Peer Review File

Article information: https://dx.doi.org/10.21037/jgo-23-627

Reviewer A

Nice paper! By being on the transparent reporting side lately, I believe it can be improved by following reporting tools for radiomics and quality scoring tools. These may help show the overall quality of this paper.

• Please try to follow a radiomic-specific reporting guideline (CLEAR checklist endorsed by ESR) not to miss any technical issues in reporting. Fill and add it to the supplements. Try to cover all the items in the methods and the open science section.

Reply 1: Thanks for your suggestion! We have conducted a step-by-step comparison and made corresponding modifications according to the radiomic-specific reporting guideline (CLEAR checklist endorsed by ESR).

• Please also try to calculate the RQS score by Lambin et al. on a percentage scale and add this to the results. Also, add the filled RQS form in the supplement.

Reply 2: Thanks for your suggestion! The RQS score by Lambin et al. was calculated and the results was added in the supplement.

Changes in the text:

(1) Furthermore, the Radiomics Quality Score (RQS) developed by Lambin et al. serves as a scoring system to evaluate the quality and reliability of radiomics studies. The RQS was calculated to evaluate the quality of this study (Lines 336-338).

(2) Moreover, we calculated the corresponding RQS points for this study to evaluate the quality of the radiomics research. The results indicated a total RQS score of 18.0. And the detailed scoring criteria were presented in Table S2 (Lines 439-442).

(3) Table S2 The result of radiomics quality score (RQS) in this study (Lines 724-725).

Id	Research Details	Research Purpose	RQS Point
1	Image protocol quality	To ensure the repeatability of the experiment.	1.0
2	Multiple segmentation	To analyze the impact of different segmentation methods on features.	1.0
3	Phantom study	To analyze the impact of different machine types on features.	0.0
4	Imaging at multiple time points	To analyze the impact of temporal heterogeneity, such as organ motion.	0.0
5	Feature reduction or adjustment for multiple testing	To prevent overfitting.	3.0
6	Multivariable analysis	To increase the clinical practicality of radiomics.	1.0
7	Biological correlates	To find the connection between radiomics and biological mechanisms.	0.0
8	Cut-off analysis	To reduce the risk of optimistic estimation.	1.0

9	Discrimination statistics	To reflect the predictive performance of the model.	2.0
10	Prospective study	To provide the highest level of evidence for radiomics research.	0.0
11	Calibration statistics	To reflect the stability of the model.	2.0
12	Validation	To increase the credibility of the model.	2.0
13	Comparison to gold standard	To demonstrate the additional value of radiomics.	0.0
14	Cost-effectiveness analysis	To report on the clinical significance of radiomics.	2.0
15	Cost-effectiveness analysis	To increase the clinical significance of radiomics.	0.0
16	Open science and data	To promote knowledge transformation and improve the repeatability of radiomics.	3.0
Total Points			18.0

Reviewer B

Thank you for conducting the research. This article is very interesting and well-written. It has good innovation regarding the use of new methods of artificial intelligence. The issue that has been investigated is globally important and prioritized due to the prevalence and widespread mortality of gastric cancer.

I have some questions and suggestions:

1. What is the relationship between the contents stated in lines 247 and 248 with reference number 22?

Reply 1: Thank you! In lines 247 and 248 of the article, we describe how to reduce the dimensionality of imaging omics features and select key features. Among them, we integrated 8 machine learning algorithms and 29 algorithm combinations, and the source and code of this method are mainly from reference 22.

2. What is the reason for choosing the 8 selected algorithms? Would it be better to explain briefly about each algorithm?

Reply 2: Thanks for your question! Here, we selected 8 different algorithms to construct the best prognostic evaluation model based on the radiomics features of gastric cancer (GC) patients. Firstly, this method was applied to improve the predictive performance of the model. Different algorithms may have different assumptions and methods. The application of different algorithms for dimensionality reduction and feature selection can help identify and select the most relevant features for the predicted target, thereby obtaining more comprehensive and accurate feature selection. Then, this process could prevent overfitting of the model to some extent. In the construction of prognostic models, overfitting is a common issue, especially when dealing with high-dimensional features. The result could reduce reliance on a specific algorithm by comparing and validating multiple algorithms to discover more robust and well-performing

predictive models. Finally, it is worth noting that diverse dimensionality reduction and feature selection algorithms often emphasize different interrelationships among features. And we could delve into a comprehensive analysis of the correlations, importance, and intricate relationships between features and the predictive target variable by comparing multiple algorithms.

