Reviewer’s report

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

Version: 1 Date: 22 August 2012

Reviewer: Glenn Heller

Reviewer’s report:

The authors state at the outset of this paper that a prognostic model should not enter clinical practice unless it demonstrates that it performs a useful role. Given the low discrimination and explained variation values in the Breast cancer derivation and validation data, I wonder if the authors believe this model should be utilized in clinical practice.

On a related point, I do not think the authors ever explicit state whether they believe the GBSC dataset validates the model derived from the Rotterdam study.

If the derivation data has low discrimination and explained variation, is it likely that a validation dataset would be acquired?

Data: Basics and reporting requirements

What do the authors mean specifically when they state that the outcome of the derivation dataset should be reported?

Check model specification/fit

The proposed test of $#*=0$ ignores the variability in the estimated PI, which is computed from the derivation dataset. Ignoring the variability will underestimate the variance of the test statistic and I would recommend its inclusion. Since the derivation and validation datasets are independent, incorporating this variability into the test statistic should be straightforward. Also, does the fact that the validation covariates are found in both the $x #*$ and PI terms result in a collinearity issue?

Continuing on this test, are the authors suggesting that the non-rejection of the null ($#*=0$) implies a good model fit? Or should a test of equivalence ($|#*|<#$) be applied for this purpose?

Measures of discrimination

Harrell’s c-index is not consistent with survival data. The authors should consider an adjusted Harrell’s c using inverse probability censoring weights.

Calibration

In the data example it is stated that calibration is “less good” for the moderate and poor risk groups. It would be helpful for the authors to give stronger guidance in determining when the level of calibration is sufficiently poor.
Level of interest: An article whose findings are important to those with closely related research interests

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