Author's response to reviews

Title: Bias Corrected Estimator for Intraclass Correlation Coefficient

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: 2 Date: 18 March 2012

Author's response to reviews: see over
March 17, 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 for the 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) Would you like to extend this study to the complex two-level data (unbalanced 1-way nested classification), the three-level data?

We thank the reviewer for this idea. In a complex multilevel setting, resampling-based approaches such as the bootstrap might prove to be useful for estimating the intraclass correlation coefficient. Derivation of analytical expression for complex scenarios does not appear to be straightforward, but we agree with the reviewer this is worth pursuing as one potential research direction in the future.

2) Please check and revise the eq (3.4).

Thank you. We now use $\hat{F}$ (not $F$) in equation (3.4) and the equation immediately after it.

3) The simulation shows bias reduction both in normal and non-normal data scenarios, but all the equations is only from the normal distribution in the appendix.

The reason we conducted simulations both under normal as well as non-normal settings is to assess the performance of our proposed method under model misspecification.
Referee #2:

The authors present a well written and well presented paper proposing a bias corrected estimator for the intraclass correlation coefficient using second order Taylor series expansion.

We thank the reviewer for the kind comment.

Major Compulsory Revisions

1) The authors stressed in the introduction that ICC plays a fundamental role when studying inter-rater reliability. In the discussion they also stressed that ICC is an important measure of reliability, consistency and agreement of measurements. However, ICC as measure of agreement have been strongly criticised in the literature. The authors should clarify in the paper in which situation the ICC is an important and useful coefficient: is it useful for agreement, for consistency or for reliability assessment?

We agree with the reviewer that confusion can arise when using and interpreting ICC because there is no one ICC that fits all situations. Different ICCs are required for different purposes, for example to assess agreement or consistency. The classic paper by Shrout and Fleiss (1979) provides six types of ICCs depending on the question of interest. We highlight this point in paragraph 4 of the “Introduction”, on page 2.

Here is a quote from Shrout and Fleiss that summarizes this crucial point succinctly:

“The guidelines for choosing the appropriate form of the ICC call for three decisions: (a) Is a one-way or two-way analysis of variance (ANOVA) appropriate for the analysis of the reliability study? (b) Are differences between the judges' mean ratings relevant to the reliability of interest? (c) Is the unit of analysis an individual rating or the mean of several ratings? The first and second decisions pertain to the appropriate statistical model for the reliability study, and the second and the third to the potential use of its results.”

Reference:

Minor Essential Revisions

2) Abstract: “Taylor serious expansion” should be “Taylor series expansion”

Done.
Discretionary Revisions

3) The inclusion of a real study could help to clarify the usefulness of ICC and consequently the usefulness of proposed bias corrected estimator for the ICC.

Unfortunately, we don’t currently have a real study that we can use for illustration. Although it would be very interesting to apply the various ICC estimators to real data sets and assess performance empirically, we believe that the simulation results provide a much useful perspective into how much bias can be adjusted, since the simulations were done under various reasonable scenarios.

For example, let us consider the illustrative example data set that Shrout and Fleiss (1979) used to demonstrate the various ICCs presented in their paper (Table 2 in the original paper and shown below). The data is based on measurements on 6 “targets” made by 4 “judges”.

<table>
<thead>
<tr>
<th>Target</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>2</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>5</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

In their Table 4, Shrout and Fleiss present 6 different types of ICC estimates that range from 0.17 (for the simple ICC that is based on one-way ANOVA and the ICC for which we proposed a bias-adjusted estimator) to 0.91.

For this data, using our bias-adjusted estimator led to an ICC of 0.17 (compared to 0.16 that is reported by Shrout and Fleiss). For this particular example, the difference is not marked, but for other data sets there might be a much more pronounced difference (as highlighted in our simulations).
Once again, we thank the reviewer for the comment. If the reviewer finds it useful to include the numerical estimate from our new ICC estimator when applied to the data shown above, we will be happy to do so.