Peer Review File

Article information: https://dx.doi.org/10.21037/atm-21-1725

Reviewer A

Comment 1: Lack of novelty. A study published in 2017 entitled "Nomograms for predicting long-term overall survival and cancer-specific survival in patients with major salivary gland cancer: a population-based study" (Li Y, et al, Oncotarget. 2017 Apr 11;8(15):24469-24482. PMID 28160551) had similarly developed and validated nomograms with c-index of 0.817 for OS and 0.829 for CSS. The authors failed to cite this work in their manuscript and made no attempt to compare their nomograms with the published studies such as the one listed here as an example. Although a greater number of patients were used in the current manuscript, the resulting nomograms do not appear to out-perform those published by Li Y et al. Notably, SEER database was also used by Li Y et al. It seems to me that the authors will need to address this directly, compare the utilities of different nomograms, and present the strengths of their new nomograms with analytical data.

Reply 1: We truly appreciated all of your valuable suggestions as well as generous comments. We read the paper written by Li Y et al. very carefully. After discussion across all authors, we believe that there are some major defects in their study. Thus, the reliability of the nomogram constructed by Li Y et al. is concerning.

(1) The study by Li Y et al. was not conducted according to the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIOPD) statement, the most important guideline for prediction model development and validation, which was published in 2015 (Collins GS et al. Ann Intern Med. 2015 Jan 6;162(1):55-63.).

(2) The primary outcome were 5-year and 8-year overall survival (OS) and cancer-specific survival (CCS) in their study. However, the median and range of follow-up were 34 months (1-119 months) and 50 months (3-120 months). Only a small portion of patients had a follow-up period more than 5 year (60 months), let alone 8 years (96 months). Predict model established using these data may lead to a biased result and characterized as the low-practical value. Although satisfied C-indexes were obtained in both internal and external validation, the results were still biased and accidental.

(3) Li Y et al. drawn calibration curves and stated that "The internal and external calibration curves approached the 45-degree ideal match straight line, indicating that the nomograms for OS and CSS in MSGC were generally well calibrated". In fact, the 95% CIs for external validation in the Figure 3 were much larger than that for internal validation in Figure 2. And the Brier score and 95% confidence intervals (CIs) was not

available. We suspect that Li Y et al. selectively reported their results, because the performance of nomogram was worser in the external validation than internal validation. (3) Li Y et al. declared in the Method section that "the final inclusion criteria were as follows: …radiation therapy…were known and exact". In the Table 1, patients were classified into 2 groups regarding radiation: Yes and None. However, patients were divided into 2 groups: Yes and No/ unknown in the SEER database (Yin X et al. Ann Transl Med. 2021 Feb; 9(4):286; Chen C et al. Exp Bio Med (Maywood). 2021 Mar; 246(6):729-739). It is impossible for them to separate patients who did not have radiation therapy from unknown, since all information was collected from the SEER database and none of the authors worked in US. The authenticity of data is doubtful.

Due to the lack of raw data, we cannot directly compare the performance of our nomogram to Li Y et al' s. However, the report of establishment and evaluation of the nomogram in the present was strictly adhered to TRIPOD statement. Well study design, rigorous methodology, and transparent report of the results make our nomogram much more reliable than Li Y et al' s. As noted by Gary Collins, PhD, chairman of steering committee for TRIPOD statement, "In the absence of detailed and transparent reporting of the key study details, it is difficult for the scientific and health care community to objectively judge the strengths and weaknesses of a prediction model study".

Changes in the text: "Li et al constructed a nomogram for survival prediction of MSGC (14). Nonetheless, improper study design and non-transparent reporting of results significantly damaged the reliability of the study."

"Finally, even though Li et al and our nomograms were constructed based on the SEER database, only the present study was conducted according to the TRIPOD statement (30). Moreover, the performance of the nomogram was evaluated all-around to ensure the reliability of the prediction tool."

Comment 2: Low quality of figures. Fig 2, 3, 5, 6 are ineligible. Will need images of much better quality with text font of appropriate sizes. In Fig 2, what is the meaning for the numbers for "Histo Type"? What type of histology do the number stand for respectively? Fig S1 is incomplete with missing labels and poorly written legend. Without eligible figures, it is impossible to evaluate the results.

Reply 2: We appreciate your accurate comment. The size of pictures, as well as text front, have been adjusted (See Figure 2-6). The number of histological type represents different subtype. The figure captions of Figure 2 and Figure S1 are rewritten (See page 27, line 560-567 and page 30, line 614-625).

