Response to reviewer’s comments

Changes in the manuscript are indicated in red font

Bacteriophage based tests for the detection of *Mycobacterium tuberculosis* in clinical specimens: a systematic review and meta-analysis

**Reviewer 1 (Carlo Mengoli)**

**General**

The article is a review, which is not currently considered among the article types to be published by BMC Infectious Diseases. The perspective is diagnostic/microbiological. Rapid and simple tests are especially suited to primary-care settings and to countries with both high TB circulation and limited laboratory resources. The conclusion formulated by the authors is appropriate. The mycobacteriophage amplification tests do not add much to the accuracy offered by microscopy, and require the level of organization and equipment of microbiological facilities capable of mycobacterial culture (reference laboratories). The turnaround time of phage amplification tests is 2 days compared to about 2 hours (microscopy) or up to 2 months (culture). The phage-based tests have to compete with the molecular methods, which are comparably time and labor-intensive. So, when TB is suspected, the algorithm of specimen processing is unlikely to change radically in view of phage-based tests.

We thank the reviewer for concurring with our conclusions. We have been informed by the BioMed Central Editorial office that the journal does consider systematic review articles and classifies them as “research articles”.

1. In general, heterogeneity of the outcome is not addressed adequately, for instance by looking for covariates in order to set up a subgroup analysis or a metaregression. Admittedly, the effort in this direction could not be rewarding. However, the attempt should be done and the outcome of the attempt reported.

We thank the reviewer for these comments. We agree that a diagnostic meta-analysis must address the issue of heterogeneity. We made several attempts to explore heterogeneity in our review. Firstly, we reported sensitivity and specificity of phage-based assays separately for commercial and in-house tests (subgroup analyses, Table 2). Secondly, we reported results stratified by smear microscopy status (Table 3). Thirdly, we reported head-to-head comparisons between smear microscopy and phage assays (Table 4). Fourthly, we refrained from pooling sensitivity and specificity estimates because of the potential difficulty in interpreting them in the presence of heterogeneity.

Although meta-regression is a useful tool in the evaluation of heterogeneity, we decided against performing metaregression because we had only 13 studies in the systematic review; with only 13 data points, it would be difficult to fit and interpret a regression model.

To address the reviewer’s concern, we have described this as a limitation in our revised manuscript (section entitled “Strengths and Limitations”, page 16)

2. The different sensitivity of the gold standard (LJ, or BACTEC, or LJ+BACTEC, or LJ+AMTD) can add to heterogeneity. See point 1.
We thank the reviewer for pointing out that the sensitivity of the reference standard can influence the heterogeneity. We have added a sentence in the discussion section (Page13, para 2) to this effect. Because grouping the studies in four sub-groups would have resulted in rather small numbers of studies in each of the category, we did not attempt statistical comparisons.

3. After separating the smear-positive subgroup from the smear-negative one, no statistical comparison is done. See point 1.

The table on sensitivity and specificity of phage assays according to smear status (Table 3) had only 5 studies. With such a small number of data points, we did not think it appropriate to perform statistical significance tests.

4. This is largely a review on FASTPlaque-TB. Ten of 13 studies are based on this commercial kit. This subgroup of studies can be compared to that based on different kits. See point 1.

We agree that ten of the 13 studies did use FASTPlaque –TB kit to evaluate the accuracy of phage-based assay for detecting M. tuberculosis. We have added a sentence in the discussion section to this effect (Page 12, discussion section, para 1). We feel that statistical comparison would not be appropriate because only 3 studies did not use FASTPlaque-TB kit.

**Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)**

The figure 1 is called where the text says two main phage-based approaches are used to detect M. tuberculosis, but depicts only the phage amplification methods.

We thank the reviewer for pointing out this error. We have replaced Figure 1 with the one that contains both phage amplification methods as well as luciferase technique.

Through the text. The expression phage-based test seems to be equivalent to phage amplification based-tests. The exclusion (or inclusion) of LRP-based tests should be stated explicitly.

