Author's response to reviews

Title: T-SPOT.TB responses during treatment of pulmonary tuberculosis

Authors:

   Samantha Ribeiro (sbrmicroimuno@yahoo.com.br)
   Kelly Dooley (kdooley1@jhmi.edu)
   Judith Hackman (hackmanj@jhmi.edu)
   Carla Loredo (loredoc@hucff.ufrj.br)
   Anne Efron (aefron@jhmi.edu)
   Richard E Chaisson (rchaiss@jhmi.edu)
   Marcus B Conde (marcusconde@hucff.ufrj.br)
   Neio Boechat (n_boechat@yahoo.com)
   Susan E Dorman (dsusan1@jhmi.edu)

Version: 2 Date: 15 December 2008

Author's response to reviews: see over
December 12, 2008

Melissa Norton, M.D.
Editor, BMC Infectious Diseases

Re: Manuscript 2139425866222112 - T-SPOT.TB responses during treatment of pulmonary tuberculosis

To the Editor:

Thank you very much for the thoughtful and detailed reviews of our manuscript 2139425866222112 entitled “T-SPOT.TB responses during treatment of pulmonary tuberculosis”. Attached (next page) for your consideration are our point-by-point responses to the reviewers’ comments, as well as a revised manuscript. Thank you very much for your consideration. Please feel free to contact me should you have any questions.

Sincerely,

Susan E. Dorman MD
Associate Professor of Medicine
From Dr. Charles von Reyn:

Major revisions

1. There is insufficient attention given in the methods, results, and discussion to the fact that the commercial test is designed to be tested on fresh cells and that the authors used frozen cells. The critical controls are missing – performing the test on fresh cells and then on a separate aliquot from the same sample after the maximum duration of freezing. It is conceivable, for example, that the baseline samples (frozen for the longest time) gave falsely low SPCs and thereby reduced the chances of detecting a fall in response with treatment. Either these controls need to be added to show the results are equivalent or the conclusion needs to be that the use of the T spot with frozen cells does not have clinical utility in this setting. That the test was performed with frozen cells should be mentioned in the title.

We agree that performing the test on frozen cells was not optimal, but was done because of high per-test-plate costs at the time when the assay was available only in a 96-well format (and not in 8-well strips, as at present). We have revised the “limitations” paragraph of the discussion to now feature more prominently this limitation, as follows: “First, we used frozen cells for cost reasons. Freeze-thaw processes may have diminished T cell reactivity, and this could account for the slightly lower-than-expected baseline sensitivity of the T-SPOT.TB assay. However, any impact of freezing-thawing on T cell reactivity would be expected to be uniform across samples and time points, and therefore not impact results related to SFC change over time. However, we cannot exclude that freezing duration could have negatively impacted T cell reactivity and thereby blunted observed differences over the course of TB treatment.”

In addition, we have revised the first sentence of the Abstract methods to clearly indicate that frozen cells were used for this study. This sentence now reads as follows: “Using the T-SPOT.TB assay and frozen peripheral blood mononuclear cells, we enumerated ESAT-6- and CFP-10-specific IFN-γ-producing T cells over time in pulmonary TB patients receiving directly observed treatment.”

2. The authors give eligibility but no consort diagram for the number of subjects screened to arrive at the final 58 subjects.

The T-SPOT.TB sub-study began after the initiation of the parent treatment trial. Once the T-SPOT.TB sub-study began, all subjects enrolled in the parent TB treatment study underwent T-SPOT.TB testing. The following information has been added to the Methods section under the Participants subheading: “Upon initiation of the T-SPOT.TB substudy, all subjects enrolled into the parent treatment trial were co-enrolled in the TSPOT-TB study until completion of the clinical treatment trial.”

With respect to how we arrived at the final 58 subjects, the following information is included in the 1st paragraph of the Results section: “During the T-SPOT.TB study period, 73 individuals were enrolled in the TB treatment trial. Among these 73 subjects, 11 had no baseline T-SPOT.TB result and 4 had only a baseline T-SPOT.TB result; these
15 subjects were excluded from the analysis. The remaining 58 patients had at least two blood samples collected, including a baseline sample, and were included in the analysis.

3. Were the isolates all drug sensitive? If not, T spot decreases would not be expected. Drug resistant cases would need to be analyzed separately. Patients with drug-resistant tuberculosis were excluded from the trial. This is noted in the first sentence of the Participants subsection of the Methods section.

Minor revisions:
4. Abstract: Should clarify this was a study in HIV-negative subjects (with one HIV positive in parentheses); it would have been better to exclude this one patient since immunosuppression may affect T spot results. If the patient is retained, the CD4 count should be given in the results.
Our a priori exclusion criteria did not include HIV seropositivity (that is, we did not exclude HIV-positive individuals), so we opted to retain this patient in the present analysis. However, the reviewer’s concerns regarding this patient’s potential immunosuppression and its effect on T spot results are well-taken. This individual had a CD4 count of 426 cells/mm$^3$ and robust TSPOT responses. The following information has been added to the 2nd paragraph of the Results section “ESAT-6 and CFP-10 spot forming counts for this patient were 14 and 99 at week 0; 4 and 77 at week 16; and 8 and 63 at week 24." The baseline CD4 count has been added to Table 1.

5. Background: “correlate more closely...’. With what?
This sentence has been modified to improve clarity. It now reads “…compared with the tuberculin skin test, the T-SPOT.TB and similar ELISPOT tests have been shown to correlate more closely with extent of M. tuberculosis exposure in contacts of TB patients.”

6. Methods: Why was there no T spot test at 8 weeks of treatment, which was the timing of the primary endpoint of the parent trial?
Funding support allowed for testing of TSPOT results at 3 time points. Given the results from previous studies, and kinetics of T cells generally, we were concerned that 8 weeks might be too early to detect TB treatment-related changes. A sentence has been added to the “study limitations” paragraph of the Discussion to point out that the lack of an 8-week sample limits our ability to evaluate T cell kinetics early in treatment.

7. Methods: As with freezing the authors should clarify if their method of defining a reactive test followed manufacturer’s instructions or represents their own method.
The manuscript has been modified to indicate that the definition of a reactive test followed then-current manufacturer’s instructions. The following information has been added to the Methods section “As per the manufacturer’s instructions, for ESAT-6 and for CFP-10, a test was scored as qualitatively reactive….”

8. Can the authors correlate quantitative microbiologic data (#AFB, #CFUs) with T spot responses? This would have been preferable to the radiologic surrogate of a presumed higher organism burden in cavitary TB.
The quantitative microbiologic data in this study were not of sufficient reliability to use as a marker of disease burden. Further, in this study, quantitative microbiologic data were collected as an ordered, categorical variable, which is analytically and statistically problematic. Instead we used cavitary disease status, which has been used widely as a covariate in TB clinical trials and is known to correlate with treatment response.

9. Was overall treatment success equal with the 2 drug regimens? If not, T spot results should be stratified by regimen.
Respectfully, we disagree. T-SPOT.TB results would need to be stratified by regimen only if the test’s ability to discriminate between those with treatment success (2-month sputum culture negativity) and those with ‘poor’ response to treatment (2-month sputum culture positivity) differs by regimen. In this study, there did not appear to be a difference in discriminatory power by regimen, but this study is underpowered to detect this interaction, especially given the high variability in results among individuals.

From Dr. Edward Graviss:
Major compulsory revisions:
1) The authors correctly attempt to exclude extra-pulmonary diseased individuals (more severe) from the data set, but it is unclear if the data set also includes individuals who are both pulm/extra-pulm infected. If not then please state. I assume there are and if so, then describe the frequency of this third “site-of-infection” strata between both treatment groups. Forty-one individuals had cavitary disease, but what proportion of these were in the pulm/extra-pulm strata?
Patients with concomitant pulmonary+extrapulmonary disease were eligible for the parent TB treatment trial and therefore for the T-SPOT.TB substudy. With Dr. Graviss’ query in mind, we have reviewed study forms for the parent treatment study, and, unfortunately, site(s) of extrapulmonary disease were not captured in the study case report forms and dataset.

However, to be study eligible, participants were required to have a Karnofsky performance status score of 70 or greater/better (score of 70 is “able to care for self”). Therefore the “sickest” TB patients would not have been included in the study. While this does not directly address the pulmonary/extrapulmonary issue as raised by Dr. Graviss, we believe it does address the question of “how sick, how severely ill, was the study population?” The following text has been added to the Methods section: “Additional inclusion criteria were documented HIV status (positive or negative), no history of prior TB treatment, and Karnofsky performance status score 70 or greater.”

A sentence has been added to the discussion indicating the need for further testing of this assay as a marker for treatment response in special populations, such as those with HIV or those with extrapulmonary disease.

2. It is not clear why the authors left a single HIV subject in the data set. The analysis should be run with the “one-less” data set.
Please see response to Dr. Charles von Reyn’s question number 4.
3. The FDA approved T-SPOT.TB assay uses a qualitative cut-point of \(\geq 8\) to determine a positive assay after subtracting out the baseline results. Does this make a difference in the results seen in this study? What proportion of the clinical specimens had results in the borderline area? The current qualitative definition of positive should be used. The T-SPOT.TB assay manufacturer’s instructions at the time of the study used a qualitative cut-point of \(\geq 6\) for a nil control of 0-5 SFCs and for a nil control of >5 SFCs, an antigen SFC count \(\geq 2\) times the nil control SFC count. The current FDA-approved T-SPOT.TB assay uses a slightly different cutoff of \(\geq 8\) spots, and spot counts of 5, 6, or 7 are considered borderline; in patients with 5, 6, or 7 SFC, retesting is recommended.

