Author's response to reviews

Title: Intracluster correlation coefficients for the Brazilian Multicenter Study on Preterm Birth (EMIP): methodological and practical implications

Authors:

Giuliane J Lajos (giuliane@unicamp.br)
Samira M Haddad (samira.m.haddad@gmail.com)
Ricardo P Tedesco (rp.tedesco@yahoo.com.br)
Renato Passini Jr (passini@caism.unicamp.br)
Tabata Z Dias (tabatazumpano@hotmail.com)
Marcelo L Nomura (mlnomura@unicamp.br)
Patricia M Rehder (pmprehder@uol.com.br)
Maria H Sousa (mhestat@ceicamp.org.br)
Jose G Cecatti (cecatti@unicamp.br)
Brazilian Multicenter Study on Preterm Birth study group (jgcecatti@terra.com.br)

Version: 2 Date: 25 March 2014

Author's response to reviews: see over
Dear Editor of the
BMC Med Res Method

First of all we would like very much to thank you and the reviewers for their important comments on our manuscript. Certainly they will be helpful in improving its quality. We are now submitting a new version of the manuscript incorporating the corrections and suggestions arising from this review. Below we are addressing all comments and suggestions (in bold red) and the correspondent modifications in the text are with tracked changes. We hope they will fit well the journal’s expectations.
Sincerely yours

Giuliane J Lajos

Title: Intracluster correlation coefficients for the Brazilian Multicenter Study on Preterm Birth (EMIP): methodological and practical implications

Reviewer's report
Version: 1
Date: 13 January 2014
Reviewer: Abhik Das

Reviewer's report:
Cluster based public health studies, including cluster or community randomized trials are increasingly common in global public health research and evaluation, and a common problem in designing such studies is the lack of published estimates of ICC from large well done studies conducted in diverse settings. The premise of this study is thus quite promising and its results can be very useful for future cluster based studies (though one limitation is that it is restricted to one specific form of clustering -- hospital based).

OK, thank you!

Major Compulsory Revision:
1. The motivation presented by the authors in the Introduction can be further strengthened, both by better explaining why this is important (specifically, there is a surprising lack of mention of cluster randomized trials, which are perhaps the most high impact form of public health research/evaluation study design that can benefit from good extant estimates of ICC) and what the ICC is (the algebra provided is perhaps less useful than an education on the concepts underlying the ICC and why clustering is important to account for).

OK, we agree and both were included in the introduction session.
2. There is no information presented in the Methods section on how the crucial ICCs were calculated and whether there was any modeling or adjustment done. Although the desire to be exhaustive is understandable, in reality ICC estimates are more helpful
only for variables that can be thought of as potential outcomes in other studies and not covariates.

OK. Now, as requested, the complete formulas we used from the software are presented in the method session. Besides determining the ICC values, the objectives of this study were also to estimate the prevalence or mean for all variables of the study, including the covariates. This is now clearly stated in the objectives.

Level of interest: An article of importance in its field
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

Reviewer's report 2
Version: 1
Date: 14 January 2014
Reviewer: Mohamed Shoukri
Reviewer's report:
Background: Cluster-based studies in health research are increasing. An important characteristic of such studies is the presence of intra-cluster correlation, typically quantified by the intra-cluster correlation coefficient (ICC), that indicate the proportion of data variability that is explained by the way of clustering. The purpose of this manuscript was to evaluate ICC of variables studied in the Brazilian Multicenter Study on Preterm Birth.

OK.

Major Compulsory revisions:
1-Confidence intervals on the ICC depend on the standard errors, which in turn depend on the type of variable being measured. Therefore, for normally distributed data we have an expression for the SE that is very different from the SE of ICC when the data are binary. The SE expression for both cases should be shown.
In fact, the variance estimate for ICC is obtained by an extensive asymptotic formula, and because this it was not showed. This has already been described in another article and we thought that these very specific details would be not so interesting and worth for the majority of the journal’s audience.

2- There are no methodology presented in the paper
In fact, the session on methods and specifically the topic on analysis was too short and some of the details regarding the formulas used were in the introduction session. Now we changed them for methods session and we also added some details regarding the processes used to get the numbers presented.

3- I took the liberty of calculating the DFF for some of the variables, and to my disappointment, many were wrong.
OK, we can understand. But in fact the calculations were performed in a much more complex way using asymptotic formula from the statistical software. Now this is better explained in the correspondent text. Be sure the estimations are not wrong.
We checked again every value and they are OK. In fact, it is not possible to recalculate these values (in the way we did) without having the full database.

4- If you plan to conduct a cluster-sampling based study, which includes many covariates, it was reported that the estimated ICC is affected by the number of covariates and the way such covariates are measured (on the cluster level, or individual within cluster level).

This was taken into consideration and the text now is clearer on this.

Discretionary revisions:
1- It might be useful to see how the cost of obtaining clinical or laboratory data affect the design effect.

We agree that this information could be interesting, however the study did not collect any information regarding costs and then this analysis cannot be performed at all.

Minor revision: None
Level of interest: An article of limited interest
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests: 'I declare that I have no competing interests'

Reviewer's report 3
Version: 1
Date: 26 January 2014
Reviewer: Martin Gulliford

Reviewer's report:
This is a generally well written and well structured paper that reports on ICC estimates from a study of pre-term births in Brazil. The data will be useful and merit reporting, with some reservations:

OK, thank you.

Major Compulsory Revisions
1. Sampling of individual participants is not explained in detail. 'Preterm' birth has not been defined; it is not explained how births were determined to be pre-term; there could be considerable misclassification in this regard that may vary between centres. The period over which data were collected is not explained.

A single-stage cluster sampling was used and this information is now included in the methods session. The definition of preterm birth and period of data collection were included in the methods session.

2. The statistical methods used to estimate ICC values and their confidence intervals are not given. Presumably, one way analysis of variance was used. The formulae given in the Introduction appear to derive from survey sampling literature, and are not those
more commonly used in respect of cluster trials, which is the application of greater relevance in terms of the literature cited.

**Now more detailed information on methods and formulas used for getting the values presented are available on methods session.**

3. The mean cluster size is presented but it may also be relevant to present the coefficient of variation of cluster size (depending on how sampling was implemented), as this may be important to include in estimation of anticipated future design effects. (Paper by Eldridge in Int J Epidemiol)

**Considering a single-stage cluster sampling was used, we consider that reporting the coefficient of variation for cluster size would not be very important and, on the other side, would perhaps to make the article more difficult for clinicians to understand.**

4. Each outcome should be distinguished as binary, ordinal or continuous. Assumptions made for ordinal variables should be stated. It is not clear that some of the variables have been appropriately analyzed. For example, number of caesarian sections with a mean of 0.3, could be better treated as binary. The majority of variables reported in this study are binary, excepting those from tables 2 and 8, where the means were calculated (also including some discrete quantitative variables).

5. The paper could be more selective in the choice of material. ICC values are generally useful for measures that might be an outcome variable in a future study. This is not the case for many of the variables presented. What is the rationale for example, if presenting ICCs for sociodemographic measures such as access to piped water? This point has already been explained before. Besides determining the ICC values, the objectives of this study were also to estimate the prevalence or mean for all variables of the study, including the covariates. This is now clearly stated in the objectives.

6. In terms of data presentation, it may be useful to present a figure showing the frequency distribution of the ICC values. For binary measures, ICC could be plotted against prevalence (up to 50%, reversing above), as prevalence has been reported to be associated with ICC magnitude in other studies (1: Gulliford MC, Adams G, Ukoumunne OC, Latinovic R, Chinn S, Campbell MJ. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. J Clin Epidemiol. 2005 Mar;58(3):246-51. PubMed PMID: 15718113.)

**We thank the suggestion. However, taking into account we opted for reporting details on a large number of variables (more than 250), we decided to present all related information in tables, with some additional details on categories of these variables in the text. We also presented results for some few continue quantitative variables.**

7. The paper needs to do more to discuss the generalisability of the findings. Would the results apply in private hospitals in Brazil? Or in a populations study? Would they be applicable in another country?
Good point. We now included in the discussion more emphasis in the fact that we have a non-probabilistic sample from the centers (hospitals). Therefore, strictly speaking, the findings cannot be generalized to other populations. However, the majority of hospitals included in the study was third level referral hospitals taking care of high risk pregnancies and preterm babies. Probably the results would be applicable to other centers with such characteristics, irrespective of being private or public, especially in middle income countries like Brazil.

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

Quality of written English: Acceptable

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

Declaration of competing interests: I declare that I have no competing interests