Reviewer's report

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

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:

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.

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.

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)

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 analysed. For example, number of caesarian sections with a mean of 0.3, could be better treated as binary.

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?

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).

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?

**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