We have made the necessary change in the manuscript (Page 9, results, para 3).

**Already in the abstract, bacteriophage tests should be introduced as (relatively) rapid tests.**

This has been changed as suggested (Page 4, para 3).

**Already in the abstract the gold standard (culture) should be indicated.**

In discussion, comparative mention should be done to molecular tests, which are the obvious competitors for phage-based tests.

We have added the following line in the discussion section (page 18, para 1): Molecular tests such as polymerase chain reaction (PCR) are competitors for phage-based assays. Of the two studies [7, 12] that reported head to head comparisons of molecular tests with phage - based assays against a common reference test (i.e. culture); one study [7] found no difference in test accuracy; the
other [12] found that phage-based assay was less sensitive (64% vs. 82%) but more specific (93% vs. 85%) compared to the polymerase chain reaction.

Reviewer 2 (John Bernando)

We thank the reviewer for positive comments.

1. Abstract; para 1. The sputum smear may be highly specific, but only in appropriate populations. This should be clarified in the text.

We have revised the sentence: Sputum microscopy, the most important conventional test for tuberculosis, is specific in populations with high burden of tuberculosis, but not sensitive. (Page 2, para 1)

2. Abstract; para 2. The authors presumably performed a review of the published literature for this study. The sentence should be completed.

This has been changed as suggested: We did a systematic review and meta-analysis of published studies to evaluate the accuracy of phage-based tests for the direct detection of M. tuberculosis in clinical specimens. (Page 4, para 2)

3. p. 5, para 2: Figure 1 only depicts only the principles of the plaque assay methodology. The designation of Figure 1, in line 1, should be moved to reflect this (perhaps to line 3).

We have replaced Figure 1 with one that also includes luciferase reporter phages (LRP)

4. p. 5, para 2: Define PhaB.

PhaB is “phage amplified biologically”. We have defined this in the manuscript (page 5, para 2)

5. p. 6, para 1: Some studies analyzed were referred (or sponsored/performed) by test manufacturers. If results can be analyzed separately, was test performance in these studies different from others?

Table 2 shows the operating characteristics of tests (commercial and in-house assays). Ten studies performed assays based on FASTPlaque test and 3 used an in-house assay. Studies that were supported by the test manufacturers [5,8,9,11,16,17] showed wide variation in sensitivity (range: 31% to 94%), as did studies that evaluated the performance of phage-based assays without support from test manufacturer [11,13,15] (range of sensitivity: 21% to 77%). Three studies [6, 7, 12] did not mention whether or not they were supported by the test manufacturer; the sensitivity ranged between 77% and 88% in these studies. The specificity of the test was almost similar among the three groups.

6. p. 10, para 2, ff. Some studies (3) analyzed sputum specimens other than sputum. In these studies, can data from these samples be separated from sputum data and analyzed separately?
Three studies (Alcaide, 2003; Shenai 2002 and Shenai 2004) analysed non-sputum specimens. Because the number of non-sputum specimens in these studies is very small, it was not possible to do a sub-group analysis.

7. Discussion: In these studies, TB disease was defined by growth in culture; clinical diagnoses are not accepted. Inclusion of clinical diagnoses may affect both sensitivity and specificity analyses, and should be mentioned in the discussion.

We are sorry that we were not able to make this point clearer. The reference standard used in all studies was culture (BACTEC 460 in 3 studies; LJ media in 6 studies, BACTEC and LJ in 3 studies and LJ and AMTD in 1 study).

8. SP is identified in several places in the manuscript as a reviewer (p. 6, para 2 and 3) author (p. 18, para 2). I assume the correct initials are SK, referring to the first author.

We have made the appropriate change.

Reviewer 3 (Ruth McNerney)

General
A clearly written article that addresses a problem important to public health in developing countries. This is a thorough review on a novel topic.

We thank the reviewer for these positive comments.

