Peer Review File

Reviewer A

The authors developed novel prognostic nomograms based on a large population-based cohort to estimate the OS and CSS for patients with PCNSL. They have conducted extensive analysis, however there are concerns.

1. Although the authors included the long time SEER database between 1975 and 2016, the standard of care in PCNSL patients has remarkably changed in recent years. So, some of the results itself are not adequate to take into consideration.

Reply 1: Thank you for raising this question! The main concern is that certain indicators, such as the treatment plan, have changed over such a long-time span. However, the indicators we included are general indicators of the basic characteristics, pathology, and treatment of patients. Although the specific treatment plan of PCNSL has undergone great changes, it is mainly about the transition from tumor resection to biopsy and the adjustment of the specific medication or dosage of chemotherapy and radiotherapy. Therefore, the indicators we have included, namely surgery, chemotherapy and radiotherapy cover these aspects.

Changes in the text: No changes in the text.

2. HIV-related lymphoma is not usually included in PCNSL. So, it is not correct to compare this model to other prediction ones.

Reply 2: Thank you for raising this question! We have to admit that there have been few previous studies on HIV-positive PCNSL. The pooled prevalence of HIV infection among PCNSL patients was 6.1%, ranging from 2.2-30.2% (please see PMID: 32956921). It can be seen that HIV infection is an important feature in PCNSL. We included HIV infection status, which allowed us to conduct a real-world study in the large cohort. The present nomogram and Deng's model were both based on SEER database. Comparing prediction models built from the same database may better show the prediction performance of the model.

Changes in the text: No changes in the text.

Reviewer B

The authors present a paper identifying the factors affecting survival in PCNSL and constructing nomograms aimed to predict outcome. Nomograms were built based on the US based SEER database, and a validation cohort of 163 patients from their local hospital used as validation cohort.

Whilst the methodology is thorough and the nomograms obtained have reasonably good C-indexes, the main caveat is the lack of specific clinical data (ie. Performance status and specific chemotherapy treatment), which will make the applicability of the nomograms very difficult in routine practice.

<u>Reply 1:</u> Thank you for raising this question! This question points out the shortcomings of database-based research very well. Since the SEER database does not provide specific clinical data (ie. Performance status and specific chemotherapy treatment), we cannot further refine the model. However, we have recognized the above problems, and we are summarizing and studying the data of our single-center PCNSL patients with detailed clinical and sequencing data. Since 2004, more than 300 patients have been treated in our hospital. Please pay attention to our follow-up research reports, thank you!

<u>Changes in the text:</u> No changes in the text.

Also, the performance of the nomograms in the local Beijing cohort is rather poor, the authors need to expand the discussion explaining the potential explanations for this performance.

<u>Reply 2:</u> Thank you for raising this question! In order to better explain the problem, we have added Table S3, which shows the heterogeneity between the Chinese cohort and the SEER database.

<u>Changes in the text:</u> We have modified our text as advised (see Page 11, line 224-227, and Page 13, line 264-266).

Some additional specific aspects come to mind:

-How reliable is the CSS form the SEER database, in other words, authors need to discuss on previous work that has used this same approach and demonstrate that the competing non-cancer survival data is accurate and reliable conclusions can be drawn from it

<u>Reply 3:</u> Thank you for raising this question! In fact, the CSS form the SEER database is widely used, and there are many articles covering this usage(please see Page 16, line 328-329). We also mentioned its advantages (please see Page 4, line 74-75) and applications (please see Page 16, line 328-329) in the article. And we believe that it is feasible to construct and apply the CSS model in a database with complete patient survival information.

<u>Changes in the text:</u> No changes in the text.

- There is no data clarifying how many patients were excused due to missing survival data and the potential impact of this missing data in the results of the models, according to methods there could be up to 1400 patients excluded because of this? This is a serious bias that needs to be discussed and acknowledged.

Reply 4: Thank you for raising this question! We actually included 6042 cases of intracranial PCNSL patients. The 7393 cases in the previous manuscript included 6042 cases of intracranial PCNSL and 1351 cases of PCNSL in other sites, such as eye, spinal cord, etc. In fact, as shown in the Figure S1, there were 73 cases excluded because of missing survival data. We apologize for the confusion caused by our inaccuracy. We have compiled our research records and reviewed the research data.