Moreover, I would introduce the 8 algorithms used in the article briefly.

• Random Survival Forest (RSF): RSF algorithm is a method based on random forests used for handling survival analysis problems. It involves constructing multiple decision trees and using randomly selected features in each tree to predict survival probabilities.

• Elastic Net (Enet): Enet algorithm is a linear regression technique that integrates the distinctive features of LASSO and ridge regression. By simultaneously considering the L1 and L2 regularization terms, it achieves both sparsity and variable selection capabilities.

• Least Absolute Shrinkage and Selection Operator (LASSO): LASSO algorithm is a linear regression method that promotes coefficient sparsity by incorporating an L1 regularization term. This enables feature selection and model construction.

• Ridge: Ridge algorithm, as a linear regression method, mitigates overfitting risks by incorporating an L2 regularization term to regulate model complexity.

• Stepwise Cox: Stepwise Cox regression algorithm is a variable selection method applied in survival analysis with Cox proportional hazards model. It constructs the optimal model by gradually adding or removing variables.

• CoxBoost: CoxBoost algorithm is a survival analysis approach that leverages the power of gradient boosting algorithm. By iteratively fitting weighted Cox regression models, it gradually enhances predictive performance, ensuring accurate and refined prognostic estimates.

• Partial Least Squares Regression for Cox (plsRcox): plsRcox algorithm is a method that employs partial least squares regression for Cox model modeling. It establishes linear relationships between features and the response variable to identify significant components, thereby reducing the number of correlated variables.

• Generalized Boosted Regression Model (GBM): GBM algorithm is an ensemble learning algorithm that enhances predictive performance by iteratively constructing multiple weak classifiers, such as decision trees.

3. Line 259, I could not understand this phrase: “In the developing cohort, 29 algorithm combinations were calculated for the high consistency features of patients, and the corresponding prediction model was constructed through cross validation.”

Reply 3: Thank you! I would explain this analysis method in detail. After completing image segmentation and feature extraction, we conducted an initial assessment of feature consistency. This process involved comparing the obtained radiomics features to evaluate the quality of the segmentation results based on the degree of consistency achieved. In this study, an experienced radiologist (Y.X.) and a general surgeon (Y.H.) drew the tumor focus randomly and independently without assessing any clinicopathological characteristics or outcomes of these patients. After that, the radiologist (Y.X.) randomly selected 50 patients and segmented their ROI again one month later to evaluate the consistency of ROI image quality. Then, the intraclass correlation coefficients (ICCs) of 1218 radiomics features from 50 patients pre- and post- segmentation were calculated to evaluate the repeatability of the data. And highly

consistent features were considered to be those with ICC values >0.75 were defined as high consistency features. Therefore, high consistency features (ICC>0.75) were used for subsequent model construction and low consistency features were discarded. Finally, we employed a combination of 29 algorithms to perform feature selection and construct predictive models.

4. In what proportion have 171 stomach cancer patients been divided into two cohorts for development and evaluation?
Reply 4: Thanks for your valuable question! We randomly divided all patients into developing and validation cohorts at a ratio of 7:3 (Lines 287-288).

5. What is meant by using 29 algorithm combinations? Is it meant to combine feature selection techniques with machine learning algorithms?
Reply 5: Thank you! When confronted with a multitude of imaging omics features whose prognostic significance remains unclear, the completion of feature selection and model construction becomes imperative. Algorithm combination refers to the utilization of Algorithm A for comprehensive variable screening, followed by Algorithm B employing all the variables obtained from A to construct the model without further dimensionality reduction. In simpler terms, within the cross-validation framework, one algorithm is employed for precise feature selection, while another algorithm is utilized for constructing the prognostic model. Notably, among the eight machine learning algorithms considered, RSF, Lasso, CoxBoost, and others exhibit dual capabilities in both feature selection and model construction, signifying their ability to independently accomplish the model construction process. Consequently, through various pairwise combinations of these algorithms, a total of 29 algorithmic compositions can be derived. In this study, we used these 29 algorithmic compositions to perform meticulous feature selection and robust model construction for radiomics features.

