more likely to reflect physician decision making, the American Thoracic Society has recommended that the NAA tests be evaluated at different levels of clinical suspicion.[30]

Several aspects of this multicenter trial advance these efforts. Patients were se-

Figure 1. Variation in Sensitivity and Specificity of Acid-Fast Bacilli Smear and Enhanced *Mycobacterium tuberculosis* Direct Test by Clinical Suspicion for Tuberculosis



Sensitivity was defined as the percentage of patients with tuberculosis with positive test results; specificity, percentage of patients without tuberculosis with negative test results.

Table 3. Numbers Used for Calculating Sensitivity, Specificity, Positive and Negative Predictive Values by Level of Clinical Suspicion for Tuberculosis (TB)*

	Clinical Suspicion Level							
	Low (n = 224)		Intermediate (n = 68)		High (n = 46)		Total (N = 338)	
	TB Present	TB Absent	TB Present	TB Absent	TB Present	TB Absent	TB Present	TB Absent
Acid-fast bacilli smear†	5	9	5	11	33	2	43	22
Enhanced *Mycobacterium tuberculosis* Direct test†	10	7	15	0	35	0	60	7
Disease frequency (%)‡	12 (5)	NA	20 (29)	NA	40 (87)	NA	72 (21)	NA

*NA indicates not applicable.
†No. of positive results; defined as at least 1 positive test in a series of 1 to 6 per patient (median, 2.5).
‡Based on comprehensive clinical diagnosis as determined by expert panel review at end-of-study (3-mo follow-up) report.

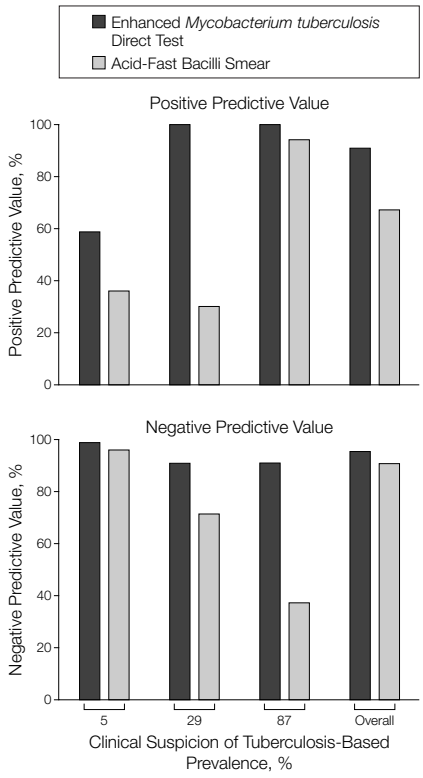
lected for evaluation by clinicians based on usual workup indications in symptomatic patients. Laboratory and clinical evidence was documented prospectively, and clinical impressions of risk were documented both quantitatively and qualitatively. A final comprehensive diagnosis as determined by an independent panel of experts was used as the diagnostic reference standard. This imposed a uniform standard of comprehensive diagnosis for analysis and was an important component of this multicenter trial. Based on this standard, a single positive culture had a sensitivity of approximately 90%. However, 38% of clinically determined TB cases had at least 1 negative culture and 10% of all cases were referred to the panel. Use of the clinical case definition solely to resolve test discrepancies with culture, as is commonly done in laboratory trials, may not reveal the discrepancy between laboratory and clinical reference standards.

In this trial, we asked enrolling physicians to quantify their degree of suspicion (CSTB) using a scale from 0% to 100%. Regression analysis suggests that these estimates were consonant with predictive signs and symptoms observed with other TB scoring systems,[10,31-35] including suggestive chest radiograph, cough, and recent weight loss. As with factor-based scoring systems, the meaning and distribution of these clinical suspicion estimates necessarily reflects the population under study, both patients and physicians, and these values are best understood as estimates of relative, not absolute, risk. Factors likely to affect physicians' estimates of relative risk include the customary prevalence of disease in the practice setting, the clinical spectrum of disease, the specialty or experience of the physician, and the quality of the medical history.[36] Further research is needed to appreciate the interior characteristics of summary risk assessments and to validate these measures in different practice settings.

Use of a CSTB is potentially informative for understanding the clinical context in which laboratory results are used. Predictive values based on clinical risk assessments like the CSTB are more likely to reflect pretest probabilities operating in the clinical setting, and hence to reflect conditions under which specific test attributes are needed. Although the average prevalence of disease in this series was 21%, prevalence ranged from 5% in the low to 87% in the high CSTB group, and 29% in the intermediate group. Characterizing the performance of the E-MTD for low, intermediate, and high CSTB groups helped to characterize the usefulness of the test in this population. When stratified for these risk levels, the sensitivity and specificity of the E-MTD were higher than AFB smear and stable, while sensitivity and specificity of the AFB smear varied considerably. For this patient series, the E-MTD appeared to offer an improvement in PPV (100%) for patients with CSTB estimates in the intermediate- or high-risk range. The high and consistent specificity of the E-MTD also appeared to be clinically valuable in excluding disease among patients with intermediate or high CSTB estimates, offering an NPV of 91%, compared with 71% and 37% for AFB smear in intermediate and high CSTB.

In this series, the CSTB was useful to document the potential for clinical spectrum bias[37-39] in the performance of the AFB smear. Approximately 40% of individuals with an intermediate CSTB were HIV-positive and nearly one third were ultimately diagnosed as having mycobacteria other than tuberculosis infections. Conventional diagnostic signs (AFB smear, chest radiograph, advanced symptoms) appeared to offer limited distinction between TB and non-TB groups in this suspicion range. Although sample size was limited for this analysis, the ex-

Figure 2. Estimated Predictive Value of Acid-Fast Bacilli Smear and Enhanced *Mycobacterium tuberculosis* Direct Test by Clinical Suspicion for Tuberculosis



Positive predictive value is percentage of patients testing positive who had tuberculosis; negative predictive value, percentage of patients testing negative who did not have tuberculosis. Stratum-specific predictive values are computed using prior risk as determined by clinical suspicion for tuberculosis group. The percentage of patients in the clinical suspicion for tuberculosis category who had tuberculosis by final clinical diagnosis are as follows: low, 5% (12/224); intermediate, 29% (20/68); and high, 87% (40/46).

Table 4. Characteristics of Patients Classified at an Intermediate Range of Physicians' Suspicion (CSTB)*

Characteristic	No. (%) of Patients		
	All	TB Present	TB Absent
Suggestive chest radiograph	39 (57)	10 (50)	29 (60)
Positive AFB smear	16 (24)	5 (25)	11 (23)
HIV positive or AIDS	27 (40)	7 (35)	20 (42)
Positive skin test	19 (28)	10 (50)	9 (19)
Contact of TB case	9 (13)	5 (25)	4 (8)
MOTT	22 (32)	NA	22 (46)
Proportion of all patients (n = 338)	68 (20)	20 (28)	48 (18)

*TB indicates tuberculosis; CSTB, clinical suspicion for TB; AFB, acid-fast bacilli; HIV, human immunodeficiency virus; AIDS, acquired immunodeficiency syndrome; MOTT, mycobacterium other than tuberculosis; and NA, not applicable.

istence of a broad and rather complex intermediate risk group (and the poor performance of the AFB smear in this range) is consistent with clinical observation. Under pressures of the HIV epidemic and new immigration patterns, clinical spectrum of disease has changed,[8,40] altering also the relative utility of many conventional diagnostic tools.[30] This has important implications for use of this test as a selection criterion in clinical and cost-utility studies.

The present observational study had design limitations that preclude it as a basis for setting clinical benchmark standards. The most important of these was a partial blind (to the NAA test only), which permitted clinicians access to some initial AFB smear results during formulation of clinical suspicion estimates. Many patients in inpatient settings were enrolled from an isolation ward, criteria for which can have included, but were not limited to, prior workup by AFB smear. Spectrum and prevalence of disease also varied by trial site, and sample sizes precluded site adjustments to test performance. With these caveats, important strengths of the present study include use of a uniform clinical diagnosis, a representative patient set, laboratory performance that is consistent with that observed in previous laboratory trials involving the NAA tests, and the clinical plausibility of findings in the experience of pulmonary physicians. More rigorously controlled, fully blinded, head-to-head studies with careful serial documentation of pretest risk are under way.

For this patient series, the CSTB helped to characterize important limitations of a standard reference test, AFB smear, and the contribution a new technology, NAA, could make. Although the study design was observational, we believe it has laid important groundwork for continuing assessment of emerging diagnostics for TB. Interdisciplinary study designs, demonstrating the performance of laboratory tests in conjunction with clinical risk assessments, are needed.

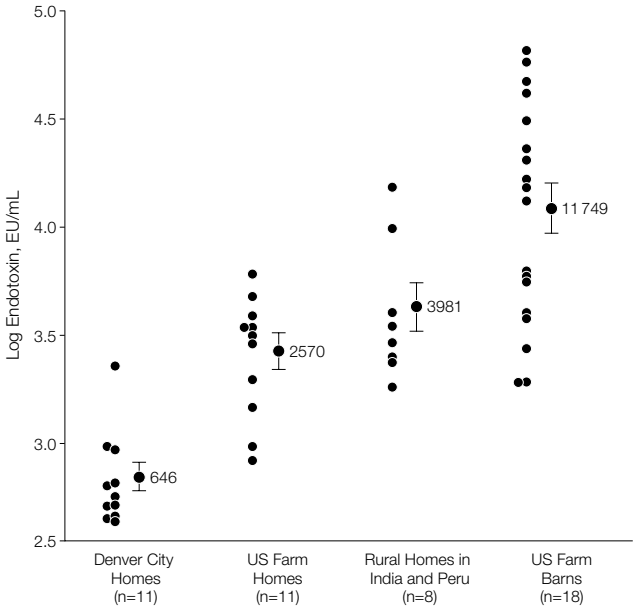
Funding/Support: Funding for this clinical trial was provided by Gen-Probe Inc.

Acknowledgment: We would like to thank the members of the expert panel, Joseph Bates, MD, John McClellan Veterans Affairs Hospital, Little Rock, Ark, Philip C. Hopewell, MD, University of California, San Francisco, and Michael Iseman, MD, National Jewish Center for Immunology and Respiratory Medicine, Denver, Colo. We also would like to acknowledge the contributions of Kathleen S. Moser, MD, MPH, director of Tuberculosis Control and Border Health, County of San Diego Health and Human Services Agency, San Diego, Calif, and Kathy Miskovish, MPH, analytical writer, University of California, San Diego.

