Author's response to reviews

Title: Sensitivity of direct versus concentrated sputum smear microscopy in HIV-infected patients suspected of having pulmonary tuberculosis

Authors:

Adithya Cattamanchi (acattamanchi@medsfgh.ucsf.edu)
David W Dowdy (david.w.dowdy@gmail.com)
J Lucian Davis (lucian.davis@ucsf.edu)
William Worodria (worodria@yahoo.com)
Samuel Yoo (yoouga@yahoo.com)
Moses Joloba (moses.joloba@case.edu)
John Matovu (johnbaptist.matovu@yahoo.ca)
Philip C Hopewell (phopewell@medsfgh.ucsf.edu)
Laurence Huang (lhuang@php.ucsf.edu)

Version: 2 Date: 17 February 2009

Author's response to reviews: see over
B.1. The categories and numbers of excluded patients in the first paragraph of the results do not coincide with those in Figure 1.

The apparent inconsistency between the text and Figure 1 results from addition of 2 patients with culture not performed to 20 with sputum not concentrated, for a total of 22 patients. We have clarified this in the revised manuscript:

“Of 388 eligible patients, 39 (10%) were unable to provide an early-morning sputum specimen (unable or unwilling to spontaneously expectorate), 20 (5%) did not have a concentrated smear performed, 48 (12%) had a contaminated sputum culture, and 2 (1%) did not have culture performed despite the availability of concentrated smear, giving a final sample size of 279 HIV-infected TB suspects (Figure 1).”

B.2. Table 1. (a) Data on mortality are missing for some participants. (b) Does the “n=241” apply to the in-hospital mortality, the mortality at two months, or both? (c) The last two p-values are not on the correct line. (d) The abbreviation TB is mentioned in the legend but is not used in the table.

(a) 38 patients were lost-to-follow-up and therefore 2-month mortality data was unavailable in these patients. (b) The “n=241” applies to mortality at 2-months. To improve clarity, we added “n=279” following “in-hospital mortality”. In addition, (c) we fixed the two p-values that were not on the correct line and (d) deleted TB from the legend.

B.3. Table 2. The sub-title “culture-positive tuberculosis” is somewhat confusing because the calculation of the specificity is based on 109 participants with a negative culture. An alternative could be: “Diagnosis of tuberculosis based on culture”. The second sub-title could then be: “Diagnosis of tuberculosis based on culture or clinical criteria”.

We appreciate this suggestion and modified the sub-titles to:
Gold Standard: Sputum Culture
Gold Standard: Sputum Culture + Clinical Criteria

B.4. Table 3. Non-Hodgkin’s lymphoma (NHL) is mentioned among the abbreviations but does not appear in the table.

We removed NHL from the abbreviations list.

C.1. The authors state that sputum concentration increases the sensitivity of smear microscopy for the diagnosis of TB in general, and find that this statement is not true in their study population of HIV-infected individuals. It would be interesting to know how the authors interpret this. Could HIV infection explain the differences in performance of the techniques? How? Or do the authors think that problems with study design might have led to wrong conclusions and recommendations, for HIV-infected as well as for HIV-seronegative subjects?

We share the referee’s interest in these questions. Certainly, HIV infection could explain differences in performance of the techniques, as patients who are HIV-infected are known to have lower bacillary burdens in general, and there may be a sizable proportion of HIV/TB-
coinfected individuals who simply do not have sufficient numbers of AFB in their sputum to detect by smear, even after concentration. However, we did not have sufficient numbers of HIV-seronegative subjects to evaluate whether there was a difference in performance of smear microscopy based on HIV-status in our study. As the referee also suggests, an alternative hypothesis is that, as the design of similar studies improves over time, we might find sputum concentration to be less effective than previously thought, even in HIV-negative patients. We limit our discussion of these topics in the manuscript as they are largely speculative, and we have no data to support one explanation over any other. Thus, we prefer to maintain our present wording, which emphasizes (p. 10, first two full sentences) the strength of the evidence presented and suggests that sputum concentration may be less effective when access to high-quality direct microscopy is assured.

C.2. Was a sample size calculated before the study started? What were the assumptions (expected sensitivity, expected difference between the methods, expected number of exclusions)?

