Author's response to reviews

Title: A Simulation Study of Sample Size for Multilevel Logistic Regression Models

Authors:

   Rahim Moineddin (rahim.moineddin@utoronto.ca)
   Flora I Matheson (mathesonf@smh.toronto.on.ca)
   Richard H Glazier (rick.glazier@ices.on.ca)

Version: 2 Date: 28 March 2007

Author's response to reviews: see over
Dear Iraxe Puebla:

We thank the reviewers for their thoughtful comments on our paper. Provided below is our response to the reviewers on our manuscript A Simulation Study of Sample Size for Multilevel Logistic Regression (MS: 2074989190130766). We hope that we have successfully addressed all concerns. I look forward to hearing from you.

Rahim Moineddin, Ph.D.
Department of Family and Community Medicine
University of Toronto
Toronto, Ontario
Rahim.moineddin@utoronto.ca
(416) 946-5860
Response to Reviewers

Reviewer: wan tang
1. This paper is totally based on simulation studies. The lack of theoretical justification of the findings in the simulation studies seriously weakens the importance of the paper.

Response: We agree with the reviewer that our results are based on simulation studies as is the case with the majority of sample size studies for multilevel models. This is indicated in our title so that it is clear to readers and is also indicated in the paper itself. Theoretical work on sample size calculation is beyond the scope of this article due to the complexity of multilevel logistic regression.

2. All the simulation studies carried out in the paper are based on very stringent assumption and this weakens its relevance and application to real study data. For example, both the individual and group explanatory variables xij and zj are generated from the standard normal, and the group random components u0j and u1j are also assumed normal. For practical purposes, various different classes of distribution other than the normal distribution should be investigated to make the results more applicable to real study designs. Even under the normal assumption, different values for the parameters should also be considered. For example, the variance of xij and zj may be different and their difference may affect bias. Without a thorough study of the behaviors of the estimators under different settings, it may be too hasty to make the suggestions and recommendations as the authors did in the conclusion of the paper.

Response: Our simulation assumptions along with our rationale for and the practicality behind choosing the number of groups, group size, and ICC were based on previous research [Maas, C. J. M. & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, 58, 127-137, & Maas, C. J. M. & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1, 85-91]. Using the standard normal distribution for generating explanatory variables is a fairly common practice in simulation studies. Varying the distributions and variances of the explanatory variables and the random components is related more to issues of robustness rather than issues around sample size calculation in multilevel logistic regression. Examining the impact of different values of the parameters - number of groups (3 choices), group size (3 choices), ICC (3 choices), the variance and distribution of the explanatory variables at the individual level (9 combinations), the variance and the distribution of explanatory variables at the group level (9 combinations), the distribution and the variance of the random intercept (9 combinations), the distribution and the variance of the random slope (9 combinations) - on the sample size would be impractical due to the extremely large number of combinations (177,147) which would result in extensive computer time..

3. The sample size recommended by the authors does not look practical. This may depend on what magnitude of bias may be considered a serious problem. For example, in many
situations, a 10% bias may well be acceptable. Such a degree of bias in general does not change the sign of the estimate, and thus can correctly suggest the association under study. In table 1, serious bias only occurs for the estimates of variance parameters ?0 and ?1 when the group size is 5, which makes it difficult to concur with the authors’ recommendation.

Response: We agree with reviewer that the acceptable level of bias may vary from one researcher or one project to another. There is no universally acceptable threshold for bias therefore we believe that our results are general enough to allow researchers to choose a sample size for their studies based on the level of bias that the researcher considers acceptable for their study. We have added a line in the discussion section of the paper about this possibility (page 13).

4. For the samples where convergences were not achieved, it is fine to discard the estimates of the variance and distribution of explanatory variables at individual, parameters (if there are any). But the nature of these samples should be investigated. In other words, why the estimation did not converge for these samples? There is no report about any investigation on this important issue in the paper.

Response: The nature of the differences between the converged and non-converged samples was evaluated and the results are included in the convergence section (page 6). We also provide some explanation of why some of the samples did not converge in the discussion (page 12).
Reviewer: Christophe Croux

Minor Essential Revisions
It would be interesting to give a kind of minimum group size and minimum number of groups that one should have before multilevel logistic regression is trustworthy. The authors already discuss this to some extent, but I think that it would be good to add a short paragraph on this.

Response. We thank the reviewer and have provided additional discussion of this issue on page 13.

Often, multilevel logistic regression models are applied on rather small data sets, where one can expect convergence problems, large biases, and inadequate statistical inference procedures ... A warning needs to be given to the practitioner.

Response: We provided a caveat to the researcher with regard to this issue (please see page 13).

The authors explain well what they are doing. I do not have many comments on this. There are only very few typo's. The authors could do a final reading of their paper again to eliminate those.

Response: Thank you for pointing this out. We have proof-read the paper for typographical errors.

The authors could add that all evidence is simulation based, and discuss in a few lines the limitations of their study.

Response: We point to the fact that our evidence is based on simulation studies and provide some of the limitations of the study on page 12.

The discussion (Section 4) is quite lengthy. In particular page 11 is too descriptive and firm enough in its conclusions.

Response: Thank you. We have shortened the discussion section (see pages 11-13).