REFERENCES

1. Koch R. Die Aetiologie der Tuberculose. *Berl Klin Wochenschr.* 1882;19:221-230.
2. Pfyffer GE, Cieslak C, Welscher HM, Kissling P, Rüsch-Gerdes S. Rapid detection of mycobacteria in clinical specimens by using the automated BACTEC 9000 MB system and comparison with radiometric and solid-culture systems. *J Clin Microbiol.* 1997;35:2229-2234.
3. Rohner P, Ninet B, Metral C, Emler S, Auckenthaler R. Evaluation of the MB/BacT system and comparison to the BACTEC 460 system and solid media for isolation of mycobacteria from clinical specimens. *J Clin Microbiol.* 1997;35:3127-3131.
4. Tortoli E, Cichero P, Chirillo MG, et al. Multicenter comparison of ESP Culture System II with BACTEC 460TB and with Lowenstein-Jensen medium for recovery of mycobacteria from different clinical specimens, including blood. *J Clin Microbiol.* 1998;36:1378-1381.
5. 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.
6. Boyd JC, Marr JJ. Decreasing reliability of acid-fast smear techniques for detection of tuberculosis. *Ann Intern Med.* 1975;82:489-492.
7. Wright PW, Wallace RJ Jr, Wright NW, et al. Sensitivity of fluorochrome microscopy for detection of *Mycobacterium tuberculosis* versus nontuberculous mycobacteria. *J Clin Microbiol.* 1998;36:1046-1049.
8. Ankobiah W, Hin M, Singh A, Sargeant C. Changing trends of mycobacteria other than tuberculosis. *Chest.* 1996;110:229S.
9. Blumberg HM, Watkins DL, Berschling JD, et al. Preventing the nosocomial transmission of tuberculosis. *Ann Intern Med.* 1995;122:658-663.
10. Bock NN, McGowan JE Jr, Ahn J, et al. Clinical predictors of tuberculosis as a guide for a respiratory isolation policy. *Am J Respir Crit Care Med.* 1996; 154:1468-1472.
11. Centers for Disease Control and Prevention. Nucleic acid amplification tests for tuberculosis. *MMWR Morb Mortal Wkly Rep.* 1996;45:950-952.
12. Eisenach KD, Sifford MD, Cave MD, Bates JH, Crawford JT. Detection of *Mycobacterium tuberculosis* in sputum samples using a polymerase chain reaction. *Am Rev Respir Dis.* 1991;144:1160-1163.
13. Crawford JT. New technologies in the diagnosis of tuberculosis. *Semin Respir Infect.* 1994;9:62-70.
14. Jonas V, Longiaru M. Detection of *Mycobacterium tuberculosis* by molecular methods. *Clin Lab Med.* 1997;17:119-128.
15. Thierry D, Cave MD, Eisenach KD, et al. IS6110, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucleic Acids Res.* 1990;18:188.
16. Clarridge JED, Shawar RM, Shinnick TM, Plikaytis BB. Large-scale use of polymerase chain reaction for detection of *Mycobacterium tuberculosis* in a routine mycobacteriology laboratory. *J Clin Microbiol.* 1993;31:2049-2056.
17. Jonas V, Alden MJ, Curry JI, et al. Detection and identification of *Mycobacterium tuberculosis* directly from sputum sediments by amplification of rRNA. *J Clin Microbiol.* 1993;31:2410-2416.
18. Pfyffer GE, Kissling P, Wirth R, Weber R. Direct detection of *Mycobacterium tuberculosis* complex in respiratory specimens by a target-amplified test system. *J Clin Microbiol.* 1994;32:918-923.
19. Bergmann JS, Woods GL. Clinical evaluation of the Roche AMPLICOR PCR *Mycobacterium tuberculosis* test for detection of *M tuberculosis* in respiratory specimens. *J Clin Microbiol.* 1996;34:1083-1085.
20. Pfyffer GE, Kissling P, Jahn EM, Welscher HM, Salfinger M, Weber R. Diagnostic performance of amplified *Mycobacterium tuberculosis* direct test with cerebrospinal fluid, other nonrespiratory, and respiratory specimens. *J Clin Microbiol.* 1996;34:834-841.
21. Wobeser WL, Krajden M, Conly J, et al. Evaluation of Roche Amplicor PCR assay for *Mycobacterium tuberculosis*. *J Clin Microbiol.* 1996;34:134-139.
22. Piersimoni C, Callegaro A, Nista D, et al. Comparative evaluation of two commercial amplification assays for direct detection of *Mycobacterium tuberculosis* complex in respiratory specimens. *J Clin Microbiol.* 1997;35:193-196.
23. Ausina V, Gamboa F, Gazapo E, et al. Evaluation of the semiautomated Abbott LCx *Mycobacterium tuberculosis* assay for direct detection of *Mycobacterium tuberculosis* in respiratory specimens. *J Clin Microbiol.* 1997;35:1996-2002.
24. Smith JH, Radcliffe G, Rigby S, et al. Performance of an automated Q-beta replicase amplification assay for *Mycobacterium tuberculosis* in a clinical trial. *J Clin Microbiol.* 1997;35:1484-1491.
25. Bradley SP, Reed SL, Catanzaro A. Clinical efficacy of the amplified *Mycobacterium tuberculosis* direct test for the diagnosis of pulmonary tuberculosis. *Am J Respir Crit Care Med.* 1996;153:1606-1610.
26. Chin DP, Yajko DM, Hadley WK, et al. Clinical utility of a commercial test based on the polymerase chain reaction for detecting *Mycobacterium tuberculosis* in respiratory specimens. *Am J Respir Crit Care Med.* 1995;151:1872-1877.
27. Cohen RA, Muzaffar S, Schwartz D, et al. Diagnosis of pulmonary tuberculosis using PCR assays on sputum collected within 24 hours of hospital admission. *Am J Respir Crit Care Med.* 1998;157:156-161.
28. American Thoracic Society Workshop. Rapid diagnostic tests for tuberculosis: what is the appropriate use? *Am J Respir Crit Care Med.* 1997;155:1804-1814.
29. Nightingale SL. From the Food and Drug Administration. *JAMA.* 1996;275:585.
30. Barnes PF. Rapid diagnostic tests for tuberculosis. *Am J Respir Crit Care Med.* 1997;155:1497-1498.
31. Barnes PF, Leedom JM, Chan LS, et al. Predictors of short-term prognosis in patients with pulmonary tuberculosis. *J Infect Dis.* 1988;158:366-371.
32. Cohen R, Muzaffar S, Capellan J, Azar H, Chinikamwala M. The validity of classic symptoms and chest radiographic configuration in predicting pulmonary tuberculosis. *Chest.* 1996;109:420-423.
33. Scott B, Schmid M, Nettleman MD. Early identification and isolation of inpatients at high risk for tuberculosis. *Arch Intern Med.* 1994;154:326-330.
34. El-Solh A, Mylotte J, Sherif S, et al. Validity of a decision tree for predicting active pulmonary tuberculosis. *Am J Respir Crit Care Med.* 1997;155:1711-1716.
35. Mylotte JM, Rodgers J, Fassl M, Seibel K, Vacanti A. Derivation and validation of a pulmonary tuberculosis prediction model. *Infect Control Hosp Epidemiol.* 1997;18:554-560.
36. Schulman KA, Escarce JJ, Eisenberg JM, et al. Assessing physicians' estimates of the probability of coronary artery disease. *Med Decis Making.* 1992;12:109-114.
37. Diamond GA. Clinical epistemology of sensitivity and specificity. *J Clin Epidemiol.* 1992;45:9-13.
38. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med.* 1978;299:926-930.
39. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research: getting better but still not good. *JAMA.* 1995;274:645-651.
40. El-Solh AA, Nopper J, Abdul-Khoudoud MR, et al. Clinical and radiographic manifestations of uncommon pulmonary nontuberculous mycobacterial disease in AIDS patients. *Chest.* 1998;114:138-145.

