CLINICAL REVIEW

Interpretation of the Tuberculin Skin Test

David N. Rose, MD, Clyde B. Schechter, MD, Jack J. Adler, MD

OBJECTIVE: **To reinterpret epidemiologic information about the tuberculin test (purified protein derivative) in terms of modern approaches to test characteristics; to clarify why different cutpoints of induration should be used to define a positive test in different populations; and to calculate test characteristics of the intermediate-strength tuberculin skin test, the probability of *Mycobacterium tuberculosis* infection at various induration sizes, the area under the receiver operating characteristic (ROC) curve, and optimal cutpoints for positivity.**

METHODS: **Standard epidemiologic assumptions were used to distinguish *M. tuberculosis*-infected from -uninfected persons; also used were data from the U.S. Navy recruit and World Health Organization tuberculosis surveys; and Bayesian analysis.**

RESULTS: **In the general U.S. population, the test's sensitivity is 0.59 to 1.0, the specificity is 0.95 to 1.0, and the positive predictive value is 0.44 to 1.0, depending on the cutpoint. Among tuberculosis patients, the sensitivity is nearly the same as in the general population; the positive predictive value is 1.0. The area under the ROC curve is 0.997. The probability of *M. tuberculosis* infection at each induration size varies widely, depending on the prevalence. The optimal cutpoint varies from 2 mm to 16 mm and is dependent on prevalence and the purpose for testing.**

CONCLUSIONS: **The operating characteristics of the tuberculin test are superior to those of nearly all commonly used screening and diagnostic tests. The tuberculin test has an excellent ability to distinguish *M. tuberculosis*-infected from -uninfected persons. Interpretation requires consideration of prevalence and the purpose for testing. These findings support the recommendation to use different cutpoints for various populations. Even more accurate information can be gotten by interpreting induration size as indicating a probability of *M. tuberculosis* infection.**

KEY WORDS: **tuberculin test; tuberculosis; nontuberculous mycobacteria.**
J GEN INTERN MED 1995;10:635–642.

What is the meaning of a positive tuberculin skin test? The answer to this question remains uncertain despite use of the test for more than a century, because there is no diagnostic "gold standard" with which to compare tuberculin reactivities. Our current understanding of tuberculin reactivity has been unchanged over the past four decades and is based on two fundamental inferences from epidemiologic data.[1,2] First, reactivity 48 to 72 hours after intradermal intermediate-strength purified protein derivative (PPD) among persons with tuberculous infection is similar to reactivity among persons with tuberculous disease. Most persons with tuberculous disease have indurations 16 to 17 mm in diameter, though indurations can range from 0 to more than 30 mm in diameter.[3,4] Second, nontuberculous mycobacterial (NTM) infection also causes tuberculin reactivity, called "cross reactions." Most of these reactions occur at 0 to 2 mm, and NTM infection produces successively fewer induration sizes up to 15 mm. The general population has a mixture of infected and uninfected persons.[1–6]

Test interpretation is made even more complex by the many factors that diminish reactivity, from technical problems in performing and reading the test to conditions that impair delayed-type hypersensitivity.[7] Furthermore, many clinicians are confused by the seemingly arbitrary choice of cutpoints (the induration sizes that separate positive tests from negative tests) and by the discordance between the definition of tuberculin positivity and the indications for isoniazid preventive therapy. Despite these complexities, the tuberculin skin test has not been subjected to modern methods of test analysis. As a result, positive tests are often interpreted as meaning the patient has *Mycobacterium tuberculosis* infection and negative tests as meaning the patient does not have *M. tuberculosis* infection. By contrast, clinicians understand that exercise electrocardiographic testing neither definitively diagnoses obstructive coronary artery disease nor indicates specific therapies. An

Received from the Departments of Community Medicine (DNR, CBS, JJA) and Medicine (DNR, JJA), Mount Sinai School of Medicine, and the AIDS Center (DNR) and Tuberculosis Elimination Program (JJA), Mount Sinai Hospital, New York, New York.

Presented in part at the 1993 American Thoracic Society International Conference, San Francisco, California, May 1993.

Supported in part by grant MH45686 from the National Institutes of Health.

Address correspondence and reprint requests to Dr. Rose: Box 1009, Mount Sinai Medical Center, New York, NY 10029.

