Reviewer's report

Title: A comparison of methods to account for centre effects in multicentre trials with a binary outcome

Version: 1 Date: 2 August 2013

Reviewer: Hester Floor Lingsma

Reviewer's report:

This paper is about the handling of treatment effect in RCTs. The authors present an example of an existing trial and perform a simulation study to compare different methods to adjust for centre effects. They conclude that GEE and random effect models the preferred methods, compared to fixed effect models and Mantel Heanzel.

The topic of the of the paper is important, as RCT results have often large implications for clinical practice and thus should be ‘correct’. The methods are decent and the paper is clear and very well written.

Major Compulsory Revisions:

1) The role of the re-analysis of the MIST trial is not completely clear. Does it serve as a case study to show what can go wrong? In that case that should be made more clear in the introduction? Or is it part of the results? It now does not come back in the discussion. Maybe change the order? Firth the ‘theoretical’ findings from the simulation flowed by the confirmation in a case-study.

2) In the methods (introducing the example of MIST2) and discussion the authors refer to the case were there is no block randomization. ‘Although centre was not used as a stratification factor in the randomisation process, trialists may wish to account for centre effects regardless.’ Is adjustment for centre also required, or recommended, when there is no stratification by centre? In other words, would you recommend a random effects model to analyze any multicentre RCT?

3) Results of the MIST2: as a comparison it would be nice to see the estimated treatment effect without adjustment for centre (as reported in het main paper of the trial?). Do we expect this to be ‘correct’ as there was no stratification for center in the randomization? (Relates to point 2)

4) The results for the ICC of 0.075 are not shown. Are they different from ICC 0.025?

5) The block size does not seem to affect the type I error and power? Can we conclude something from that? Would be interesting to mention in the results.

6) The authors recommend to use either random effect or GEE. Can they also recommend one method? Or else under which circumstances they would choose GEE and when random effects?
7) In my experience many clinicians and (conservative) trial statisticians are reluctant to use ‘complicated’ methods. They seem to be afraid of being accused of ‘datamassage’ or making the interpretation of the trial too complicated. Do the authors have an opinion on how to ‘sell’ the approaches they studied to trialists? Maybe they can say something about that in the discussion.

8) Related to that: many trials require pre specification of the analysis, which also makes it difficult to use models with specific assumption. The authors write: Random-effects models generally make the assumption that the centre-effects follow a normal distribution (although other distributions could be used). This may not be known in advance, and may be an unrealistic assumption. Would they recommend to test this assumption? And use another method if it is not met? I do not think I would but I am interested in the opinion of the authors.

Minor Essential Revisions:

9) ‘We consider the endpoint need for surgical intervention up to 90 days, a major secondary outcome.’ Why not the primary endpoint of the trial?

10) We set the treatment effect to 0 to evaluate the type I error rate, and set the treatment effect to give 80% power based on the sample size and event rate in order to evaluate power.

The second part of this sentence is hard to understand. Suggest to make two sentences and to formulate a bit clearer.

11) We used ICC values of 0.025 and 0.075

Where are these values based on?

Discretionary Revisions:

12) Paragraph 1.2: ‘successes or failures’ Change in ‘with or without the outcome’?

13) You use the word ‘nominal’ for the true(?) expected(?) type I error rate. This word is not familiar to me. I have no direct suggestions but explain it or use another word or at least indicate the ‘nominal’ value (5)

Level of interest: An article of importance 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 have no competing interests