Reviewer’s report

**Title:** A straightforward approach to designing a scoring system for predicting outcome of critical patients

**Version:** 2  **Date:** 24 July 2012

**Reviewer:** Thomas L Higgins

**Reviewer’s report:**

This study employed a naive Bayes approach to develop a simple scoring system to predict prolonged LOS in the ICU after heart surgery. In addition to developing the score (and an excellent AuROC), the authors provide a proof-of-concept for rapid updates of scoring systems. This is an area of concern. I believe there is a need for standardized systems to benchmark results across institutions, and over time, while acknowledging that models age and require updating at regular intervals (perhaps 5 to 10 years). While a model that can be fine-tuned to a local environment (for internal benchmarking) is useful, there needs to be some objective standard for cross-institutional comparisons. I would look to SAPS-III as an example of a model that is generally useful, but also accommodates regional differences.

**Major compulsory revisions:**

Abstract: "model designed in a specific institution is easily modified" - I'm not sure that this is a good goal. I would rephrase it as a method for adjusting models over time, or for internal purposes only. There is, however, value in being able to modify a LOS model for a specific hospital, when the model is being used primarily for internal purposes such as OR scheduling.

Thus, I also disagree with the statement at the bottom of page 3 that "Any scoring system should therefore be locally adapted". The three references provided are from the same group as the authors; this is not a widely-held viewpoint.

**Minor essential revisions:**

Are "low postoperative cardiac output" and "suitable postoperative cardiac output" truly independent predictors? I agree that the former is a risk factor, and the latter is protective, but it seems like the two variables are opposite sides of the same coin.

Page 3 - I don't think models are used for "withdrawal of treatment" very often, although the data might inform a family discussion.

Page 3 - Most published models are not simple - they are MLR’s (APACHE 4, SAPS 3, MPM3, some cardiac surgical models such as Cleveland Clinic, STS) I'd change the sentence to read "In many cases, these models are.."

Page 3 - First para of background, last 2 sentences - could be moved to discussion.
Page 12 - Bayes often, but not always outperforms more sophisticated models. MLR incorporating interaction terms is likely to deliver better performance.

Background (page 3) Suggest 2nd sentence read "One of their purposes is to aid...." Models are more often used for benchmarking/retrospective analysis than in clinical decision-making. The exception might be LOS models which can be used to predict resource utilization and modify scheduling.

Optional revisions:
On page 3, the "significant loss in model performance" should be referenced - the Murphy-Filkins reference #10 would be appropriate.
The cutpoint of 120 hours for LOS (page 8) could be referenced to other studies that have chosen this; I don't believe there is a standard, however.
The SPSS and MATLAB versions and city of manufacture should be cited.
I am not familiar with the term "confusion matrix" (page 10), but more conventionally this is called 2x2 matrix or decision matrix.
Page 13 - I'd like more detail on the "result was obtained by use when we used the Cleveland scoring system" What result? This sentence is not clear as it stands.
Table 3 - how do "Valve substituion" and "Aortic valve substitution" differ? Is the former any valve (mitral, tricuspid, pulmonic, aortic)?
Table 3 - addition of blood to the CPB circuit is somewhat provider dependent unless there are clear indications for transfusion.
Table 4 - I would note that ECC time and Aortic clamping time are not truly independent variables, although this may not matter with the Bayes technique.

Level of interest: An article whose findings are important to those with closely related research interests

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.