We did estimate sample size before the study started. We calculated that 262 patients would be needed to provide 90% power to detect a difference between concentrated and direct smear microscopy, assuming a 2-sided alpha of 0.05, phi of 0.5, 50% sensitivity of direct smear, and a matched-pair odds ratio of 2.25 (equivalent to 60% sensitivity of concentrated smear). We projected a 20% dropout rate, and thus set a target of $262/0.8 = 328$ initial patients evaluated. We ultimately evaluated 329 patients (Figure 1). We have included this information in the updated manuscript (p.6, final paragraph).

C.3. The proportion of positive culture results among TB suspects is very high in this study (61%), and the patients appear to have severe forms of TB with high mortality. How could this influence the findings? Could the advantage of the concentration technique in terms of sensitivity be higher among patients with mild or early forms of TB?

The referee raises an interesting possibility. We explored this issue in our comparison of the density of AFB in direct and concentrated smears. This analysis did not suggest that concentrated microscopy was more effective at picking up paucibacillary specimens. We did not highlight this aspect of our analysis in the discussion due to space limitations.

C.4. Results, p 7, third paragraph: “As seen in Table 2” could be replaced by “As shown in Table 2 (…)”.

The text was changed as suggested by the referee.

C.5. The authors use the word “sex” in text and tables, except for the second paragraph of the results where “gender” is used.

The text was modified. The word “gender” is now used throughout.
1. The target sample is not indicated making it difficult to establish how a sample of 388 patients was achieved.

Please see our response to Referee 1, item C.2, which we paste below for this Referee’s convenience. Our target sample was based on the number of concentrated smears performed (n = 329), but we included the total number of patients evaluated (n = 388) in the manuscript for increased clarity to readers.

Response to Referee 1, item C.2:
We did estimate sample size before the study started. We calculated that 262 patients would be needed to provide 90% power to detect a difference between concentrated and direct smear microscopy, assuming a 2-sided alpha of 0.05, phi of 0.5, 50% sensitivity of direct smear, and a matched-pair odds ratio of 2.25 (equivalent to 60% sensitivity of concentrated smear). We projected a 20% dropout rate, and thus set a target of 262/0.8 = 328 initial patients evaluated. We ultimately evaluated 329 patients (Figure 1). We have included this information in the updated manuscript (p.6, final paragraph).

2. Since the focus of the study was sputum samples from HIV infected persons, it would have been important to indicate the quality of the sputum.

This has been a concern of ours throughout the study as well. We collected data on sputum quality during the study, but did not initially report this data due to its subjective nature. However, given the concerns raised here and by Referee 3 (below), we have included this data in the results section of the revised manuscript:

“Regarding sputum quality, of the 279 specimens, 33 (12%) were described as salivary, 191 (68%) mucoid, 41 (15%) purulent, and 14 (5%) bloody. Exclusion of the salivary specimens, or restriction to mucoid specimens, reduced the sample size but did not materially affect results.”

3. The second limitation needs further clarification.

The text was modified as follows, to respond to this concern and to that of Referee 3 (item 7):

“Second, this study was conducted in a national reference laboratory using NALC-NaOH on early-morning specimens collected from a population of HIV-infected, hospitalized patients. Thus, our results may not fully generalize to other settings (e.g., peripheral laboratories, laboratories using alternative sputum processing methods, non-HIV populations, healthier outpatient populations). Estimates of diagnostic performance are known to vary between ambulatory and hospital settings. However, the choice of study population is less likely to impact a comparison between two diagnostic techniques. In addition, given the rigorous training required of microscopists at the Uganda NTRL, it is unlikely that laboratory inexperience explains the results of the present study.”

4. Explanations for the last limitation are irrelevant since findings do not warrant change in policy!
Based on the referee’s suggestion, the following sentence was deleted from the text:

“Lastly, we did not collect data on the operational implications (e.g., cost-effectiveness) of our findings, which would more fully inform policy decisions regarding the optimal algorithm for sputum smear microscopy in HIV-infected individuals.”

5. Since findings from this study are contrary to majority of previous studies, it may be important to indicate that this was a unique population and therefore, the Title should be revised to include ‘...hospitalised HIV-infected......’.

As discussed in Comment 3, we believe that the study setting – inpatient vs. outpatient – is a critical factor when determining point estimates of diagnostic performance. However, as both direct and concentrated smears were performed in the same study population, the comparison between the two tests should not be affected. We prefer the study title as written but will defer to the editor’s judgment regarding whether or not it should be modified as suggested by the referee.

REFEREE 3
1. In Methods, under “Study population”, could the authors precise if any consecutive eligible patients were enrolled?

Consecutive eligible patients were enrolled. The first sentence of the Methods section was modified as follows:

“Consecutive HIV-infected patients admitted to the medical wards of Mulago Hospital (Kampala, Uganda) between September 2007 and April 2008 for respiratory illness with cough of at least 2 weeks’ duration were eligible for the study.”

2. In Methods, under “Patient evaluation”, could the authors describe the criteria used by the treating ward physician to initiate TB treatment?

According to the two chest consultants at Mulago Hospital (W.W. and S.Y.), ward physicians generally follow WHO and Uganda NTP guidelines, which recommend initiation of TB treatment in the following settings:

a. sputum smear positive for AFBs
b. sputum smear negative for AFBs but with suggestive radiological findings (chest x-ray and/or abdominal ultrasound scan) AND no response to antibiotics AND without an alternative explanation for the disease.
c. neither of the above but evidence of TB from another site (e.g. TB meningitis, TB pleuritis, TB adenitis).

However, these criteria are largely designed for treating a healthier outpatient population; many hospitalized patients appropriately receive empiric therapy for TB at the time that clinical specimens are obtained and without waiting for response to antibiotics. Thus, for ethical reasons and in order to better replicate the actual clinical setting, we did not require ward physicians to follow specific criteria for initiating treatment. Therefore, in the manuscript, we state that treatment decisions were under the discretion of ward physicians and do not elaborate further, as we feel it would be misleading to suggest that strict treatment criteria were in place. This is a key reason that we selected a culture-based “gold standard” (i.e., one that does not
rely on clinical treatment decisions) as our primary outcome. We report the TB definition that uses clinical criteria only as a secondary analysis.

3. Under “Laboratory methods”, the level of performance of direct smear microscopy should be better documented with results of the internal and external quality controls during the study period. As clearly pointed out by the authors, the quality of the direct smear microscopy influence the results of the comparison with concentrated smear microscopy.

The study design compared both direct and concentrated smear to a “gold standard” of TB culture. Although repeated examinations of individual specimens were not performed during the study period (in order to better replicate “real-world” conditions), we do not expect that this would affect direct smear microscopy without also affecting concentrated smear microscopy. Furthermore, as mentioned in the text, since 2005, the Uganda NTRL has participated in a semi-annual external quality assurance program for smear microscopy administered by the World Health Organization and has passed all quality assurance assessments. Finally, we note that the sensitivity of direct smear in the present study was higher than in five other studies evaluating direct vs. concentrated smear in HIV-infected subjects (refs. 21-25 in the text), suggesting that false-negative results were likely minimized by the highly trained NTRL staff. Nevertheless, we share the referee’s concern and have added the following sentences to our limitations to this effect:

“Finally, in order to better replicate actual test conditions, internal quality assurance was not performed during the study period. Though we would not expect lack of reliability to differentially affect direct versus concentrated sputum smear results, we were unable to quantify inter-reader and intra-reader agreement.”

4. Under “Outcome definition”, could the authors precise the criteria use to define patients’ clinical improvement, mostly when the evaluation was conducted through telephone interview.

The criteria used to assess clinical improvement included resolution of symptoms (fevers, night sweats, cough), decreased fatigue, and weight gain. In the text, we summarized this as follows:

“…. we broadened the definition of TB to also include patients who improved clinically on empiric TB therapy, as documented by a study medical officer and a chest consultant (W.W. or S.Y.) between two and four months after hospital discharge.”

We acknowledge that clinical improvement is somewhat subjective. For this reason, we reported diagnostic estimates based on a clinical gold standard only as a secondary analysis. Given this subjectivity, we believe that including a more specific definition of clinical improvement provides readers with a false sense that such criteria are more objective, but we are happy to include such detail if the editor feels it appropriate.

