Evolution and Current Use of the Tuberculin Test

Elsie Lee and Robert S. Holzman
Department of Medicine, New York University School of Medicine, New York

Since it was first introduced in the late 1800s, the tuberculin test has undergone continual refinement in its formulation, standardization, and dosage, as well as its interpretation and indications for use. New guidelines have replaced universal screening with targeted testing and rigid definitions of positivity with individualized criteria formulated from a Bayesian approach. This review summarizes the evolution of the test and provides information to help gauge its predictive value.

In 1890, Robert Koch reported to the Tenth International Congress of Medicine in Berlin that *Mycobacterium tuberculosis*–infected guinea pigs evinced a destructive inflammatory reaction at the site of injection of heat-killed tubercle bacilli and extracts prepared from them [1]. Although not the cure he prematurely announced [2], the description of the Koch phenomenon and tuberculin hypersensitivity led ultimately to the practical tuberculosis screening tests (TSTs), first introduced by von Pirquet in 1909 [3] and now in common use today [4].

**EVOLUTION OF TUBERCULOPROTEINS**

Koch's “old tuberculin” (OT) was prepared by dissolving, in a glycerin-containing solvent, the residue from cultures of *M. tuberculosis* heated for several hours at 100°C and concentrated 10-fold by evaporation. Because of its impurity, toxicity, nonspecificity, and inadequate standardization, Koch's OT and several similar products are not used in TSTs in the United States. OT was quickly adopted as a screening test for active tuberculosis and subsequently for the screening of apparently healthy people for “inactive” or latent infection. A total of 4 or 5 sequential injections of graded strengths were used until 1932, when D'Arcy Hart [5] demonstrated that a 1:10 dilution of OT was the highest concentration needed and that some reactions to higher doses were probably not due to *M. tuberculosis*.

Fenger et al. [6] found that 92% of tubercular patients reacted to 0.01 mg of OT, an amount equivalent to 0.0005 mg of a more purified material, MA-100. MA-100 preparations from bovine, avian, and botanical mycobacteria elicited reactions from uninfected persons, again suggesting that reactions to large doses of OT resulted from sensitization to nontuberculous mycobacteria. There was a clear need for a purer tuberculin with less propensity to cross-react.

Florence B. Seibert at the Henry Phipps Institute in Philadelphia established that OT preparations contained varying amounts of proteins and polysaccharides and that the proteins were the relevant TST antigens. She produced purified protein derivative (PPD), a more consistent and standardizable material than OT, by steaming cultures of *M. tuberculosis* in an Arnold sterilizer and purifying the proteins by repeated precipitation with neutral ammonium sulfate [7].

In 1939, Seibert [8] prepared a large lot (PPD lot 49608) that became the reference standard for the US Public Health Service’s Bureau of Biologics Standards. In 1944 this lot was renamed PPD-S (“S” meaning “standard”), and in 1952 PPD-S was adopted as an international standard by the World Health Organization. By convention, 5 test units (TU) is the bioassayable skin test activity contained in 0.0001 mg of PPD-S [8].

By 1940, standard TSTs used 2 sequential injections: 1 TU (“first-strength”) and 250 TU (“second-strength”). In 1942, Furcolow et al. [9] recommended substitution of a single 5 TU (“intermediate-strength”) injection, on the basis of the fact that 5 TU caused reactions in 99.6% of patients with active tuberculosis. They noted that even infants <6 months of age, and therefore unlikely to be infected, would react if sufficiently high doses were given. Goddard et al. [10] contemporaneously noted that intermediate-strength PPD was highly efficient for identifying persons with tuberculous pulmonary infiltrates or calcifications.
In 1969, Grzybowski et al. [11], and in 1971, Holden et al. [12] reported inconsistencies when marketed TSTs were compared with PPD-S. Holden et al. [12] found the cause to be alterations in the delivered dose of PPD due to varying degrees of adsorption by surfaces of syringes and containers. In 1972, the Division of Biologics Standards mandated that stabilized liquid formulations replace tableted PPD formulations (which lost more activity as the time between adding of diluent and use increased). In addition, all commercially produced, stabilized, PPDs were required to be bioassayed and shown equivalent to 5 TU of PPD-S before marketing. Since then, master lots of PPD products marketed in the United States have been standardized by bioassay in both tuberculin-sensitized guinea pigs and skin test–positive humans.