Reviewer C

1) First of all, my major concern regarding this paper is the unclear focus of this study, as indicated in the title and elsewhere of this paper. Because radiomics-based PRS is only one of the predictors in the nomorgram, the title is not accurate. Further, in the data analytic strategies, the authors need to report the predictive accuracy with and without PRS, to indicate the critical role of PRS. The current title also did not indicate the clinical research design of this study, i.e., the development and validation of the prediction model based on a retrospective cohort.
Reply 1: Thanks for your suggestion! The purpose of this article is to construct a stable prognosis prediction model for GC patients based on enhanced CT images. As a novel biomarker, the radiomics risk score (RRS) could effectively and accurately predict the clinical outcome, disease recurrence, and the benefits of postoperative chemotherapy. The RRS combined with clinicopathological features could further promote the practicability of the clinical prediction model. Therefore, the construction of Nomogram is only one of the application directions of RRS. Here, we have made corresponding modifications in the article. Firstly, we have added the research design of this study in the title, which is a retrospective

study. Next, we described the evaluation method of RRS prediction ability in the research method.

Changes in the text:

(1) Radiomics based on machine learning algorithms could predict prognosis and postoperative chemotherapy benefits of patients with gastric cancer: a retrospective cohort (Lines 4-5).

(2) The timeROC package (version 0.40) was utilized for plotting ROC curves and calculating AUC values to assess the accuracy of RRS predictions for prognosis (Lines 334-335).

2) Second, the abstract needs some revisions. The background did not the clinical needs for this research focus and why machine learning algorithms and radiomics parameters could accurately predict the prognosis. The methods need to describe the inclusion of subjects, the assessment of baseline clinical factors, how the patients were followed up, and the measurements of prognosis outcomes. The results need to summarize the baseline clinical characteristics of the patient cohort. The authors need to tone down the current conclusion because of the 0.733 AUC for DFS, which is lower than 0.75.

Reply 2: Thanks for your suggestion! We have made corresponding modifications in the Abstract. First, we have added the content of clinical needs for this research focus in the background. Then, we further described the collection of clinical features and follow-up data in the methods.

Changes in the text:

(1) Traditional clinical characteristics have certain limitations in evaluating cancer prognosis. The radiomics features provide information on tumor morphology, tissue texture, hemodynamics, and other aspects, which can accurately reflect personalized predictions (Lines 47-50).

(2) The general information, pathological characteristics, and postoperative chemotherapy information were then collected. And patients were also monitored through telephone interviews or outpatient treatment (Lines 56-59).

3) Third, in the introduction of the main text, the authors need to explain the benefits of the accurate prediction of the prognosis in GC and its clinical significance. The authors need to review available prediction models for the GC patients and have comments on their limitations including their accuracy, as well as the analysis on the adoption of machine learning algorithms and radiomics parameters to indicate their potential to improve the predictive accuracy.

Reply 3: Thanks for your suggestion! We have made corresponding content in the Introduction. We explained the benefits and clinical significance of predicting the prognosis of GC patients. And the limitations of previously reported prognosis models for GC patients have been compared in the discussion section. “Shin et al. used a radiomics model to predict the relapse-free survival (RFS) of patients with advanced GC (26). In the training cohort, the C-index of the prediction model for RFS was 0.719 (95% CI, 0.674–0.764). Li et al. built a radiomics signature based on intratumoral and peritumoral regions to predict the DFS of patients (27). The results showed that the radiomics signature was an independent risk factor for the prognosis of patients. Moreover, Zhang et al. further built a radiomics prediction model for early recurrence in patients with advanced gastric cancer (28). However, these models only used single or traditional machine learning algorithms to search for key radiomics features, which may miss some important features (Lines 458-467).”

Changes in the text:
(1) Therefore, precise and personalized prognosis prediction could better assist physicians in comprehending the risks associated with disease progression and provide them with a solid scientific foundation for devising optimal treatment strategies (Lines 147-150).
(2) Therefore, models constructed based on radiomics features could make personalized predictions based on the unique features and tumor manifestations of each patient. This helps to determine the survival risk of patients and provides a basis for developing personalized treatment plans more accurately (Lines 166-169).

