Reviewer's report

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

Version: 1 Date: 7 June 2011

Reviewer: Dolf Boer

Reviewer's report:

General comments
The focus of the article is an instrument to measure organizational context and measuring organizational context is important to advance research into determinants of research utilisation as the authors claim, but also institutional performance. In addition indices of organizational context may also be useful for assessing the impact of policy changes focussed on elements of organizational context. As such, this material is interesting and worthy of publication. The authors explain the setup, methodology, statistics and results of the study at length and the paper is generally well written. Nevertheless, I found some parts of the methodology difficult to follow although I do work with multilevel models regularly. In addition, it should be noted that although the methodology used seems appropriate, it is quite complicated as well and will be difficult to read for a large part of the audience.

Major compulsory revisions
The authors present their results as evidence for the validity of the Alberta Context Tool (ACT) (page 20). 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? And are any alternative assessments of validity worth exploring? 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. 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 and the authors should also discuss whether they expect that the number of health care professionals per unit – other than nurses – is sufficient for the purposes of the ACT.

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.

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 still many differences between individuals when ICC(1) =0.1. Accordingly, it would appear that much information is lost by specifying such variables as unit-level variables. 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.

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 significance of the variance at the unit-level. Further, it is unclear whether the caterpillar plots are based on the results of the ANOVA or on the results of the multi-level analyses. The plots should be based on posterior means from the multi-level model (also known as empirical Bayes estimates or shrinkage estimators). In addition, the authors should mention how the confidence intervals in the caterpillar plots were calculated. Goldstein and Healy suggest smaller confidence intervals than 95%-intervals for the purpose of graphical representation (see: http://www.bristol.ac.uk/cmm/team/hg/full-publications/1995/graphical-presentation-of-a-collection-of-means.pdf). The authors are free to choose 95%-intervals or the smaller version suggested by Goldstein and Healy, as long as they specify their choice.

Minor essential revisions
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.

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 must admit that I do not regularly assess formulas but nevertheless, some more clarification on these issues may be helpful to the
Minor discretionary revisions

At page 14 $E_i$ 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.

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.

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.

In table 7, the reference groups of categorical predictors are not specified.

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%.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests