Author’s response to reviews

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

Authors:

Hui Miao (ephmh@nus.edu.sg)
Mikael Hartman (mikael_hartman@nuhs.edu.sg)
Helena Verkooijen (H.M.Verkooijen@umcutrecht.nl)
Nur Aishah Taib (nuraish@gmail.com)
Hoong-Seam Wong (whseam@yahoo.com)
Shridevi Subramaniam (shridevi@crc.gov.my)
Cheng-Har Yip (chenghar.yip@gmail.com)
Ern-Yu Tan (ern_yu_tan@ttsh.com.sg)
Patrick Chan (patrick_chan@ttsh.com.sg)
Soo-Chin Lee (csilsc@nus.edu.sg)
Nirmala Bhoo-Pathy (ovenjjay@gmail.com)

Version: 2 Date: 06 Sep 2016

Author’s response to reviews:

Dear Editor and Reviewers,

We would like to thank you for considering our manuscript for publication. The manuscript has certainly benefited from insightful comments from the reviewers. Below you will find our response to the reviewers’ comments.
Reviewer #2: This is a well written paper on the validation of CancerMath, an existing web based prognostic tool, posted by a research group in the US.

The authors did a thorough evaluation of the CancerMath tool using Singapore Malaysia hospital based breast cancer registry data.

Their finding is valid: CancerMath is suitable in general, but not suitable for patients with poor prognosis.

However, we know that CancerMath was built up using data from the US SEER registry, and the majority of the people in the SEER database were not Asian, but Caucasian or African American. So it won't be a surprise to find that CancerMath is not totally suitable for a different ethnic group, such as those in Southeast Asia. I would suggest that the authors add more background information on the percentage of different ethnic groups in the SEER database, and that is the main reason why CancerMath is not completely suitable for Asian people.

We thank the reviewer for this comment and compliment. As the review suggested, we have added information on percentage of different ethnic groups in the SEER database in the discussion section.

Reviewer #3: This study validates the CancerMath prognostic tool for breast cancer. There are two types of validation, the first for the (binary) outcome nodal status and the second for survival. I discuss each separately.

The researchers validate the nodal status calculator by constructing a calibration plot and calculating the AUC. This approach is appropriate but I think it would be improved if the researchers also presented the ROC curve. Also, the researchers should quantify or comment on the under-estimation in the calibration plot. A comment regarding the 'calibration slope' (see Miller, Statistics in Medicine, 1991) would also be helpful.

We agree with the reviewer that ‘calibration slope’ should be discussed to provide a greater understanding of the model and added this under the discussion section in the manuscript. Our manuscript focuses on discrimination and calibration of the model. We think that ROC curve which illustrates sensitivity and specificity is not necessary as we interested in neither comparing multiple ROC curves nor determining the optimal cut-off point for nodal status from the ROC analysis. Thus area under the ROC curve (AUC) is sufficient to demonstrate discrimination.
The researchers validate the survival calculators in the same way as nodal status which I found surprising since these methods do not take account of censoring. I was further confused by the researchers’ responses to Reviewer #1. They say that only patients from 2007 and earlier were included in the validation of the 5-year survival calculator and that no individual had a follow-up length shorter than 5 years. This is unusual as there is almost always loss to follow-up (censoring) in such studies; omitting censored patients would lead to the observed survival estimates being biased downwards (towards worse survival). I also note that the omission of more recent patients (2008 onwards) is not necessary as such patients could be censored at the end of their follow-up. Similar comments apply to the validation of the other survival calculators. A related point concerns the choice of measures for the survival validation. I would expect to see Kaplan-Meier curves to illustrate observed survival and the c-index (not AUC) to quantify discrimination. The use of the c-index is alluded to but its calculation is not described correctly.

We thank the reviewer for detailed comments. Death information were obtained from national death registries on 1st March 2013 for UMMC, 31st July 2013 for NUH, and 1st October 2012 for TTSH, and these dates were treated as date of last follow-up. If death was not registered, patients were censored on the date of last follow-up. As only citizens and permanent residents in Singapore (NUH and TTSH) were selected for analysis and the death registry is very reliable, we assume all patients from NUH and TTSH have complete follow-up information. Only 12 patients from UMMC had loss of follow-up (censoring before 1st March 2013), among which 6 patients had less than 5 years of follow up and 10 patients had less than 10 years of follow up. We acknowledge that excluding these cases may lead to biased estimates but the impact is minimal due to our large sample size (N=4517 for 5-year survival and N=1649 for 10-year survival).

We agree that omission of more recent cases is not necessary for estimating observed survival using Kaplan-Meier method. However we want to make sure table 2, 3 and calibration plots use the same subsets as AUC analysis so that results on 5-year survival and 10-year survival can be clearly interpreted. For AUC calculation, a binary outcome is needed instead of time-to-event data. Therefore, when outcome is time-to-event data, we have to specific the length of follow up (5-year and 10-year) to dichotomize survival outcome. For patients diagnosed recently (2008 onwards), their death status is unknown at exact 5th year after diagnosed. Therefore we have to exclude them from AUC analysis for 5-year predicted survival. To take account of censoring, C-statistic developed by Harrell et al [ref 27] was used to quantify discrimination using time-to-event data. It is calculated by considering all possible pairs of patients, at least one of whom has died. If the predicted survival is larger for the patient who lived longer, the predictions for that pair are said to be concordant with the outcomes. We have amended the description of Harrell’s C-statistic under method section in the manuscript. Since AUC is commonly reported in all other validation studies for comparison, we would like to focus more on AUC instead of C-statistic.
As suggested by reviewer, Kaplan-Meier curve of overall survival (new Figure 2) was added to illustrate observed survival of the study population.

I have a few comments regarding the Tables. Please add the denominators ('N') to Table 1. I note that the 'Observed deaths' in Tables 2-4 are not adjusted for censoring and hence the Mortality Ratio is likely to be misleading.

Many thanks for the comments. Denominators were added to Table 1 as suggested. As we explained earlier, all patients have complete follow-up information except 12 patients from UMMC. Excluding these patients would not affect the mortality ratio significantly due to the large sample size.

Yours Sincerely,

Hui Miao