Changes in the text: We have added the Figure S1, showing the flow diagram of patients' selection, with numbers of included and excluded patients (see Page 5, line 87, and the supplementary Figure S1).

- I sincerely doubt the importance of the marital status on the prognosis of PCNSL, the authors speculated with some ideas regarding social support network, but this not adequately supported with evidence.

Reply 5: Thank you for raising this question! According to reports in the literature, marital status was related to the prognosis of a variety of malignant tumors. There were even literature suggestions that it was necessary to proactively provide single and divorced/separated/widowed individuals with appropriate social and psychological support that would benefit them (please see PMID: 33686220). PCNSL was no exception, divorced/single/widowed were all risk factors for early mortality in patients with PCNSL (please see PMID: 32251807)

Changes in the text: No changes in the text.

- It is seriously concerning that 25% of PCNSL patients underwent resection, are these patients adequately classified, is this an error when indexing diagnosis, is this standard practice in the US. This goes against all the current paradigms of PCNSL treatment and hence, should be more carefully explained and discussed.

Reply 6: Thank you for raising this question! We have reviewed the research data. We believe that there were 1006 patients who clearly underwent STR, GTR, Partial lobectomy and Lobectomy in the SEER data set. There were 500 patients undergoing Unknown-type surgery, most of which may undergo biopsy. It can also be seen from the univariate analysis that there was no difference between the prognosis of patients undergoing biopsy and that of patients undergoing surgery of unknown-type, but the difference among other groups was significantly (please see Table 2). In other words, there were a little more than 17.0% of patients underwent resection in the SEER database, which was about the same as 20.2% of that in our external validation set. As we all know, the standard practice in the treatment of PCNSL has changed a lot, from surgery and radiotherapy to the current chemotherapy, and the chemotherapy regimen has also undergone great changes. Most patients undergoing surgical resection were diagnosed in an earlier year, or because they were misdiagnosed as gliomas or

metastases before surgery. However, when we re-examined these data, we found that surgical resection combined with chemotherapy regimens may benefit patients.

Changes in the text: We have modified our text as advised (see Page 8, line 146-147).

- If the above is true, benefit of resection is a provocative idea, but the conclusions regarding survival advantage of total resections needs to be softened, these were in the end a minority of the patients and there could have been many unknown factors influencing the outcomes.

Reply 7: Thank you for raising this question! We fully agree with the question you mentioned. The prognosis of cancer has its own characteristics, which are nothing more than the characteristics of the tumor itself, the clinical characteristics of the patient, and the treatment strategy. We have been trying hard to find out the mechanism and the correlation between prognosis and these characteristics. Especially for rare diseases such as PCNSL, there is still no ideal tumor staging or prognostic model that can predict the prognosis of PCNSL well. Therefore, in this manuscript we applied relatively large data to try to solve this problem. As far as we know, this manuscript is currently the only one to build a model in a large data set and validate the model in an external data set.

Changes in the text: No changes in the text.

- Same principle applies for the analysis of HIV negative/positive patients. Most of HIV patients are likely to have been treated in the earlier years of the observation period, as adequate ART therapy have significantly reduced prevalence of HIV-associated PCNSL. This needs to be discussed, as HIV positive patients treated in the later years with better protocols and better supportive care could actually do very well, hence, statement that HIV significantly affects survival in the current era could be flawed.

Reply 8: Thank you for raising this question! Although the efficacy of adequate ART therapy has made great progress, HIV infection makes the condition of PCNSL patients more complicated. Data analysis did show that, compared to HIV-negative PCNSL patients, HIV-positive PCNSL patients had a worse prognosis.

Changes in the text: We have modified our text as advised (see Page 15, line 315-317).

- Figures and figure legends need to be significantly improved, they fail to convey a clear message, too much information is condensed on them and the curves are not easy to discern.

Reply 9: Thank you for raising this question! We have modified accordingly.

Changes in the text: We have modified our text as advised (see "Supplementary information")

Reviewer C

In this study the authors developed and validated models predicting 1-, 3- and 5-year overall and cancer-specific survival in a cohort of men with primary CNS lymphoma. While it is commending that the authors have attempted to address several important issues such as the presence of competing event(s), treatment and generalizability of the model, there are several issues with how they have done so that need to be considered. Major issues:

1. The outcome definition

While the authors define two outcomes:

i) Overall survival – time from the diagnosis to death from any cause and

ii) Cause-specific survival – time from diagnosis to death from PCNSL.

they do not really explain what is the estimand (or better yet, predictimand) that they are actually trying to estimate (please see PMID: 32445007).

Reply 1: Thank you for raising this question! In the title of the manuscript and the Background part of the abstract, we pointed out the purpose of this research, which was to estimate the OS and CSS of patients with PCNSL. The events of interest were death from any cause and death from PCNSL. According to the description in PMID: 32445007, our research strategy belongs to the "Ignore treatment" strategy. This research strategy is applied to most clinical prediction model research.

Changes in the text: No changes in the text.

Typically, when the outcome is defined as the time from diagnosis to event, we are interested in the natural history of the disease. The goal is to estimate the probability of the outcome if treatment was never started (please see PMID: 32445007). This probability is then useful to guide primary treatment decision. Indeed, in the discussion (and only there) the authors state that prognostic models should include, not only tumor and patient related characteristics, but also treatment related characteristics (line 233). Including the treatment in the model is one way of the ways of estimating the probability of the outcome as if treatment was never started (please see PMID: 27045189), but there are many other ways described (please see: https://doi.org/10.1111/stan.12193). However, this only works if among untreated and treated patients one can observe all covariate patterns. Otherwise, the coefficients estimated when the whole population is used are not transportable to untreated patients.

Reply 2: Thank you for raising this question! In our prognostic model, the event of interest is death from any cause and death from PCNSL. The treatment, including surgery, chemotherapy, and radiotherapy, was included in the model. After taking all the covariates into account, the risk factors were obtained through univariate and multivariate analysis, and they were then included in the model. As shown in Table S2,

in the estimation of survival probability, no chemotherapy or no radiotherapy was used as a reference, and the risk coefficient is 0. This is a common method for clinical prognosis models.

Changes in the text: No changes in the text.

The proportion of patients in this study which has not received any treatment is not clear from Table 1, and is thus difficult to say if estimating the probability of the outcome as if treatment was never started is possible. However, if this is what the authors were attempting to do, then the outcome should be clearly defined, and it should also be clear that the treatment is not modelled as a prognostic factor but was included in the model only to be able to estimate the predictimand of interest.

Reply 3: Thank you for raising this question! The treatment, including surgery, chemotherapy, and radiotherapy, was included in the model. In fact, every patient diagnosed with PCNSL at least received surgery, that was, biopsy, or tumor resection. We were not interested in the probability of the outcome as if treatment was never started. As mentioned in "Reply2", what we want to know is the effect of the presence or absence of a certain treatment on the probability of survival.

Changes in the text: No changes in the text.

In populations where most of the patients have been treated, however, what the authors could estimate instead, is the probability of the outcome after the treatment with radiotherapy and/or chemotherapy and/or surgery. In this case, the follow-up should start at the time of treatment. The intended use of such prognostic model would obviously be different.

Reply 4: Thank you for raising this question! As we know, SEER database is updated regularly every year. Since the information of cancer patients is registered, the follow-up data is updated every year. Therefore, the SEER database meets the needs of our research.

Changes in the text: No changes in the text.

Furthermore, it should be clearly stated that the outcome of interest is 1-, 3- and 5-year OS/CSS, and the follow-up used for the estimation of these models should be clearly defined. From the paper it seems that the authors have used all available fallow-up for model development, but they could instead consider administratively censoring follow-up at 5 years.

Reply 5: Thank you for raising this question! The survival time of PCNSL patients varies greatly, with a median time of about 27 months, with a short period of several months, and a long period of up to several decades. Therefore, we predicted the 1,3,5-year survival probability. This research strategy can also be seen in many literatures on

prognostic models.

Changes in the text: No changes in the text.