4) Fourth, in the methodology of the main text, please accurately describe the clinical research design, sample size estimation, and how the training and validation samples were generated. In statistics, please provide the threshold value of AUC for a good prediction model and indicate the P value for statistical significance.
Reply 4: Thanks for your suggestion! In the Methods, we have described the detail of the clinical research design according to the Fig. 1A. (1) "GC patients who underwent radical surgery in the First Affiliated Hospital of Wenzhou Medical University from January 2014 to December 2016 were included retrospectively (Lines 187-188)." (2) "The inclusion criteria were: (I) age >18 years old; (II) abdominal enhanced CT scan before operation; (III) gastric adenocarcinoma confirmed by surgical treatment and pathological examination; and (IV) agreement to participate in this study. The exclusion criteria were as follows: (I) patients who did not agree to participate in this study; (II) patients who received neoadjuvant chemotherapy; (III) patients with confirmed incurable cancer metastasis during surgery; (IV) patients with other tumors or other serious organic diseases; and (V) patients with no abdominal enhanced CT examination or imaging data before the operation or who underwent preoperative abdominal CT imaging in other hospitals (Lines 194-209)." (3) "To screen the prognostic imaging characteristics of gastric cancer, we randomly divided all patients into developing and validation cohorts at a ratio of 7:3 (Lines 287-288)." And we further used the timeROC package to calculate the threshold value of AUC of RRS for predicting the survival status of GC patients. The threshold value of AUC was 0.10. "According to the best cutoff value of the RRS of GC patients in the developing cohort (RRS = 0.10), we divided all patients into high-risk and low-risk groups (Lines 309-313)."

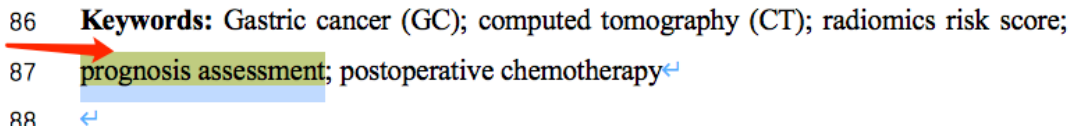
5) Finally, some related papers need to be reviewed and cited: 1. Li Z, Chen L, Song Y, Dai G, Duan L, Luo Y, Wang G, Xiao Q, Li G, Bai S. Predictive value of magnetic resonance imaging radiomics-based machine learning for disease progression in patients with high-grade glioma. Quant Imaging Med Surg 2023;13(1):224-236. doi: 10.21037/qims-22-459. 2. Ma T, Cui J, Wang L, Li H, Ye Z, Gao X. A CT-based radiomics signature for prediction of HER2 overexpression and treatment efficacy of trastuzumab in advanced gastric cancer. Transl Cancer Res 2022;11(12):4326-4337. doi: 10.21037/tcr-22-1690. 3. Dong Z, Liu G, Tu L, Su X, Yu Y. Establishment of a prediction model of postoperative infection complications in patients with gastric cancer and its impact on prognosis. J Gastrointest Oncol 2023;14(3):1250-1258. doi: 10.21037/jgo-23-231.
Reply 5: Thank you! The corresponding references have been cited in the article.
Changes in the text:

Several studies have shown that radiology has a good predictive ability in subtype classification (10,11), staging evaluation (12), clinical outcome (13-16), and treatment response of tumor patients (17,18). Moreover, machine learning has unique advantages in processing high-dimensional data and finding feature variables (19-21) (Lines 169-173).

11. Ma T, Cui J, Wang L, et al. A ct-based radiomics signature for prediction of HER2 overexpression and treatment efficacy of trastuzumab in advanced gastric cancer. Transl Cancer Res, 2022, 11: 4326-4337 (Lines 582-584)
16. Dong Z, Liu G, Tu L, et al. Establishment of a prediction model of postoperative infection complications in patients with gastric cancer and its impact on prognosis. J Gastrointest Oncol, 2023, 14: 1250-1258 (Lines 597-599)
21. Li Z, Chen L, Song Y, et al. Predictive value of magnetic resonance imaging radiomics-based machine learning for disease progression in patients with high-grade glioma. Quant Imaging Med Surg, 2023, 13: 224-236 (Lines 612-614)

Reviewer D

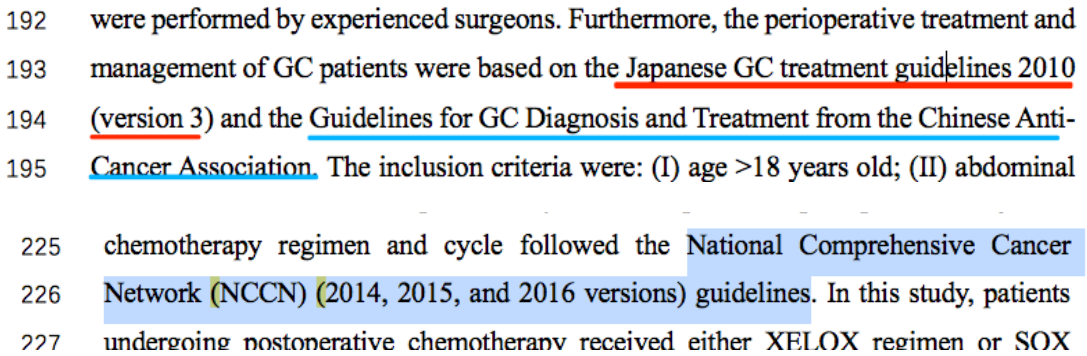
1. This **Keyword** cannot be found in Abstract or main text. Please revise.


86 **Keywords:** Gastric cancer (GC); computed tomography (CT); radiomics risk score;
87 prognosis assessment; postoperative chemotherapy
88

Reply: Thank you! We have revised the Keyword.

2. Reference/citation

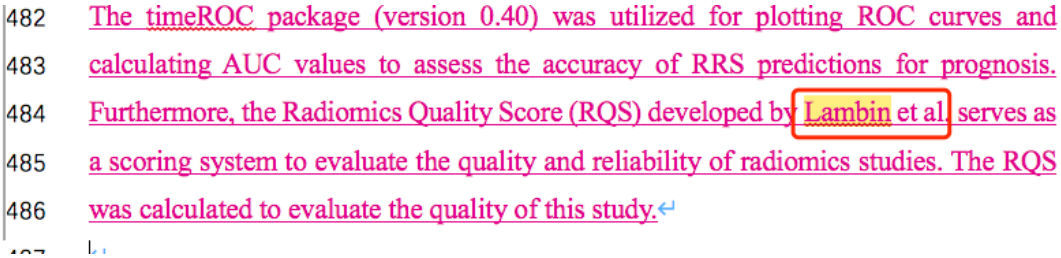
a. Please confirm if references should be cited for the guidelines.


192 were performed by experienced surgeons. Furthermore, the perioperative treatment and
193 management of GC patients were based on the Japanese GC treatment guidelines 2010
194 (version 3) and the Guidelines for GC Diagnosis and Treatment from the Chinese Anti-
195 Cancer Association. The inclusion criteria were: (I) age >18 years old; (II) abdominal

225 chemotherapy regimen and cycle followed the National Comprehensive Cancer
226 Network (NCCN) (2014, 2015, and 2016 versions) guidelines. In this study, patients
227 undergoing postoperative chemotherapy received either XELOX regimen or SOX

Reply: Thank you! We

b. Reference of "Lambin et al." should be cited here. Please revise.

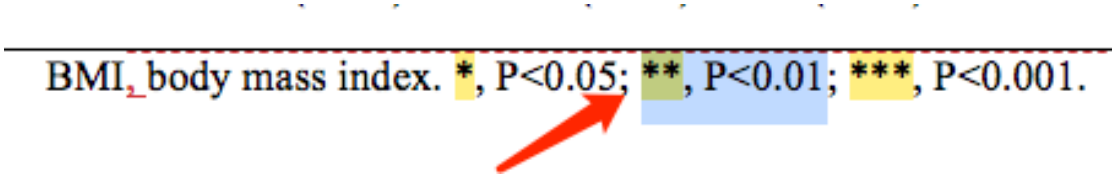

482 The timeROC package (version 0.40) was utilized for plotting ROC curves and
483 calculating AUC values to assess the accuracy of RRS predictions for prognosis.
484 Furthermore, the Radiomics Quality Score (RQS) developed by Lambin et al. serves as
485 a scoring system to evaluate the quality and reliability of radiomics studies. The RQS
486 was calculated to evaluate the quality of this study.

3. Table 1

a. Please add the unit of BMI.

Reply: Thank you! We have added the unit of BMI.

