Reviewer's report

Title: Clustering in surgical trials - database of intra-cluster correlations

Version: 1 Date: 30 September 2011

Reviewer: Allan Donner

Reviewer's report:

Major compulsory revisions

i) The nested structure of the clustering, surgeons within trials, is ignored in these calculations. For example calculating the ICC’s independently for surgeons and trials implies that for the ICC for centres also contains within-surgeon clustering effects, an approach which is potentially misleading. A more natural approach that exploits the nesting would compute the ICC within trials (between surgeons) and the ICC within surgeons. This is perhaps why it is stated at the top of page 6 that “ICC estimates appeared to be generally slightly smaller for surgeon than centre levels. One would expect intuitively that it should be the reverse—that is, one would naturally expect greater similarity among patients of the same surgeon than among patients merely in the same centre.

ii) For some ICC’s it is stated that if bootstrapped ICCs were not calculable (with possible reasons for this not given) a normal distribution for the ICC was used. Aside from the fact that the variance of this distribution is not specified by the authors, it is well-known that the distribution of the ICC is severely non-normal, which is why a transformation proposed by Fisher is commonly recommended for continuous outcomes. A resulting problem regarding the interpretation of these results is that the CI’s for the “normal” approach and the bootstrapping approach presented in Table 2 are not comparable. This is reflected in the observed results, where the widths for the former approach are usually narrower than the widths for the latter approach (not surprising given the parametric nature of the “normal” approach).

iii) The ICC’s and the lower limits of all the CI’s presented have been truncated at 0. Some justification for the decision to truncate should be given beyond the brief statement given at the bottom of page 6, since as the authors acknowledge, the resulting point estimates and corresponding upper CI limits will be inflated.

iv) The expressions for the design effects of the different designs presented briefly at the bottom of Table 4 should be presented in the text, with some accompanying explanation. Also the same symbol is used at the bottom of page 4 to denote an ICC within clusters and the correlation between observations in different clusters (relevant to expertise-based an stratified trials).

v) The authors state that a common ICC was assumed across intervention groups for the purpose of their calculations. They acknowledge, however, that differential
clustering effects may exist across interventions. A resulting problem is that the ICCs for those trials characterized a strong intervention effect exists will tend to be inflated (due to increased between-cluster variation) as compared to trials with weaker intervention effects. Thus it would be more appropriate to compute ICC’s that are pooled across intervention groups.

Minor essential revisions

1. On both pages 3 and 6 it is stated that for a continuous outcome the failure to account in the analysis for the stratification in a stratified within-surgeon design leads to a loss in precision. But surely this applies to any outcome, not just a continuous one.

2. The authors appear to assume that in multicenter surgical trials the centre is invariably modeled as a random effect. But in most multicenter trials centre is modeled as a fixed effect, in which case clustering does not arise as an issue. This should at least be discussed.

3. The average number of surgeons per trial should be given in the Methods section along with the other information provided describing the features of the database.

Level of interest: An article of limited interest

Quality of written English: Not suitable for publication unless extensively edited

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

Declaration of competing interests:

I declare that I have no competing interests.