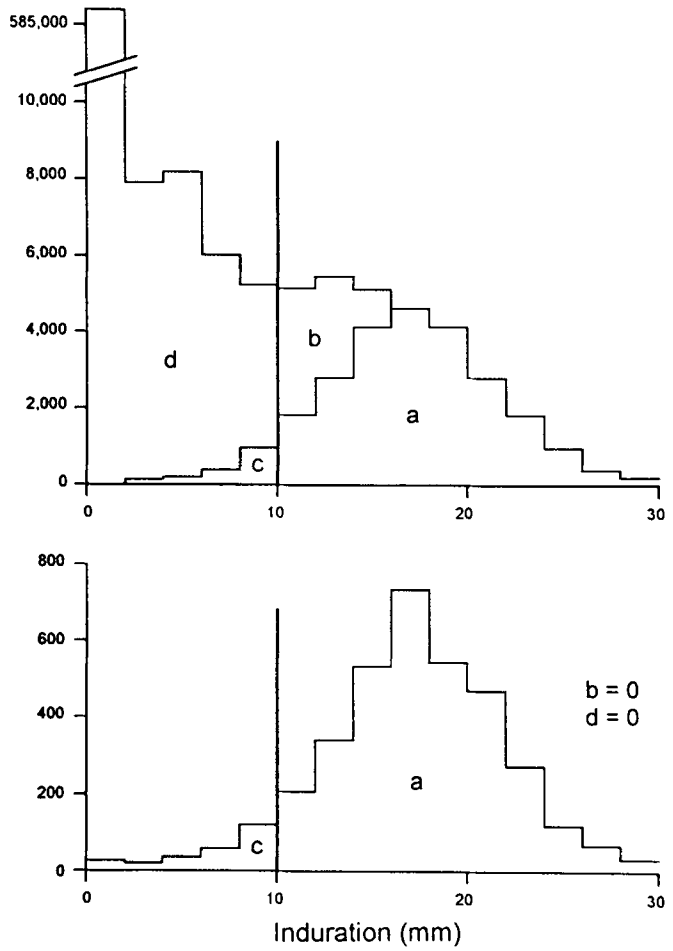

FIGURE 1. Frequency of tuberculin reactions in 643,694 U.S. Navy recruits (*top graph*) and 3,826 tuberculosis patients (*bottom graph*). Tuberculin reactions are classified as true positives (a), false positives (b), false negatives (c), or true negatives (d). Among the navy recruits, the classification is made by distinguishing *Mycobacterium tuberculosis*-infected persons from uninfected persons and by the choice of cutpoint (*vertical line*). Among the tuberculosis patients, the classification is made by the choice of cutpoint (*vertical line*); there is no uninfected patient.

extensive literature relates the findings of exercise testing to the probability that a patient has coronary heart disease. Our purpose here is to apply similar methods of test analysis to the PPD test.

We used information from two well-regarded studies to calculate several kinds of test characteristics at various induration sizes: the probability of *M. tuberculosis* infection; standard test characteristics (sensitivity, specificity, positive predictive value, and negative predictive value); and the receiver operating characteristic (ROC) curve. We then explored the consequences of different cutpoints used for various purposes of tuberculin testing. This article is not intended to be a comprehensive review, which is available elsewhere.[7–11] We present instead a quantitative analysis of test characteristics and the implications for interpretation of the tuberculin test.

METHODS

We studied reactions to intradermal (Mantoux), intermediate-strength (5 tuberculin units [TU], or 0.0001 mg) PPD. We reanalyzed the same studies that inform the currently recommended interpretation of the tuberculin test.[7, 8] The first is a study by the U.S. Public Health Service and the U.S. Navy of tuberculin reactions in 643,694 male navy recruits (all races, ages 17 to 21 years) between 1958 and 1964.[12, 13] This is the largest tuberculin survey conducted and is used to describe tuberculin reactivity in the general population.[7] The survey also measured reactions to intradermal antigens made from an NTM agent, and therefore also documented the geographic distribution of that infection. The second study, conducted by the U.S. Public Health Service and the World Health Organization in the early 1950s, measured tuberculin reactions in 3,826 hospitalized active tuberculosis patients (ages not stated) in eight countries.[14] In both studies, the frequency of reactions was reported only for even-numbered induration sizes in millimeters. We interpreted each of these frequencies as including reactors who had the next higher odd-numbered induration size, i.e., we assumed the frequency of reactions at 2 mm included persons who had 3-mm reactions.

We used standard epidemiologic assumptions to distinguish *M. tuberculosis* infection from NTM infection.[8] The major obstacle to evaluating the characteristics of the tuberculin test is the lack of a definitive reference criterion (gold standard). We therefore followed the recommendation of the American Thoracic Society and the Centers for Disease Control and Prevention (CDC) in comparing the distribution of tuberculin responses in a population of known cases of tuberculosis with that of a population having an unidentified mixture of infected and uninfected persons.[8] The uppermost contour of the top graph in Figure 1 shows the frequency distribution of induration responses to 5-TU PPD among 643,694 U.S. Navy recruits. This sample is a mixture of people uninfected by any mycobacterium species, people infected with an NTM species, and people infected with *M. tuberculosis*. The frequency distribution of tuberculin reactions of 3,826 patients with known tuberculosis is shown in the bottom graph in Figure 1. The latter curve approximates a bell-shaped distribution centered at about 17 mm and its right half resembles the shape of the portion of the upper curve to the right of 17 mm. Because it is plausible that the number of PPD responses 17 mm or larger among people not infected with *M. tuberculosis* is negligible, it is reasonable to identify that portion of the upper graph as representing responses of people with *M. tuberculosis* infection. Relying on the symmetry of the bottom graph of Figure 1, we then mirrored the right-most end of the upper curve around a vertical axis at 17 mm to "cut out" the portion of the upper curve attributable to people with *M. tuberculosis*

infection. This approach assumes that people asymptomatically infected with *M. tuberculosis* have a tuberculin reactivity that is similar to the reactivity of people with known tuberculosis.

