PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Prognostic value of different lymph node staging methods for node-positive cardia gastric cancer: a register-based retrospective cohort study
AUTHORS	Wang, Xiao-Qing; Bao, Min; Zhang, Cheng

VERSION 1 – REVIEW

REVIEWER	Zhong, Rui Southwest Medical University
REVIEW RETURNED	04-Mar-2021

GENERAL COMMENTS	1.The continuous factors evaluated, LNR, age, tumour size, LODDS were all categorized and then included in the model. Categorization results in substantial loss of statistical power and reduced interpretability. For instance, by selecting the cutoff of age at 65, the interpretation is that a 64 year old is the same as a 30 year old, but the 64 year old is different than a 65 year old. This makes no biological sense. Further, although the authors state that the 'best optimal cutoff' was selected (which is a form of data dredging), it is questioned that the best optimal cutoff would result in cutoffs as they did (i.e. age=65, tumour size=25, etc). The best option would be to leave the continuous variables as continuous for modeling purposes. 2.What selection process was used to select factors for inclusion in the multivariable model? Please explain whether a single factor regression analysis was performed before the variables entered the multivariate COX regression. In addition, please explain how this process addresses the potential effects of confounding or collinearity. 3.Please consider augmenting the discussion of the findings concerning LNR. LNR association with survival is an exciting aspect of cardia gastric cancer that is currently emerging and may be clinically meaningful. 4.Why did the author use the x-titile software for the cutoff value to choose the third quartile instead of the binary or interquartile range? 5.As the author said whether LNR or LODDS based staging system outperforms TNM 8th edition needs to be further investigated. We want to know whether all the stagings can be corrected to the eighth edition based on the existing fields of the seer databases. 6.Given the importance of the Cox PH model for the development of the nomogram, it would appropriate to include validation that the assumptions of a Cox PH model are met. Please include plots (in the main text or a supplemental figure) of the Schoenfeld, Martingale, and Deviance residuals.

	7.ECOG/Karnofsky performance scores not utilized. Please comment on why these were not utilized as they serve as significant reference points for PC treatment. If possible, this would be a great thing to include in this analysis or the analyses suggested above. 8.altered cutoff values would be an essential factor to analyze during the further optimization of this model. This needs to be further explained in the discussion. 9.Please explain why the N stages do not appear in this nomogram. For example, they do not show significance in the multivariate COX regression? 10.We believe that the nomogram established by the author should be compared with traditional models, such as ROC curve and Decision Curve Analysis.

REVIEWER	Zarnegar , Rasa Weill Cornell Medical College, Surgery
REVIEW RETURNED	14-Mar-2021

GENERAL COMMENTS	I think this is a nice paper with 2 nomograms for the determination for CGA. I think there are some revisions that would make this paper better. The SEER DB has access to total number of nodes analyzed and the number of positive nodes. It is important to use the current guidelines 8th Edition for this analysis even though the data was from prior to implementation of the 8th. The concept should still hold and validate the rigor of the study based on current guidelines. 2. Raw data on the patient population No of nodes harvested and total positive is required to determine the frequency of low yield in the study design. 3. There should be discussion and data on neoadjuvant therapy as this likely impact survival and the number of patient that received therapy. 4. I suggest picking one nomogram. I think its important to send a clear message and by presenting 2 nomograms the authors are creating confusion and an unclear message. I think by being more focused on one approach would allow for improved implementation. The authors may want to compare whichever they select with conventional lymphadenectomy is so desired.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Rui Zhong, Southwest Medical University

Comments to the Author:
1.The continuous factors evaluated, LNR, age, tumour size, LODDS were all categorized and then included in the model. Categorization results in substantial loss of statistical power and reduced interpretability. For instance, by selecting the cutoff of age at 65, the interpretation is that a 64 year old is the same as a 30 year old, but the 64 year old is different than a 65 year old. This makes no biological sense. Further, although the authors state that the ‘best optimal cutoff’ was selected (which

is a form of data dredging), it is questioned that the best optimal cutoff would result in cutoffs as they did (i.e. age=65, tumour size=25, etc). The best option would be to leave the continuous variables as continuous for modeling purposes.
Response: Thank you for your suggestion. According to the coding rule of SEER database, age and tumor size are not always in numeric format (e.g. year of 85+, or tumor size of >990 mm), nevertheless you kindly remind us that stratification by cutoff values will result in substantial loss of statistical power. Therefore the revised manuscript transforms age and tumor size into categorized variables that contain much more strata. Age is transformed as a variable of 9 levels, with the lowest of 0-10 and the highest of 80+ (interval of 10 years). However univariate analysis shows that age is not associated with survival (see Suppl. Table 1.), so it is not included in the final model. Likewise tumor size is transformed as a variable of 6 levels, with the lowest of less than 1cm and the highest of over 5cm (interval of 1cm).
LNR and LODDS are still kept as trichotomous factors because they are used as alternative indicators for AJCC N stage.

