Reviewer's report

**Title:** Survival and health status of a cohort of tuberculosis patients in rural Lao PDR

**Version:** 3 Date: 15 January 2010

**Reviewer:** Giorgio Bedogni

**Reviewer's report:**

I was asked to review this paper as BMC statistical referee. I read the revised manuscript (MS) together with the replies to the 2 other reviewers. Although I fully understand the difficulty of performing an epidemiological study in a developing country, I believe that the MS would still benefit from a substantial revision. I am fully convinced of the importance of creating a culture of public health epidemiology in developing countries and this study, despite its limitations, is clearly a step in the right direction. This is the spirit with which my suggestions are to be interpreted.

**Major Compulsory Revisions**

**GENERAL** I would use the term “cohort” *just* for the prospective part of your study. Strictly speaking your “cohort” starts with the first home assessment and as such it is “followed up” for 8 months at most (Figure 2). I would say something like: “Our patients were all those seen at our Center in the years 2002-2004 and were followed-up for X months (range: Xlow – Xhigh)”, where X is the median, and Xlow and Xhigh the minimum and maximum values of follow-up time. By reading the abstract one gets the wrong impression of a longitudinal study protracted for 3 years. Of course, by putting together the years 2002-2004, you are supposing that the year of admission to the hospital is not a confounder of the relationship(s) you are interested into.

P5 Did all sputum-negative patients undergo chest x-ray? Is this the standard procedure in your Center? A spectrum bias will always be present as more severe patients are more likely to reach your Hospital but here I am interested to know whether this is the diagnostic algorithm *always* employed at your Center.

P6 Please, be very clear in the definition of the outcomes, predictors and regression models employed. You mention for instance the log-rank test and the footnote of Table 4 says “multivariate Cox model” so I suppose that here you use a time-to-event outcome. However, be this a time-to-event outcome or not, there is no operational definition of it (besides a generic “non-compliance” label in table 4).

Table 1 Apparently, you used omnibus tests here as there is just one p-value. Which tests did you use for continuous and categorical variables? Are you not interested to know which between-group differences are responsible for the “significant” findings? You may wish to use exact tests for post-hoc analysis here,
especially for some categorical comparisons (e.g. just 1 child out of 13 subjects in the extra-pulmonary TB group).

Figure 2 The problem in interpreting the sputum-time relationship is that sputum examinations are not available for most patients right from time 0. Also, the number of sputum examinations decreases during the study. You should provide some data showing that the patients who undergo sputum examinations do not differ from those who not undergo it. The great risk here is that the patients who undergo the sputum examinations are not representative of your whole set of patients.

Figure 2 I strongly suggest to replace this graph and other tables with a table:

<table>
<thead>
<tr>
<th>T0</th>
<th>T2</th>
<th>T5</th>
<th>T8</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive sputum (n,%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and so on...

N = number of patients at time T
Value = descriptive statistics

P19 Table 3 and *elsewhere* How did you handle the fact that these measures are correlated, i.e. they are made on the same individuals?. Were these 72 patients selected on the basis of having both 2 and 8 month data? If the overall missingness rate is not too high (say < 15%) you may consider studying the relationship using a mixed model which can handle missing data. The risk again is that these 72 patients are not representative of the whole cohort.

P20 Table 4 Please specify where do these 300 patients come from. Give also details on the model (see above): time-to-event outcome? I suppose that all predictors are dichotomous (yes vs. no?). Please specify.

MINOR COMMENTS

GENERAL The title focuses on “survival”, the running title on “efficiency” and the abstract on “feasibility”. I would try to harmonize these part of the MS.

P1 Please, give the denominator of the per capita income (USD 500 per year, I suppose).

P1 I would write “including a minority (X%) of HIV patients” with X% being the percentage of HIV patients with TB in Lao PDR.

P1 I would write “drugs” instead of “medicine”

P6 Strictly speaking your “sampling units” are patients and not “villages”. Thus, I would avoid writing “in a subsample of villages”. Because you are referring to a previous study, you may write: “In a previous study we found that, in a random
sample of 10 villages of the same geographical area covered by the present study…” or something alike.

P6 Did you mean that you used the Epidata double-entry feature? Other?

Table 2 Age and weight were obtained at admission but was the weight the same at admission and dismissal? I suppose that BMI was obtained by measuring both weight and height at home. Do you confirm?

Table 2 Please, specify the tests used. The footnote says mean and SD but age has 3 parameters. Why?

**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