Reviewer's report

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

Version: 3 Date: 16 July 2013

Reviewer: Thomas L Higgins

Reviewer's report:

This manuscript presents an approach to developing and updating a model for length-of-stay. While I still object to the concept of individual institutions customizing mortality benchmarks, there is some merit to being able to customize a length-of-stay predictive model as it can be applied to an institution’s patient flow needs.

Major compulsory revisions:

The title should be changed to "A Straightforward approach to designing a scoring system for predicting length-of-stay in cardiac surgical patients" While LOS is indeed an "outcome", similar literature generally considers mortality as the primary outcome. The abstract and introduction talk about outcome prediction without mentioning that a customizable model is most useful for internal purposes (although this has been added to the background section of the abstract).

I still think that low postoperative cardiac output and suitable postoperative cardiac output are two sides of the same coin, and it’s redundant to say one is a risk factor and the other is protective. If different cut-points for the same continuous variable are being used, then these should be defined in the text. (Typically, <2.2 L/min/m2 is considered a low cardiac index)

The second paragraph of "background" more properly belongs in the discussion. I think a distinction should be made between scoring systems for mortality and scoring systems for LOS. One needs stable benchmarks to conclude if high-quality care is being delivered across institutions. For LOS, one can make the argument that a customized model can be useful for internal purposes. The authors have added a line about operating room scheduling (appropriately) but there is a need for more clarity on which types of models should be customized. Standardization has its benefits when doing cross-institutional comparisons.

Minor points:

120 hours (5 days) would be an extreme outlier for cardiac surgical postop ICU LOS. One might argue that a linear model might be more useful than a logistic regression to a single endpoint if the goal is predicting bed utilization.

Page 12, discussion. The sentence on "Scoring systems are therefore generally used....developed" could use a reference. (Kramer editorial in CCM on "Models
are not like fine wine”, for example, or the MPM-III or APACHE-IV models both mention this). BTW, MPM-III and APACHE-IV are both sophisticated models that incorporate interaction terms. (same page, later)

Figure 1 - I wonder if the same information might better be displayed with an ROC curve (Sensitivity vs. 1-Specificity) which is the more traditional method of displaying discrimination of a model.

**Level of interest:** An article of limited interest

**Quality of written English:** Needs some language corrections before being published

**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.