Nontuberculous mycobacteria contain proteins analogous to those of M. tuberculosis, and considerable cross-reactivity exists. PPD-B (from Mycobacterium intracellulare), PPD-G (Mycobacterium scrofulaceum), and PPD-Y (Mycobacterium kansasii) have been helpful in epidemiological studies and in clarifying the interpretation of modest degrees of reactivity to the PPD of M. tuberculosis. Their clinical role is limited [13], because no standard comparable to PPD-S exists and they are not commercially available.

Two commercial TSTs are marketed in the United States: Aplisol (Parkdale Pharmaceuticals) and Tubersol (Connaught). Although each is tested for bioequivalence to PPD-S, numerous reports state that Aplisol elicits more or larger reactions [14]. Other reports document equivalence [15], although this has been challenged [16]. It is best to use one of these products consistently and not shift from one to the other.

Outside of the United States, another purified protein derivative, RT-23, is used. Introduced in 1958 by the World Health Organization, the master batch was prepared at the Statens Serum Institute and standardized by bioassay [17]. The standard skin test dose for RT-23 is 2 TU, and human testing indicated that reactions in sensitive persons, tested simultaneously, averaged 16.8 mm for 2 TU of RT-23 and 18.7 mm for 5 TU of Tubersol [18]. Field studies have questioned whether preparations of RT-23 have lost potency over time [19, 20] and, in response, Statens Serum Institute has published its quality control data and information suggesting that reported declines in RT-23 potency result from dilution and handling at local sites and not from a change in the product as manufactured [18]. Additional studies suggest comparable results with the two preparations [21, 22].

It has been >60 years since Seibert produced PPD-S. To anticipate the eventual depletion of this primary standard, Villarino et al. [23] have developed a new batch of PPD, PPD-S2, and compared the old and new standards in a randomized, double-blinded clinical trial to establish relative potency.

**TESTING TECHNIQUES**

Several techniques have been used for tuberculin skin testing. Multiple puncture tests, such as the Heaf test, tine test, and MONO-VACC test, used concentrated tuberculin, but the varying dose of PPD or OT contained on each tine results in inadequate sensitivity and specificity. Intracutaneous injections may be given by the jet injector or by the Mantoux test, but only the Mantoux test is recommended for clinical assessment or screening. It is more sensitive and specific than the other methods [24]. In the Mantoux test, 0.1 mL of PPD solution is injected intracutaneously with a #27 steel (or #26 platinum) needle attached to a 1.0-mL syringe. Creation of a visible wheal is crucial; subcutaneous administration will result in a false-negative test. After 48 h, the diameter of induration transverse to the long axis of the forearm is measured. A number of specialized rulers and other aids to assessment of reaction size have been advocated, such as drawing a ballpoint pen across the skin to mark the border, but accurate measurement requires experience.

The tuberculin reaction, which is read after 48 h, is a delayed-type hypersensitivity reaction. Sensitivity develops 2–12 weeks after infection with M. tuberculosis [25]. Macroscopically, redness and induration at the site peaks at 48–72 h and slowly wanes over several days. Intense reactivity may be associated with fever, more general swelling of the limb, regional lymphadenopathy, or rarely lymphangitis (H. S. Lawrence, personal communication) [26]. Microscopically the test site shows edema and a dense infiltration of the dermis by mononuclear cells, especially around small blood vessels.

Immediate or accelerated reactions to TSTs occur occasionally. Urticarial reactions may occur within 20 min of administration, possibly because of reactivity with polysaccharides in the test material. Accelerated reactions may represent Arthus reactivity, generally peaking within the first 24 h and blending with an evolving delayed-type reaction. Immediate and hypersensitivity reactions to constituents of the diluent can also occur. These reactions occur within 24 h and should not be confused with delayed-type hypersensitivity [27]. Repeated tuberculin skin testing will not cause a truly tuberculin-negative person, who has never been infected with M. tuberculosis or sensitized to other mycobacteria, to become TST-positive [28].

In some M. tuberculosis–infected persons, the ability to react to TST diminishes over time. Administration of a TST to them can restore reactivity, boosting the response to future TSTs [29]. Believed to result from recall of waned, cell-mediated immunity, boosting is common in persons >55 years of age and in those born outside the United States who have been vaccinated with BCG. Two-step tests are used to avoid interpreting the boost as a new infection. If the reaction to the first TST is negative, a TST is repeated in 1–3 weeks. The second test is interpreted as measuring the degree of reactivity. Two-step testing is es-
induration, with a diameter of 10 mm. Fifty-four (3.7%) had a ≥6-mm increase in induration, with a diameter of >10 mm on the second test. The average increase was 13 mm [30].

