Reviewer's report

Title: Ongoing monitoring of data clustering in multicentre studies

Version: 2 Date: 14 September 2011

Reviewer: Emilie Vierron

Reviewer's report:

This study uses ICC as a useful tool evaluating uniformity and quality of measurement methods in multicenter studies. This is an original work, clearly presented and written.

Major Compulsory Revisions

1) ICCs were periodically assessed: 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? This point might be added to discussions about points of the protocol that had not been correctly explained to pediatricians.

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 latter, all along the result section.

3) Authors do not describe the software that was used to perform ICC estimations in the method section, could it be provided? Moreover, it would be interesting to associate confidence intervals (CI) to these estimations. For example, Ukoumunne provides some methods of CI estimations in A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials. Stat Med 2002;21:3757-3774.

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?

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.

7) Concerning this second axe, 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.

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. Vierren E, Halimi JM, Tichet J, Balkau B, Cogneau J, Giraudseau 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%).

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)?

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