Reviewer's report

Title: Stratification of the severity of critically ill patients with classification trees

Version: 1 Date: 27 March 2009

Reviewer: Michael Racz

Reviewer's report:

Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

1) In Figure 1, continuous variables were put into at most 2 categories. The two other CTs had categorical variables in 2 or 3 categories. This is just an observation which you can comment on if any of the tree algorithms have any rules about the maximum number of categories these variables can be divided into.

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

1) In the abstract it’s stated the C4.5 CT has 7 variables. Aren’t there only 6?

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

1) This is a paper which compares the results of three methodologies for classification trees (CTs) and compares them to a logistic regression model and three severity scores. Although the utility in the ease and simplicity of the CTs is appreciated, the comparisons may not be the best way to justify them. It’s clear from Table 5 that the external scores (APACHE II, SAPS II and MPM II-24) need to be re-calibrated and I don’t think they should be compared to internally developed models and CTs. I don’t understand the usefulness of taking the classical scores and applying them to both the development and in the validation sets. A suggestion might be a separate table which, when applying the severity scores to the entire database, show the measures of discrimination, calibration, etc.

2) The process of developing a model or CT on a development set and then testing on a validation set takes place for one logistic model and 3 CTs. An obvious concern that’s made apparent by Figure 4 plots and the SMR calculations on the validation data set is how much logistic under predicts. One issue in making this comparison is that the continuous variables in the CTs are all converted to categorical. There may be some continuous outlier points which are influencing the predicted outcome when these are entered as continuous in logistic regression. I think another
logistic model should be added where the continuous variables are entered as categorical sub-groups minus the reference group. Also, you might emphasize more the variables selected by logistic regression that were not selected by any CTs and vice-versa.

3) This data is over ten years and there may be a hidden time component. Are there certain types of patients, for example, that had a much higher mortality rate in 1997 than in 2006? Please address this. A plot over time might be suggested or adding a time variable to the model and/or CTs.

4) If multiple CTs are proposed, which is the one to choose? This might be difficult to answer, but some attempt should be made to see how similar the probabilities are that they produce. Please construct a correlation matrix of the probabilities assigned to each patient for each of the CTs for the development and validation data sets. Adding how they correlate with the logistic regression probabilities might also be helpful.