We then calculated test characteristics (Table 1) at eight cutpoints between 2-mm and 16-mm indurations for the navy recruits and the tuberculosis patients. The cutpoints separate tuberculin reactions into true-positive, false-positive, true-negative, and false-negative reactions, as shown in Figure 1. A test's sensitivity is the probability of a positive test among infected persons. The specificity is the probability of a negative test among uninfected persons. The positive predictive value is the probability of infection among persons with a positive test. The negative predictive value is the probability of no infection among persons with a negative test. The true-positive rate is the same as the sensitivity. The false-positive rate is the probability of a positive test among uninfected persons (it is also 1 minus the specificity).

We then constructed a ROC curve for the tuberculin test based on the navy recruit data. ROC curves plot a test's true-positive rate against its false positive rate. The area under the curve reflects how well the test discriminates between infected and uninfected persons. The area under the curve of a perfect test (a 1.0 true-positive rate and a 0 false-positive rate at all cutpoints) is 1.0, and describes a test that discriminates perfectly between infected and uninfected persons. The area under the curve for a useless test is 0.5, and describes a test that does not discriminate between infected and uninfected persons. We calculated the area under the curve by fitting a binormal model to the frequencies of tuberculin reaction sizes imputed to infected and uninfected navy recruits.[15] The standard method of calculating the area under the ROC curve for a binormal test was then applied (details available from the authors on request).

Based on our allocation of tuberculin reactions to putatively infected and uninfected persons (Fig. 1), we calculated the prevalence of *M. tuberculosis* infection among the navy recruits to be 4%. Because the predictive value of a test varies directly with the prevalence of infection, we used the sensitivity and specificity of the tuberculin test among the navy recruits to calculate positive predictive values for population groups with other prevalences of *M. tuberculosis* infection.

We also calculated the probability of *M. tuberculosis* infection for a variety of prevalences and at eight tuberculin reactivity sizes, from 2–3 to 16–17 mm. The frequency with which an uninfected person would exhibit a tuberculin reaction in a specific range, P(SIZE|uninfected), was calculated as the false-positive rate for the lower end of the range minus the false-positive rate for the upper end. Similarly, the frequency of observing a tuberculin reaction in that range for an infected person, P(SIZE|infected), was the difference in the corresponding true-positive rates. The likelihood ratio for that range of sizes of tuberculin reaction, LR(SIZE), was then defined as:

$$\mathrm{LR(SIZE)} = \frac{\mathrm{P(SIZE|infected)}}{\mathrm{P(SIZE|uninfected)}}$$

The likelihood ratio form of Bayes' theorem was used to calculate the probability of infection for a person from a population of given prevalence:

$$\mathrm{Odds(infection\,|\,SIZE)} = \mathrm{LR(SIZE)} \times \frac{\mathrm{Prevalence}}{1 - \mathrm{Prevalence}}$$

and

$$\mathrm{P(infection|SIZE)} = \frac{\mathrm{Odds(infection|SIZE)}}{1 + \mathrm{Odds(infection|SIZE)}}$$

Optimal cutpoints for each population group were then calculated for three illustrative situations using three methods for evaluating tuberculin reactivity: 1) the

Table 1
Standard Test Characteristics Calculated for the Tuberculin Skin Test at Each Cutpoint

		Mycobacterium tuberculosis Infection		
		Yes	No	
Test	Positive (induration ≥ cutpoint)	a	b	a + b
	Negative (induration < cutpoint)	c	d	c + d
		a + c	b + d	

True positives = a.
False positives = b.
False negatives = c.
True negatives = d.
Sensitivity (also true-positive rate) = a/(a + c).
Specificity = d/(b + d).
Positive predictive value = a/(a + b).
Negative predictive value = d/(c + d).
False-positive rate = b/(b + d).

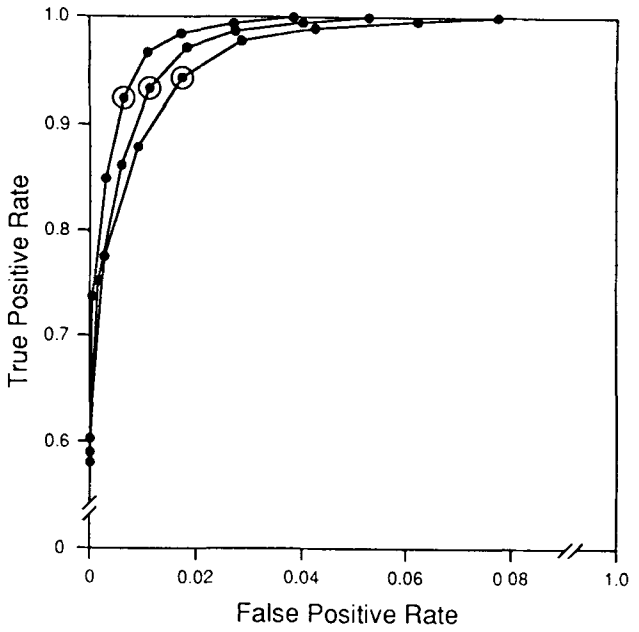

FIGURE 2. Receiver operating characteristic (ROC) curves of the tuberculin test in all 643,694 U.S. Navy recruits (*middle curve*), in the 193,656 recruits from 21 Southern states (*lower curve*), and in the recruits from other states (*upper curve*). The circled points represent the test characteristics using the 10-mm cutpoint. On each curve, points above and to the right of the circled point represent test characteristics of smaller cutpoints, and points below and to the left represent test characteristics of larger cutpoints.

cutpoint that minimizes the sum of false-positive and false-negative tests (i.e., minimizes the total number of incorrect test results), 2) the cutpoint that identifies 95% of the *M. tuberculosis*-infected persons in each group, and 3) the cutpoints that indicate a 0.75 and a 0.95 probability of *M. tuberculosis* infection.

RESULTS

Test Characteristics in the General Population and in Tuberculosis Patients

We calculate that 25,629 (4.0%) of the 643,694 navy recruits had tuberculin reactivity 2 mm or larger caused by *M. tuberculosis* infection and an additional 32,682 (5.1%) had tuberculin reactivity 2 mm or larger caused by NTM infection. The test's sensitivity to detect *M. tuberculosis* infection varied from 0.59 at a 16-mm cutpoint to 1.0 at a 2-mm cutpoint, and the specificity ranged from 0.95 at a 2-mm cutpoint to 1.0 at a 14-mm cutpoint (Table 2). Although the prevalence of *M. tuberculosis* infection varied by region, the sensitivity varied little by region. The specificity was slightly lower in Southern states, where the survey also revealed high prevalences of NTM infection. The positive predictive value ranged from a 0.44 probability of *M. tuberculosis* infection at a 2-mm cutpoint to a 1.0 probability at a 16-mm cutpoint. The positive predictive values were substantially lower

Table 2
Test Characteristics of the Tuberculin Skin Test in the General Population and among Tuberculosis Patients

	Cutpoint (mm)							
	2	4	6	8	10	12	14	16
General population								
All navy recruits								
Sensitivity	1.00	0.99	0.99	0.97	0.93	0.86	0.75	0.59
Specificity	0.95	0.96	0.97	0.98	0.99	0.99	1.00	1.00
Positive predictive value	0.44	0.51	0.60	0.69	0.77	0.86	0.95	1.00
Negative predictive value	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98
Recruits from states with high prevalences of NTM* infection†								
Sensitivity	1.00	1.00	0.99	0.98	0.94	0.88	0.77	0.60
Specificity	0.92	0.94	0.96	0.97	0.98	0.99	1.00	1.00
Positive predictive value	0.32	0.36	0.45	0.55	0.66	0.78	0.91	1.00
Negative predictive value	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99
Recruits from all other states								
Sensitivity	1.00	0.99	0.98	0.97	0.92	0.85	0.74	0.58
Specificity	0.96	0.97	0.98	0.99	0.99	1.00	1.00	1.00
Positive predictive value	0.52	0.61	0.71	0.79	0.86	0.92	0.98	1.00
Negative predictive value	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98
Tuberculosis patients								
Sensitivity	0.99	0.99	0.98	0.96	0.93	0.87	0.78	0.63
Specificity	—	—	—	—	—	—	—	—
Positive predictive value	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Negative predictive value	—	—	—	—	—	—	—	—

*NTM = nontuberculous mycobacterial.

†Alabama, Arkansas, Delaware, Florida, Georgia, Hawaii, Iowa, Kansas, Kentucky, Louisiana, Maryland, Mississippi, Missouri, Nevada, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Utah, and Virginia.

Table 3
Positive Predictive Value* of a Positive Tuberculin Test

Prevalence of *Mycobacterium tuberculosis* Infection	Cutpoint (mm)							
	2	4	6	8	10	12	14	16
1%	0.16	0.20	0.27	0.35	0.45	0.60	0.82	1.00
4%	0.44	0.51	0.60	0.69	0.77	0.86	0.95	1.00
10%	0.68	0.73	0.80	0.86	0.90	0.94	0.98	1.00
20%	0.83	0.86	0.90	0.93	0.95	0.97	0.99	1.00
30%	0.89	0.91	0.94	0.96	0.97	0.98	1.00	1.00
60%	0.97	0.97	0.98	0.99	0.99	1.00	1.00	1.00
80%	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00
100%	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Probability of infection for persons with these induration sizes and larger.

Table 4
Probability* of *Mycobacterium tuberculosis* Infection

Prevalence of *Mycobacterium tuberculosis* Infection	Tuberculin Reactivity (mm)							
	2–3	4–5	6–7	8–9	10–11	12–13	14–15	16–17
1%	0.00	0.01	0.02	0.05	0.12	0.21	0.50	1.00
4%	0.02	0.02	0.07	0.18	0.36	0.52	0.81	1.00
10%	0.05	0.06	0.16	0.38	0.60	0.74	0.92	1.00
20%	0.10	0.13	0.30	0.58	0.77	0.86	0.96	1.00
30%	0.16	0.20	0.42	0.70	0.85	0.92	0.98	1.00
60%	0.39	0.47	0.72	0.89	0.95	0.97	0.99	1.00
80%	0.63	0.70	0.87	0.96	0.98	0.99	1.00	1.00
100%	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Probability of infection for persons with these induration sizes.

in Southern states, where NTM infections were highly prevalent. Among 193,656 recruits from the 21 states with the highest ratio of nontuberculous to tuberculous mycobacterial infection, the positive predictive value of a 2-mm cutpoint was 0.32. The positive predictive values were correspondingly higher in other states where NTM infections were less prevalent; at the 2 mm cutpoint, it was 0.52. These regional differences are gradually eliminated as the cutpoint increases from 2 mm to 16 mm. The negative predictive values in contrast were uniformly very high in all regions.

Among the 3,826 tuberculosis patients, the test's sensitivity was nearly identical to the sensitivity found among the general population (Table 2). This finding is consistent with the idea that sensitivity and specificity do not depend on variations in prevalence. Because all these patients had tuberculosis, tuberculin reactivity at any cutpoint indicated a certainty of *M. tuberculosis* infection (positive predictive value 1.0). The specificity and negative predictive value were not calculated because there was no uninfected patient.

ROC Curves

The ROC curve for all the navy recruits is the middle curve in Fig. 2. The area under the curve is 0.997. Also shown are the ROC curves for the recruits from the 21 Southern states (the lower curve) and for all other recruits (the upper curve). The areas under their curves are 0.994 and 0.998, respectively.

Positive Predictive Value and Probability of Infection for Various Populations

The positive predictive values of various cutpoints for populations with different prevalences of *M. tuberculosis* infection are shown in Table 3. These values ranged from 0.16 at the 2-mm cutpoint for persons from a population with a 1% prevalence of infection to 1.0 for all persons with tuberculosis, all persons with 16-mm or greater tuberculin reactivity, many persons from a population with an 80% prevalence of infection, and fewer persons from populations with 30% to 60% prevalence.

Table 4 shows the probability of *M. tuberculosis* infection at each tuberculin reactivity size. These values ranged from 0 at the 2–3-mm induration size for persons from a population with a 1% prevalence of infection to 1.0 for all persons with tuberculosis, all persons with 16-mm or greater tuberculin reactivity, and persons with 14-mm or greater tuberculin reactivity from a population with an 80% prevalence of infection.

Table 5
Optimal Cutpoint (mm) by Various Methods of Evaluation

Prevalence of *Mycobacterium tuberculosis* Infection	Minimizes False Positives and False Negatives	Identifies ≥95% of Infected Persons	≥75% of Positives Are Infected	≥95% of Positives Are Infected
1%	14	8	14	16
4%	12	8	10	14
10%	10	8	6	14
20%	8	8	2	10
30%	8	8	2	8
60%	6	8	2	2
80%	2	8	2	2
100%	2	8	2	2

Optimal Cutpoint by Various Methods of Evaluation

The optimal cutpoints for populations with different prevalences of *M. tuberculosis* infection are shown in Table 5. Using the method of minimizing the sum of false-positive and false-negative tests, the optimal cutpoint ranged from 2-mm induration for persons from a population with an 80% prevalence of infection to 14-mm induration for persons from a population with a 1% prevalence of infection. Using the method of identifying 95% of the infected persons, the optimal cutpoint was 8 mm for all persons (the sensitivity does not vary with prevalence). Using the method of identifying persons for whom a positive test means a 0.75 or greater probability of infection, the optimal cutpoint ranged from 2 mm for persons from populations with a 20% or higher prevalence to 14 mm for persons from a population with a 1% prevalence of infection. Using the method of identifying persons for whom a positive test means a 0.95 or greater probability of infection, the optimal cutpoint ranged from 2 mm for persons from populations with a 60% or higher prevalence to 16 mm for persons from a population with a 1% prevalence of infection.

