Comparison of Tuberculosis Tests: Finding Truth or Confirming Prejudice?

Str—Bellete et al. [1] and the accompanying editorial by Nadal [2] report on the whole blood interferon-γ release assay (IGRA; QuantiFERON-TB; Cellestis) and its performance relative to the tuberculosis skin test (TST), but they appear selective in using methods and data that favor the TST. Although Nadal [2] refers in passing to problems with the TST, both articles consider differences between the IGRA and the TST as a negative factor, tacitly accepting that the TST is the gold standard and overlooking facts to the contrary. We suggest their conclusions are unsound and biased in favor of the TST; as developers of the IGRA that they used, we admit to having a commercial bias, but at least you know where we stand.

Selectivity is inherent in the article by Bellete et al. [1], as it includes data from a small subset (<15%) of subjects in a multicenter trial already reported by Mazurek et al. [3]. Nadal’s editorial [2] infers that we accept the conclusions drawn from the data for the 175-subject Baltimore subset, not those from the data for the total group of 1234 subjects studied by Mazurek et al. [3]—a statistical absurdity. Although we take issue with many aspects of the article by Bellete et al. [1], mostly because of its overinterpretation of a small and biased sample, it is the uncritical acceptance of the TST in Nadal [2] that we consider to fall short of reviewing standards.

Nadal’s bias is clear when he contends that the single patient contact required for the IGRA is no advantage, as all subjects have to return for an additional visit to be told their IGRA result, even if the result is negative [2]. As is the case for most tests, receiving results via a telephone call can save money and considerable inconvenience for the majority of subjects whose test results are negative. In many instances, subjects do not return for a TST reading; still other subjects do not return within the 48–72 h window. In addition, there is often a considerable logistical cost involved in return visits. The IGRA allows medical staff to focus on individuals whose test results are positive and to devote time to ensuring patient compliance with anti-TB prophylaxis rather than to chasing subjects with negative TST results.

There was little discussion by Nadal [2] of the deeper scientific aspects of the article by Bellete et al. [1]; in particular, the highly debatable proposal that previously treated TB is a model of latent tuberculosis infection (LTBI). Although both TST and IFN-γ responses can be diminished in cases of untreated active TB, it is well known that patients with TB show rapid conversion to a positive TST result following treatment [4–7], which indicates a change in the patient’s immunological status. However, underlying IFN-γ responses often remain low both during and after treatment for TB [6, 7]—a fact acknowledged by Bellete et al. [1], who fail to contrast this with the IFN-γ responses of healthy LTBI subjects, which are usually high. Thus, for both the TST and IFN-γ assays, it is clearly inappropriate to test individuals with previously treated TB to determine the sensitivities of the assays when used to test subjects with either untreated active TB or LTBI.

Mazurek et al. [3] reported digit preference in reading the TST result (e.g., rounding measures of TST induration to the nearest multiple of 5 mm) at some sites. Nadal [2] sees this as an inadequate control for TST subjectivity, inferring that digit preference lowers the validity of the data reported by Mazurek et al. [3]. This gratuitous slight on Mazurek and his co-authors contains a damning criticism of the TST. Nadal [2] is suggesting that the Centers for Disease Control and Prevention and 5 of the most respected TB clinics in the United States cannot read TST results accurately. So what hope is there for the rest of us? An irony of this criticism is that the Baltimore site discussed by Bellete et al. [1] displayed significant digit preference in the multicenter study.

The conclusion by Bellete et al. [1] regarding IGRA reproducibility is made on the basis of a very small number of individuals (8) for whom additional testing was performed. This sample is inherently biased, as all of the subjects who were tested again had discordant TST/IGRA results when first tested. Both size and bias mean that neither the conclusion of Bellete et al. [1] nor the reiteration by Nadal [2] have statistical validity. Even ignoring this point, their conclusion is unreasonable given the data provided. Both of their articles ignore the possibility that a recent TST may have boosted or altered IGRA responses. The results of most of the additional IGRA tests were boosted by the prior TST, and all of the changes in the diagnostic interpretation of the IGRA (which was done using the US Food and Drug Administration [FDA]–approved method) accorded with boosting. Performing the IGRA soon after a TST is not recommended by the FDA or the manufacturer. Nadal [2] and Bellete et al. [1] made sweeping negative comments on the IGRA reproducibility, despite the fact that the IGRA was used against current recommendations in a biased sample of minimal amounts of data—data, moreover, that fail to support their contention. A fair reviewer should be aware of the product instructions and should refer to more comprehensive data on IGRA reproducibility, such as those available at the FDA Web site [8].

The specificity of the IGRA is reported as “high” in the FDA summary—information that was freely available to Nadal. Bellete et al. [1] estimated IGRA specificity on the basis of data for 52 people recruited at Baltimore into a low TB-risk group, of whom only 18 individuals were found by Mazurek et al. [3] to be truly at low risk. Using the FDA–approved cutoff for low risk individuals (not that used by Bellete et al. [1] or Mazurek et al. [3]), the specificity of the IGRA was 98% in >1500 low risk individuals tested [8].

The original Ethiopian study protocol recognized TST inadequacies and sought to use chest radiograph evidence of minimal TB to clarify test outcomes. This ap-
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-\(\gamma\) 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-\(\gamma\) 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-\(\gamma\) 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


Detection of Mycobacterium tuberculosis Infection by Whole-Blood Interferon-\(\gamma\) 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-\(\gamma\) 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 \(\pm 2.6\%\) 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 \(\geq 5\) 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 \(\geq 5\) 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 \(\geq 5\) mm but <10 mm. The IGRA results appear not to have been interpreted according to the manufacturer’s specifications, in that IFN-\(\gamma\) responses to human PPD of \(\geq 15\%\) (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-\(\gamma\) responses of \(\geq 15\%\) 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-