Figure. House-Dust Endotoxin Concentrations From Homes in Denver, US Farms, and Rural Areas of Peru and India



House-dust endotoxin levels are reported in endotoxin units (EU)/mL, using reference standard endotoxin provided by the US Food and Drug Administration. Log transformation of endotoxin values normalized the distribution of the data. Data points with error bars and associated numbers indicate the geometric mean values and SEM for each location subgroup. Wilcoxon rank sum test, $P<.001$. Farm barns outnumber farm homes because some homes had more than 1 associated barn and also because some barns were sampled with permission from workers when farm homeowners were not home.

Results. Urban homes had significantly lower house-dust endotoxin levels than farm homes and rural homes in developing countries ($P<.001$, Wilcoxon rank sum test) (FIGURE). Farm barns had significantly higher endotoxin levels compared with both farm homes ($P<.001$, t test) and rural homes in developing countries ($P=.03$, t test). Farm home and associated barn endotoxin levels were significantly correlated (Spearman $r=0.67$, $P=.02$).

Comment. Greater levels of exposure to environmental endotoxin from early childhood, especially in rural areas of developing countries and in farming communities, may help explain the low prevalence of asthma and allergies observed in children raised in these environments. If environmental endotoxin exposure in early life has an atopy-protective effect, then T_H1-type activity may be induced separately from exposure to serious infections, thus suggesting a possible strategy for allergy and asthma prevention.

Jose E. Gereda, MD
Donald Y. M. Leung, MD, PhD
Andrew H. Liu, MD
National Jewish Medical and Research Center and University of Colorado Health Sciences Center
Denver

1. Beasley R, Crane J, Lai CK, Pearce N. Prevalence and etiology of asthma. *J Allergy Clin Immunol*. 2000;105:466-472.
2. von Mutius E. The environmental predictors of allergic disease. *J Allergy Clin Immunol*. 2000;105:9-19.
3. Weinberg EG. Urbanization and childhood asthma: an African perspective. *J Allergy Clin Immunol*. 2000;105:224-231.
4. Martinez FD, Holt PG. Role of microbial burden in aetiology of allergy and asthma. *Lancet*. 1999;354(suppl 2):SII12-SII15.
5. Ball TM, Castro-Rodriguez JA, Griffth KA, Holberg CJ, Martinez FD, Wright AL. Siblings, day-care attendance, and risk of asthma and wheezing during childhood. *N Engl J Med*. 2000;343:538-543.
6. Gereda JE, Leung DY, Thatayatikom A, et al. Relation between house-dust endotoxin exposure, type 1 T-cell development, and allergen sensitisation in infants at high risk of asthma. *Lancet*. 2000;355:1680-1683.

CORRECTION

Incorrect Negative Predictive Value: In the Original Contribution entitled "The Role of Clinical Suspicion in Evaluating a New Diagnostic Test for Active Tuberculosis: Results of a Multicenter Prospective Trial" published in the February 2, 2000, issue of THE JOURNAL (2000;283:639-645), there were errors in the reporting of 1 negative predictive value. On page 639, in the abstract, under "Results," the last sentence should read "Corresponding negative predictive values were 99%, 91%, and 55% (E-MTD test) vs 96%, 71%, and 37% (AFB smear)." On page 643, the second to last sentence in the first paragraph under the heading "Estimating Clinical Utility" should read "However, the expected NPV of the AFB smear was only 37%, compared with 55% for the E-MTD." On page 644, in Figure 2, the Enhanced *Mycobacterium tuberculosis* Direct Test bar for negative predictive value of high suspicion for tuberculosis (87) should be 55% instead of 91%. Also on page 644, in column 3, the last sentence in the first paragraph should read "The high and consistent specificity of the E-MTD also appeared to be clinically valuable in excluding disease among patients with intermediate CSTB estimates, offering an intermediate NPV of 91%, compared with 71% for AFB smear."