2. Model validation

To properly validate a prognostic model (internally and externally) the exact information from the original model should be used (both the baseline survival/CIF and the linear predictor). The model discrimination and calibration should be based on the original model.

While the authors correctly calculate the risk scores (presumably linear predictor) in the validation dataset, they then also refit the score in the validation dataset(s) (line 189, 190). This is essentially model updating. Model performance is expected to be improved by model updating.

Minor issues:

Methods:

1. The SEER is known for its poor coverage. It would be useful to state the coverage of SEER for primary central nervous system lymphoma. Furthermore, a flow-chart of patients' selection (with numbers and percentages of excluded patients) would greatly facilitate the understanding of generalizability of the results.

Reply 6: Thank you for raising this question! As we know, SEER currently collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 35 percent of the U.S. population, from 18 registries. The SEER database is a good database for studying rare diseases such as PCNSL.

Changes in the text: We have added the Figure S1, showing the flow diagram of patients' selection, as advised (see Page 5, line 87, and the supplementary Figure S1). And we have also adjusted the order of Figure legends accordingly.

2. It is unclear which method was used to evaluate discrimination for the cancer-specific model, where Harrell's C-index is not an appropriate measure? Exact methods, rather than papers applying them, should be cited.

Reply 7: Thank you for raising this question! In fact, in the "Statistical analysis", we mentioned the method used to evaluate discrimination for the cancer-specific model was Fine and Gray model, where Harrell's C-index was an appropriate measure.

Changes in the text: We have modified our text as advised (see Page 6-7, line 119, 123-124).

3. It is unclear how the non-parametric CIFs were calculated. A citation to the method would be useful.

Reply 8: Thank you for raising this question!

Changes in the text: we have modified our text as advised (see Page 6, line 118)

Results – Tables:

4. Table 1 - It is unclear what are the P-values comparing. Are these P-values for the comparison of the OS curves, cancer-specific CIFs and competing event CIFs over time? While, to some extent, it makes sense to see these P-values in the Figure 1 (see also comments 8 and 9), in Table 1 they seem out of place.

Reply 9: Thank you for raising this question! P-values in Table 1 were derived from the comparison of the overall survival, cancer-specific death and competing death in different subgroups over time. P-values in the Figure 1 were derived from the comparison of difference between different subgroups when the OS survival rate is 50%. These two P values were different.

Changes in the text: No changes in the text.

5. There is no table summarizing the baseline characteristics of the external validation cohort

Reply 10: Thank you for raising this question! We have summarized the baseline characteristics of the external validation cohort in the Table S3.

Changes in the text: We have added supplementary Table S3 summarizing the baseline characteristics of the external validation cohort (please see supplementary Table S3)

6. Table 2 - Are the authors showing HRs or SHRs for the cancer-specific survival models?

Reply 11: Thank you for raising this question! We showed the HRs for both the OS and CSS models in the Figure S4 (the number of the pictures changed due to the addition of pictures in the newly submitted manuscript).

Changes in the text: No changes in the text.

7. Table 2 – Given the chosen variable selection strategy (the much-criticized backward selection), it is unclear why the authors are showing multivariable models. If the backward selected models are the final models, should they not be presented in the manuscript (instead of the multivariable models), rather than in the supplement? (Figure S3 would be obsolete in that case).

Reply 12: Thank you for raising this question! The univariate and multivariate analyses were used for OS model, and Fine and Gray model for CSS model. The backward method was used to select the independent prognostic factors, and to simplify the model.

Changes in the text: No changes in the text.

Results – Figures:

8. Figure 1. - For some of the predictors, e.g., gender, race, tumor site, there seems to

be little to no difference in the survival curves for the period under the study (1-, 3- and 5-year survival). It would be much more informative to focus on the survival curves up to 5-years (see also the major issue 1). Furthermore, panel A of Figure 1 seems out of place. Panel A, if I understood correctly, refers to the observed survival in the risk groups defined using the predicted probabilities from the model for OS. Such mixing of descriptive information and post-modelling information can be confusing.

