Reviewer's report

Title: External validation of a Cox prognostic model: principles and methods

Version: 2 Date: 26 November 2012

Reviewer: Frank Harrell

Reviewer's report:

Major Compulsory Revisions

The authors have chosen to perpetuate the problematic practice of categorizing continuous risk predictions. They cling to the notion that patients and physicians want this, which is not bolstered by data or practice. To make this point clear consider an example. A patient is seeing her physician who tells her that her risk of an untoward event is between 10% and 20%. The patient asks the physician "What is the best estimate of my actual risk?" to which the categorizing physician replies "As I said, between 10% and 20%". The patient persists, asking "Did anyone estimate the risk before it was grouped into an interval? My other doctor tells me my blood pressure, not what range it's in." The physician now says "All right, your estimated risk is 10%, we just are used to grouping risks between 10 and 20% together.". The patient says "Now I understand. Your grouping is not my grouping. I consider 10% much different from 20% and I will not undergo surgery."

Risk groups do not aid in decision making; they get in the way of it. When and if grouping is useful it is done post-publication.

It is not difficult to gauge accuracy of risk predictions using completely continuous measures and graphics.

The statement in the authors' responses that "half the patients have a similar, middling risk and a quarter have a relative high or a relatively low risk" would be true if the risk groups each have a width of 0.01, i.e., if they were homogeneous. This is not the case.

Regarding PIs having an approximately Gaussian distribution I don't see this happening in a significant proportion of my practice. It depends entirely on the nature and number of predictors.

The revised manuscript persists in not emphasizing a formal test that the slope of the linear predictor is 1.0 in the validation sample.

Minor Essential Revisions

The authors understate the problems of stepwise variable selection in large samples.
Discretionary Revisions

**Level of interest:** An article of importance in its field

**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.