Sensitivities of the Commercially Available Tuberculin Skin Test Reagents in Persons with Recent Tuberculosis

Sir—The tuberculin skin test is the only diagnostic tool available for identifying tuberculous infection and is the basis for selecting individuals who need to receive antituberculous prophylactic therapy and for monitoring the prevalence or incidence of tuberculous infection in various populations. Therefore, the selection of a tuberculin reagent, its administration, and the reading of the test itself should be done with great care. Recently, Duchin et al. [1] compared the two commercially available tuberculin skin test reagents in persons with recent tuberculosis. The editorial on this study by Sbarbaro and Iseman [2] emphasized the limitations of the results of the study; these limitations were due to the selection of patients with previously culture-proven tuberculosis. We would like to address some methodological issues that may further compromise the results of this study.

The analyses that are presented by Duchin et al. [1] are mainly descriptive. The authors should have provided the readers with more conventional global agreement statistics, namely intraclass correlation coefficients [3]. Moreover, it would have been more informative to perform an analysis such as the one recommended by Bland and Altman [4]. This method is based on the construction of a residual-like plot of the difference between the measurements against their mean, which allows the detection of important lack of individual agreement and the unmasking of a relationship between the differences and the means (i.e., the fact that the discrepancies between the two reagents possibly depend on the diameter of the induration). Such phenomena could not be highlighted on the basis of the presentation of Duchin et al. [1]. Indeed, the computation from the data in figure 2 gives a rather wide 95% confidence interval of −6.8 mm to +8 mm around the mean difference of 0.6 mm. This means that 5% of the time, the difference between two readings could be at least 6.8 mm less to 8 mm more, evidencing low agreement.

In addition, we believe that the study by Duchin et al. [1], although carefully designed to ensure blinding of the administration of either reagent, could not avoid a recall and unconscious bias during the reading of the test itself, which may have favored the reliability. Indeed, to avoid interobserver lack of reliability, all readings were performed by the same investigator. As a consequence, for a given patient, knowledge of the first measurement may have influenced the result of the second reading in the contralateral forearm. Furthermore, selection of patients with culture-proven tuberculosis and awareness of such patients’ selection for the study by the investigator who performed the measurements may have introduced a similar bias and may explain why all results for the two reagents were concordant. Such biases could be suspected by the width of the computed confidence interval of the differences. In a recent study, we demonstrated that reading of tuberculin skin tests may in fact frequently result in misclassification, especially when the results of the measure are close to the cutoff value that separates negative from positive results [5].

Pending the availability of new tests with better diagnostic capability, we must continue to rely on the tuberculin skin test to diagnose tuberculous infection. However, great caution should be exercised in the administration and reading of the test; in addition, the reagent should be selected carefully as Duchin et al. [1] deservedly addressed.

Jacques Pouchot, Anne Grasland, Joël Coste, and Philippe Vinceneux
Service de Médecine Interne V, Hôpital Louis Mourier, Colombes; and Département de Biostatistiques, Hôpital Cochin, Paris, France

References

Reply

Sir—We thank Pouchot et al. for their thoughtful comments regarding our study demonstrating equivalence of Aplisol (Parke-Davis, Morris Plains, NJ) and Tubersol (Connaught, Swiftwater, PA) [1]. Although the study was potentially limited in statistical power by the relatively small sample size, we found reaction-size distributions and median tuberculin skin test results that were indistinguishable for the reagents. Figure 1 shows a plot of the empirical cumulative distributions of skin test–reaction sizes, which demonstrates that the distributions were very close. The Kolmogorov-Smirnov test of the equivalence of these distributions yields a P value of .41, indicating no evidence of a difference. On the basis of this finding, one may regard these antigens as indistinguishable when administered to this study population.

The variability inherent in tuberculin skin testing has been demonstrated previously [2–4]. We confirmed this substantial and incompletely explained variability. Table 1 lists the prevalence-adjusted measure of agreement between measurements (i.e., Kappa). After adjusting for chance agreements in measurements of the reaction sizes for Aplisol and Tubersol, there appears to be