EVOLUTION OF CURRENT PRACTICES AND GUIDELINES

During much of the past 50 years, TSTs were used for universal screening of the general population and periodic screening of high-risk populations [27, 31]. Current guidelines suggest that low-risk persons in the US general population need not receive routine screening [32]. The anticipated yield is low, and there is an increasing risk of false-positive results due to the low prevalence of latent tuberculous infection but high prevalence of sensitization to nontuberculous mycobacteria. Groups of high-risk adults and children in whom screening is likely to be productive have been defined, and consensus recommendations of the American Thoracic Society (ATS) [32] and Centers for Disease Control and Prevention (CDC) [33] can be viewed on the Internet at http://www.thoracic.org/statements/. They have been endorsed by the Infectious Diseases Society of America and the American Academy of Pediatrics.

ATS/CDC guidelines emphasize that administering a TST implies a commitment to administer therapy if latent infection is diagnosed. Over the last half of the 20th century, there has also been an evolution in the definition of what degree of reactivity is accepted as diagnostic of latent infection. In 1952, Palmer and Bates [34] established that among hospitalized patients with (predominantly pulmonary) tuberculosis, reactions averaged a mean (±SD) of 18.3 ± 5.3 mm. Evidence that similar reactivity occurred in those with active and latent tuberculosis was obtained in a 1962 study of Alaskan Inuit, among whom tuberculosis was prevalent but exposure to nontuberculous mycobacteria was absent [35]. The data showed a bimodal distribution of reactions with peaks at 0 and 18 mm and few reactions between 2 and 5 mm. Because cross-reactivity was not an issue, the authors concluded that Inuit with ≥5 mm of induration should be considered to be infected with M. tuberculosis.

Between 1958 and 1964, a total of 643,694 male navy recruits were enrolled in a large study conducted by the US Public Health Service [36]. This included persons sensitized to nontuberculous mycobacteria, M. tuberculosis, or both. In contrast to the sharply separated bimodal distribution seen in the Inuit, there was much greater overlap. By subtracting the imputed distribution of the M. tuberculosis–infected, it was possible to infer the shape of the distribution of those reactions due to nontuberculous mycobacteria. At ≥5 mm, virtually all reactivity could be so ascribed (“negative”). Above 10 mm, almost all reactivity could be ascribed to M. tuberculosis (“positive”). Between 6 mm and 10 mm, however, reactivity might well be due to either (“indeterminate”). Figure 1, redrawn from [37], contrasts the distributions of PPD reactions seen among the recruits with that seen among the Inuit.

Over the last 2 decades, a Bayesian approach to the interpretation of TSTs has replaced rigid criteria for positive, negative, and indeterminate tests. Under ATS/CDC guidelines, the criterion for a positive test (i.e., a test indicative of tuberculous infection) is determined by assigning the testee to 1 of 3 pretest risk levels, on the basis of both the prevalence of infection in similar testees and an assessment of risk of the progression of latent to active infection. Reactions of ≥5 mm of induration are considered positive for those at highest risk, ≥10 mm for those at intermediate risk, and ≥15 mm for those at low risk and who would not be otherwise targeted for a TST (table 1).

Rose et al. [38] calculated the sensitivity and specificity of the Mantoux test from the data provided by the navy recruits. Sensitivity was 1.00 at a 2-mm cutoff, falling to 0.59 at a 16-mm cutoff. Specificity was 1.00 at a 2-mm cutoff, falling to 0.95 at a 2-mm cutoff. The area under the receiver operating characteristic curves was 0.994–0.998, indicating that the test discriminated well between M. tuberculosis–infected and uninfected persons. Results of a second analysis of a different test population showed similar results and demonstrated that sensitivity and specificity did not depend on the prevalence of tuberculosis.

We used data from Rose et al. [38] to calculate, for each of

Figure 1. Frequency distribution of tuberculin purified protein derivative reactions observed in 643,694 navy recruits and 865 Inuit (redrawn from [37]).
the currently recommended cutoffs, the predictive value (or posttest probability) of a positive and a negative TST result as a function of the prevalence of tuberculous infection (or pretest probability). Figure 2 should be helpful in approximating the reduction in uncertainty afforded by a positive or negative test. To provide perspective, some relevant approximate estimates of prevalence include ∼5%–10% in the total US population, and ∼1% in US children entering school, ∼5%–10% in the total US population, and ∼25% among contacts of a new case of tuberculosis [23].