DISCUSSION

We found the tuberculin skin test to have excellent test characteristics. The area under the ROC curve, from 0.994 to 0.998, indicates that the test can discriminate well between *M. tuberculosis*-infected and -uninfected persons. It compares favorably with other widely accepted tests. For example, the area under the ROC curve for magnetic resonance imaging in the diagnosis of multiple sclerosis is 0.82[16]; for Doppler velocity measurements in the diagnosis of carotid artery stenosis it is 0.78 to 0.94[17]; for serum ferritin concentration in the diagnosis of iron-deficiency anemia it is 0.95[18]; for prostate-specific antigen measurement in the diagnosis of prostatic cancer it is 0.90[19]; and for the CAGE questionnaire in the diagnosis of alcoholism it is 0.89.[20] In an age of sophisticated diagnostics, it is easy to disparage a quaint test first used in the 19th century. Yet new tests for diagnosing *M. tuberculosis* infection must compare favorably to the tuberculin test's excellent test characteristics. Nevertheless, most tests with high sensitivity and specificity will still be found to have low positive predictive values when used in low prevalence populations, for example, when the tuberculin skin test is used for screening.

These excellent test characteristics persist despite the many factors that may increase the false-negative rate. These factors include improper storage of the solution, improper injection method, an inexperienced reader, and the presence of conditions that diminish delayed-type hypersensitivity to tuberculin, including many viral, bacterial, and fungal infections, live virus vaccinations, chronic renal failure, protein malnutrition, immunosuppression drugs, HIV infection, and overwhelming active tuberculosis.[7-10, 21]

Others have used these methods to estimate the tuberculin test's characteristics.[7, 8, 10, 11] Rust and Thomas used data from the tuberculin survey of U.S. Navy recruits but a different method to calculate the prevalence of *M. tuberculosis* infection at various induration sizes. Their findings were similar to ours, though they found higher prevalences at small induration sizes and lower prevalences at large induration sizes.[22] Edwards and Palmer also used the U.S. Navy recruit data, but combined these data with the results of skin tests made from an NTM agent, the Battey bacillus. Based on the relative induration sizes of each recruit's two skin tests, Edwards and Palmer were able to predict the recruits' tuberculosis morbidity in subsequent years better than by using the tuberculin reaction alone.[23]

What is the relationship between the size of a tuberculin reaction and the risk of reactivation disease? The size of a tuberculin reaction in animal models is unrelated to the number of infecting tubercle bacilli.[24] Also, epidemiologic studies have demonstrated that the risks of infection are different from the risks of reactivation among infected persons.[25, 26] The risks of infection are a function of the duration and intensity of exposure to airborne droplet nuclei and are higher among poor, minority populations, residents of congregate living facilities, and immigrants from high-prevalence

countries. The risks of activation are a function of the recency of infection, age (reactivation is especially high for adolescents and the elderly), life expectancy (the annual risks are additive), and the presence of conditions that suppress cell-mediated immunity. Therefore, the probability of reactivation disease is a product of the probability of infection and the probability of reactivation among infected persons.

The choice of an optimal cutpoint for a test depends on the prevalence of disease as well as the value of the consequences from the actions taken for positive and negative test results. These parameters differ from population to population and with the purpose of carrying out testing. For example, if a screening survey were conducted solely for reporting and planning purposes, with no intervention for those found to be positive, then the optimal cutpoint could be the one that minimizes the total number of false-positive and false-negative findings. On the other hand, if testing were done on admission to a congregate residence with a high risk of transmission among residents, sensitivity might be more important than specificity. A smaller induration size would be chosen as a cutpoint: the specific size would depend on the prevalence of infection and the number of false positives one would willingly trade off against one false negative.

Although the tuberculin test is the definitive test for latent *M. tuberculosis* infection, it provides only one piece of information for decision making about isoniazid preventive therapy. That decision depends on a complex calculation that weighs the risks and benefits of preventive therapy. It is based on not only the probability of infection, but also the risk of activation among infected people, their life expectancy, the case-fatality rate of tuberculosis, the effectiveness of preventive therapy, and the risks of isoniazid hepatotoxicity and other adverse reactions. As a result, some persons with a high probability of infection may not benefit from isoniazid preventive therapy, while others with lower probabilities of infection may benefit. Therefore, recommended cutpoint choices for identifying infection are not the same as those for giving preventive therapy.

