Peer Review File

Reviewer A

The manuscript aims to compare 3 different measures of axillary lymph node involvement in mastectomy female. It uses data from the SEER, claims validation on an Asian population of patients from the authors' institution, and argues the utility of nomograms.

Overall concern:

1) The manuscript analyzes and provides results almost exclusively using SEER data.

Reply 1：Thank you for your advice. In this article, two data sets were used, one with the SEER public database to build the model, and the other from the First Affiliated Hospital of China Medical University for external validation.

2) To this reviewer, the originality and possible added value of this manuscript would be to provide detailed observation of their patients' characteristics and outcomes. There is only one table giving characteristics of their patients thrown far out of the manuscript in a supplementary table, not much else.

Reply 2：Thank you for your valuable suggestion. We've added information related to survival in Supplementary table 1.

Changes in the text: we have modified our text as advised (see Supplementary table 1)

3) I suggest trying for more balance. Whenever a Table or a Figure is done, should be mirrored in their patients. If unfeasible, for example too small groups, or too many results, then something is wrong, that would require organizing the data differently or focusing on what is most informative.

Reply 3: Thank you for your advice. The classifications of multiple variables are very detailed in this article. Although the final results undergo external validation using the Asian people, the efficiency of validation in this model is good, which may be related to the detailed classification. This is more beneficial to find the most important influencing factor. In the future, we hope this problem to be fixed.

4) Having an external validation set of patients is an asset of the paper. Unfortunately results are limited to a meager mention of a C-index. Should show plots of the models (the nomograms developed from the SEER) applied to authors' patients: plot actual survival of patients vs. predicted survival based on SEER models. Published examples of plot actual vs. predicted abound, for example Vos et al, External validation ... Journal of Medical Virology, 27 Aug 2019, 91(12):2117-2124,

van de Laar et al, External validation of three prognostic models... British Journal of Cancer volume 110, pages42–48(2014) (elegant presentation of results and plots).

Reply 4：Thank you for your suggestion, I enjoyed it very much. According to your suggestion, we have supplemented the calibration plot, which also indicates that the model has good applicability to Asian population.

Changes in the text: we have modified our text as advised (see figure 12)

Authors could also benefit from looking at Harrell's book Regression modeling strategies and his website.

Reply 5：Thank you for your advice. We will study the relevant content carefully. I believe we will have great harvest.

Minor concerns:

Title:

Needs to give some additional info, for example "mastectomy".

Reply 6: We appreciate this kind suggestion very much. According to your suggestion, the title has been modified. Thank you!

Changes in the text: we have modified our text as advised (see page 1, line 2).

Abstract:

Background: unspecific, clarify what subset of breast cancer is of interest.

Reply 7:We appreciate this kind suggestion very much. In the assessment of the prognosis of breast cancer, the stage of axillary lymph nodes plays an important role in the assessment of the prognosis of patients regardless of the surgical mode of breast cancer. That is why it is not mentioned in the background section of the abstract.

Methods: A total of 9616 BC patients from the Surveillance, Epidemiology, and End Results (SEER)

database and 675 patients from the First Affiliated Hospital of China Medical University were reviewed. This is opaque. Specify selection criteria. Multivariate analysis: what type? what endpoint?

Reply 8: We appreciate this kind suggestion very much. According to your suggestion, We have made a few modifications in the methodology part of the abstract, please read it.

Changes in the text: we have modified our text as advised (see page 2, line 9-13.)

Results: Readers would be more interested to have some info on the survival. The C-indexes and AIC's: would suffice to say which factor has lowest AIC, which highest C. Also, consider consider consistency of the number of digits (C and CI). LNR, LODDS: first time mentioned, would be more appropriate to abbreviate in Methods.

Reply 9: We appreciate this kind suggestion very much. It is very useful for us. Since this manuscript involves many risk factors and involves multiple models, it may be difficult to describe too much information such as survival and confidence interval in the abstract, so this information is put in the results section and supplementary materials table 2. In addition, according to your suggestion, the number of digits in the text has been unified, thank you!

Changes in the text: we have modified our text as advised (see page 3, line5-10. Page14 line1-6.Page 16,line 18).

Main text.

Introduction:

Missing LODDS in full letter first time it is mentioned.

Reply 10: We appreciate this kind suggestion very much. And we've already added to that in the body part.

Changes in the text: we have modified our text as advised (see page 5, line 10.)

Lines 22-23: "Our results indicate... ": this belongs to Discussion or Conclusion.

