Reviewer's report

Title: Comparing methods to estimate treatment effects on a continuous outcome in multicentre randomized controlled trials: A simulation study

Version: 1 Date: 18 August 2010

Reviewer: Catherine Legrand

Reviewer's report:

This is a nicely written paper addressing a question of upmost interest in a clear and relatively concise way. The authors performed Monte Carlo simulations to compare 6 different methods to adjust for center when analysing multicenter RCTs under different scenarios regarding distribution of patients and centers. We may have to regret that the less realistic scenarios (equal number of patients per centers) are discussed in length while the more realistic scenarios (different number of patients per center) are more shortly discussed. The authors also present the results obtained when applying the six methods to their motivating application. It would have been interesting to have at least one of the scenarios that coincide with this real example.

Major Compulsory Revisions:

1) In most clinical trials, the number of patients per center is highly variable and we often have to face (very) small centers. It would be interesting to discuss the impact of these small centers according to the 6 different approaches (both from a theoretical point of view and in the simulations). Also in these small centers it may happen that all patients are actually randomized to the same treatment group, how well do the various approaches (or in fact does not) take information available from these centers in account?

The authors wrote (page 14) that their last scenario is “designed so that the centre composition and degree of allocation balance were analogous to the COMPETE II trial: ...”, however they consider then only 17 clusters for a total of 180 patients while the COMPETE trial includes apparently 511 patients from 116 practitioners! It would be really interesting to have the results of the simulations for this scenario also. From a general point of view, different center size is the more interesting situation (as the more realistic one) and thus would deserve more simulation results.

2) My impression is that the GEE methods (which probably the one which is less known and understood by non-specialists) may require more explanations considering the targeted audience of this journal. Also, the approach of center-level random effects would need more clarification. In particular what do the authors call “heterogeneity” in Figure 2 and how is this estimated?

3) The fact that model D had problems to converge (page 20 – line -4) seem to be a major problem of this approach. This would need to be discussed further (in which proportion did it happen? In which situation? Why?) and this should also
be addressed as a drawback in the discussion and conclusions.

4) The conclusion of the text is not fully in accordance with the one of the abstract. It put the fixed and random effects at the same level, while abstract (and results) seem more in favour of the random effect approach. Conclusions on the text should be re-worked and more assertive.

Minor Essential Revisions:

1) Authors should check the references lists, I'm not sure they are all cited in the text and ref of Brown and Kempton appears two times (10 and 26). Check also is all figures are referenced in the text.
2) Abstract, second line of "methods": there is a coma missing

Discretionary Revisions:

1) It would be interesting to address the further points of discussions in the paper:
   - What are the actual recommendations in actual guidelines? How does the results obtained here and the recommendations made by the authors fit into these "official" recommendations?
   - Reviewing the work based on this question, it would be interesting to also add the work that has been done when considering survival endpoint (e.g. oncology trials)
   - Link between the analysis of meta-analysis seems quite obvious (where we then take trials as cluster into account) and how to handle "trials clustering" has already been subject to several publications. This should be discussed to, highlighting similarities and dissimilarities between the two situations and considering which information obtained in the meta-analysis setting can be transferred to this situation.
2) "patient-level random effect" seems to me to be quite confusing way to refer to model C (as we include random effect for center and not for patient), so I would avoid it.
3) The various levels of ICC are a bit difficult to interpret, especially for medical readers, and it would be interesting to translate this value in something easier to interpret, considering for example the heterogeneity introduced between centers by the different value of sigma (cf for example the work of Duchateau and Janssen (American Statistician, 2005).
4) The section “design with equal centre size and chance imbalance” could be shorten to just mention what are different results from the previous section. On the other hand the next session “design with unequal centre sizes and chance imbalance” is the most interesting one and could be further extended to other scenarios, investigating for example the impact or very small and very large sample.
5) Some of the sentences might eventually need to be clarified:
   - Page 6 - line 9-10: "Randomization was stratified by family physician to obtain sufficient enrolment to make the trial feasible" (what makes the trial feasible is to
run it as multi-physician, stratification has nothing to do with feasibility?)
- Page 7 line 9-10: "Fixed-effects regression takes patient-level sampling variation and stratification by centre into consideration" - maybe to be further explained
- Page 9 - 1st line section Methods: I'm not sure "to pooling" is the appropriate verb
- Page 11 - line -7: "Under a commonly ..." - long sentence!
- Page 23 – line 1-3: “The fixed-effects model had extremely similar performance compared with the fixed-effects model balanced design, …” – not clear
4) According to me, Figure 1 is not necessary and could be deleted.

**Level of interest:** An article of importance in its field

**Quality of written English:** Needs some language corrections before being published

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

**Declaration of competing interests:**
I declare that I have no competing interests.