In our study, only three individuals had a baseline SFC count in this borderline range (each had a SFC of 4 in one well and 5 in the other). Because the manufacturer’s instructions at the time classified these individuals as negative, we did not retest to determine their status. Using the current, FDA-approved qualitative classification system, one of the three had negative time 16 and 24 tests, one of the three had an indeterminate time 16 and a positive week 24 test, and one of the three had positive week 16 and week 24 tests. Analysis using the new criteria does not substantively change our results or conclusions. From our perspective, it seems scientifically appropriate to interpret the test according to the manufacturer’s recommendations in place at that time. However, the discussion section has been modified as follows to alert the reader to the change in the assay qualitative test result instructions, to indicate that only 3 individuals would be re-classified as borderline, and that this did not change our study conclusions: “Of note, current manufacturer instructions in the U.S. FDA-approved T-SPOT.TB assay classify those with spot counts of 5, 6, or 7 as borderline and recommend retesting. Retesting was not done for three individuals in our study who had spot counts of 5 at baseline, as manufacturer instructions at the time of the study classified them as negative. Using the current classification system, two of the three had positive results upon completion of treatment while one had a negative result, resulting in no substantive change in our overall results or conclusions.”

Minor essential revisions

4. In the statistical analysis section the authors mention, “An independent correlation structure with robust standard errors was used for the GEE models, …” What were these robust standard errors, and how large were they?

When employing generalized estimating equations using longitudinal data, it is important to attempt to determine the correlation structure that best fits the data. Even after carefully choosing the best correlation structure and most appropriate model, to be conservative, it is recommended that one use a robust standard error in the analysis (this is included as part of the command line and is calculated by the statistical program). This helps minimize the risk of a Type I error. The robust variance estimator compensates for any broken model assumptions, the nonasymptotic nature of data, or the misspecification of correlation structure.

5. Several studies have shown that BMI is a better marker of severity than weight. Can BMI be gleaned from the data set or only weight?
We agree with the reviewer, and the manuscript has been revised to include BMI rather than weight as a marker of severity of disease.

6. In the discussion, the authors correctly assess that their results “indicate that the tests utility in evaluating or predicting treatment response in individual patients appears poor”, this seems appropriate, but the overall sample size is relatively small and the issue regarding pulm vs extra-pulm has not been addressed. Any new analysis will need to control for these other factors such as HIV status and or pulm/extra-pulm status. We agree that subsequent analyses that take into account HIV status and extrapulmonary status will be informative and interesting. A sentence has been added to the discussion describing the need for further testing of TSPOT-TB in HIV-seropositive individuals or patients with extrapulmonary disease to determine its utility in these populations.

From Dr. Cynthia Bin-Eng Chee
Major Compulsory Revisions
1. I have concerns regarding the validity of the T-SPOT assay results, as the samples were frozen for as long as 28 weeks (for time 0 samples) prior to the performance of the assays. Where were these assays performed and what were the transportation conditions like? Do the authors have evidence that T-cell reactivity is not diminished by prolonged storage? This could account for the low baseline sensitivity of the T-SPOT (72.4%, vs other published studies which report sensitivities of around 90%) in this cohort of mostly immunocompetent, non-elderly, adult pulmonary TB patients? What was the failure/indeterminate rate at each time point?

Please see our response to Dr. Charles von Reyn’s Question #1 which addresses the issue of frozen cells.

With respect to test failures, there were no assay failures; all results were interpretable based on manufacturer’s criteria. With respect to indeterminate rate, please see our response to Dr. Graviss’ Question #3, which addresses the issue of indeterminate results using the interpretation criteria per the current U.S. FDA-approved test kit.

Minor Essential Revisions:
2. How many sputum TB cultures samples per subject were performed at the 8-week time point?

Samples were collected once weekly for eight weeks, then monthly. This information has been added to the Methods section, Mycobacterial Culture subsection.

3. Analysis could also be performed for only subjects with T-SPOT samples available at all three time points to see if the conclusion is any different.

We re-analyzed the data, excluding patients with any missing data points (6 individuals). The results of the analysis did not differ, but the confidence intervals were overall slightly wider. Specifically, the relationship between week 8 culture results and spot counts and between disease severity and spot counts did not change.

4. Was there any difference in culture conversion at 8 weeks and T-SPOT results between the two treatment regimens?
There was no significant difference in T-SPOT.TB results by assigned treatment regimen.

From Dr. Keertan Dheda

1. The discussion needs to be expanded and improved. Specifically, there is selective quotation of the literature (paragraph 1 of the literature). There are several studies which show that responses do not decline or even become more robust during the course of treatment. There may be several putative reasons. Firstly there seems to be a dichotomy between high and low burden countries (bacillary load, strain (see de Jong BC et al, JID, 2006 and 2008), exposure to ESAT-6 producing homologues of NTMs e.g. M. marinum, HIV, malnutrition etc). There is also the issue of peptides vs. proteins as stimulants and incubation times. I have found (Dheda et al; J Infection, 2007; also used the T SPOT assay) a useful reference that summarizes the relevant studies and associated factors; these need to be discussed in more detail here.

We appreciate this comment, and the discussion section has been expanded and improved as per the reviewer’s suggestions.

2. These data also support the notion that the ELISPOT assay is detecting ‘infection’ rather than ‘exposure’ i.e. TB infection rather than memory of exposure (important as there is no gold standard for the detection of LTBI and it is controversial what this assay is really detecting). So, the data are useful in this respect.

This comment is well-taken, and we agree with this scientific concept. However, since all of our participants had active TB (and none had latent TB), we believe that our study really does not directly address this topic, which is outside of the scope of our study. Therefore we have not addressed this point in the manuscript text.

3. Frozen cells; a differential effect may have occurred with ‘low’ ESAT responders being more affected by freeze thaw. This should be stated; the viability of the recovered cells should be mentioned.

Please see responses to Dr. Charles von Reyn’s Question #1 which addresses the issue of frozen cells.

4. Excluded patient’s characteristics…..Were these similar to the cohort analyzed e.g. more or less extensive disease?

The T-SPOT.TB substudy began after the initiation of the parent treatment trial. Once the T-SPOT.TB substudy began, all subjects enrolled in the parent TB treatment study underwent T-SPOT.TB testing; none were excluded from the T-SPOT.TB study. The following information has been added to the Methods section under the Participants subheading: “Upon initiation of the T-SPOT.TB substudy, all subjects enrolled into the parent treatment trial were co-enrolled in the T-SPOT.TB study until completion of the treatment trial.”

With respect to how we arrived at the final 58 subjects, the following information is included in the 1st paragraph of the Results section: “During the T-SPOT.TB study period, 73 individuals were enrolled in the TB treatment trial. Among these 73 subjects, 11 had no baseline T-SPOT.TB result and 4 had only a baseline T-SPOT.TB result; these
15 subjects were excluded from the analysis. The remaining 58 patients had at least two blood samples collected, including a baseline sample, and were included in the analysis.

5. *HIV negative patients were evaluated and this may not be generalisable to high HIV prevalence settings.*

We agree, and the text has been modified to suggest that results in special populations, such as HIV-seropositive individuals or those with only extrapulmonary disease, may differ from the results in our study population and that these differences remain to be studied.

6. 70% were smear +, which is not reflective of clinical cohorts in high burden settings (40 to 50% smear + in Africa). How and who reported the cavitation on chest radiographs?

All participants in the treatment trial (and therefore in the T-SPOT.TB substudy) were smear positive, as this was a component of the study eligibility criteria for the parent study (please see Methods section, Participants subsection). In this study, as in a number of other contemporary phase 2 TB clinical trials, this was done to maximize the likelihood of TB growth in culture such that the primary outcome of the parent study (2 month culture status) could be observed and interpreted. We agree with Dr. Dheda that this does not necessarily reflect the population of TB suspects or TB patients seen in clinical practice in any setting. After careful consideration we have elected not to further address this in the manuscript text, given manuscript length considerations, the fact that we have included in the text the relevant eligibility criteria, and taking into consideration that (hopefully) most readers are familiar with the idea that a study population may not have exactly the same attributes as a population seen in routine practice.

With respect to cavitation, this was assessed by a study radiologist.

7. Did the investigators look at change in weight from 0 to 2 months and correlate this with the ELISPOT data?

ELISPOT data were not collected at 2 months, precluding us from performing this analysis.

8. It should probably be mentioned that only one biomarker was evaluated here (IFN-g). It is entirely possible that other biomarkers e.g. IP-10 or including a combination, may be more useful in predicting response to treatment.

We agree, and the following sentence has been added to the text (4th paragraph of the Discussion) to address this comment. “The utility of an RD1-based interferon gamma release ELISPOT assay in combination with other immunologic tests to evaluate treatment response also remains to be studied.”

9. The culture + group and CFP-10 shows a transient increase and then decrease; could this be a compartmentalization effect? It is possible that the response 1st increase and then decrease but the timepoints used here cannot detect this change. This should be mentioned.
The kinetics of T cell responses early in treatment could not be assessed because funding limitations allowed for only 3 collection timepoints, and we could not collect an 8-week sample. A sentence has been added to the Discussion section to describe this limitation.