Reviewer's report

Title: Choosing algorithms for TB screening: a modelling study to compare yield, predictive value and diagnostic burden.

Version: 2  Date: 11 May 2014

Reviewer: Lucie Blok

Reviewer's report:

- Major Compulsory Revisions:

1. Figure 4 as currently presented cannot be interpreted without a clear heading and labels for the y-axes. It seems that the second graph depict NPV (or the % of cases being diagnosed) for different algorithms (100% for an algorithm with 100% sensitive screen and 100% sensitive confirmatory test). If that is indeed what the graph presents, the outcomes do not seem logical (e.g. not very logical that an algorithm using cough >2 weeks as first screen and XP as confirmation would perform better than screening for any symptom followed by XP). A better description and explanation of these graphs is needed.

2. One statements in the results section is not logical and does not seem to be supported by the data presented in the tables. I refer to: “The NPV of all algorithms drops if the proportion of persons receiving clinical diagnosis after a negative confirmatory test would increase or a clinical diagnostic algorithm with lower specificity would be applied.” Comparing the NPVs for different algorithms with or without clinical diagnosis seems to lead in most cases to a slight improvement in NPV. However more importantly the PPV seems to drop in most cases and this is an important finding, but not stated here.

- Minor Essential Revisions:

3. Several parts of the text are unclear and several sentences seem incomplete or grammatically inconsistent. Especially the results section is difficult to read and the key results seem to get lost. The authors should revise the manuscript and consider asking help of an English editor. Here are a few examples (no complete list):

3a Last paragraph of the abstract needs editing.
3c Text box 1. Explanatory note § has duplicated “as”
3d The definition of NPV in the last paragraph before ‘screening methods’ in the methods section seems to be wrong. Please revise.
3e The second sentence under screening methods is unclear and needs editing
3f In the screening methods 4-7 symptoms suggestive of TB are mentioned followed by a list of 5: what are the other two?
3g The last sentence under screening methods is not clear: “The screening algorithms are composed of […] accuracy of any CXR abnormality”. ???
3h Last sentence of the paragraph entitled confirmatory test needs revision.
3i First sentence under results is incomplete.
3j Discussion: the fore last paragraph -which starting with “Our analysis points to …” – is difficult to understand. Please consider editing.

4. All headings and labels in tables, graphs and figures need to be carefully checked.
4a Table in annex has a heading that suggests that costs of tests and diagnostic algorithm is included while it is not.
4b Figure 4 has no proper headings and axis labels
4c In several graphs labels are missing for x- and or y-axes
4d Table 1 is not easy to read, consider revising.
4e Text box 2 is difficult to read in particular having the text of column two ( “uncertainty in”) running over two lines makes understanding of its relation to numbers in the first column and to the bullets in the next column not very clear

- Discretionary Revisions
5. Text box 1: The inclusion of CD (clinical diagnosis) as a last step is somewhat confusing and could be better explained in the note or in the text.

6. Why use two categories with same name but different definitions (both for True Negative as well as False Negative). Consider using distinctive names if it is felt important to review these as two separate categories.

7. In the methods section the second paragraph after the description of the six outcome algorithm categories seems a repetition. Although these are very important considerations there seems to be more emphasis on the potential negative consequences of screening than the potential positive consequences. Why repeating the negative consequences that are already mentioned under each category if not also repeating the potential benefits.

8. It is not clear why the point predicted value for GXP is taken rather than a range (based on 95% credibility interval), while taking the range of published sensitivity and specificity for SMS?

9. An assumption is made that a negative confirmatory test is followed by a CD in a % of cases that equals (1-NPV)*10. In practice this means that the authors believe that in those situations where 10% of people with positive first screen would be falsely categorized as ‘non-TB’, that 100% would then be diagnosed as TB on clinical grounds. What do the authors base this assumption on? This seems not realistic and is not current practice as far as I have observed in screening settings. In most situations a negative sputum test (in particular when using GXP) is usually accepted as final screening outcome in an active screening setting as opposed to passive case detection.

10. In the discussion the following statement is made: “CXR for screening should
ideally not be combined with SSM because of [...]the higher risk of false positives, especially if among SSM negatives CXR is also used for clinical diagnosis.” This statement seem to assume that people with abnormal CXR suggestive of TB but without bacteriological proof would be false positives. This may not necessarily always be the case. The referenced study takes culture (Lowenstein-Jensen) as gold standard. What proof is there that none of these culture negative cases are active TB cases (not yet detectable) that could nevertheless benefit from treatment?

11. One additional limitation of this model is the implicit assumption that people that agree to be screened are a representative sample of all the specific risk group. While practice has shown (own observation) that people that consider themselves at higher risk (due to pre-existing symptoms or having been in contact with a case) seem more willing to be screened and to follow the entire algorithm.

12. The uncertainty of level of accuracy of culture as a gold standard poses another limitation on the assumptions regarding unnecessary treatment of ‘false positive’ cases. The authors may consider highlighting this.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests