Reviewer's report

Title: Screening for hypoglycemia at the bedside in the neonatal intensive care unit (NICU) with the Abbott PCx glucose meter

Version: 2  Date: 15 August 2006

Reviewer: Marek Brabec

Reviewer's report:

General
This is an interesting paper, whose significance is very much clear and whose practical implications are certainly important. The more so that it uses real clinical data analysis to reach general conclusions about precision of PCx meters as compared to the lab analyzer results (as to a reference method). There are methodological problems here, however. Basically all regression results that are really shown in the paper seem to be based on ordinary least squares (OLS), without correction for correlation among the repeated measurements taken on the same individual. Importance of the (intraclass) correlation is underlined by the unbalance of the data (e.g. Table 1 shows that number of measurements ranges from 1 to 23 !). It is hard to be sure about the actual method used because the model used for the analysis is never exposed in a formal way (e.g. as a regression equation with explanation of the assumptions on the error term), one can only guess from several comments and from the text appearing in the paragraph â€œEffect of repeated measures on subjects and difference in operatorsâ€. From there, one can also guess that the authors are somewhat aware of the well known fact that there tends to be a correlation among measurements taken on the same unit and that its presence might spoil the statistical inference severely, but it seems that this awareness did not press them to use appropriate statistical tools. If OLS was used, it is clearly inadequate and does not follow current approaches to the repeated measures (or longitudinal data). For a popular exposition, see e.g. Diggle, Liang, Zeger (1999): Analysis of longitudinal data. Oxford. If something else than OLS was used to produce the results (Table 2, Table 3, etc.), it should be explained carefully in the methods section (in an commented equation form).

All this is in sharp contrasts to authorsâ€™ previous terse response to Reviewer 4 question â€œ... I would suggest a statistician review the methods and results, ...â€, stated as: â€œThe statistical analyses were performed by ... a statistician.â€. After a careful reading, it really seems that an experienced statistician did some calculations in line with the modern approaches to repeated measurements data (e.g. those based on â€œmultilevel regression modelingâ€ quoted in â€œStatistical analysisâ€ part of the â€œMethodsâ€ chapter), but for some strange reason, the results did not find the way to the tables, figures and to text of the paper. Instead, only general (and fairly vague) comments about the effects of adjustments for repeated measures being taken on subjects appear in a very short section â€œEffect of repeated measures on subjects and difference in operatorsâ€. No figures are stated there. Further, there is no description of the model used to assess the effects or to do the adjustments, so that the reader has to believe (or not) that bold statements like â€œThe same variables were significant in all comparisons with values very similar to the adjusted model (table not shown).â€ This is not a very appealing approach. Instead, I would suggest that the model should be stated fully (including assumptions on error and its covariance structure) and that the now-common longitudinal data analysis in one form or other should be applied and its results displayed in the paper (in place of the OLS results). The more so that it seems that the results are sitting somewhere, they are just not being utilized. Resulting coefficient estimates will be probably quite similar to those obtained by OLS, but the inference might be substantially different. Larger standard errors and larger p-values are to be expected when appropriate methodology will be used. One only hopes that the interesting and practically important nice effect of hematocrit will remain significant!

The phrase â€œmultivariate regressionâ€ is used many times throughout the text, where just â€œmultiple regressionâ€ is meant almost surely.

One of the previous reviewers suggested to split the data into two parts (for lower and higher glucose concentration) below and above 4.2 mmol/L to see the stability of the results in region of the most practical relevance. The suggestion is very reasonable and it is nice that the authors followed it in a revised version of the paper. Namely, the values of the coefficients in the regression model with all the data and the low sub-part should be compared carefully. On the other hand, the formal inference for less than 4.2 mmol/L (as represented by p-values and standard errors in Tables 2, 3) is very suspicious and can be invalid easily if OLS (or even GLS) was used. This is because the low-data are probably quite skewed and the underlying assumption of normality (or at least â€œreasonable symmetryâ€) might not fulfilled even approximately. Appropriate analysis would not be all that easy (as it would involve both truncation and correlation in the
data). Therefore, I would recommend not to report the inference-related quantities for the artificial sub-group analyses. Just the coefficients can (and should) be listed and discussed. Alternatively, this interesting statistical problem can be attacked by maximum likelihood (with the need to formulate not-quite-standard model and probably to program it into a piece of software).

The Passing-Bablok regression equations are mentioned without quoting any reference formally. It seems that there is a misprint on page 8: The regression equation listed in the sentence starting with “The comparison between...” should probably read RN PCx=0.92 (LAB PCx) -0.58, shouldn’t it?

Figures are rather busy and can probably be made a bit more clear (perhaps through a different choice of symbols). It would be nice to have a nonparametric smooth (e.g. loess or something similar in nature) superimposed in Figure 1 and Figure 2 (nonparametric fit would be done separately for RN PCx and LAB PCx). Especially Figure 3 is very hard to read in detail. Perhaps, magnified left part (which is most important from practical point of view) can be shown in an additional figure. Also the legend to the Figure 3 (on page 20) is hard to read. It seems that there might be even some typos there. Should not be “Vitros 950” replaced somehow (as the Vitros corresponds to the reference method to which RN PCx and LAB PCx are compared)?

It is nice that the authors tried to figure out the cut off value based on comparison with the reference method. The number of cases below the important 2.6 mmol/L was rather small (only 25, fortunately for the patients), however. Larger number from several centers would be needed for having greater faith in the resulting recommendations.

As a statistician, I concentrated almost exclusively on statistical methods. Certainly, the paper is interesting for subject matter and practical connotations and should be suitable (after some revisions, clarifications and corrections) for publication.

-------------------------------------------------------------------------------
Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

-------------------------------------------------------------------------------
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

-------------------------------------------------------------------------------
Discretionary Revisions (which the author can choose to ignore)