Reviewer's report

Title: Assessing and Reporting Heterogeneity in Treatment Effects in Clinical Trials: A Proposal

Version: 1 Date: 31 May 2010

Reviewer: Gilda Piaggio

Reviewer's report:

Comments of reviewer

This article contains a proposal for a framework that addresses multiplicity of subgroup analysis in clinical trials, prioritizing the analysis and reporting of multivariate risk-based heterogeneity in treatment effects. Caveats of subgroup analysis have been addressed by many investigators, but this article is innovative in that it proposes a strategy. The recommendations #1 to #5 are clear and helpful. This initiative should be welcome by the triallist community and the authors are commended for developing this framework.

Major Compulsory Revisions

Tables 1 and 2 are not clear. I would suggest a few modifications.

Table 1:
1) An additional column showing the Experimental Event Rate (EER), after the CER column, would help the average reader to clearly see where RRR and ARR come from.
2) The CER column shows a rate (e.g., 8 per 1000 patients over 5 years follow-up, for overall). However, the ARR column looks like a difference in proportions. If the EER is 6 for overall (as derived from CER=8 and RRR=0.25), then is ARR=Abs(EER-CER)=0.02 a difference in rates?

Table 2:
1) Same comments as for Table 1.
2) The typical patient (I understand these are the ‘Average risk subjects’) receives no benefit or harm in Scenario #1 (RRR=ARR=0), but not in Scenario #2, since in the latter RRR=0.14 and ARR=0.0125. Therefore the title should take this into account.
3) The main column heading encompassing the four columns CER, RRR, ARR and NNT reads ‘Treatment reduces baseline risk by 25% but a cost of 2 serious treatment-related adverse events per year’. I don’t see where the 25% and the 2 come from. Also, wording is missing something (should it be ‘at a cost’?)
4) Text referring to Table 2 in middle paragraph of page 5 is not clear (where do the 25% and the 2 come from?).

Discretionary Revisions
1) It would help the reader if the measures of treatment effect RRR and ARR are briefly defined.

2) Most trials in the references would use rates as outcome measures, but proportions are also relevant, as well as continuous outcome measures. Perhaps some clarification about this is worth. Tables 5A and 5B show proportions, which are denoted as ‘rates’, this could be confusing, since some investigators use ‘rates’ for events/persons-time.

3) Should triallists power trials for risk-based analysis or just for the overall result?

4) Could something be said about how external tools were developed? Or are these using a variety of different techniques?

5) It would be useful if something more is said about internally-developed models. The mention on first paragraph under Recommendation #2 to regression analysis of the data using all treatment arms is very vague. Even though external tools are available for many of the primary outcomes used in clinical trials, investigators in other areas need to know how to proceed, in spite of the over-fitting problem. Also, some explanation about how over-fitting can exaggerate the degree of risk heterogeneity would be useful.

6) How to interpret the multivariable continuous risk measure, as shown in Table 5A? Is this measure on a scale from 0 to 100? See also second paragraph under Recommendation #1, reference to “Table 1”: how to interpret the statistics mean, SD, median, for predicted baseline risk?

7) Recommendation #5, last paragraph: it would help to specify how to adjust for multiple testing, and how to adjust for multiple inference when using confidence intervals (eg, use 99% CIs instead of 95%?).

8) Table 4: checklist is very adequate. However, how to overcome the space limitations required by editors?

Minor Essential Revisions

1) Typo in first paragraph, line 5 under ‘Why the overall result...clinical practice’: ‘more benefit than’ (not ‘then’).

2) ‘Why the overall result...clinical practice’, 2nd paragraph, first line: add ‘adverse’ before ‘outcomes’.

3) ‘Why the overall result...clinical practice’, 2nd paragraph, last lines referring to Scenario #2: which is the difference in Table 2, Scenario #2, that is ‘slightly greater’ (please clarify)?

4) ‘Why risk stratified analyses...feasible’, 2nd paragraph, 6th line from bottom: add ‘with respect to a single variable’ after ‘in terms of risk’?

5) ‘A proposal of reporting...effects (HTE)’, last paragraph, line 2 from bottom: reference 27 is repeated.

6) ‘Recommendation #1’, first line: ‘over a decade ago’ is not right because reference 7 is from 2004 (or the reference is not right).

7) ‘Recommendation #2...’, first paragraph, line 6: it should be ‘effect’, not ‘affect’. 
Level of interest: An article of outstanding merit and interest in its field

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.