Major Compulsory Revisions
Assay characteristics: the in-house assay did not use NALC-NaOH decontamination method.

We apologize for this error. We have revised the text. (Page 10, para 1, line 2)

Discussion:

1. The authors have not commented on the bias introduced through the exclusion of samples due to microbiological contamination. Smear microscopy results may be obtained for samples found contaminated by culture or phage.

We thank the reviewer for the valid comment. We have added the following text in the discussion section. (page 13, para 2)

The specificity of phage-assays for detecting acid fast bacilli is likely to decrease in settings in which infection with mycobacteria other than *M. tuberculosis* are common. Sodium hydroxide, used to decontaminate the specimens, may damage the acid fast bacilli and can reduce the sensitivity of phage-based assays for detecting tuberculosis. Use of gentler decontamination techniques can reduce the specificity of the test: this approach protects the acid fast bacilli but fails to prevent contamination from other microorganisms.

Contamination of LJ slants with micro-organisms is a common problem in certain settings (40.4% in Zambia [13] and 18.2% in Pakistan [6]). The exclusion of contaminated results may result in biased
estimates of sensitivity and specificity. This approach (partial verification) can lead to bias if systematically more abnormal than normal test results are subjected to the reference standard. Although no study reported use of smear instead of contaminated cultures, such a strategy could weaken the reference standard and result in differential verification bias.

2. Assessment of accuracy has been made in comparison to a single smear test. Current WHO recommendations are for three sputum samples to be examined for each patient. To assess diagnostic accuracy the phage tests should be compared to three smears. The authors have not addressed this issue in their discussion.

We thank the reviewer for this important comment. We have added the following sentences- and cited a reference - in the discussion section. (Page 14, para 4)

To evaluate the accuracy of smear microscopy for detecting *M. tuberculosis*, except one study [5] in which patients provided two sputum specimens each, all studies used only a single sputum sample. This approach differs from the current WHO and IUATLD recommendations, which state that at least three sputum samples must be examined for each patient. Use of a single sample might make the smear microscopy look less sensitive than what it actually it is because the sensitivity may go up with greater number of smears. It would be interesting to evaluate the additional yield from repeated sputum examinations by microscopy and culture and do a head-do-head comparison of 3 sputum smears with phage-based assays. To our knowledge, this has not been done in any of the available studies.

New reference cited: (reference 25)


3. Whilst reporting that phages tests may produce false positive results due to non-tuberculosis mycobacteria the authors do not comment on the bias caused by sampling in settings with high (or low) incidence of TB.

We have added the following sentence in the discussion section: (page 15, para 2)

The effect of non-tuberculous mycobacteria (NTM) might be minimal in places where true *M. tuberculosis* is very common (high incidence settings). In settings with low incidence, NTM might have a greater impact.

4. The authors state a need to verify that the sputum found positive in the phage test contains *M. tuberculosis* but do not suggest how this might be undertaken. A requirement for a confirmatory test would affect the cost/time of diagnosis and thus affect the usefulness of the technology. The authors should comment on this.

We have added the following sentence in the discussion section (page 15, para 2):
A requirement for a confirmatory test would make the overall testing strategy more expensive, delay the diagnosis, and result in an additional patient visit to the laboratory. Clearly, these factors could adversely impact the practical applicability of the test in resource–limited settings with high burden of tuberculosis.

Discretionary Revisions (which the author can choose to ignore)

Background (page 1), regarding costs of liquid culture in developing countries. This has recently reduced through an agreement with WHO and Becton Dickinson.

We have revised the sentence to this effect (Page 4, para 2)

Fig 1. Though readable the quality of reproduction of the diagram is poor. The quality of reproduction of figures 4, 5 and 6 should be improved. Figure 5 does not fit on the printed page.

We have added the legend to the figure. We have re-formatted the table. We used the maximum resolution allowed by Meta-Disc software to improve the quality of figures 4, 5 and 6.