Comparable Specificity of 2 Commercial Tuberculin Reagents in Persons at Low Risk for Tuberculous Infection

Margarita E. Villarino, MD, MPH
William Burman, MD
Yong-Chen Wang, PhD
Linda Lundergan, MD, MPH
Antonino Catanzaro, MD
Naomi Bock, MD, MS
Crystal Jones, MD
Charles Nolan, MD

THE DIAGNOSIS OF LATENT TUBERCULOSIS infection is the basis of preventive therapy and a key indicator of tuberculosis transmission. The tuberculin skin test, which is the intradermal injection of a purified protein derivative (PPD) from broth culture of Mycobacterium tuberculosis, remains the only validated method for diagnosing latent tuberculosis. Parkdale Pharmaceuticals, Rochester, Mich (Aplisol), and Pasteur Mérieux Connaught USA, Swiftwater, Pa (Tubersol), are the 2 companies that manufacture tuberculin in the United States. Despite regulations for standardization of tuberculin manufacturing and testing, clusters of suspected false-positive results involving both products have been reported. We performed a randomized, double-blind study of the 2 commercial reagents and PPD-S1 (the standard) in 2 populations: (1) persons at low risk for latent tuberculosis infection and (2) patients with culture-positive tuberculosis.

METHODS

Subjects were enrolled in Denver, Colo; Marion County, Indianapolis, Ind; Atlanta, Ga; San Diego, Calif; Seattle, Wash; and Tucson, Ariz. Eligibility criteria included no risk factors for tuberculosis exposure (by questionnaire, available on request), no prior BCG immunization, no known immunodeficiency, age 18 to 50 years, and birth in the United States or Canada. To confirm the immunogenicity of the tuberculin skin test reagents used, we also studied patients with culture-positive tuberculosis diagnosed within 5 years and a favorable clinical response to 2 months or more of therapy. All participants gave written informed consent.

Skin test placement and reading were performed by experienced personnel using a standard protocol. The reagents used were Tubersol (lot numbers 2443-11 and 2458-11), Aplisol (lot numbers 01206p and 00417p), PPD-S1, and PPD-S2. All reagents were injected using 1-mL insulin syringes.

Context One or both commercial tuberculin skin test reagents (Aplisol and Tubersol) may have a high rate of false-positive reactions.

Objective To compare the reaction size and specificity of skin testing with Aplisol, Tubersol, and the standard purified protein derivative (PPD-S1).

Design Double-blind trial, conducted between May 14, 1997, and October 28, 1997, in which each individual received 4 tuberculin skin reagents at sites assigned at random.

Setting Health departments and universities in 6 US cities.

Participants A total of 1555 persons at low risk of latent tuberculosis infection.

Intervention Simultaneous skin tests with Aplisol, Tubersol, PPD-S1, and either a second PPD-S1 or PPD-S2 (a proposed new standard).

Main Outcome Measure Reaction size at each injection site measured by 2 investigators blinded to type of reagent.

Results Aplisol produced slightly larger reactions than Tubersol, but this difference did not significantly change skin test interpretation. The mean \(\pm SD\) reaction sizes were 3.4 \(\pm\) 4.2 mm with Aplisol, 2.1 \(\pm\) 3.2 mm with Tubersol, and 2.5 \(\pm\) 3.6 mm with PPD-S1. Assuming that all participants were uninfected and using a 10-mm cutoff, the specificities of the tests were high: Aplisol, 98.2%; Tubersol, 99.2%; and PPD-S1, 98.9%. Significant variability was not detected in interobserver, host, and lot-to-lot reagent comparisons.

Conclusion Using a cutoff of at least 10 mm, testing with 3 different PPD reagents resulted in similar numbers of uninfected persons being correctly classified.
SKIN TESTING WITH COMMERCIAL TUBERCULIN REAGENTS

**Table 1. Demographic Characteristics of Subjects at Low Risk for Latent Tuberculous Infection Who Were Included in the Analysis**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No. (%) of Subjects (N = 1555)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (range), y</td>
<td>26 (18-50)</td>
</tr>
<tr>
<td>Men</td>
<td>590 (38)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1069 (69)</td>
</tr>
<tr>
<td>Black</td>
<td>209 (13)</td>
</tr>
<tr>
<td>Hispanic*</td>
<td>180 (12)</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>50 (3)</td>
</tr>
<tr>
<td>American Indian/Alaska native</td>
<td>19 (1)</td>
</tr>
<tr>
<td>Unspecified</td>
<td>28 (2)</td>
</tr>
<tr>
<td>Place of birth</td>
<td></td>
</tr>
<tr>
<td>West§</td>
<td>717 (46)</td>
</tr>
<tr>
<td>Central</td>
<td>540 (35)</td>
</tr>
<tr>
<td>East§</td>
<td>298 (19)</td>
</tr>
<tr>
<td>Student</td>
<td>760 (49)</td>
</tr>
</tbody>
</table>

*Includes all races.
‡Includes North Dakota, South Dakota, Minnesota, Wisconsin, Michigan, Nebraska, Kansas, Iowa, Missouri, Oklahoma, Arkansas, Illinois, Indiana, Ohio, Tennessee, Kentucky, Mississippi, Alabama, Texas, and Louisiana.

lin syringes (Becton Dickinson, Franklin Lakes, NJ), and participants returned to the study site for reading of test 48 to 72 hours later. Randomization lists were prepared for each of the 6 study sites using randomized blocks of antigen sequences for groups of either 3, 6, or 9 patients. Sequences were randomized by antigen and injection site. Separate randomization schedules were configured for the low risk and the tuberculosis-patient study groups. Three fourths of subjects were randomized to receive Aplisol, Tubersol, PPD-S1, and PPD-S2; one fourth received Aplisol, Tubersol, and 2 injections of PPD-S1. Injections were placed on 2 sites of the flexor surface of each forearm, 5 and 10 cm below the elbow. The 2 investigators reading results, who were blinded to the identity of the test reagent and the other person’s readings, recorded the reactions in millimeters of induration in the transverse diameter.

If the false-positivity rate of tuberculin skin testing is 4%, detecting a 2% difference between false-positive rates with 80% power and 95% certainty requires a sample size of 1146. To account for losses to follow-up, we planned to enroll 1500 low-risk participants. We evaluated 3 potential sources of variability. First, we assessed interobserver variability, the difference between 2 independent readings of the same skin test, by grouping the results into 3 categories (0-4 mm, 5-9 mm, and ≥10 mm) and evaluating the paired agreement with the $k$ statistic. Second, we assessed host variability, the difference between 2 PPD-S1 tests in the same participant, by comparing differences in skin test interpretation. Third, we assessed the variability between different reagents and between different lots of the same reagent. For these comparisons we evaluated the difference between reaction-size means in subjects who had at least 1 skin test reading greater than 0 mm, using nonparametric analyses of variance (Friedman test if repeated measures and Wilcoxon signed rank tests if nonrepeated measures), and pairwise comparisons. We also compared (using Wilcoxon signed rank tests adjusted for multiple comparisons) the mean reaction sizes by study site, age, sex, and race. The results of testing with PPD-S2 will be presented separately.

We calculated test specificity in 2 ways. First, all low-risk subjects were assumed to be uninfected; therefore, specificity equals 1 minus the rate of reactions measuring 10 or 15 mm or more (false-positive reactions). Second, subjects having reactions of 10 mm or more to PPD-S1 were assumed to be infected and eliminated from the specificity calculations. Among patients with culture-positive tuberculosis, we compared the mean skin test reaction sizes and the rate of false-negative reactions (<10 mm).

**RESULTS**

**Study Population**

Between May 14, 1997, and October 28, 1997, we enrolled 1596 low-risk subjects, 41 of whom were excluded from analysis for various reasons, and 99 persons with histories of culture-positive tuberculosis. Demographic characteristics of the remaining 1555 low-risk participants are shown in Table 1. There were no clinically significant adverse reactions to skin testing. Of the 1555 low-risk subjects, 366 (23.5%) received 3 unique antigens and 2 injections of PPD-S1; the rest received 4 unique antigens. Of the 99 patients with TB, 30 (30.3%) received 3 unique antigens and 2 injections of PPD-S1; the rest received 4 unique antigens.

**Interobserver and Host Variability**

Of the 1555 low-risk subjects, 127 (8.2%) had a PPD-S1 reading greater than 0 mm. Among these 127 subjects, the differences between the 2 readers were small in most cases and only equaled or exceeded 5 mm in 18 instances (14%). There was a 69% probability ($k$ statistic × 100) that the agreement between 2 readers of the same PPD-S1 test was not by chance alone. Thirty-six (9.8%) of 366 persons who received 2 PPD-S1 injections had at least 1 of these tests read as greater than 0 mm. Among these 36 subjects, the differences between the 2 PPD-S1 tests equaled or exceeded 5 mm in 4 cases (11%). Using a 10-mm cutoff, the difference in the readings would result in a difference in skin test interpretation in only 2 subjects (0.5%).

