Author's response to reviews

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

Authors:

Shinsaku Sakurada (ssakura@mti.biglobe.ne.jp)
Nguyen TL Hang (lehang0310@gmail.com)
Naoki Ishizuka (Naoki.Ishizuka@sanofi-aventis.com)
Emiko Toyota (etoyota-in@tokyo-hosp.jp)
Le D Hung (ledinhung.hth@gmail.com)
Pham T Chuc (phamthechuc.hth@gmail.com)
Luu T Lien (luuthilien.hth@gmail.com)
Pham H Thuong (phamhuuthuong.hth@gmail.com)
Pham TN Bich (phamngocbich2010@gmail.com)
Naoto Keicho (nkeicho-ty@umin.ac.jp)
Nobuyuki Kobayashi (nkobayas@hosp.ncgm.go.jp)

Version: 3 Date: 12 December 2011

Author's response to reviews: see over
Dear Editors,

We appreciate all the reviewers' criticisms to improve our manuscript which we have meanwhile revised in accordance with your observations. All changes we have made are described below. We also carefully edited our manuscript and revised parts in the text are shown in blue.

We are hoping that our manuscript will now have been revised satisfactorily for publication in BMC Infectious Diseases.

Faithfully yours,

Naoto Keicho.
National Center for Global Health and Medicine, Tokyo, Japan.

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 “suspicous 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)?

  ➜ Thank you for the reviewer’s reasonable comment.
  Previous studies suggest that variability in chest X-ray interpretation among raters is attributed to subjective reading accompanied by insufficient experience or different professional background of the raters [Abubakar I, et al. Eur Respir J 2010; Brealey S et al.. Br J Radiol 2007 and others.]. However, the relationship between agreement levels and relevant factors that may cause disagreement, particularly influence of medical background including different national origins has not been characterized.

  ➜ We have thus expanded the corresponding parts of the Background by adding the above description quoting references in the revised version of the manuscript.

• 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.

  ➜ The major objective of this study was to highlight inter-rater agreement between raters with different medical backgrounds. As analytical tools, two different types of coding
systems were used. One was previously reported by another group and the other was newly developed in this study.

- Considering a registration form used in a public payment system for TB treatment expenses in Japan and reading practice in Viet Nam together, we developed a simple coding system.
- We have supplemented the above information in the Background and in the revised paragraph of chest X-ray coding systems and reading of films in the Methods section.

• 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?
  
  ➔ All 11,624 individuals who were in the prevalence survey provided informed consent. We clarified this point in the section of ethics approval in the revised manuscript.

• 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).

  ➔ Thank you for the reviewer’s useful suggestion. This section was re-titled as “Study population” as the reviewer recommended. More detailed descriptions regarding the prevalence survey, such as participants, methods of screening etc were also supplemented: “A population-based TB prevalence survey of 11,624 people aged 15 and over was conducted in Hanoi in 2003 as reported previously. Briefly, subjects suspected of having active TB based on chest X-ray or on symptoms underwent sputum smear microscopy and/or mycobacterial culture. Details of HIV status were not obtained from the study subjects. According to the report of World Health Organization during this period, estimated prevalence of HIV co-infection in new TB patients aged 15-49 was relatively low (2.8%) in Viet Nam.”

  ➔ We have also re-located the description of X-ray films under the subheading of “Study population” to facilitate the readers’ understanding on the relationship between the prevalence and the follow-up surveys and this X-ray-reading study.

• 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.

- Thank you for the reviewer’s comment. We supplemented the following description in the section “CXR coding systems and reading of films” of the revised manuscript: “Additionally, CRRS classifies nodules based on their size and calcification, whereas JVCS separately records nodules and calcification.”

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).

- Thank you for the reviewer’s recommendation. We revised this figure 2 to a vertical format and removed the boxes on the left.

• 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.

- We agree to the reviewer’s suggestion. We removed the data indicating - / - / + / + / + / + from this table.

Discretionary Revisions

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

- Although we thank the reviewer’s suggestion, we would like to keep Table 3 and Figure 3, if the reviewer permits. Table 3 showed the level and details of disagreement on overall assessment of chest X-ray after 3 years between two raters. On the other hand, the Figure 3 presented the change of infiltrates after 3 years assessed by each reader. The difference in assessment of infiltrates between raters shown in Figure 3 may explain the disagreement on overall assessment shown in Table 3.

- We showed the difference between Table 3 and Figure 3 more clearly by adding the word "infiltrates" inside Figure 3 and revising figure legends.

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.

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: 29 September 2011

Reviewer: Videlis Nzioka Nduba

Reviewer's report:

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

Shinsaku Sakurada, Nguyen TL Hang, Naoki Ishizuka, Emiko Toyota, Le D Hung, Pham T Chuc, Luu T Lien, Pham H Thuong, Pham TN Bich, Naoto Keicho and Nobuyuki Kobayashi BMC Infectious Diseases Research article

1. Is the question posed by the authors well defined? Yes, the question posed by the authors is very clearly defined

2. Are the methods appropriate and well described? The methods are quite appropriate, they wanted to do interrater agreement assessment and methodology chosen is quite acceptable

3. Are the data sound? The data are quite sound; the have presented multiple comparisons that show
the same trend using different methods of assessing CXRays

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
Yes, the manuscript is well structured, the tables and figures are very clear

5. Are the discussion and conclusions well balanced and adequately supported by the data?
Yes the discussions and conclusions are well balanced and supported by the data. It makes sense that a simpler approach and creating uniformity and understanding different issues and interpretations that might vary due to training backgrounds is quite useful

6. Are limitations of the work clearly stated?
The limitations are clearly stated

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
The authors have provided adequate references and clearly compared the simplified system they developed and the already published CRRS system of CXRay interpretation

8. Do the title and abstract accurately convey what has been found?
Yes they accurately convey the results

9. Is the writing acceptable?
Yes the writing is acceptable

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

There are minor inconsistencies to the 317 individuals with abnormal CXRays where active TB was not excluded at baseline. The abstract: Methods make it clear 258 individuals were evaluated since their baseline CXR films didn’t exclude active TB
The Methods: The last paragraph of the Methods section has an error; it states “The CXR films analysed in this study were those in which active TB had not been radiographically excluded during the prevalence survey and were those taken during the second follow-up in 2006.
I think they meant to say the films analysed were both the baseline prevalence survey films (258) and follow up films (93); this is made clear in figure 2. I would suggest the authors make this minor correction for consistency and smooth flow of the article

➤ Thank you for the reviewer’s comment.
➤ We changed the flow diagram (figure 2) to a vertical format and then revised the text in a more straightforward manner to facilitate readers’ understanding:
   In the methods section, we have re-located the description of X-ray films under the subheading of “Study population”.
   We also added one sentence in this paragraph: “In total, 258 of the 317 films in the prevalence survey and 93 follow-up films were available at the time of analysis in this study”.

**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