5. Could the authors precise the reasons why 10% of patients did not submit a morning specimen, and why 6% had no concentrated smear. There is an inconsistence between the text and figure1 regarding the number of sputum non-concentrated.
The apparent inconsistency between the text and Figure 1 results from addition of 2 patients with culture not performed to 20 with sputum not concentrated, for a total of 22 patients. We have clarified this in the revised manuscript:

“Of 388 eligible patients, 39 (10%) were unable to provide an early-morning sputum specimen (unable or unwilling to spontaneously expectorate), 20 (5%) did not have a concentrated smear performed, 48 (12%) had a contaminated sputum culture, and 2 (1%) did not have culture performed despite the availability of concentrated smear, giving a final sample size of 279 HIV-infected TB suspects (Figure 1). The majority of exclusions (other than for contamination) occurred during a two-week period when the NTRL lacked sufficient staffing to process research samples in addition to their routine clinical work.”

6. Could the authors present the reasons of patients’ hospitalisation and the baseline clinical presentation? This could be added in table 1.

All patients were admitted for respiratory illness, including a cough of at least two weeks’ duration without suspected heart failure. We have clarified this in the first sentence of page 4, and based on the referee’s suggestion, we have also added the proportion of patients with fever, night sweats, weight loss, and hemoptysis to Table 1.

7. The authors should further address the study limitations discussed above in General Comments (study population, concentration method, morning specimen).

The referee raises important issues and each is addressed below:

Study population: We agree that “the level of TB suspicion and the clinical presentation of the study patients are likely to be different compared with outpatient TB suspects at a peripheral setting.” We also agree with the referee and Steingart et al (as referenced by the referee) that sputum concentration may perform differently in such settings. However, we note that Steingart et al argue (and we agree) that “greater increases in sensitivity after sputum processing were found in the research laboratories [as compared to peripheral centers].” Thus, our finding of no increase in sensitivity in a reference-laboratory setting is arguably more compelling, as one would expect sputum concentration to be less effective in peripheral laboratories. In our limitations section, we clearly state that our results may not fully generalize to peripheral laboratories serving healthier outpatient populations. Nonetheless, we feel that a study of hospitalized patients with advanced HIV-infection (who tend to be sicker and arguably in greater need of highly sensitive diagnostic testing) in a reference laboratory (where sputum concentration is arguably most effective) is still a valuable contribution to the existing body of literature.

Concentration method: We agree that there are many methods of sputum decontamination and concentration reported in the literature; unfortunately, in any study, one method must be chosen, and this method is unlikely to generalize to all centers. We chose the NALC-NaOH method as it is the standard method recommended by the Centers for Disease Control and is one of the methods recommended by the World Health Organization. Among the studies evaluated by Steingart et al (Table 1 in that publication), NALC-NaOH appeared to offer the greatest benefit in terms of added sensitivity, compared to the NaOH (Petroff) method mentioned by the referee, or the sodium hypochlorite (bleach) method also commonly studied. Thus, while we agree with the referee that NALC is not available in all centers, we find it difficult to argue that failure to use NALC would have increased the sensitivity of sputum concentration.
in the present study. Nevertheless, we modified our second limitation as follows to clarify that
the present study may not generalize to those centers where NALC-NaOH is unavailable:

“Second, this study was conducted in a national reference laboratory using NALC-NaOH on
early-morning specimens collected from a population of HIV-infected, hospitalized patients.
Thus, our results may not fully generalize to other settings (e.g., peripheral laboratories,
laboratories using alternative sputum processing methods, non-HIV populations, healthier
outpatient populations). Estimates of diagnostic performance are known to vary between
ambulatory and hospital settings. However, the choice of study population is less likely to
impact a comparison between two diagnostic techniques.”

Morning specimen: We agree that some studies have shown differences in diagnostic
performance of smear microscopy in morning versus spot sputum specimens. However, these
performance differences have been shown for both direct and concentrated smear microscopy.
We limited the study to the morning specimen as we wanted to compare the diagnostic
performance of direct and concentrated microscopy under ideal field conditions (i.e., morning
specimen, highly experienced microscopists at a national reference laboratory). Our intent was
not to compare a sputum collection strategy (i.e., spot-spot vs. spot-morning). Had sputum
concentration been shown to be better than direct microscopy, our next step would have been
to compare sputum collection strategies when using concentrated smear microscopy. Based on
our results, we believe comparing sputum collection strategies when using direct smear
microscopy is more relevant in peripheral field laboratories and such studies are currently
ongoing.

8. The authors should further discuss the reasons, which could explain that 19 culture positive
negative concentrated smears were positive with >10AFB/100 HPF using direct smear
microscopy.

We share the referee’s interest in this question. We suspect that bacilli may have been
destroyed during decontamination or sheared during centrifugation, as has been previously
suggested although never proven (e.g., Toman’s tuberculosis, WHO, 2004; Wilkinson D, Trans
reasons are speculative and this was not the primary focus of our manuscript, we did not
expand on this finding in the discussion section.

9. In Methods, under “Laboratory methods”, the sample collection from the concentrated
specimen for the smear preparation should be further described.

We have clarified the “Laboratory methods” section to specify that “both direct and concentrated
smears were prepared from the same specimen.” As currently stated in the text, the direct
smear was prepared first. The specimen was then decontaminated and concentrated by
centrifugation. A portion of the concentrated specimen was used to prepare a concentrated
smear and to perform mycobacterial culture.

10. Could the author add the numbers used for the calculation of the overall concordance? I
found an overall agreement of 80% (224/279) when I include both culture positive and negative
results. 75% is limited to culture positive results.
The referee is correct that the 75% agreement reported is limited to culture positive results. This was done because we were focused on the sensitivity of direct plus concentrated smears, which is based on the number of culture positive results. However, we agree that overall concordance should also be reported. The text was modified accordingly and now includes both concordance overall and concordance limited to culture-positive results:

“However, concordance between the two methods was only fair, both among all specimens (80% overall agreement; unweighted kappa 0.56, 95% CI: [0.44, 0.68]) and among culture-positive specimens (75% overall agreement; unweighted kappa 0.51, 95% CI: [0.35, 0.66]).”

11. In table 3, the scanty result of patient 9 is not at the right place. It should be in the column “concentrated” of smear result.

We apologize for this confusion and have clarified the Table by adding the following footnote to the result for patient 9:

“Direct smear alone was positive; for other patients, the concentrated smear alone was positive.”

12. Table 4 presents AFB density for direct and concentrated smears regardless of the results (negative or positive). The first sentence of the chapter “Density of Acid-Fast Bacilli” should be revised accordingly.

Based on the referee’s suggestion, the sentence was modified as follows:

“Table 4 compares the density of AFB on specimens from patients with culture-confirmed TB.”

13. In the 2nd chapter of the discussion, it would be interesting that the authors compare also their results with studies using similar in-patients population and similar concentration method (NALC-NaOH and centrifugation).

We agree that comparing our results to studies using similar populations and concentration method would have been interesting. We identified 8 studies that reported comparing direct microscopy with microscopy following processing with the NALC-NaOH method. Three of these studies included both inpatients and outpatients, but none reported HIV status. In the discussion, we chose to focus on comparison of our results to previous studies in HIV-infected populations. Further studies are needed to confirm whether alternative concentration methods account for the differences observed between our study and prior studies in HIV-infected patients.

14. Page 11, in the last chapter of the discussion, the authors should also discuss the operational benefit of using direct + concentrated smear of 1 single specimen compared to direct smear on 2 specimens in term of the potential reduction of the number of visit to the health centre and the delay to initiate treatment. I agree that this should be further evaluated in a cost effectiveness study.
We share the referee’s sentiments regarding the operational benefits of performing multiple smears on a single specimen. We added the following sentence to the discussion summarizing these benefits:

“Collection of a single sputum specimen would reduce patient burden in terms of number of visits to health care centers, and could also decrease the delay between clinical presentation and initiation of treatment.”

15. In the conclusion, could the authors precise “HIV-infected hospitalised adult population” in the 1st sentence.

We modified the first sentence of the conclusion as suggested by the referee.

16. The description of the quality of the morning specimen is missing. This information would be very useful to maybe understand why there was no difference between direct and concentrated smear microscopy in this study. Indeed, if most of the morning specimen were purulent, then the benefit of the concentration method is likely to be minimal.

We paste below our response to Referee 2, Comment 2, which addresses the same concern:

This has been a concern of ours throughout the study as well. We collected data on sputum quality during the study, but did not initially report this data due to its subjective nature. However, given the concerns raised here and by Referee 3, we have included this data in the results section of the revised manuscript:

“Regarding sputum quality, of the 279 specimens, 33 (12%) were described as salivary, 191 (68%) mucoid, 41 (15%) purulent, and 14 (5%) bloody. Exclusion of the salivary specimens, or restriction to mucoid specimens, reduced the sample size but did not materially affect results.”