2.What selection process was used to select factors for inclusion in the multivariable model? Please explain whether a single factor regression analysis was performed before the variables entered the multivariate COX regression. In addition, please explain how this process addresses the potential effects of confounding or collinearity.
Response: In previous manuscript, we selected factors due to clinical significance (for example, higher stage or large tumor size indicates unfavorable outcome). Inspired by your question, we consider both univariate model results and clinical significance in the revised manuscript, and finally include tumor size, race, grade, T stage, M stage and low nodes yield as adjusted variables for stepwise Cox regression model. The results of univariate analyses are listed in the supplementary table 1. So the confounding effect is addressed by multivariable analysis that incorporates with potential confounders.

3.Please consider augmenting the discussion of the findings concerning LNR. LNR association with survival is an exciting aspect of cardia gastric cancer that is currently emerging and may be clinically meaningful.
Response: Thank you for your suggestion. In the previous paper, we focused on the advantage of LNR when LN harvest is inadequate. In the revised paper, we further discuss the clinical meaning of LNR. We are very grateful for the comment that helps us improve this study.

4.Why did the author use the x-titile software for the cutoff value to choose the third quartile instead of the binary or interquartile range?
Response: We transformed LNR and LODDS into trichotomous variables and selected two cut-off points that represented the greatest group difference of CSS probability according to the minimum p-value method. Thus LNR was re-coded as R1, R2 and R3; LODDS was re-coded as L1, L2 and L3, which was similar with the trichotomous AJCC N staging (N1, N2 and N3). Since only node-positive patients were enrolled, no N0 patients existed here. We address this issue in the revised manuscript in order to make it clear and understandable for readers. Thank you for your question.

5.As the author said whether LNR or LODDS based staging system outperforms TNM 8th edition needs to be further investigated. We want to know whether all the stagings can be corrected to the eighth edition based on the existing fields of the seer databases.
Response: Inspired by your suggestion, all the stagings are now corrected to the 8th edition. Accordingly, the results have been modified. Thank you.

6.Given the importance of the Cox PH model for the development of the nomogram, it would appropriate to include validation that the assumptions of a Cox PH model are met. Please include

plots (in the main text or a supplemental figure) of the Schoenfeld, Martingale, and Deviance residuals.
Response: The plots for PH model validation are included in the supplementary figure 1. Plots of Schoenfeld, Martingale, and Deviance residuals for models that incorporate N stage, LNR and LODDS are all presented. The tests show that PH assumption is met in all models (P>0.05). Thank you for the advice.

7.ECOG/Karnofsky performance scores not utilized. Please comment on why these were not utilized as they serve as significant reference points for PC treatment. If possible, this would be a great thing to include in this analysis or the analyses suggested above.
Response: Unfortunately, variables that reflect general status are not available in SEER database, so ECOG or KPS was not utilized. We address this issue as a limitation in the Discussion Section of the revised paper. Thank you for the comment.
"…One limitation of this study is that some important factors that are associated with survival are not considered in the model due to unavailable data source. For example, ECOG/KPS score is commonly taken into account in survival analysis due to its remarkable relationship with general status and prognosis. Unfortunately the SEER 18 database does not record the score at diagnosis, so the impact of it is not considered in this analysis……To overcome this limitation, a database that provides with fully detailed medical records is needed for analysis."

8.altered cutoff values would be an essential factor to analyze during the further optimization of this model. This needs to be further explained in the discussion.
Response: Thank you for your suggestion. In order to cut LNR and LODDS into trichotomous factor, the previous study selected two points by using the minimal p-value method via X-tile software. Unlike one cutoff point selection, ROC and maximally selected rank statistics cannot be applied in two-point selection. Therefore the revised manuscript attempts to generate trichotomous factors using P25/P75, construct regression model and compare discrimination ability with X-tile based cutoff values. As a result, the X-tile based values have higher power and are finally included in further analysis. We address this issue in the revised paper and present the process in Suppl. Table 2. Although the previous results remain, this step is very crucial for optimizing the models and improving the study reliability. We deeply appreciate this comment.
"…For model optimization, LNR and LODDS were also categorized into trichotomous factors using cut-off values of P25 and P75. The discrimination ability of the model based on interquartile was lower (Suppl. Table 2), so this model was not further analyzed…."

9.Please explain why the N stages do not appear in this nomogram. For example, they do not show significance in the multivariate COX regression?
Response: In the present study, we attempt to construct a new alternative indicator for N stage, because the current N stage classification may not perform well in cardia gastric cancer, so we presented the nomograms that cooperated with LNR or LODDS, other than N staging. In other words, N stages DO appear in the nomogram, but in the form of LNR or LODDS. In addition, the other reviewer suggested us to pick one plot to avoid confusion and unclear message; we only show one nomogram that cooperates with LNR in the main text, and show the other plot in the supplementary file. We hope the revision is acceptable.

10.We believe that the nomogram established by the author should be compared with traditional models, such as ROC curve and Decision Curve Analysis.
Response: In the revised paper, ROC and DCA curve are made to compare the prognostic powers between the new nomogram models based on LNR and LODDS and the traditional model based on N stage. The results indicate that the new nomogram models are better. For details, please see the supplementary figure 3. We really thank you for the suggestion that further confirms our results.

##Reviewer: 2
Dr. Rasa Zarnegar , Weill Cornell Medical College

Comments to the Author:
I think this is a nice paper with 2 nomograms for the determination for CGA.

I think there are some revisions that would make this paper better. The SEER DB has access to total number of nodes analyzed and the number of positive nodes. It is important to use the current guidelines 8th Edition for this analysis even though the data was from prior to implementation of the 8th. The concept should still hold and validate the rigor of the study based on current guidelines.
Response: All the 7th stagings are now corrected to the 8th edition. Thank you for your brilliant suggestion.

2. Raw data on the patient population No of nodes harvested and total positive is required to determine the frequency of low yield in the study design.
Response: In the revised paper, we describe the number of nodes harvested and total positive, and the frequency of low nodes yield (please see Table 1). The frequency of low yield is also in the multivariable model because it is associated with survival in univariate analysis (Suppl. Table 2).

3. There should be discussion and data on neoadjuvant therapy as this likely impact survival and the number of patient that received therapy.
Response: Neoadjuvant therapy is likely to influence survival; unfortunately the database, Incidence - SEER Research Data, 18 Registries, Nov 2019 Sub 2000-2017 (SEER 18 database), does not provides with information about chemotherapy, therefore it is unavailable for this study. We treat it as a limitation and discuss the impact of neoadjuvant therapy in the Discussion Section.
"…Treatment mode is also associated with clinical outcome. This study enrolled patients who received gastric resection; however other information about chemo- or radiotherapy is not available in SEER 18 database. Randomized clinical trial demonstrates that compared with surgery alone, preoperative administration of carboplatin and paclitaxel with concurrent radiotherapy significantly improved overall survival among patients with esophageal or GEJ cancer (HR = 0.657) (31). The NCCN clinical practice guidelines for GEJ cancer recommend preoperative chemoradiation or perioperative chemotherapy due to substantial survival benefit compared with surgery alone (32). To overcome this limitation, a database that provides with fully detailed medical records is needed for analysis…."

4. I suggest picking one nomogram. I think its important to send a clear message and by presenting 2 nomograms the authors are creating confusion and an unclear message. I think by being more focused on one approach would allow for improved implementation. The authors may want to compare whichever they select with conventional lymphadenectomy is so desired.
Response: Thanks for the advice. We pick the nomogram that incorporates with LNR because the calculation of LNR is easier than LODDS, which is more convenient in clinical practice. The nomogram and calibration curves based on LODDS are shown in supplementary figure 2. In addition, for more clear presentation, we put the calibration curves of training (red) and validating set (blue) into one plot.

VERSION 2 – REVIEW

REVIEWER	Zhong, Rui Southwest Medical University
REVIEW RETURNED	01-Jul-2021

GENERAL COMMENTS	1.The author's verification of the Cox PH model was correct, because the variables included in the histogram were classified variables, so we believe that only the KM curve can be used for verification. 2.We noticed that the clinical usability of the nomogram was no better than that of other models in decision curve analysis, which was a problem. In other words, the decision curve analysis does not show the advantages of the nomogram. 3.The author has made very detailed modifications and answers to the other questions.

REVIEWER	Zarnegar , Rasa Weill Cornell Medical College, Surgery
REVIEW RETURNED	07-Jul-2021

GENERAL COMMENTS	Nice revisions. No further issues

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Rui Zhong, Southwest Medical University

Comments to the Author:

1.The author's verification of the Cox PH model was correct, because the variables included in the histogram were classified variables, so we believe that only the KM curve can be used for verification.

Response: According to the current suggestion, we have removed the plot of the Schoenfeld, Martingale, and Deviance residuals. The KM curves were newly plotted for PH model observation, and were presented in Suppl. Fig 1. In addition, we mentioned that the statistical test (cox.zph package in R software) also showed that PH assumption held for each variable in the model.

"…all the variables met proportional hazard assumption (Suppl. Figure 1, all P values > 0.05)…"

2.We noticed that the clinical usability of the nomogram was no better than that of other models in decision curve analysis, which was a problem. In other words, the decision curve analysis does not show the advantages of the nomogram.

Response: We have re-phrased the relevant context and caution readers that the decision curve did not show advantage of the nomogram. We further discussed it in the limitation section.

Line 175-176 "…However the DCA does not show advantage over the nomogram (Suppl. Figure 3)"

Line 257-261 "…The third limitation is clinical usability. The DCA result is proposed for assessing the potential clinical impact of risk models for recommending treatment or intervention, and the suggested clinical usability of the nomogram may be poorer than that of other models. In this regard, although this model may have some merits regarding outcome prediction, its use for guiding clinical decisions should be further studied."

3.The author has made very detailed modifications and answers to the other questions.

Response: We really thank you for your suggestions that help us improve this study.

Reviewer: 2

Dr. Rasa Zarnegar , Weill Cornell Medical College

Comments to the Author:

Nice revisions. No further issues

Reviewer: 1

Competing interests of Reviewer: No

Reviewer: 2

Competing interests of Reviewer: None

VERSION 3 – REVIEW

REVIEWER	Zhong, Rui Southwest Medical University
REVIEW RETURNED	30-Jul-2021

GENERAL COMMENTS	Nice revisions. No further issues.