Reply 10: We appreciate this kind suggestion very much. And we've already modified this part.

Changes in the text: we have modified our text as advised (see page 5, line 10-12.)

Methods:

"The Training cohort of a total of 9616 female BC... ": Place the total patients *after* the inclusion criteria, and preferably at start of Results.

Reply 11: We appreciate this kind suggestion very much. And we have migrated this section to the start of Results.

Changes in the text: we have modified our text as advised (see page 9, line3-8.)

Variables filtering: *race*: paper's interest is on Asian women. SEER has info on Asians: why white, black only in the paper, but not Asian?

Reply 12: We appreciate this kind suggestion very much. Under the inclusion criteria, fewer Asians can be screened out, so we simplified the population here to white, black and other groups, and it is indeed a deficiency of the article.

R: but no mention of the packages used.

Reply 13: We appreciate this kind suggestion very much. And we have added relevant content in the Statistical Analysis section of the article.

Changes in the text: we have modified our text as advised (see page 8, line13.)

Results:

Detailed %survival info is given on the SEER patients. But missing survival info on the 675 patients from authors' institution. Consider for example modifying Table 1 & 2, column SEER OS, and provide column Institution OS.

Reply 14: Thank you for your valuable advice. The purpose of Table1 is to compare the differences between groups of LNR based on SEER database in a large sample population, and to select an appropriate cutoff value of LNR. Finally, the continuous variable LNR is converted into a categorical variable and applied to model construction. The existence of external validation data added here may easily lead to readers' unclear grasp of the problems to be explained in this table. I would like to ask for your valuable opinions again. If necessary, we will add this part in time. Thank you!

Many nomograms are shown. Models underlying the nomograms are essential, but the models are missing.

The various LNR and LODDS tables indicate categorization. If the same categorization has been used for the nomograms, then should mention so. If not, clarify.

Reply 15: Thank you for your valuable advice. Yes, the same LNR or LODDS classification was used in the model construction.

Changes in the text: we have modified our text as advised (see page 11, line 13,14.)

Discussion:

Background, US women only: what of China women, or rest of world?

Reply 16: Thank you for your suggestions. I have made corresponding modifications in the Discussion section.

Changes in the text: we have modified our text as advised (see page 14, line 10.)

Page 13 (11) Line 5-6: "Until now, no study concerning the prognostic value of LODDS for breast cancers has been reported." ??? Contradicted by the references.

Reply 17: Thank you for your suggestions. Sorry for our mistake. This should be changed to "Until now, no study comparing the three lymph node staging methods for predicting outcome in breast cancer patients with mastectomy has been reported."
Changes in the text: we have modified our text as advised (see page 15, line 13,14.)

Page 13 (11) Line 12: "our cubic splines revealed... ": where does that cubic splines come from? What was the usage?
Reply 18: In medical research, we often construct regression models to analyze the relationship between independent variables and dependent variables. In fact, most regression models have an important assumption that independent variables and dependent variables are linearly dependent, which is difficult to satisfy in practice. A common solution is to classify continuous variables, but the selection of the number of categories and node locations is often subjective, and classification tends to lose information. Therefore, a better solution is to fit the nonlinear relationship between independent variables and dependent variables. Restrictive cubic spline is one of the most commonly used methods to analyze nonlinear relations. It is essentially a piecewise polynomial, and the relationship between independent variables and dependent variables can be clearly described by restrictive cubic spline functions. In fact, restrictive cubic splines can be applied to a wide range of situations, and can be added to regression models when you want to describe the relationship between independent and dependent variables.
Fig 3-4, Fig 5-8, Table 1-2: could specify overall survival and mortality hazard ratio.
Nomograms Fig 9-12: clarity could benefit from using full names instead of abbreviations. Why "AD" and a lot of space, instead of Age, "RR" instead of Race, etc? Also, specify Age units, years.
Reply 19: Thank you for your valuable comments. We have modified the relevant images.
Changes in the text: we have modified our text as advised (see fig.9,10,11.)

Supplements:
Supplementary Table 1: Interesting. Should put at the front and comment the differences. Worthwhile to do a subset comparison with the "other" race, almost same number of patients. Do the "other" race have same age ER PR etc distribution or not?
Reply 20: Thank you for your suggestion. In the Supplementary table 1, the difference between the training set and the validation set should come from the other population, which includes not only Asian population, but also a considerable number of people of multiple races around the Pacific Ocean, such as American Indian/AK Native, Asian/Pacific Islander. According to your suggestion, I conducted a stratified analysis of the training set and the validation set of "other" population (Supplement table 3), and found that significant differences still exist, which further indicates that there may be not only Asian population but also other population in the "other" population. It also suggests that the model has certain efficacy for different populations.
Changes in the text: we have modified our text as advised (see page 9, line 12-18.Page 10, line 1,2).

Supplementary Table 2: "record", or "recode"? Lacks mirror table for the validation 685 patients.

Reply 21: Thank you very much for your insight, after re-checking, it should be "recode". I'm very sorry that I still don't know the meaning of mirror table despite a lot of references, could you please tell me what it is?

Changes in the text: we have modified our text as advised (see Supplementary table 1，Supplementary table 2)

References:

Several references are misformated.

Reply 22: Thank you for your advice. We have corrected the overall references.

the paper is highly technical and loses the audience

if the intended audience is breast surgeons and medical oncologists must present the results in a more transparent matter. what is currently done? How should we modify it. what will the difference be?

Reply 1: Thank you for your question. Draw a line straight up to the Points axis to determine how many points that patient with this characteristic received. Repeat this process for each of the remaining axes. Sum the points for each predictor and locate this sum on the Total Points axis. Draw a straight line from the Total Points axis until it intersects with the horizontal line drawn from the 5-/10-y overall survival. Corresponding to the 5-/10-y overall survival rate. By incorporating multiple risk factors and new lymph node evaluation indexes, the model was further optimized on the basis of the existing lymph node evaluation system, which had better efficacy in predicting the overall survival of patients. In the future, with the progress of biochemical and immune index detection technology, further optimization can be carried out on the basis of this model.

The authors described a study that compared 3 lymph node staging methods for predicting outcome in patients with breast cancer. They performed univariate and multivariate analysis and concluded that LODDS had the best predictive ability. They also externally validated the performance of the nomogram incorporating LODDS on a different dataset. The manuscript was well written. My major comments are listed below.

1. The conclusion may not apply to the entire population of breast cancer patients, since certain breast cancer patents (e.g., patients with fewer than five nodes collected) were excluded.
Reply 1: Thank you for your comments. In this study, since the main research content is lymph node evaluation method, which involves the ratio of positive lymph nodes and the minimum number of acquired lymph nodes, this study needs to ensure a certain number of acquired lymph nodes, and future studies may further study patients with low acquired lymph nodes. Based on this, we have redrawn the title.
Changes in the text: we have modified our text as advised (see page 1, line 1，2.)

2. The authors claimed that the nomogram with LODDS had the best performance among all systems, while its C-index appears to be lower than that of nomogram with LNR. An explanation is needed. (My intuition is that a better performance corresponds to a higher C-index)
Reply 2: Thank you very much for your question, as you said, the higher the C-index, on behalf of the better performance of the model, but when the pN is equal to zero, LNR is also equal to zero, unable to differentiated the overall survival in patients with zero lymph node metastasis, whereas LODDS can do differentiated the overall survival in patients with pN = 0, as shown in fig. 7, as a result, although LNR and LODDS have the same C-index, but LODDS performance is superior.

3. The authors claimed that LODDS could help select the appropriate treatment. However, it is not clear how the developed system could be used to do that.
Reply 3: Thank you for your question. In the using of the model, draw a line straight up to the Points axis to determine how many points that patient with this characteristic received. Repeat this process for each of the remaining axes. Sum the points for each predictor and locate this sum on the Total Points axis. Draw a straight line from the Total Points axis until it intersects with the horizontal line drawn from the 5-/10-y overall survival. Corresponding to the 5-/10-y overall survival rate. One of the features of LODDS is the ability to distinguish total survival in patients without lymph node metastasis. The model constructed by combining LODDS with traditional risk factors can achieve better predictive efficacy, which was also verified by external validation. Based on this model, the treatment cycle and medication intensity of patients can be better determined.

4. Please compare the finding in this manuscript with the finding in the paper

"Creating prognostic systems for cancer patients: A demonstration using breast cancer, Hueman et al" , which used c-index, SEER data, and machine learning to develop prognostic system for breast cancer.

Reply 4：After carefully reading the paper, we found that this text is mainly described by running the EACCD to produce dendrograms and cutting dendrograms to generate prognostic groups according to the C- index and to generate the system of the group to create a cancer prognosis system method, which is not the same with us. We mainly use SEER public database to identify which methods of lymph node staging can effectively predict the prognosis. Based on this method of lymph node staging, we further build the prognostic model and finally validate these results by external data.