Reviewer’s report

**Title:** Inter-rater agreement in the assessment of abnormal chest X-ray findings for tuberculosis between two Asian countries

**Version:** 2  **Date:** 13 October 2011

**Reviewer:** Sarita Shah

**Reviewer’s report:**

**General Comments:**

The authors present results of a study of inter-rater agreement using two x-ray scoring systems: one developed by Den Boon, et al, and one developed by the researchers in this study. Four physicians (2 Japanese pulmonary physicians, 2 Vietnamese radiologists) read 258 x-rays using both scoring systems. Two physicians (1 from Japan, 1 from Vietnam) also read 93 follow-up x-rays using both scoring systems. Physicians were blinded to others’ readings and to clinical information. Physicians were trained in both scoring systems using a sample set of 10 x-rays prior to the study. X-rays used in this study were taken from a previous prevalence survey and were limited to x-rays that were “suspicious for TB” (i.e., normal or not TB x-rays were excluded).

Overall, the paper is well-written and the methods are generally well-described. The study is of importance to the field and provides additional data on the subject of x-ray readings in TB. The authors find that x-ray readings vary markedly between readers, regardless of which scoring system used. The Discussion presents possible explanations for the disagreement between readers and identify differences in definitions used by Japanese vs Vietnamese readers (despite the pre-study training). They further postulate that differences in clinical practice and TB epidemiology between Japan and Vietnam may result in differing clinical skills for interpreting x-rays. However, even between raters from the same country, agreement was low; reasons for this are not fully explored, but are not unexpected based on published literature.

**Major Compulsory Revisions**

- Lines 61-64: This section should be expanded and references provided for sentence 2 and sentence 3. The primary aim of this study (as stated in lines 68-69) is to assess differences in x-ray reader agreement by medical background. Only one line is devoted to this central thesis, which is too little. What prior data exist that led the authors to consider this hypothesis? Why is this an important issue (e.g., are x-rays read by differing types of physicians globally)?

- Lines 66-71: The authors should tie in the aim of this study, the gap in the current system, and the new scoring system they have developed. How did they derive this scoring system? Was it piloted prior to this study? This should be
explained in the Background and Methods.

• Lines 77-78: Please clarify which participants provided informed consent. Was it all 11,624 who were in the prevalence survey or just the 317 with radiographic suspicion of TB?

• Lines 80-82: This section would be better titled as “Study Population” or “Study Design”. The section should start with a clearer description of the study population, e.g., who was included in the prevalence survey (anyone, just people with TB symptoms, etc), was x-ray the primary means of screening (this is not the standard approach to TB prevalence surveys, so should be clarified), was the survey of just adults or also children, was HIV status known or obtained? These factors may affect x-ray readings and are important to clarify and, possibly, stratify results by (e.g., HIV+ vs. HIV-, adults vs children).

• Lines 102-103: The authors propose a new x-ray scoring system that is simpler than CRRS. The only area where the difference is explained is in lines 102-103. Are there other differences between JVCS and CRRS, other than these? This should be explained further in the Methods section.

Minor Essential Revisions

• Figure 2 is difficult to follow in the current format. Please consider revising this to a more vertical format and removing the boxes on the left (active TB cases).

• Table 2 is difficult to read given that there are a lot of numbers in a small space. Consider removing the data indicating - - / - + / + - / + + from this table.

Discretionary Revisions

• Figure 3 is redundant with Table 3, so could be removed.

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

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I declare that I have no competing interests.