For many years, 10 mm was the recommended cutpoint for identifying infection.[27] Despite its simplicity, however, that cutpoint resulted in too many false positives in the general population and too many false negatives in higher prevalence populations.[7] For this reason, the CDC and the American Thoracic Society recently redefined tuberculin positivity as 5, 10, or 15 mm, depending on the presence or absence of additional risk factors for infection.[9] Our study demonstrates how important it is to interpret tuberculin reactivity in light of the prevalence of infection in the patient's population group (that is, the prior probability of infection). For persons with a high probability of infection, such as urban, minority patients with upper-lobe infiltrates, small cutpoints (2 mm) provide good confirmatory information. The traditional 10-mm cutpoint for active tuberculosis patients is inappropriate and results in confusing conclusions, such as the finding that 19% of these patients have negative tuberculosis skin tests.[28]

How do a patient's age and geographic residence influence interpretation of his or her tuberculin reactivity? Other than decreased delayed-type hypersensitivity in newborns and the elderly,[8, 29] the biologic response to intradermal tuberculin does not vary with age or by geography.[1, 2, 14] Therefore, the test characteristics are determined by the relative prevalences of *M. tuberculosis* and NTM infections. The relative prevalences vary by geography: there are some settings in which one infection is highly prevalent and the other is not.[1, 7, 12, 13] These differences nevertheless have only moderate influence on the test characteristics (Table 2). While the prevalence of *M. tuberculosis* infection is known to increase with age,[1, 24, 25, 30–32] the relationship between age and the prevalence of NTM infection is less substantiated. Small tuberculin reactions, presumably caused by NTM infections, increase with age,[1, 30] but NTM disease peaks in childhood and middle age.[33] Tuberculin test characteristics, nevertheless, probably do not vary significantly by age. However, age should influence interpretation of the test result because the prevalence of *M. tuberculosis* infection rises with age and therefore so does the positive predictive value.[10]

There are insufficient data to calculate test characteristics for HIV-infected tuberculin reactors or those who have other conditions that diminish cell-mediated immunity. HIV-infected persons are more likely to be anergic and less likely to have positive tuberculin tests than are HIV-uninfected persons.[34] One could assume that the most frequent tuberculin reaction is a smaller induration size than the 16 to 17-mm mode found in immunocompetent persons. This finding has not been documented, however, and unpublished data suggest otherwise. HIV-infected patients with tuberculosis in CDC studies demonstrate a mean tuberculin reaction of 15.3 mm, not different from the mean reactions of 16 mm found in HIV-uninfected tuberculosis patients, suggesting that an individual's tuberculin reactivity is an "all or nothing" phenomenon.[35] Nevertheless, a study of tuberculin reactions among intravenous drug users suggests that the 2-mm cutpoint identifies the same proportion of persons infected with *M. tuberculosis* among HIV-infected persons as is identified using the 10-mm cutpoint among HIV-uninfected persons.[34] Perhaps the best information about an HIV-infected person's risk of *M. tuberculosis* infection is the prevalence of infection in the patient's community.

Many clinicians consider a positive tuberculin test to mean that the patient is infected with *M. tuberculosis*. Given this perspective, the optimal cutpoint should be one that provides a high positive predictive value, such as 0.75 to 0.95. This use of the test supports the decision to classify positivity by cutpoints that vary by the prev-

alence of infection. However, more accurate information can be gotten by interpreting induration size as indicating a probability of *M. tuberculosis* infection,[9] rather than classifying the test result as positive or negative. For example, persons with 10- to 12-mm reactions and no additional risk factor for infection are currently classified as tuberculin-negative, yet they have a 0.12 to 0.74 probability of infection, depending on the prevalence of infection in their communities. In contrast, HIV-infected persons with 6-mm reactions are currently classified as tuberculin-positive, yet their probability of infection is unknown. HIV-uninfected persons with this reactivity and no additional risk factor have a 0.02 to 0.16 probability of infection. We encourage readers to interpret tuberculin reactivity as a probability of infection, because doing so will lead to more thoughtful treatment decisions.

