Reviewer’s report

Title: Validation of the CancerMath prognostic tool for breast cancer in Southeast Asia

Version: 0 Date: 09 Feb 2016

Reviewer: Sanjib Basu

Reviewer's report:

The objective of this manuscript is to compare the individual subject level nodal status and survival predictions from CancerMath for breast cancer cases with 7064 observed cases of stage I to III breast cancer in Singapore Malaysia Hospital-Based Breast Cancer Registry.

For either nodal status or survival, the data in question thus involve two separate lists (or distributions), one set of subject level predictions from CancerMath and another set of subject level observed values. What is of interest is the "match" between these two values with the criterion for "match" appropriately defined (here I am using the term "match" in a rather general sense). The interest is thus in the joint performance (in terms of matching) of the predicted and observed lists, rather than on the individual lists.

Many of the "match" results described in the manuscript are however based on two distributions (predicted and observed) separately, rather than their joint distributions. For example "Nodal status calculator predicted 40.6% of patients to be node positive which was lower than the observed 43.6%", in fact, reports two numbers calculated separately from the CancerMath predicted distribution and observed values and does not consider the joint distribution. While this reported information is certainly of substantial value, it is nor based on "match" at the individual subject level. Similarly, "Cancermath predicted and observed overall survival probabilities were 87.3% vs 83.4% at 5 years after diagnosis and 75.3% vs 70.4% at 10 years after diagnosis" is again based on two separate sets of calculations, one using the CancerMath predicted distribution and the other using the observed values.
On the other hand, the calibration plot and the ROC curves ("The calibration plot showed underestimation for most of the groups. The AUC was 0.71 (95% CI, 0.70-0.72)") are, in fact, based on the joint distributions of the predicted and the observed values and considers the "match" at the individual subject levels. I find these to be of more value and I feel that the manuscript needs to focus more on these results and report more results based on the joint distributions.

In usual statistical analysis, the prediction method/model is (trained) based on the observed data and hence the question of statistical variability comes from that aspect. In this case however, the prediction model of CancerMath pre-exists and is not based on the observed data and hence there is no source of statistical variability from that aspect. The authors have been careful in their statistical analysis from this respect and I congratulate the authors.

In terms of prediction, I would think CancerMath predicts a probability (such as 30%) of being node positive for each subject (and similarly a probability (such as 70%) of 5 year survival and another probability of 10 year survival and so on). Thus, the cancerMath predictions of Node positive for all the subjects is list of such predicted probabilities (based on the individual CancerMath inputs for each subject). The reported 40.6% (prediction of node positive) is the median of these individual predicted probabilities. While the authors have been quite careful in describing the 40.6% as "median predicted probability" within the manuscript, I feel that the statement in the abstract of "Nodal status calculator predicted 40.6% of patients to be node positive" is NOT appropriate.

The issue with the 40.6% goes actually a bit beyond. I feel that even if it is correctly described as the "median predicted probability" in the manuscript, a probable common interpretation would be "Nodal status calculator predicted 40.6% of patients to be node positive". As illustrated, even (some of) the authors interpreted it that way.

The other issue with the 40.6% vs the observed 43.6% is that these are based on the predicted and observed distributions separately and does not provide us with any idea on the sensitivity and specificity. In fact, equally important here are the positive and negative predictive values. The issue with these is that a cutoff on the predictive probabilities is needed to evaluate this. But I feel that this is the question of real interest and the authors need to report this based on a cutoff (such as default of 0.5) and also should show the ROC curve obtained by changing the cutoff.
Another general comment is that the authors report "median of the predicted probabilities" and median of the predicted survivals. I feel that from statistical viewpoint, the mean (of the predicted probabilities, survivals) is more appropriate. For example, I would think that the idea of the calibration plot is taken from the Hosmer-Lemeshow goodness of fit test and the mean is actually used in that test.

Details are needed for the calibration plot in figure 2. How is the observed survival calculated. Is it the Kaplan-Meier estimate of 5 year survival for the subgroup of patients in that decile group? As before, I suggest to use average predicted survival rather than the median.

In Table 2, the n=4517. Are these who were not censored by 5 years? Are the observed and predicted number of deaths out of these subgroup only? Same question for Table 3.

I did not follow how the C-statistics is calculated "Outcome calculator was further evaluated using concordance statistics (C-statistics) for the entire dataset regardless of follow-up time"?

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics
Quality of written English
Please indicate the quality of language in the manuscript:
Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests
I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal