Author's response to reviews

Title: Bias Corrected Estimator for Intraclass Correlation Coefficient in the Balanced One-Way Random Effects Model

Authors:

Eshetu G Atenafu (e.atenafu@utoronto.ca)
Jemila S Hamid (jhamid@mcmaster.ca)
Teresa To (teresa.to@sickkids.ca)
Andrew R Willan (andy@andywillan.com)
Brian M Feldman (brian.feldman@sickkids.ca)
Joseph Beyene (beyene@mcmaster.ca)

Version: 3 Date: 12 June 2012

Author's response to reviews: see over
June 11, 2012

Adrian Aldcroft
Editor
BMC Medical Research Methodology

Re: **MS: 7207078545857472 - Bias Corrected Estimator for Intraclass Correlation Coefficient** (Eshetu G Atenafu, Jemila S Hamid, Teresa To, Andrew R Willan, Brian M Feldman and Joseph Beyene)

Dear Mr. Aldcroft,

Thank you very much once again for the second round of referee reports of our manuscript entitled “**Bias Corrected Estimator for Intraclass Correlation Coefficient**”.

Please find an itemized list of responses along with the revised manuscript. We used *italics* for the referees’ comments (verbatim) and regular font for our responses.

Thank you once again for the opportunity to submit this manuscript to BMC Medical Research Methodology.

Sincerely,

Joseph Beyene, PhD
Associate Professor of Biostatistics
John D. Cameron Endowed Chair in the Genetic Determinants of Chronic Diseases,
Program in Population Genomics, Department of Clinical Epidemiology & Biostatistics,
Faculty of Health Sciences, McMaster University,
1280 Main Street West, MDCL 3208, Hamilton, ON L8S 4K1, Canada
Referee #1:

1) The authors agree that it is hard to extend the study to the complex two-level data (unbalanced 1-way nested classification), the three-level data, etc. Therefore I wonder whether the authors should change the title to match the content, for example, like "Bias Corrected Estimator for Intraclass Correlation Coefficient of the balanced 1-way nested classification".

We agree with the reviewer and have changed the title of our manuscript to: “Bias Corrected Estimator for Intraclass Correlation Coefficient in the Balanced One-Way Random Effects Model”
Referee #2:

I cannot agree with the authors when they claim that “Different ICCs are required for different purposes, for example to assess agreement or consistency.” In fact, ICC has been often used as measure of agreement; however, ICC is a measure of reliability (1). Actually, no type of ICC can be used as measure of agreement.

An important limitation of ICC (any type) as measure of agreement is that it is strongly influenced by the variance of the trait in the population in which it is assessed (2,3). An article about ICC cannot ignore this well-known ICC limitation as measure of agreement.


We thank the reviewer for the above references. We have examined these references carefully and also took time to review various interpretations and uses of ICC as a measure of agreement and reliability. We share the reviewer’s concern that this particular index might be misinterpreted, and we have now added this point in our discussion and cited the last two of the above references as well as two more references cited below (please see last sentence of our discussion in the revised manuscript). Having said this, we would also like to highlight the following key points:

1) The main contribution of our manuscript was investigation of the statistical performance and operating characteristics of a new bias-adjusted estimator compared to existing and widely used estimators for one special case (ICC based on one way random-effects ANOVA model). As we have showed through extensive simulations, the proposed estimator improves performance by reducing the known bias that the standard method exhibits;

2) For example, the 2nd paper cited above (Muller and Buttner, 1994) correctly points to the inappropriateness of the Pearson correlation coefficient as a measure of agreement. Note that we cited this paper in our original submission. The use of correlation coefficient as a measure of agreement is one of the most common misinterpretations/misuses of statistics that one might encounter in medical research. However, Muller and Buttner also argue that ICC is a more suitable and widely used measure of agreement. They acknowledge that the term “agreement”
can be ambiguous and they propose to replace it by “consistency and conformity”.

3) We refer the reviewer to another interesting review paper by Barnhart et al 2007 
(Barnhart HX, Haber MJ, Lin LI. An overview on assessing agreement with 
paper, the authors discuss measure of agreement in detail and discuss ICC along 
with concordance correlation coefficient (CCC). They note “reliability assessed 
by ICC is a scaled agreement index under ANOVA assumptions”.

4) The same group (Chen C-C, Barnhart HX. Comparison of ICC and CCC for 
assessing agreement for data without and with replications. Computational 
Statistics and Data Analysis, 2008, 53: 554-564) stresses that “ICC is widely used 
for assessing agreement between observers, which is defined in terms of variance 
component with respect to the specific ANOVA model”. Like we did in our 
manuscript, they note that particular ICC can only measure consistency rather 
than agreement (please see second to last paragraph, page 555 of their 
manuscript).

5) Finally, we note that the paper by the reviewer (Costa-Santos et al., J Clin Epi 
2011) is an interesting empirical evaluation of uses and interpretations of ICC 
compared with the well-known limits of agreement by Bland and Altman in the 
context of one clinical area, i.e., neonatology. We believe that this contribution 
brings up an important opportunity to debate diverse views on the foundational 
methodological, statistical and interpretation aspects of ICC as a measure of 
agreement. As we stated above, the main focus of our manuscript was on the 
statistical properties of a well-established index. Having said this, if the reviewer 
strongly feels that ICC should not be used as a measure of agreement in all 
circumstances, we are happy to propose to the editor for an opportunity to have an 
open debate where diverse views and opinions can be expressed by experts in this 
area.