Author's response to reviews

Title: Ongoing monitoring of data clustering in multicentre studies

Authors:

Lauren B Guthrie (lauren_guthrie@hphc.org)
Emily Oken (emily_oken@hphc.org)
Jonathan AC Sterne (jonathan.sterne@bristol.ac.uk)
Matthew W Gillman (matthew_gillman@harvardpilgrim.org)
Rita Patel (rita.patel@bristol.ac.uk)
Konstantin Vilchuck (sevenhos@mail.belpak.by)
Natalia Bogdanovich (magicienne@tut.by)
Michael S Kramer (michael.kramer@mcgill.ca)
Richard M Martin (richard.martin@bristol.ac.uk)

Version: 3 Date: 18 November 2011

Author's response to reviews: see over
November 8, 2011

To the Editors,

Many thanks for the opportunity to resubmit our paper, “Ongoing monitoring of data clustering in multicenter studies.” We also appreciate the detailed reviews, which have strengthened the manuscript. We have addressed all reviewers’ comments point-by-point below.

All authors have again reviewed and approved the manuscript and take full responsibility for the manuscript. None of the authors have any conflicts of interest or financial disclosures to declare. None of the manuscript content has been previously published and there are no prior submissions with any overlapping information. The data used for the analyses presented in this manuscript have not been published previously. This manuscript is not and will not be submitted to any other journal while under consideration by BMC Medical Research Methodology.

Thank you for handling our submission and for considering our manuscript. We look forward to hearing your response.

Sincerely,

Emily Oken, MD, MPH
Response to Reviewers:

**Reviewer 1: Emilie Vierron**

1) Do the authors think that there can also be a learning curve, necessary for standardizing the measurement, thus inducing a center effect at the beginning (as a center=a pediatrician for most of centers) and, once this learning period is finished, inducing a decrease in systematic measurement errors?

We agree with the reviewer that there likely is a learning curve in measurements that may improve precision over time. However, it is uncertain whether this learning curve will result in better accuracy and consistency of measurement across sites. We have now addressed this question on pp 10-11.

2) Data were collected during the most recent follow-up in 2008-2010 and ICCs were calculated all along this follow-up but the different dates of correction and recalculation of ICCs are not detailed in the method section (they just appear in table 1). More generally, it would be interesting, to understand the methods used, to have information on the moments chosen to question and train pediatrician, the number and dates of ICC estimation, the number of centers included in each estimation etc… Even if some of these subjects are detailed later, all along the result section.

We have now included this information in pages 6-7.

3) Authors do not describe the software that was used to perform ICC estimations in the method section, could it be provided?

We now report that we used SAS version 9.2 (page 6)

Moreover, it would be interesting to associate confidence intervals (CI) to these estimations. Thank you for this suggestion. We have calculated CI’s using the multivariate delta method, and include some sample CI’s in the text (page 4, page . We have opted not to include them in the table as we thought it would be too busy and not provide much additional information.

4) At the beginning of the result section, authors say: “we checked for implausible values or extreme outliers that might represent transcription or data entry errors”. What has been decided for these extreme values? Was it impossible to check in the patient reported forms to correct these data and avoid any transcription or data entry error?

We reported all implausible values to the data entry team in Minsk, who made the appropriate corrections (page 6).

5) The first center effect authors describe is not a real center effect as it comes from the wrong use of instrument measuring sitting height, then inducing a translation of data of 50 cm. This example should be shortened to put the emphasis on other variables such as SBP or waist or head circumferences. 6) Same remark: in March 2009, authors observed elevated ICCs for sitting height for the same reasons as detailed in the previous paragraph. This is not a real center effect and this point does not bring more information on the subject. Here, ICC can be used to help the data management, (by adjusting data of some centers by 50 cm or 10 cm) and this point should
be exposed in the manuscript in a single paragraph. Moreover, a single box plot graph, as figure 1a, is sufficient to describe the problem encountered. Figure 1c does not bring any information as we already know that ICC in October 2009 was very low. Another subject of interest is to know how ICC estimation can be used as a useful tool for standardize data collection. Results could be segmented according these two axes of work.

We have now deleted Figure 1c. However, we do feel it is useful to highlight the example of sitting height, as similar situations may result in smaller absolute differences that may not be so readily apparent in other studies, as we discuss on page 10.

7) Concerning this second axis, for example, ICC of waist circumference was elevated in 09/08. Here it should be interesting to see the box plot graph showing variations in all centers.

We have included this figure as Figure 2.

8) As well described, authors identified some sources of systematic measurement errors and pediatricians were trained to standardize measurement. However, some ICCs, as for systolic blood pressure or triceps skinfold thickness, remain high (more than 10%) at the end of the study and there is too few explanations for that. It should be interesting to discuss such result, to suggest reasons, to compare them with others papers (for example Center effect on ankle-brachial index measurement when using the reference method (Doppler and manometer): results from a large cohort study. Vierron E, Halimi JM, Tichet J, Balkau B, Cogneau J, Giraudelau B; DESIR Study Group. Am J Hypertens. 2009 Jul;22(7):718-22) because it is surprising to observe such high ICCs with semi-automated device as the Omron. Moreover, some measurements may be very difficult to standardize across centers, then it seems essential that statistical analyses take account of this for variables having a high ICC (around 10%).

We agree and have added a discussion of the importance of accounting for ICC in the analysis, even when the value is not so high (page 11), and have discussed the different findings in our paper compared with the one cited above.

Minor Essential Revisions

9) The sentence “a box-plot graph of SBP suggested some clustering by clinic” do not provide any information since we know that the ICC is elevated. Do the authors mean that there are variations between all centers (by opposition to one or two centers very different from others, as for sitting height)?

We have now clarified that a handful of polyclinics had higher blood pressure measurements than the others.

Reviewer 2: Mirjam Moerbeek

1) In the statistical literature the term multicentre studies is often used for trials that randomize patients within centres to treatments and all treatments are available within one centre. Cluster randomize trials randomize complete clusters to treatments and one treatment is available per cluster. The authors use the term multicentre trials in the title of their paper but use a cluster randomized trial as an example on page 4.
We have now clarified that we include a cluster-randomized trial as an example, but that this approach would be relevant for all multicenter studies (p3-4).

2) It is obvious that measurement error can lead to biased conclusions with respect to the effect of treatment. Would it be possible to show this on basis of an actual data analysis? The authors could perform two analyses for the outcome variable sitting height: one with and another one without corrections and then compare the results of the analyses. We now include results of the suggested analysis on page 12.

3) As variables are almost always measured with some error an ICC reflects true differences between centres plus measurement error. On page 4 the authors write “A high ICC indicates higher than expected variation due to clustered measurement”. So to use the ICC to indicate measurement error it might be very helpful if a benchmark value is available to judge if an ICC estimate is indeed “high”. The authors were able to compare the ICC for sitting height, head circumference and triceps skinfold thickness to a previous follow up measurement. Also they were able to compare the ICC for waist circumference to those for hip and head circumference that were measured using the same instrument (measuring tape). However they did not have a benchmark for systolic blood pressure but apparently they considered a value of 0.18 large enough to warrant further investigation. What would they have done if they observed a value of, say, 0.15 or 0.12? It appears as if some experience and subject matter knowledge is required. Could the authors give some general guidelines on how to proceed if a benchmark is not available? We have included a paragraph addressing this comment on page 11.

4) Related to the previous comment: it could be helpful if references to papers that list ICC estimates could be given to provide the reader with at least some information of ICC values. A useful paper for this purpose appeared in 2004 in American Journal of Public Health and was written by Murray and co-authors. We have now referenced this useful paper, thanks very much!

Reviewer 3: Brent Burch

Major Compulsory Revisions

1. The authors state that an intracllass correlation coefficient (ICC) value of zero indicates that the data is not clustered by centre which implies that ICC = 0 is desired for a multicentre study. This is certainly true in a geographic region where the physical characteristic under consideration is for a homogeneous population. If different racial/ethnic groups live in areas served by just a few clinics, it is possible that ICC values of certain physical characteristics may not be close to zero. The authors should address or rule out this possibility. We have addressed this comment on page 11.

2. Continued monitoring of the data enabled the authors to correct for procedural discrepancies for many but not all of the continuous measurements. What would the authors do differently (or propose for the future) to address these unresolved issues?
While it is optimal to perform measurements centrally, have quick data entry and reporting, and rapid correction, the practical challenges of conducting a study in a country such as Belarus, and in fact in most distributed settings, make some measurement issues unavoidable. We have addressed this comment on page 11.

Minor Essential Revisions

1. Methods section of Abstract, second sentence. The sentence is a bit terse so consider “An ICC of 0 indicates that the variance in the data is not due to variation between the centres and thus the data is not clustered by centre.”
We have made the suggested change to the abstract.

2. Quantifying Clustering – The ICC section of Methods, first paragraph. Change “observed variance” to “variance” in three occasions. #b2 and #w2 are parameters which can be estimated but not observed.
Done

3. Quantifying Clustering – The ICC section of Methods, second paragraph. Consider replacing the third and fourth sentences with “Suppose the centres are in the process of measuring a specific physical characteristic of a homogeneous population. A high ICC would indicate higher than expected variation due to clusters of measurements between centres. If detected early, the reasons for the clustered measurements may be corrected.”
We have made the suggested changes.

4. The Promotion of Breastfeeding Intervention Trail (PROBIT) section, last word in paragraph two. Change “technique.” to “techniques.”
Done

Discretionary Revisions

1. Because there are many ways to estimate the ICC, the authors could cite a reference that contains the estimating equation. For example, the following reference is useful. Donner A: A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. International Statistical Review 1986,54:67-82.
Thank you for this suggestion, we have now included this reference on page 4.