Author's response to reviews

Title: Assessment of variation in the Alberta Context Tool: The contribution of unit level contextual factors and practice specialty in Canadian pediatric acute care settings

Authors:

Carole A Estabrooks (carole.estabrooks@ualberta.ca)
Janet E Squires (jasquires@ohri.ca)
Alison M Hutchinson (alison.hutchinson@deakin.edu.au)
Shannon Scott (shannon.scott@ualberta.ca)
Greta G Cummings (greta.cummings@ualberta.ca)
Sung Hyun Kang (sunghyun.kang@nurs.ualberta.ca)
William Midodzi (wmidodzi@epicore.ualberta.ca)
Bonnie Stevens (bonnie.Stevens@sickkids.ca)

Version: 2 Date: 17 August 2011

Author's response to reviews: see over
16 August 2011

Dr. Sudha Xirasagar
Journal Editorial Office
BioMed Central

RE: Assessment of variation in the Alberta Context Tool: The contribution of unit level contextual factors and specialty in Canadian pediatric acute care settings

(MS: 1046690238504274)

Dear Dr Xirasagar:

Thank you for the opportunity to revise and resubmit this manuscript. We have addressed the comments requested. Each comment, along with our response and reference to where the issue is addressed in the revised manuscript, is itemized in the table below. We have maintained ‘tracked changes’ in the manuscript as requested for the revisions. If you require further clarification please do not hesitate to contact me.

Thank you.

Sincerely,
Janet E Squires RN, PhD
Ottawa Hospital Research Institute

cc
Carole A. Estabrooks
Alison Hutchinson
Shannon Scott
Greta G Cummings
Sung Hyun Kang
William Midodzi
Bonnie Stevens
**Response to Reviewers**

<table>
<thead>
<tr>
<th>Reviewer Dolf Boer</th>
<th>Major Compulsory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reviewer comment</strong></td>
<td><strong>Response</strong></td>
</tr>
<tr>
<td>The assessment of validity is explained in statistical terms, but lacks clarification in conceptual terms. For example, at page 11, it is stated that larger values of n² and w² indicate stronger validity. Why is this the case and which type(s) of validity do these measures address?</td>
<td>We have expanded conceptually on our explanation of validity. We have specifically addressed both questions posed in this comment regarding: (1) why larger values indicate stronger validity (see page 12-13) and (2) type of validity. The type of validity is ‘relations to other variables’ validity evidence in the “Standards for Educational and Psychological Testing framework” which contributes to construct validity (see page 6 for description of the Standards for Educational and Psychological Testing framework and page 13 for answer to the question of the “type(s) of validity do these measures address”.</td>
</tr>
<tr>
<td>And are any alternative assessments of validity worth exploring?</td>
<td>The measures referenced above represent the standard aggregation indices that assess reliability and validity of aggregated data. In addition to aggregation, there are many other ‘validation’ assessments. We have previously conducted both a traditional validity assessment of the ACT (which was presented in the background and used the same data of the current paper, focus was on principal components analysis). Since the initial submission of the current paper, we have also conducted and published a second psychometric assessment of the ACT (in long-term care) using confirmatory factor analyses. In this paper we focus on an additional component (not part of either of these previous assessments) – assessment of unit-level discrimination ability of the ACT. We have added a section to the current paper summarizing the two past reports and references to the papers (pages 5-7).</td>
</tr>
<tr>
<td>I also understand that the between-unit variance is essential in this context, but again the authors do not seem to explicitly explain or justify this.</td>
<td>We have added a sentence to page 8 following the study purpose and objectives, explaining why between-unit variance is essential in this context. [“While assessment of between-unit discrimination and variance is not a traditional form of validity testing, it is essential to understanding the construct validity of instruments like the ACT that collect data at the individual (respondent) level with the purpose of aggregating those responses to obtain higher (e.g., unit) level estimates.”]</td>
</tr>
</tbody>
</table>
Further, the validity of the ACT is only demonstrated for nurses as sample sizes for other health care professionals were limited (page 8). This issue deserves consideration in the discussion section.

We have added a new section to the discussion entitled “Aggregation of the ACT concepts” addressing this comment. See page 24 in the Discussion.

The most widespread form of the ICC is: unit-level variance / (unit-level variance individual-level variance). The authors employ other forms of the ICC and it would be helpful if they address the differences of the ICCs employed with the more widespread form.

We have added a paragraph under ‘Analytic Approach’ (Reliability and Validity of aggregated data at the unit level) addressing this comment and explaining our rational for choosing the approach we took (see page 13 starting with “There are multiple methods for calculating intraclass correlations (ICC).”). We have also attached a second additional File (Additional File 2) that shows the similarity in estimates obtained between the random coefficient (multi-level) model approach (mentioned in the reviewer’s comment) and the one-way random-effects ANOVA model that we used at that phase of the analysis.

