Peer Review File

Article information: http://dx.doi.org/10.21037/tcr-22-1238

Reviewer A

1. **Comment 1:** In lines 21-22 (introduction) the authors use GLOBOCAN 2018 as a reference, please use the most recent data (2022).

Reply 1: We have made correction according to the Reviewer's suggestion.

Changes in the text: see Page 3, line 3.

2. **Comment 2:** Please insert reference(s) to the statement "Patients often present with distant metastasis, poor performance status, and have an unfavorable outcome even with aggressive interventions".

Reply 2: We have inserted relevant reference in the manuscript as advised.

Changes in the text: see Page 3, line 5.

3. **Comment 3:** Please specify which version of SEER database was used.

Reply 3: We collected data between 1975 and 2017 using SEER*Stat software (v 8.3.5). We have modified our text as advised.

Changes in the text: see Page 4, line 10.

4. **Comment 4:** Have the assumptions for the cox proportional hazard models been tested? If yes, how? Please describe in the manuscript.

Reply 4: First, log-rank test was used to analyze univariate factors. Second, factors with a p value < 0.05 in univariate analysis were carried into a multivariate COX Proportional Hazard Regression analysis to obtain the hazard ratio (HR) and corresponding 95% confidential interval (CI). Independent prognostic variables, identified by univariate and multivariate analyses, were used for generating nomograms to predict the 2- and 5-year OS and CSS rates. Our models were further tested by C-index, AUC values and calibration curves. We have revised the statistical

analysis section of our manuscript as suggested.

Changes in the text: see Page 5, line 6-9.

5. **Comment 5:** What is the methodology for inserting variables in multivariate models? Backward selection, forward selection, clinical decision, p value? Please describe in the manuscript.

Reply 5: Factors with a p value < 0.05 in univariate analysis were carried into a multivariate COX Proportional Hazard Regression analysis to obtain the hazard ratio (HR) and corresponding 95% confidential interval (CI). We have revised the statistical analysis section of our manuscript as suggested.

Changes in the text: see Page 5, line 6-9.

6. **Comment 6:** In the discussion, I suggest some lines about the clinical impact and what is the suggested management of patients based on the findings.

Reply 6: In page 9/ line 5-6, we suggested that early detection and intervention might be critical for the long-term survival in patients with esophageal SRCCs. In page 10/ line 1-4, we suggested that chemotherapy was a significant prognostic factor and future subgroup analyses according to different chemotherapy regimens should be carried out to validate its role in esophageal SRCC.

"Moreover, during patient-clinician communication, our novel nomograms comprised of a few easily accessible clinical variables can help clinicians accurately estimate individual prognosis and thereby design appropriate treatment strategies for different patients."

"Additionally, our study confirmed the feasibility of nomogram in generating a numerical probability of survival in esophageal SRCC patients and provided a direction for future multicentre, large-scale cohort studies with adequate follow-up time."

We have also added these comments to the discussion section according to the Reviewer's suggestion.

Changes in the text: see Page 10/ line 21 and Page 11/ line 1, 19-21.

Reviewer B

1. **Comment 1:** According to the data use agreement, authors are required to provide details about which SEER database they used. This would also help reviewers understand how better to help them. The choice of the database is usually a choice between years of coverage (back to 1975) versus the detail required for the analyses. Typically, unless the authors are looking for trends over time, the more recent data (17 or 18 registries for the 2020 and 2021 submissions) are the best.

Reply 1: We collected data between 1975 and 2017 using SEER*Stat software (v 8.3.5). We have modified our text as advised.

Changes in the text: see Page 4, line 10.

2. **Comment 2:** The SEER variables for chemotherapy and for radiation therapy have significant limitations which were highlighted when signing the data use agreement to receive the data. According to the SEER data use agreement, the authors are required to highlight the limitations of the chemotherapy and radiation therapy variables in their manuscript.

Reply 2: We have modified the limitations of our study in the discussion part as advised.

"First, our study was limited to retrospective data collection, which may lead to inevitable bias. The variables of chemotherapy and radiotherapy were classified as "yes" or "no/unknown" in the SEER dataset, we cannot accurately distinguish between "no treatment" and "unknown" if patients received treatment. Second, the sequence of treatment variables was not considered. Since tumor recurrence and progression were not recorded in the dataset, we had to use treatment as a baseline variable rather than a time-dependent covariable. In the absence of an exact time of treatment, we assume that the exact combination of treatments is determined and given at the time of diagnosis. This assumption is necessary to incorporate treatment information into the model."

Changes in the text: see Page 11/ line 4-14.

3. **Comment 3:** AJCC 7th edition staging is only available for select years – it is not available back to 1975. Authors should examine the year of diagnosis in the data to see what years are actually captured. For each period of time in SEER, a different stage variable may be required. If earlier staging is specific enough to capture relevant patients, this may increase the sample size. For example, Derived AJCC 7th Edition is only available for 2010-2015. Derived SEER Combined stage group is available for 2016-2017. To get all patients of a particular stage from 2010-2017 would require that both stage definitions be used to select patients. It may be helpful to watch some of the webinars available to understand how to properly use the SEER data: https://seer.cancer.gov/news/seerstat-webinars.html#past-4.

Reply 3: We adopted AJCC 7th Edition in our study.

"Tumors were classified in accordance with the 7th AJCC Tumor-Node-Metastasis (TNM) staging manual, which was published in 2010, and so the years allowed for diagnosis ranged from 2010 to 2015".(see Page 4/line 21-22 and Page 5/ line 1)

4. **Comment 4:** In addition to the limitations of the treatment variables, treatment can't be part of a predictive model that starts at diagnosis because of immortal time bias. That is, patients who are treated, by definition, have survived long enough to be treated. In other words, they cannot die between diagnosis and treatment. There is a SEER variable for time from diagnosis to treatment. It is up to the authors as to whether their prediction model is most relevant to treated patients or to all diagnosed patients. Generally, the only way to include treatment variables to predict outcomes from diagnosis is by using time-dependent covariates for treatment (i.e., patients are counted as "not treated" until they are "treated"). Time-dependent covariates make prediction models very challenging.

   Along these lines, if the authors would prefer a model that predicts outcomes starting from diagnosis (i.e., not from treatment initiation) patients who die in the month after diagnosis should not be removed. It would be better to use a half month (0.5) for their time to death since the SEER time variable is defined in

months.

Reply 4: I have confirmed with the co-author of the data analysis that he did not exclude cases that died within a month of diagnosis. We are sorry for our miswriting. We have deleted the sentence from the manuscript. In addition, we addressed the issue and revised the limitations of our study in the discussion part.

"Second, the sequence of treatment variables was not considered. Since tumor recurrence and progression were not recorded in the dataset, we had to use treatment as a baseline variable rather than a time-dependent covariable. In the absence of an exact time of treatment, we assume that it is determined and given at the time of diagnosis. This assumption is necessary to incorporate treatment information into the model.**"**

Changes in the text: see Page 11/ line 10-14.