Reviewer’s report

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

Version: 1 Date: 17 November 2010

Reviewer: Alexis Turgeon

Reviewer’s report:

This study, which compares different imputation strategies for CRTs using simulated missing data is of high interest for anyone involved in research using CRTs. The manuscript is well written, but is very long. In order to improve focus on the main message, the introduction, the methods and the results section should all be shortened by 50%. A final version of the manuscript of about 15 pages should be the objective. Several specific comments on the manuscript should be considered prior to publication.

- Major Compulsory Revisions

1. The introduction, should end with a brief sentence summarizing the study design. The last paragraph of the introduction should be moved to the methods section.

2. At the beginning of the methods section, a paragraph should be added to briefly present the database used to perform the comparative analyses (2-3 sentences) followed by a paragraph on the comparative methods studied (3-4 sentences), similar to the last paragraph of the introduction section (would benefit to be moved here, as stated above).

3. In the results, the section on MCAR should be shortened by 50% to avoid redundancy with the tables and figures which are clear, concise, and speaks for themselves. The paragraph on CD is a good example.

4. The decision to use the kappa statistics to compare the different imputation methods is questionable. Multiple imputation is intended to be accurate for generating unbiased parameter estimates but are not supposed to be accurate on an individual level. In addition, as kappa statistics for the simulation with 10% missing data are calculated on 90% identical data, the results aren’t very surprising.

5. The presentation of OR and CI for each simulation make it difficult to compare results. Estimates of standardized bias and coverage would be more informative and would also provide a quantitative guideline to assess the adequacy of imputes – standardized bias of >40% is considered to have potentially adverse effects on bias and coverage.[1]

7. The discussion section is of adequate length, but would benefit from a more...
concise summary of the main findings followed by a comparison of study findings with the current literature, and a discussion of study’s main limitations.

8. Finally, the conclusion section of the manuscript does not really represent the main findings. The conclusion states: ‘Results from the simulation study show that MI strategies which take into account the intra-cluster correlation are more appropriate when the amount of missingness or the intra-cluster correlation are relatively large’. In the current study, MI was tested under a fixed (not very large) ICC. This section should discuss whether one of the tested methods might be preferable.

- Minor Essential Revisions

1. There are frequent grammatical errors in the text, especially in the Discussion section.

2. Abstract: the background section of the abstract would benefit from being shortened by half; three sentences would be sufficient. The term ‘RCT’ was probably mistaken for ‘CRT’ at the end of the background section. Numeric results should be given in the Results section of the Abstract.

3. Did you try just adding dummy variables representing the clusters in the imputation model? This has been suggested as a method for addressing cluster data and is a lot simpler than those proposed.

4. MI not accounting for clusters would be expected to give unbiased effect estimates but to underestimate variance. In addition, MI estimates should have wider CIs than those calculated on the complete data set due to accounting for the uncertainty of missing data. From what you observed, across cluster imputation with RE logistic regression leads to a potentially biased estimate. This is unexpected and should be discussed in the manuscript.

5. It is true that the cut-off of 50% does not work well for binary variables with proportions close to 0 or 1. What was the proportion of the outcome variable in this study?

6. The adequacy of imputation methods that do not account for clustered data does depend on the ICC, which was quite low in this study (0.055). Was the cluster effect in the random-effects model statistically significant? Could you comment on this in the Discussion?

7. Which resampling method was used to perform the 1000 simulation runs?

8. Can you give information on the chains and priors used for MCMC and how convergence was verified? It may also be useful to know the relative efficiency of 5 imputes and intra versus inter-imputation variance.

9. It would be important to mention that the results cannot be generalized to MAR or NMAR data, which are much more common. Moreover, the best option under the MCAR and CD scenarios tested in this study is clearly to do nothing!
10. You state that MI is the best technique (Intro, 4th paragraph), but the ‘best’ method might depend on the data and the analysis to be performed with the data. For example, maximum likelihood estimation of parameters in the presence of missing data is also a good technique but is currently restricted to linear models.

11. In the methods, we suggest replacing ‘No imputation approach’ with ‘complete case analysis’

12. The within cluster strategy leads to much wider CI than the across cluster strategy. What are your thoughts on this? May the former lead to too much variation as it only uses a fraction of the data?

13. You imputed for missing data on individual patients. Any comments on missing data for entire clusters as mentioned in the Intro/Discussion? Do you think your results apply?

14. There appears to be a contradiction between two sentences in the Discussion: ‘Moreover, among the two across-cluster MI strategies, the RE logistic regression method provided much wider the 95% CIs for the treatment effects compare to the propensity score method.’ and ‘Fifth, the logistic regression model, and MCMC method were slightly better than propensity score methods since they yielded larger kappa statistics and precise estimated treatment effects’

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:

'We declare that we have no competing interests'