The authors state that when the ICC(1) of an independent variable exceeds 0.1, they specified this independent variable as a unit-level variable on the grounds that there is sufficient agreement between patients within groups (page 12). This surprised me as the ICC typically ranges from 0 – 1, suggesting that there are too many differences between individuals when ICC(1) =0.1. ….

In addition, although some variables are clearly intended to measure unit-level variables, they are measured at the level of individuals, and individuals may differ in their assessment of these variables. As such, specifying these variables as unit-level variables seems odd and requires further justification.

We have addressed this comment in the paper on page 12 under Analytic Approach (Reliability and Validity of aggregated data at the unit level) with respect to a discussion on ICC(1). Specifically, we offer increased support from the literature with respect to the ICC(1) values that are usually found in applied research studies. Additionally, 0.10 is the accepted standard in applied organizational research for justifying aggregation of individual level data (we have clarified this in-text).

Most of the analyses consisted of multi-level models. However, whether there were differences between units in dependent variables was assessed using ANOVA (page 17). In addition, the ICC’s are based on an ANOVA model rather than a multi-level model (page 10). The authors do not justify why they did not perform these analyses within a multi-level framework, for example by comparing the two-level model with a model without levels or by reporting the

We have explained in the paper our rationale for this (see Results - page 19). The findings referred to by the reviewer are preliminary to the multi-level analysis and are meant to be descriptive and used to inform the multi-level modeling. We did report the significance of the unit-level error variance in the second phase of our analysis (i.e., in the multilevel analysis, See Table 5).
Further, it is unclear whether the caterpillar plots are based on the results of the ANOVA or on the results of the multilevel analyses. The authors are free to choose 95%-intervals or the smaller version suggested by Goldstein and Healy, as long as they specify their choice.

We have clarified this in the paper (see Results - Variability in the Dependent Variables, page 20-21). [“The caterpillar plots (Figure 1) were generated using the null hierarchical linear models and 95% confidence intervals; the MLwiN 2.12 program was used to generate these plots.”]

<table>
<thead>
<tr>
<th>Minor Essential</th>
</tr>
</thead>
<tbody>
<tr>
<td>The authors link their research to other publications in the discussion section. However, this does not include findings regarding the dependent variables, possibly because the authors claim that “there are few validated measures of organizational context”. Although this may be true, it would still be interesting if the findings regarding the dependent variables – particularly concerning ICC’s and discrimination between units – could be linked to the few other validated measures or to measures that have not been formally validated.</td>
</tr>
<tr>
<td>We have addressed this in the manuscript by adding two new paragraphs in the discussion under the section titled 'Discrimination between patient care units’ (see pages 25-27).</td>
</tr>
</tbody>
</table>

At page 18 in the second paragraph, the effect of predictor variables is assessed in terms of the between-individual variance. This, I did not understand. When interested in measuring organizational context it seems to make sense to focus on the effect of predictor variables on unit-level variance or the ICC, rather than between-individual variance.

Also, the description of the formulas below table 6 seem incomplete to me. Tau0 and tau1 appear to refer to unit-level variance of the null model and model 1. From the description, I also derive that sigmap is the same as tau1, but this is probably a misunderstanding because why would the authors introduce different symbols for the same things? Further, it seems as if sigma0 is not explained. I

We have reworded that section for increased clarity (see Results - Analysis of individual predictors (Model 1), page 21). |

We have corrected this (See table 6). |
must admit that I do no regularly assess formulas but nevertheless, some more clarification on these issues may be helpful to the reader.

**Discretionary**

<table>
<thead>
<tr>
<th>At page 14 Eij is described as the individual offset from the grand mean. Presumably, the authors mean the individual offset from the unit mean or unit offset, since the individual offset off the grand mean would also include the unit offset, which is already captured in another term of the model.</th>
<th>We have corrected this (See page 17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>At page 18, the authors report the ICC’s as a coefficient in the tables, but as a percentage in the text (first paragraph). This confused me initially; it would be better to be consistent in this regard.</td>
<td>We have corrected this so that all text appears as coefficients to be consistent with the tables (see pages 21-23).</td>
</tr>
<tr>
<td>In the second paragraph at page 18, the authors refer to table 5, but since they report data from table 6 in that paragraph, they presumably mean table 6.</td>
<td>This is a report of data is Table 5. We have therefore not changed this. It reflects data in Column 4 of Table 5 and we have now specified that in the paper, see page 21.</td>
</tr>
<tr>
<td>In table 7, the reference groups of categorical predictors are not specified.</td>
<td>We have added the reference groups to Table 7.</td>
</tr>
<tr>
<td>In table 4 the percentages of the age category 20-24 years seem incorrect for the categories surgical, medical and critical. These are no way near the overall percentage. Further, the percentages of the overall category do not seem to add up to 100%.</td>
<td>This was a mistype. The percentage for age category (20-24 years) for overall is 10% (not 1%). It has been corrected.</td>
</tr>
</tbody>
</table>