proach was abandoned, which left an eclectic collection of methods and results in use, and Bellete et al. [1] used TST as the standard. No attempt was made to examine which test was more accurate, and there was no analysis of factors such as BCG response or exposure risk. Overall, too few data are provided on the Ethiopian cohort to enable a useful comparison of the TST and the IGRA. The Ethiopian sample includes an HIV-positive group, in whom the agreement between the TST and the IGRA results is less and IFN-γ levels are lower than they are in HIV-negative subjects. Bellete et al. [1] quotes articles by Kimura et al. [9] and Converse et al. [10], both of which also report lower IFN-γ levels in this group. However, the article by Bellete et al. [1] fails to mention the major finding of Converse et al. [10]—namely, that measurement of IFN-γ release appeared to be a more sensitive assay for the HIV-positive group.

To improve on the TST, a new test must have discordant results. The article by Mazurek et al. [3] did not make the mistake of using the TST as the gold standard and studied a population of sufficient size to draw solid statistical conclusions and find possible reasons for IGRA/TST discordance. Both Bellete and colleagues and Nadal could learn from this approach and avoid using the inexact TST or people with previously treated TB as their standards in the future.

James S. Rothel and Anthony J. Radford
Cellestis Limited, Carnegie, Victoria, Australia

References


Clinical Infectious Diseases 2003:36:1206–7 © 2003 by the Infectious Diseases Society of America. All rights reserved. 1058-4838/2003/3606-0033$15.00

Detection of Mycobacterium tuberculosis Infection by Whole- Blood Interferon-γ Release Assay

Sir—We read with interest the article by Bellete et al. [1] comparing the performance of the tuberculin skin test (TST) with that of the whole blood IFN-γ release assay (IGRA; QuantiFERON-TB; Cellestis), among a subset (14%) of individuals from our larger multicenter study [2] and among subjects in Ethiopia. The authors selectively report a subset of data from a larger study that was specifically designed to measure agreement between TST and IGRA results with a margin of error of ±2.6% and to identify factors associated with differences between results of these tests for a diverse population. By limiting their analysis to this subset, Bellete et al. [1] have presented results with greater margins of error than, and conclusions that conflict with, those of the previously published multicenter study [2].

The methods and assumptions described by Bellete et al. [1] differ from those used in the multicenter study [2]. For example, 34 (65%) of the 52 subjects classified as being at low risk for tuberculosis infection by Bellete et al. [1] were classified as high risk in the multicenter study because they reported working or living in a correctional facility, homeless shelter, health care facility, or country where tuberculosis is prevalent [2]. In their report, Bellete et al. [1] did not use these risk factors for tuberculosis infection and, by classifying these people as being at low risk, arrived at a lower estimate of the specificity of the IGRA. In addition, they used different breakpoints in interpreting TST results. Indurations with diameters of ≥5 mm but <10 mm were not considered a positive result among subjects from the Baltimore cohort suspected of being infected with tuberculosis or with culture-confirmed tuberculosis; however, indurations of ≥5 mm were interpreted as a positive result among the Ethiopian subjects regardless of risk classification. It is interesting that, according to figure 2 in Bellete et al. [1], a relatively large number of Ethiopian subjects had TST reactions with diameters of ≥5 mm but <10 mm. The IGRA results appear not to have been interpreted according to the manufacturer’s specifications, in that IFN-γ responses to human PPD of ≥15% (as defined in the Methods section in Bellete et al. [1]) were considered indicative of tuberculosis infection regardless of the response to avian PPD. Sixteen percent of the Baltimore subjects with IFN-γ responses of ≥15% to human PPD had significantly greater responses to avian PPD (i.e., the difference between responses to human and avian PPD was greater than −10%). In addition, reporting “agree-
ment beyond chance” in the abstract (i.e., $\kappa$ values for the comparisons of 68% among subjects from Baltimore and 35% among those from Ethiopia) without reporting the actual agreement (i.e., 80% and 68%) can be misleading [1].

We observed another inconsistency between the assumptions made in the article by Bellete et al. [1] and those made in the multicenter study [2]. In defining sensitivity, Bellete et al. [1] deemed the test results for 21 individuals who had been treated for tuberculosis disease as “true-positives” and considered these individuals to be immunologically representative of people with tuberculosis infection. In their discussion, however, they note that Hirsch et al. [3] demonstrated that IFN-\(\gamma\) responses can be depressed in individuals treated for active tuberculosis, a finding recently supported by Pathan et al. [4]. In contrast, it is well documented that the proportion of people with a positive TST response increases rapidly following the initiation of treatment [5]. Data from our multicenter study [2] shows that the percentage of people with a positive IGRA response decrease in proportion to the length of treatment, whereas the percentage of people with a positive TST response increases. Thus, people previously treated for tuberculosis infection do not appear to be immunologically representative of people with latent tuberculosis infection or active untreated tuberculosis disease and are not good surrogates for comparing the sensitivity of tests for tuberculosis infection.

The conclusion by Bellete et al. [1] that the test results have poor reproducibility is based on a small nonrepresentative sub-sample from our larger multicenter study and is not supported by the data. Subjects in our multi-center study were asked to return to undergo an additional IGRA and TST if follow-up testing was required because of recent exposure to tuberculosis or if the results of the initial tests were discordant [2]. However, we did not report on the reproducibility of the tests because results for both IGRA and TST were available for only 18% of eligible subjects, methods did not account for the effect of boosting by the initial TST, few contacts were enrolled, and no other group without a bias due to initial discordance had undergone additional testing. With one exception, changes in the interpretations of both the TST and the IGRA results for the Baltimore subjects were consistent with boosting by the prior TST. Moreover, the interpretation of the IGRA result for the subject who was the exception would not have changed if the test had been interpreted in accordance with the recommendations approved by the FDA [6]. As a laboratory-based assay with multiple controls, IGRA is likely to have highly reproducible results, and this is supported by data from reproducibility studies submitted to the FDA [6].

When compared to the multicenter study [2], the report by Bellete et al. [1] illustrates the power of multicenter studies in controlling for site-specific variations and local bias. Compared to the parent study [2], a larger percentage of Baltimore subjects had factors that put them at risk for HIV infection (27% vs. 20%) and a larger percentage of subjects enrolled in Baltimore were excluded due to collection of insufficient blood, incubator failure, failure to place or read the TST as specified in the protocol, or inability to document mycobacterial culture results (30% vs. 9%). The prevalence of digit preference also varied between the study sites, and this contributed to differences in the results reported by the 2 studies.

In contrast to Bellete et al. [1], we found good agreement between TST and IGRA results (83.1% agreement overall; $\kappa = 0.60$) [2]. Additional studies are needed to clarify the potential role and usefulness of IGRA. Although alternative methods of interpreting these tests may be investigated, it is critical that any comparison between the 2 tests use results obtained by agreed-upon methods.

References


G. H. Mazurek 1,2 P. A. LoBue 1,2 C. L. Daley 4 J. Bernardo 5 A. A. Lardizabal 6 and M. F. Iademarco 1,2

1 Division of Tuberculosis Elimination, Centers for Disease Control and Prevention and 2Pulmonary and Critical Care Division, Department of Medicine, Emory University School of Medicine, Atlanta, Georgia; 3Pulmonary & Critical Care Division, University of California at San Diego, and Pulmonary & Critical Care Division, University of California at San Francisco, San Francisco, California; 4Pulmonary Center, Boston University School of Medicine, Boston, Massachusetts; and 5Tuberculosis Center, New Jersey Medical School National Tuberculosis Center, Newark, New Jersey