Reviewer’s report

Title: Predicting mortality in sick African children: a clinical bedside risk score from the FEAST trial.

Version: 1  Date: 27 April 2015

Reviewer: Jeeva Sankar

Reviewer’s report:

Major revisions

1. The need for a new score has not been brought out well in ‘Introduction’. What are the pitfalls in the existing scores and why do we need another score – to simplify, make it more feasible to record, to have less number of variables, or to ensure better discrimination? The authors cite two reasons – lack of validation for some of the scores and the need for a “practical risk score based only on clinical bedside measures that can be easily and quickly identified…” The first does not warrant development of a new score; one can easily validate the existing scores and identify the best among them. The authors themselves undermine the second reason by adding four laboratory variables to the ‘clinical’ score.

2. The original study – FEAST trial – from which the score was derived, showed a significantly higher risk of mortality in the intervention (bolus) arm. In that case, it does not seem to be a good idea to combine the data from both the arms – intervention and control – to develop a new score. The authors did try to address this issue partially by adjusting for the intervention arm and looking for an interaction effect. But these are not ‘fool-proof’ techniques that can reliably exclude a possibility of small but meaningful interaction. Using the data from only the control arm of the trial would have been ideal (of course it would reduce the available sample size). The authors can still run the analysis using only one group and examine if the results are any different (‘sensitivity analysis’).

3. It is not clear why Cox proportional hazards regression model was preferred over a simple stepwise logistic regression model. Cox model is used for a time-to-event dependent variable. If the variable to be predicted is mortality at 48 hours after admission, there is no need to use it as a time-to-event variable in the analysis (we are not interested in that). Unlike logistic regression that estimates the odds, Cox proportional hazards regression model estimates the hazard – instantaneous event rate for an individual who has already survived to time t –, which is difficult to interpret. The authors did censor the data at 48 hours or at the time of leaving hospital. But the need for censoring is questionable because the number of children who left the hospital by 48 hours is very small (n=11). Given the facts, a simple stepwise logistic regression would have been a better choice.

4. The regression coefficients obtained in the multivariate model for different predictor variables do not differ much – they ranged from 0.63 to a maximum of 1.53. It would be good to compare the results of the univariate analysis (of the predictor variables) with that of the multivariate model.
5. The final score used 8 clinical variables. The authors excluded only two predictor variables that were significant from the final model. Ideally, a risk score should be derived from a parsimonious model with the least number of variables whose discriminatory ability is almost the same as that of the full model. It is not clear if the authors tried to exclude more variables and examined its effect on the discriminatory ability.

6. The final model includes ‘lung crepitations’ as one of the predictor variable. There are at least three pitfalls in including this variable. First, it requires training to correctly identify lung crepitations. Interobserver variability is likely to be high even among trained pediatricians. It is not clear how other health providers can accurately pick up this sign in the emergency room. Second, having both respiratory distress and lung crepitations in the final model does not seem to be a good idea. There must be some collinearity between these two variables. It would be interesting to omit ‘lung crepitations’ from the model and examine how the discriminatory ability changes. Third, the variable could have emerged significant because of inclusion of participants from the intervention (bolus) arm. The predictor variables were measured at or within 1 hour of randomization. It is possible to have lung crepitations because of fast fluid boluses in the intervention arm.

7. Having laboratory features in the final model would make it difficult to use at the bedside. The complex methods of imputation techniques (net reclassification index; NRI) do not necessarily make it appropriate to include in the final model. The authors can restrict to a simple clinical tool with the least number of variables that can be easily identified by any health worker.

8. The final score had a good discriminatory ability among general pediatric admissions (AUROC 0.86) but not among admissions to a high dependency ward (AUROC 0.82). Given that the score was developed from a population of sick children with shock and life threatening infections, one would have expected a good discriminatory ability among the admissions to HDU and not among general population. Does this discrepancy suggest that the score is likely to perform well in only those with a very low risk of mortality? Ideally, triaging using a risk score would be needed in emergency rooms or HDUs and not in relatively healthy children being admitted to general wards from an outpatient setting.

Minor revisions
1. Definitions for many predictor variables have not been provided in the manuscript.

Discretionary revisions
1. Discussion needs to be trimmed.

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests:

I declare that I have no competing interests