**Variability Among Reagents**

There were no significant differences between the mean reaction sizes of the 2 lots of each commercial reagent (means for the Aplisol lots: 3.43 and 3.43 mm, $P = .95$; means for the Tubersol lots: 2.50 and 1.71 mm, $P = .19$). However, there were differences between the mean (±SD) reaction sizes of Aplisol (3.4 ± 4.2 mm), Tubersol (2.1 ± 3.2 mm), and PPD-S1 (2.5 ± 3.6 mm) ($P = .001$ by analysis of variance and $P < .05$ for all 3 pairwise comparisons). Mean reaction sizes were significantly larger at the San Diego site compared with other sites; however, most of these reactions were small (263 [91%] of 288 ranged from 1-9 mm). Excluding participants from San Diego does not significantly change the results of this analysis (data not shown). There were no significant differences in mean reaction sizes by age, sex, or race.

**Test Specificity Results**

The first scenario assumed that all reactions greater than the cutoff value (10 or 15 mm) were false-positive. At either cutoff value, with any of the 3 skin test re-
agents, the number of persons with positive reactions was small and the corresponding specificities were all greater than 98% (Table 2). At the 10-mm cutoff there was a significant difference in specificity between Aplisol and Tubersol, but neither commercial reagent differed from PPD-S1. In the second scenario, all subjects with reactions of 10 mm or greater to PPD-S1 were defined as latently infected and eliminated from the analysis (Table 2). In this scenario, there were no significant differences between the specificities of Aplisol and Tubersol using either a 10- or 15-mm cutoff.

### Immunogenicity of Study Antigens

The mean (±SD) reaction sizes in the persons with culture-positive tuberculosis were 16.3 ± 5.6 mm for Aplisol, 14.9 ± 6.0 mm for Tubersol, and 16.2 ± 6.4 mm for PPD-S1 (P = .006 by analysis of variance). In pairwise comparison, the differences between Tubersol and PPD-S1 (P = .008) and between Tubersol and Aplisol (P < .001) were statistically significant, whereas there was no difference between Aplisol and PPD-S1 (P = .84). Thirteen persons (13%) had false-negative test results with either PPD-S1 or Tubersol, and 11 (11%) had false-negative test results with Aplisol.

### Comment

This study demonstrates that the results of skin testing with the 2 commercial reagents, Aplisol and Tubersol, are quite comparable with that of the standard tuberculin preparation, PPD-S1. Tubersol produced slightly smaller reactions, and Aplisol slightly larger reactions, than did PPD-S1, but these differences in reaction sizes did not result in significant differences in skin test interpretation; the specificities of both commercial reagents were high (>98%) and similar to that of PPD-S1.

We explored several potential sources of variability in tuberculin skin testing. Interobserver agreement was similar than that previously reported. Host variability was quite low; differences between the reaction sizes of 2 simultaneous PPD-S1 tests would have resulted in a discordance in skin test interpretation in only 0.5% of those tested. We detected an association between reaction size and enrollment in San Diego. The readers at this site had previously evaluated skin tests (other than tuberculin) having expected reaction sizes of less than 10 mm. We suspect that readers from other sites were not trained to detect small skin test reactions, leading to a tendency to record such reactions as 0 mm of induration.

Clusters of unexpected positive tuberculin skin test results have been previously reported, often in groups of low-risk persons tested with Aplisol that, on subsequent testing with Tubersol, were believed to be clusters of false-positive reactions. None of these reports involved testing with 2 commercial products simultaneously and thus cannot exclude the possibility of false-negative reactions associated with Tubersol, or another kind of error associated with tuberculin skin tests not performed under the same conditions. Our study included simultaneous testing of both commercial reagents, as well as the standard tuberculin, in a large sample of well-characterized subjects. A limitation of our study is that we only evaluated 2 lots of commercial tuberculin manufactured in the same period. It is possible that variations in manufacturing processes over time may have produced some of the reported differences in false-positive rates.

Skin test variation related to human factors can be controlled only to a finite degree. In clinical practice, these factors cannot be eliminated completely and should always be recognized as potential sources of false-positive tuberculin skin test results. However, our study demonstrates that both Aplisol and Tubersol will correctly classify comparable numbers of persons not infected with M tuberculosis and that the choice of product used for skin testing has little effect on test performance.

### Funding/Support

This work was supported by the US Public Health Service, Centers for Disease Control and Prevention, Atlanta, Ga.

### References


©1999 American Medical Association. All rights reserved.