Reviewer's report

Title: Reporting of loss to follow-up information in randomised controlled trials with time-to-event outcomes: a literature survey

Version: 1  Date: 21 March 2011

Reviewer: Eleanor Pullenayegum

Reviewer's report:

Standards for reporting of RCTs and other studies are receiving increasing attention, and this manuscript addresses the analysis of time-to-event outcomes, which are harder to report than many other outcomes. The manuscript is well-written but would benefit from clarification around some of the definitions.

Major compulsory revisions

1. Abstract, conclusions: Here it is stated that the lack of assessable LTFU information calls the validity of the results and the conclusions of the surveyed articles into question. I think this may be a stretch. There is an important distinction between poor reporting (often caused by stringent word counts) and poor conduct of the analysis. Moreover, the sensitivity analyses did not convince me that the validity of the results should be questioned (see comment below).

2. Last line of page 5/first line of page 6: It seems misleading to classify articles with LTFU information on the plot but not on the text as “not consistent” – if information is in just one place, how can it be inconsistent?

3. Page 6, last paragraph of methods: This mentions “one time point before minimum follow-up” – how was this point chosen, as presumably there could be more than one option? Was it the closest point to minimum follow-up at which there was LTFU information reported?

4. Page 6, last paragraph of methods: I was confused by the definition of “robust”. When you say the p-value remained either significant or non-significant after imputing missing data, do you mean compared to a log-rank test done on the Kaplan-Meier plot, or compared to no imputation, or for equal case as compared to worst case? Also, if you picked a time point before minimum follow-up, then as stated on page 5 no patient should be censored so why would you have censored data to impute?

5. Page 6, 2nd last paragraph: It’s not too surprising that best-case and worst-case scenarios should lead to different results if censoring is substantial.

6. Figure 1: I was confused as to why this was classified as “consistent”. The calculation says you should have 17 in the prednisolone group and 6 in the control group; the table says there were 16 and 6. This is stated to be consistent because one patient was lost to follow-up, however on page 5 you stated that no
patient should be censored before the minimum follow-up time, so the numbers should match exactly.

Minor compulsory revisions

7. Background, page 4, last sentence of first paragraph: The condition for censoring to be non-informative is not quite stated correctly. As stated, survival probabilities must be the same for those who are censored and those who are not censored, however, since patients who survive for longer have a greater probability of being censored, this will not hold even if censoring is completely at random. Non-informative censoring requires that at any given point in time, the survival distributions amongst those censored at that time point and those not censored at that time point be the same.

8. Page 5, third paragraph: Did you collect items 2(a)-2(d) separately for each group? Did you collect items 2(c)-2(d) for more than one time point? For item (3), if the planned and actual minimum duration of follow-up were different, which one did you record?

9. Page 5, first line of fourth paragraph: What sort of details on LTFU can be inferred? % lost to FU, or time to LTFU or median FU time?

10. Page 6, 2nd paragraph: With the exception of articles classified as “not consistent” articles were reviewed by a single reviewer. It would strengthen the study methodologically if at least a subset of the articles could be reviewed by a second reviewer in order to quantify agreement between reviewers.

Discretionary revisions

11. Abstract: I wonder whether when quantifying the percentage of articles presenting consistent or non-consistent LTFU information, it would make more sense to make the denominator the 187 assessable articles rather than the 319 eligible articles? The reasoning behind this would be that the classification of “consistent” or “not consistent” was made only for assessable articles.

12. Page 5: It would be helpful to give the search strategy in a web appendix.

13. Page 5: What was the logic behind preferring the “all-cause mortality” or other mortality-related outcome rather than simply the primary outcome? Wouldn’t reporting likely be better for the primary outcome than for secondary outcomes?

14. Figures 1 and 2: It would be helpful to state what the numbers below the two figures mean – I assume these are the numbers of patients at risk in each group at each time

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

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests