Reviewer’s report

Title: Imputation strategies for missing binary outcomes in cluster randomized controlled trials

Version: 1 Date: 16 November 2010

Reviewer: Michael Campbell

Reviewer’s report:

Review

This paper investigates the issue of missing binary values in cluster randomised studies. In particular it considers multiple imputation of the missing values. It considers three within cluster and two across cluster strategies. The issue of missing values in cluster trials has not been considered extensively and so this paper could be a welcome addition to the literature.

The problem is well defined and the authors have made a contribution to its solution.

Comments

Major Compulsory Revisions

1) The paper comes to a fairly unsurprising conclusion that one should take into account intra-cluster correlation when imputing missing binary values. However it seems to suggest that the point estimates are little affected even when there is 50% missing, with the curious exception of the RE logistic regression (either GEE or RE). The CIs are somewhat too narrow although again rather oddly the no-imputation GEE method does not do too badly even for 50% missing. This needs further exploration and explanation.

2) The main difficulty in this paper is that they use real data, which they progressively censor. The CHAT study had relatively large clusters (55 patients in each cluster). Thus even with 50% missing, the chances of missing an entire cluster are not high. Because of intra-cluster correlation, one would not expect missing values to have to much impact if a large proportion of a cluster is still present. Thus I am not sure to what extent one can generalise these results to other situations. This needs to be spelled out more clearly.

3) P19 line 14 It is not clear from the way this is written that the fact that the CIs becomes narrower is a poor result and (usually having narrower CIs is a good thing) and clearly they are biased.

4) P22 Conclusions only spell out what was done, and are very weak. They should be much clearer.

Discretionary Revisions
1) I don’t think the way they present Tables 3 and 4 is particularly clear. It may be better to express the entries as percentages of the values obtain without the imposed missingness (allowing for whether they used GEE or RE). The entry RE logistic regression and then GEE as an analysis model is a bit confusing.

2) I am not sure we need 6 columns of data to show the effects. I think 3 would do (say 5%, 20% and 50%).

3) Although their literature review is quite extensive they omitted three recent useful reviews from this side of the Atlantic [1-3]

4) P11 line-7 Is cluster added as a fixed effect? If so, is this inconsistent with the random effect model?

5) P11 line 11, line -2 It would be useful in an appendix to actually have the SAS code for these procedures.

6) P14 section 3.2 Why were these risks for missingness chosen?

7) P15 Were the kappa statistics averaged over the 5 imputations and if so how?

Minor Essential Revisions

1) Spell out the acronym ABB (P7, P9)

2) Minor English problems (P9 line 12,13)’ _The_ propensity score.’ P15 line -7 ‘Rubin’ P18 line -7 “Kappa’s” P20 l-9 ‘effect is consisted’ P21 line -4 ‘shall’=>’should’

References


Level of interest: An article whose findings are important to those with closely related research interests

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