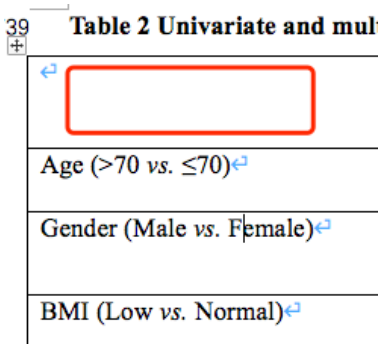
b. No "**" was marked in table 1, but it is explained in footnote. Please check and revise.

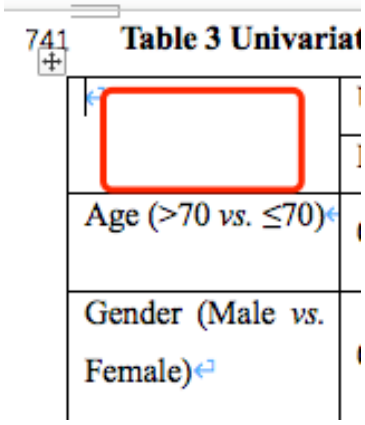


Reply: Thank you! We have deleted "*, **" in the table 1.

4. Table 2 and Table 3

a. Please add a table header.





Reply: Thank you! We have added the table header.

b. There are some empty boxes in tables 2 and 3. Please confirm if data are missing in them.

Table 2 Univariate and multivariate Cox analyses for overall survival (OS) of GC patients in the developing cohort.

	Univariate Cox analysis				Multivariate Cox analysis			
	HR	HR.95L	HR.95H	P value	HR	HR.95L	HR.95H	P value
Age (>70 *vs.* ≤70)	1.11	0.61	2.03	0.741				
Gender (Male *vs.* Female)	0.80	0.42	1.55	0.510				
BMI (Low *vs.* Normal)	0.70	0.17	2.93	0.629				
BMI (High *vs.* Normal)	0.86	0.43	1.73	0.671				
Anemia (Yes *vs.* No)	1.42	0.80	2.51	0.233				
Charlson score (≥2 *vs.* <2)	3.38	1.87	6.10	P<0.001***	2.38	1.18	4.82	0.016*
Location (Body *vs.* Cardia)	0.99	0.37	2.66	0.983				
Location (Pylorus *vs.* Cardia)	1.30	0.56	2.99	0.539				
Location (All *vs.* Cardia)	2.24	0.71	7.08	0.168				
Tumor size (≥3 *vs.* <3)	11.24	3.48	36.25	P<0.001***	4.47	1.24	16.07	0.022*
Nerve invasion (Yes *vs.* No)	2.10	1.18	3.76	0.012*	1.16	0.58	2.32	0.670
Vascular invasion (Yes *vs.* No)	1.51	0.82	2.79	0.187				
Stage (Stage II *vs.* Stage I)	1.10	0.21	5.69	0.906	0.96	0.18	5.18	0.966

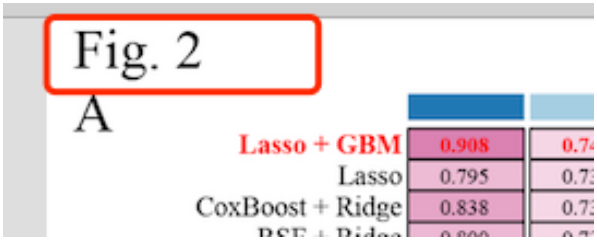
Reply: Thank you! We have confirmed the content in tables 2 and 3. Since these variables had a p-value greater than 0.05 in the univariate Cox analysis, they were not further included in subsequent multivariate Cox analysis.

5. **When using abbreviations** in table/figure or table/figure description, please mention the entire expression in a footnote below the corresponding table/figure. Please check and revise. Such as: GC, DFS, RRS, AIC, (in figure 5); RRS (in table 1); GC, HR, L, H (in table 2); RRS, GC, HR, L, H (in table 3); GC (in table S1); etc.

Reply: Thank you! We have added the entire expression in a footnote below the corresponding table/figure.

6. Figures 1-7

Please remove the wording "Fig. XX" from figures.



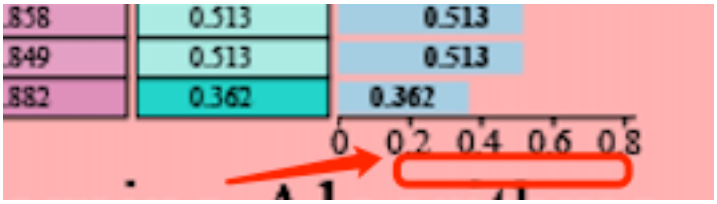
Reply: Thank you! We have removed the wording "Fig. XX" from figures.

7. Figure 1

a. To better prioritize your manuscript for copy-editing, it would be much appreciated if you could kindly provide an **editable** form for **figure 1A** in **DOC/PPT format**, which should not be as a picture.

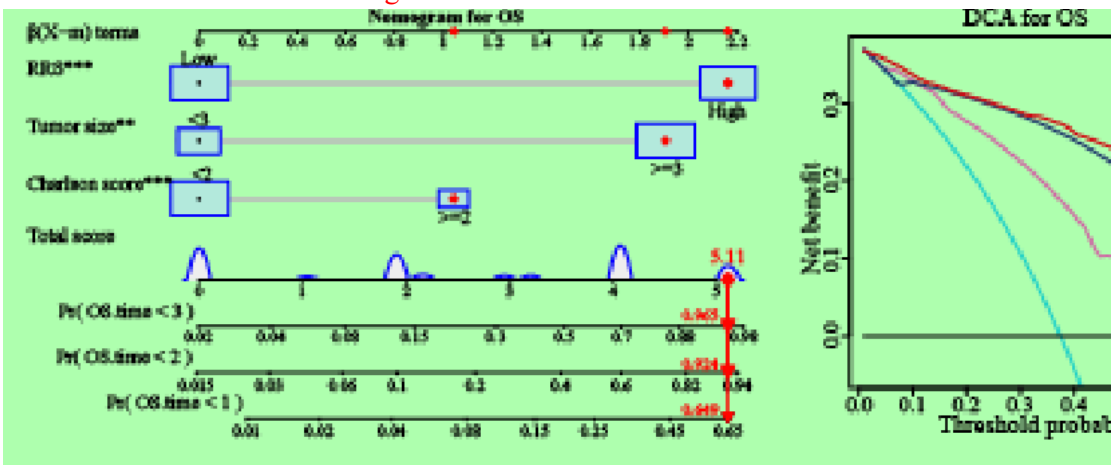
Reply: Thank you! We have provided an editable form for Figure 1 in Adobe Illustrator 2020 format.

b. Please add description for the X-axis in figure 1B.



Reply: Thank you! We have added the description for the X-axis in the Figure 1B.

c. Please send us **Figure 1B** with higher resolution as a separate file in JPG/TIFF, as the current one is not clear enough.

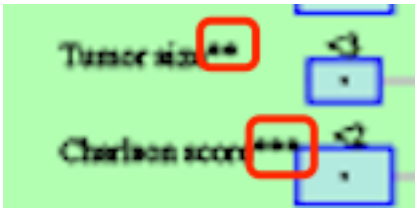


Reply: Thank you! We have provided an editable form for Figure 1 in Adobe Illustrator 2020 format and the picture file in TIFF with 600 DPI.

d. Please confirm if this picture was taken by the authors. If it is not original, please remove it.

Reply: Thank you! We have removed this picture in the Figure 1B.

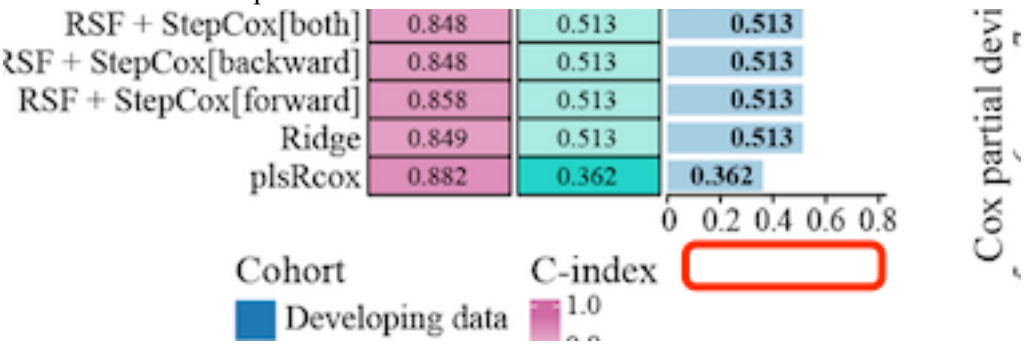
e. Please include definitions of any special symbols (**, ***) used in the figure legend.



Reply: Thank you! We have added the description in the Figure 1B.

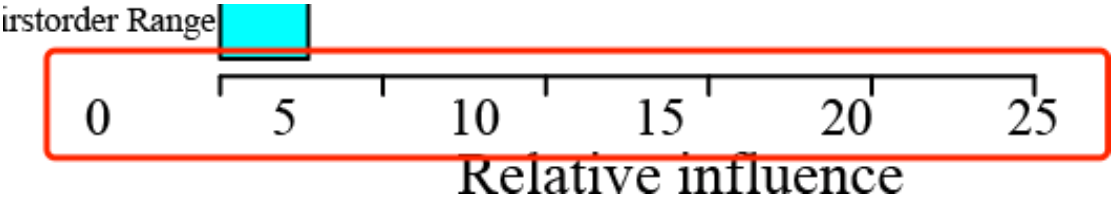
8. Figure 2

a. Please add description for the X-axis.



Reply: Thank you! We have added the description for the X-axis in the Figure 2A.

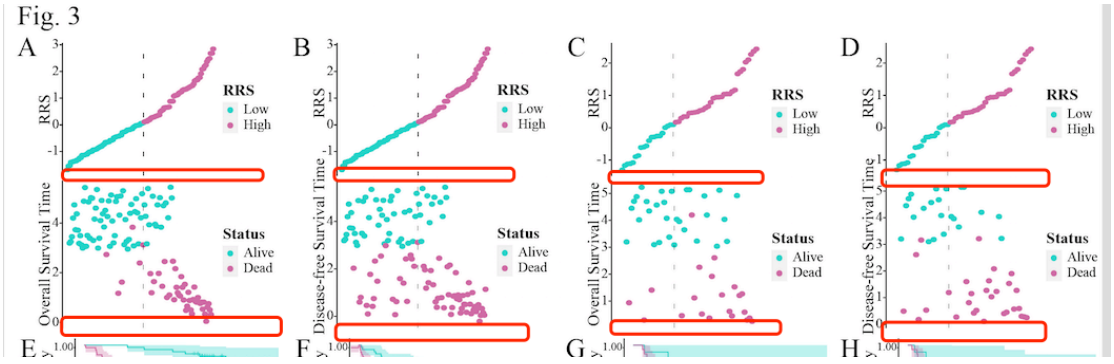
b. Please check and revise the X-axis.



Reply: Thank you! We have revied the X-axis in the Figure 2E.

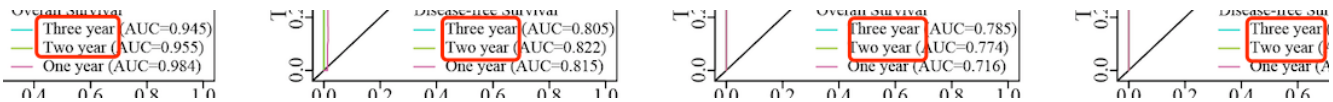
9. Figure 3

a. It seems that X-axis is missing in figure 3A-D. Please check and revise.



Reply: Thank you! We have added the X-axis in the Figure 3A-D.

b. Figure 3I-L: please chose to revise them to "Three-year" and "Two-year"; **or** "Three years" and "Two years".



Reply: Thank you! We have made corresponding modifications in the Figure 3I-L.

10. Figure 4

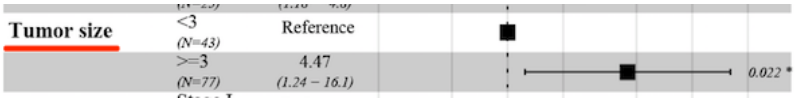
No "*, **" are marked in figure 4, but they were explained in figure legend. Please check and revise.

706 receive postoperative chemotherapy. *, P<0.05; **, P<0.01; ***, P<0.001.

Reply: Thank you! We have deleted "*, **" in the figure legend.

11. Figure 5

Please add unit for the tumor size.



Reply: Thank you! We have added unit for the tumor size in the Figure-5-revised.