REFERENCES

1. Edwards PQ, Edwards LB. Story of the tuberculin test from an epidemiologic viewpoint. Am Rev Respir Dis. 1960;81(1 pt 2):1–47.
2. Palmer CE, Edwards LB. Tuberculin test in retrospect and prospect. Arch Environ Health. 1967;15:792–808.
3. Palmer CE, Bates LE. Tuberculin sensitivity of tuberculous patients. Bull World Health Organ. 1952;7:171–88.
4. Palmer CE. Tuberculin sensitivity and contact with tuberculosis: further evidence of non-specific sensitivity. Am Rev Tuberc. 1953;68:678–94.
5. Palmer CE, Edwards LB, Hopwood L, Edwards PQ. Experimental and epidemiological basis for the interpretation of tuberculin sensitivity. J Pediatr. 1959;55:413–29.
6. Edwards LB, Comstock GW, Palmer CE. Contributions of northern populations to the understanding of tuberculin sensitivity. Arch Environ Health. 1968;17:507–16.
7. Snider DE Jr. The tuberculin skin test. Am Rev Respir Dis. 1982;125(no 3, pt 2):108–18.
8. American Thoracic Society/Centers for Disease Control. The tuberculin skin test. Am Rev Respir Dis. 1981;124:356–63.
9. American Thoracic Society/Centers for Disease Control. Diagnostic standards and classification of tuberculosis. Am Rev Respir Dis. 1990;142:725–35.
10. Huebner RE, Schein MF, Bass JB Jr. The tuberculin skin test. Clin Infect Dis. 1993;17:968–75.
11. Bass JB Jr. The tuberculin test. In: Reichman LB, Hershfield ES (eds). Tuberculosis: A Comprehensive International Approach. New York: Marcel Dekker, 1993;139–48.
12. Edwards LB, Acquiviva FA, Livesay VT, Cross FW, Palmer CE. An atlas of sensitivity to tuberculin, PPD-B, and histoplasmin in the United States. Am Rev Respir Dis. 1969;99(pt 2):1–132.
13. Edwards LB, Palmer CE. Tuberculous infection. In: Lowell AM (ed). Tuberculosis. Cambridge: Harvard University Press, 1969; 123–202.
14. WHO Tuberculosis Research Office. Further studies of geographic variation in naturally acquired tuberculin sensitivity. Bull World Health Organ. 1955;12:63–83.
15. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. J Math Psychol. 1975;12:387–415.
16. Mushlin AI, Detsky AS, Phelps CE, et al. The accuracy of magnetic resonance imaging in patients with suspected multiple sclerosis. JAMA. 1993;269:3146–51.
17. Hunink MG, Polak JF, Barlan MM, O'Leary DH. Detection and quantification of carotid artery stenosis: efficacy of various Doppler velocity parameters. AJR Am J Roentgenol. 1993;160:619–25.
18. Guyatt GH, Oxman AD, Ali M, Willan A, McIlroy W, Patterson C. Laboratory diagnosis of iron-deficiency anemia: an overview. J Gen Intern Med. 1992;7:145–53.
19. Pagani F, Zambolin T, Bonora R, Panteghini M. Diagnostic value of prostatic acid phosphatase and prostate-specific antigen in patients with prostatic cancer. J Nucl Med Allied Sci. 1990;34:85–7.
20. Buchsbaum DG, Buchanan RG, Centor RM, Schnoll SH, Lawton MJ. Screening for alcohol abuse using CAGE scores and likelihood ratios. Ann Intern Med. 1991;115:774–7.
21. Chaparas SD, Mac Vandiviere H, Melvin I, Koch G, Becker C. Tuberculin test: variability with the Mantoux procedure. Am Rev Respir Dis. 1985;132:175–7.
22. Rust P, Thomas J. A method for estimating the prevalence of tuberculous infection. Am J Epidemiol. 1975;101:311–22.
23. Edwards LB, Palmer CE. Identification of the tuberculous-infected by skin tests. Ann N Y Acad Sci. 1968;154:140–8.
24. Dannenberg AM Jr. Immune mechanisms in the pathogenesis of pulmonary tuberculosis. Rev Infect Dis. 1989;11(suppl 2):S369–S378.
25. Comstock GW. Frost revisited: the modern epidemiology of tuberculosis. Am J Epidemiol. 1975;101:363–82.
26. Rieder HL, Cauthen GM, Comstock GW, Snider DE Jr. Epidemiology of tuberculosis in the United States. Epidemiol Rev. 1989;11:79–98.
27. National Tuberculosis and Respiratory Disease Association. Diagnostic standards and classification of tuberculosis. New York, 1969.
28. Holden M, Dubin MR, Diamond PH. Frequency of negative intermediate-strength tuberculin sensitivity in patients with active tuberculosis. N Engl J Med. 1971;285:1506–9.
29. Thompson NJ, Glassroth JL, Snider DE Jr, Farer LS. The booster phenomenon in serial tuberculin testing. Am Rev Respir Dis. 1979;119:587–97.
30. Kuemmerer JM, Comstock GW. Sociologic concomitants of tuberculin sensitivity. Am Rev Respir Dis. 1967;96:885–92.
31. Smith DT. Diagnostic and prognostic signficance of the quantitative tuberculin tests. Ann Intern Med. 1967;67:919–46.
32. Trump DH, Hyams KC, Cross ER, Struewing JP. Tuberculosis infection among adults entering the US Navy in 1990. Arch Intern Med. 1993;153:211–6.
33. O'Brien RJ, Geiter LJ, Snider DE Jr. The epidemiology of nontuberculous mycobacterial diseases in the United States: results from a national survey. Am Rev Respir Dis. 1987;135:1007–14.
34. Graham NMH, Nelson KE, Solomon L, et al. Prevalence of tuberculin positivity and skin test anergy in HIV-1-seropositive and -seronegative intravenous drug users. JAMA. 1992;267:369–73.
35. Huebner RE, Villarino ME, Snider DE Jr. Tuberculin skin testing and the HIV epidemic. JAMA. 1992;267:409–10.