Reviewer’s report

Title: A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: A simulation study

Version: 0 Date: 06 Apr 2017

Reviewer: Henry Mwambi

Reviewer’s report:

A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: A simulation study

By: Anurika Priyanjali De Silva, Margarita Moreno-Betancur, Alysha Madhu De Livera, Katherine Jane Lee, and Julie Anne Simpson

Based on this paper, I request that the corresponding authors should consider the comments below and possibly effect the corrections where necessary. Other than this the paper is well written, clear message and useful to researchers faced with the problem of missing data in longitudinal studies.

1. On page 3, line 48 to 50. Studies had been conducted to investigate the performance of two-fold FCS and MI methods for handling missing data in a time-varying covariate. What you did was to re-visit the methods.

2. On page 5, line 96. MI is a two-stage process. It should be: MI is a three-stage process: Imputation, analysis and combining.

3. On page 5, line 97-98. Stage 1 the incomplete dataset is replicated multiple times and missing values are replaced by values drawn from a suitable imputation model using the available. The sentence is the same. It should be: In stage 1, each of the missing observation is replaced by M > 1 plausible values according to the imputation model.

4. On page 5, line 98 to 101. In stage 2 the target analysis is performed on each of the imputed datasets with the resulting parameter estimates and corresponding standard errors of each dataset, combined into a single estimate (and standard error). The statement contains information for both stage 2 and 3. It should be: In stage 2, the complete datasets are analyzed using standard statistical methods such as GEE for longitudinal data. Stage 3, The results of the M analyses are combined into a single inference (parameter estimates and standard errors).

5. On page 6, line 199. When there are a large number of variables and time points, simulation studies have shown that both MVNI and FCS in their standard form face convergence issues. Comment: can you specify how large the number variables?
6. On page 17, line 342. The simulations didn't reveal too large biases for any of the approaches in any of the scenarios. Check pages 29 and 30. Is the relative bias under the complete case of MAR (weak and strong) not too large?

7. Page 18, line 356. They reported unbiased and more precise estimates…. Who are they? Please change the word 'they' or re-write to improve clarity.

8. In eqn (2): Explain the inclusion of random effects which leads to the mixed effects formulation. What I mean is the model a random sintercept, random slope, random intercept and slope or others.

9. In line 257: Correct notation to be \( \nu_{0,j} \) and \( \omega_{0,j} \) and not \( \nu_{0,i} \) and \( \omega_{0,i} \).

10. When you say number of imputations equal to percentage of individuals with missing data. Do you mean that if the percentage is 30\%, then the number of imputations is 30? Please clarify.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal