Reviewer’s report

Title: Risk Prediction for Breast Cancer in Han Chinese Women Based on a Cause-Specific Hazard Model

Version: 0 Date: 10 Oct 2018

Reviewer: Robert J. MacInnis

Reviewer's report:

This paper outlines a new risk model for breast cancer developed on a Han Chinese population. There are few if any breast cancer models developed to show applicability to Chinese populations and the need is certainly there. There are several issues that I have identified, however, that need to be addressed by the authors before being suitable for publication.

Background

1. "…the mammographic screening participation rate is only 21.7% in China…". Does this apply for all ages? In what time-period?

2. Discussion about the Gail model "…showed poor effect when directly applied in Han Chinese women." Can the authors cite reference(s) to back this claim?

3. Also with the Gail model "…the use of number of previous breast biopsies as a predictor also make it difficult for application." I don't agree with this statement as this does not disqualify its ability to be used as it is can simply be marked "unknown".

4. The IBIS model has not (yet) incorporated genetic variants explicitly, though some of the effects will likely be captured in the unknown dominant gene component.

5. The sentence about adding 7 SNPs is very much out of date as there are now over 200 SNPs associated with breast cancer. Anyhow, a 0.02 gain can make substantial difference to discrimination.

6. In discussing ref 9, the mutation rates of BRCA1/2 genes are lower in Chinese populations, but the authors of that publication state they are still the most important genetic indicators of risk.
Methods

7. In the case-control study, when were the breast cancer patients selected? Was it at diagnosis or sometime after? If it’s sometime after, then there may be the possibility that some would have been missed due to dying before being recruited. Also, does the case group include in situ breast cancer cases?

8. What was the response rate for recruitment of cases and controls?

9. 5365/18541 women were lost to follow-up is a large proportion, and it greatly affects the validity of the results. How many were known to have died? Those who are known to have died should not be excluded from the validation dataset. There should also be a comparison of baseline characteristics of those who were lost to follow-up with the rest of the cohort. Preferably, women lost to follow-up (but not known to have died) should be censored at last known date of contact.

10. The authors need to define "family history of breast cancer". Is it confined to first-degree relatives or does it include distant relatives?

11. How was the cut point of 13 chosen for the life satisfaction score?

12. Age range groups go from <20 years up to 85+ years, but they limited the case-control study to 25-70 years. Is it valid to apply the model to all ages when it was developed on a narrower range?

Results

13. The number of incident cases in the Taixing study is small (n=34), which will give low precision to the validation estimates. Cell counts in certain combinations equal zero. This needs to be mentioned as a limitation of the study.

14. It would help if the authors could state explicitly in the results section that diabetes drops from the multivariable model (it is later mentioned in the discussion).

15. Likewise with the attributable risk, should state (maybe better in the methods) that it applies only to the Taixing province.

16. A 95% CI should be reported for the E/O ratio. Given the small number of incident cases, though, the CI will include 1.
Discussion

17. I have several issues about the following statement: "Compared with other breast cancer prediction model, our model showed better E/O ratio and AUC". Firstly, it depends on the population one is assessing the model on. Secondly, as long as a E/O ratio from a model includes 1 and that there is reasonable precision around the estimate (which based on 34 incident cases this study does not really have), then one cannot truly say one model is "better" than the other, except perhaps for certain subgroups within the population. Thirdly, comparing AUCs between different studies is fraught with numerous hazards. The underlying structure of the validation population (such as age, stronger family history, etc) can have a huge impact on the AUCs. Basically, comparing model AUCs can only really be done within the same study population.

18. Further to the previous point, it would be better to compare the results of the Gail model (probably should use Asian American reference rates) using the same validation cohort. Just because biopsy tests were not collected doesn't preclude one from using the model. Sure, the AUC might not be as good, and this can be mentioned as a limitation. Could also compare with Yuan et al model as well, but again, not all variables might not be available. The point I'm making is unless they can show there is a clear improvement over the existing models, then what benefit is there in developing yet another model?

19. Notwithstanding the difficulties of comparing AUCs, stating that their "model showed narrower confidence interval" is totally meaningless.

20. The sentence comparing E/O ratios of your model with the Gail model is also totally meaningless. The authors only need to be concerned about whether their model calibrates with the external population they're testing the model on.

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

No

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.

No
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:

Not suitable for publication unless extensively edited

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