Reply 13: Thank you for raising this question! The survival time of PCNSL patients varies greatly, with a median time of about 27 months, with a short period of several months, and a long period of up to several decades. Therefore, we predicted the 1,3,5-year survival probability. And the free software for individualized survival prediction provided in http://www.pcnsl-survivalprediction.cn, which will be available in a few days, presents more.

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

9. Figure 1 and Supplementary Figure S1 are too small and too busy to be useful. The information of interest, 1-, 3- and 5-year survival/CIF is barely visible as figures show the curves for the entire observed follow-up time. The log-rank tests and likely the Gray's test (which is not shown in the Figure S1 though), seem to be largely driven be the differences in the curves at later points in time.

Instead of showing all these curves, which are anyway only descriptive, it would be much more useful to present this information in a separate table. This is already done in Table 1, however, the authors could consider showing the distribution of the baseline characteristics and the survival/CIFs in two separate tables. The P-values corresponding to log-rank test and Grey test could then refer to comparison of the curves up to 1-, 3- and 5-years.

If the authors believe that the survival/CIF curves still could be informative for some readers, they could at least consider moving Figure 1 to the supplementary material.

Reply 14: Thank you for raising this question! As mentioned in "Reply 9", Figure 1 and Table 1 provided different information. Both can help us understand the composition of model construction. We think it is inappropriate to delete or move to the supplementary material.

Changes in the text: No changes in the text.

10. Figure S2 could be seen as a sort of collinearity check, however, this does not seem to be what the authors use it for. This information was also not used for model development, so the figure does not seem to add anything to the paper.

Reply 15: Thank you for raising this question! Figure S2 provides the correlation between different factors. It can be seen intuitively that over time, the number of radiotherapy patients has decreased, the number of chemotherapy patients has increased

significantly, and the number of HIV-positive patients has also decreased significantly. These can help us better understand the characteristics of the rare disease of PCNSL
Changes in the text: No changes in the text.

Model validation:
11. Figure 3 – The authors are comparing the average predicted and observed survivals/CIFs in only 3 groups. This could be somewhat misleading. Typically, 10 groups are used.
Reply 16: Thank you for raising this question! In most published articles, the prognostic prediction model was created in a data set, and then was validated in the same data set. A more complete article will verify the model in one or two external data sets. As far as we know, this manuscript is currently the only one to build a model in a large data set and validate the model in an external data set.
Changes in the text: No changes in the text.

12. Figure 4 and 5 – ROC curves are uninformative. A table with the C-indices at 1-, 3- and 5-years in the training, testing and external validation dataset would be more informative.
Reply 17: Thank you for raising this question! In fact, 1, 3 and 5-year were just three of the time nodes, which were used to test the effectiveness of the model. You can find the survival probability for each year from <1 year to 40 years in the web tools we provided, http://www.pcnsl-survivalprediction.cn, which will be available in a few days.
Changes in the text: No changes in the text.

13. The discrimination and calibration of the Deng's nomogram (or any other model such as MSKCC, IPI, IELSG or the Taipei Score which are only briefly mentioned in the discussion (line 228-232)) should be compared to the C-index and calibration plots for the present model in the internal and external validation data (once correctly estimated, see major issues 2). The discrimination and calibration of the present model which is estimated in the training data is likely overly optimistic and the comparisons such as the one in Figure 5 could give a false impression of the "superior" performance.
Reply 18: Thank you for raising this question! Due to the different risk factors evaluated and the insufficient indicators provided by the SEER database itself, we could not fully evaluate and compare other models such as MSKCC, IPI, IELSG in the training set, which was a bit regrettable. We will summarize the data of about 300 cases in our single center since 2004 for further analysis and research. Please pay attention to our follow-up research reports, thank you!
The present nomogram and Deng's model were both based on SEER database. Comparing prediction models built from the same database may better show the

prediction performance of the model. As shown in Figure 5, the performance of the present nomogram is indeed better than Deng's model.

Changes in the text: No changes in the text.

General comment:

14. Some of the references seem to refer to random method application papers, rather than to the methodological papers which introduce the methods used by the authors (e.g., references 11, 13 and 15).

Reply 19: Thank you for raising this question!

Changes in the text: We have modified our text as advised (see Page 6-7, line 119, 123-124).

information as advised. Please check it, thank you!