### ANERGY TESTING

Many studies attest that some with tuberculosis will not react or have only small reactions, especially those with disseminated or advanced tuberculosis, severe malnutrition, and immune deficiency due to HIV or immunosuppressive chemotherapy. To assess whether someone with suspected tuberculosis but a negative TST result was capable of mounting a cellular immune reaction, it became popular to test with a battery of ubiquitous antigens, such as streptokinase, Candida extract, Trichophyton extract, and mumps antigen. In addition, a multiple-puncture antigen test was developed by Merieux Diagnostics. The combination of a positive anergy panel result and a negative TST result for a patient with pulmonary disease consistent with tuberculosis was to be taken as evidence against the diagnosis [39]. Anergy testing was also advocated to assess the degree of immune suppression of persons with HIV infection.

Several problems attended the interpretation of anergy tests. Most of the antigens were actually marketed for other purposes, such as thrombolysis, immediate-type allergy testing, or desensitization therapy, and modified by the user for Mantoux testing. Skin reactions could be produced in most persons by suitably increasing the doses of the antigens, and there was no standardization as to what degree of reactivity to the anergy panel would predict tuberculin reactions of any given size. Despite this, anergy testing became popular in the last quarter of the 20th century and it was only in 1997 that CDC guidelines recommended that an anergy panel should not be done in conjunction with a TST to assess an HIV-infected person’s risk for tuberculous infection [40].

### SPECIAL POPULATIONS

**Recipients of BCG.** BCG vaccine has never been recommended for routine use in the United States. Aside from efficacy issues, the major argument against its use has been the desirability of preserving the utility of the TST by keeping the prevalence of positive tuberculin tests low.

Postvaccination BCG-induced tuberculin reactivity ranges from no induration to induration of 15 mm at the skin test site [41]. Twelve weeks after vaccination, >90% of BCG vaccinees will have developed tuberculin reactions of ≥10 mm. This reaction wanes over the subsequent decade [42, 43] but may be maintained by boosting from repetitive TSTs or if concomitant infection with *M. tuberculosis* occurs [44].

Current ATS/CDC recommendations are that BCG vaccinees receive a TST when warranted by increased risk for recent infection or medical condition. Under the guidelines, the cri-
teria for test positivity are similar to those for unvaccinated persons. Because the prevalence of boosting is higher among BCG recipients than among unvaccinated persons, we believe that it is prudent to use a 2-step test to evaluate those vaccinated >5 years before the TST.

**Infants and children.** As in adults, a Mantoux test with 5 TU of PPD is used for tuberculin skin tests in infants and children. Universal testing of children, including screening in school-based programs, has a low yield of positive results and a large proportion of false-positive results [45, 46], resulting in the inefficient use of health care resources. ATS/CDC guidelines advocate the replacement of universal screening with targeted skin testing of selected groups at increased risk. A child can receive a TST simultaneously with immunizations, including live virus vaccines. Prior BCG vaccination is not a contraindication to an indicated TST.

The interpretation of TST results in children is based on a Bayesian approach analogous to that in adults. ATS/CDC guidelines should be consulted for specific criteria.

**HIV-infected persons.** A TST reaction of ≥5 mm in persons who are infected with HIV is considered “positive” (table 1), but if the HIV-positive person is known to have been exposed to a case and there is any palpable reaction whatsoever, the possibility that the reaction represents tuberculous infection should be considered. With the success of effective antiretroviral therapy, repeat TSTs may be indicated for HIV-infected persons who were TST-negative on initial evaluation and whose general health has been restored.

**Recent immigrants from high-prevalence countries.** In 1986, there were 4925 cases among foreign-born persons in the United States (22% of total) and in 1997 there were 7702 (39% of total). ATS/CDC guidelines are that those who immigrated to the US from high prevalence countries within the 5 years before assessment be considered TST-positive if the reaction is ≥10 mm.

**CONCLUSION**

Over the last century, the decline in tuberculosis and the evolution of tuberculin testing have resulted in the creation of a very useful test for latent *M. tuberculosis* infection. Eradication of tuberculosis from the United States will require identification of such cases and the administration of effective therapy to prevent reactivation.

**References**

37. Reichman LB. Tuberculin skin testing. the state of the art. Chest 1979; 76:764–70.