Changes in the text: "For histological type, the number 1 to 19 represents: (1) unclassified carcinoma; (2) secondary carcinoma; (3) large cell carcinoma; (4)

undifferentiated carcinoma; (5) small cell carcinoma; (6) squamous cell carcinoma; (7) lymphoepithelial carcinoma; (8) adenocarcinoma; (9) basal cell adenocarcinoma; (10) adenoid cystic carcinoma; (11) neuroendocrine carcinoma; (12) oxyphilic adenocarcinoma; (13) mucoepidermoid carcinoma; (14) duct carcinoma; (15) acinar cell carcinoma; (16) epithelial-myoepithelial carcinoma; (17) carcinoma in pleomorphic adenoma; (18) malignant myoepithelioma; (19) others. ”

Figure S1. Prognostic factor selection using LASSO regression model. The LASSO coefficient profiles of the 14 prognostic factors for OS (A) and CSS (C). The coefficient profile plots were produced against the log lambda. Numbers represent prognostic factors, i.e.: (1) sex; (2) surgery; (3) removal of lymph nodes; (4) marital status; (5) M stage; (6) race; (7) age; (8) site; (9) differentiation grade; (10) histological type; (11) tumor size; (12) AJCC stage; (13) N stage; (14) T stage. Tuning parameter (Lambda) selection in the LASSO model used tenfold cross-validation for OS (B) and CSS (D). The partial likelihood deviance curve was plotted versus log lambda. The two vertical dashed lines represent the minimum value and one standard deviation from the minimum value. The minimum value of lambda, 0.00136 and 0.00080, was chosen for OS and CSS, respectively.

Comment 3: Excluding M stage. Distant metastasis is arguably one of the most important predicators for long-term outcomes. Consistently, this parameter was shown to be a powerful parameter in Li Y et al’s study and was used for nomograms. However, in the current manuscript the authors found that M stage was not significant and therefore excluded it from the model. In Discussion, they “suspected that this was because the majority of patients who occurred distant metastasis were diagnosed with adenoid cystic carcinoma (ACC), which was featured with a slow clinical course. It was reported that ACC patients may survive for years even after metastasis and the 5-year OS of ACC could reach 75% (30). In addition, most of patients (n = 10,082, 89%) in this cohort were classified as M0 stage. All these contributed to the less effective of M stage”. I found these arguments failed to provide plausible explanations. This is because 1) the same SEER database was used, albeit consisting of different years; 2) ACCs are relatively more aggressive cancer compared to others, and certainly not “slow” compared to others; 3) M1 stage has 573 patients (5%) in this manuscript, comparable to 188 patients (4.9%) in Li Y et al. The authors are strongly encouraged to re-examine their statistical analysis.

Reply 3: Thank you for your valuable comment. After re-examination of statistical analysis, the predict models are established according to the best performance screened using LASSO regression. Sex, age, race, material status, stie, grade, AJCC stage,

T/N/M stage, tumor size, surgery, and histological type are incorporated in the Cox regression for OS prediction, and sex, age, race, grade, AJCC stage, T/N/M stage, tumor size, surgery, removal of lymph nodes, and histological type were selected for CSS prediction. The evaluation of performance of nomograms are re-performed (See Figure 2-6 and Figure S2-4). The discussion section related to predictor selection is also rewritten (See page 12, line222-223 and page 18, line 342-345).

Changes in the text: For OS, only one variable (removal of lymph nodes) was not incorporated in the model. A total of 12 predictors, except for marital status and site, were selected for CSS prediction.

In the present study, the M stage showed a significant association with OS but not CSS (P>0.05). The important role of distant metastasis in MGSC has been well-recognized (35, 36). Considering the clinical significance, the M stage was finally included.

Comment 4: Some kind of author services are strongly suggested in order to improve the overall structure, clarity of expression, and check the grammar, punctuation and phrasing.

Reply 4: We agree your comments. The manuscript has been polished by MedSci, which is a professional service provider of language editing (See certificate). Some examples of changes are listed as follow, and all changes are present in the main text (See page4 line 57, 59-61, 62):

Changes in the text:

"Patients pathologically diagnosed with MSGC were recruited"

"Univariate, multivariate Cox proportional hazard models, and LASSO regression were adopted for the selection of risk factors"

"The model performance was evaluated by…"

Reviewer B

Comment 1: This manuscript has grammatical errors which need to be fixed. eg on like 90-91 " It was reported that the 5-year overall survival…' should be 'It is reported that the 5-year overall survival…", line 108 " … the sample size of development cohort is limited." Should be … the sample size of development cohort was limited.", line 167 " The majority of the patients was male…" should be " The majority of patients were male…" There are several such grammatical errors which need to be rectified.

Reply 1: We deeply appreciated all of your thoughtful suggestions that helped improve the manuscript. We trust that all of your comments have been addressed accordingly in

a revised manuscript. Thank you very much for your effort. We have modified our text as advised (see page 6 line 98-99, page 7 line 121-122, page 11 line 198-199).

Changes in the text: "It is reported that the 5-year overall survival…"

"…however, the sample size of the development cohort was limited…"

" The majority of patients were male…"

Comment 2: Sections of the discussion seem to be a repetition of the results section eg paragraph from lines 245 to 253

Reply 2: Thank you for your constructive comments. The repetition part of discussion section has been deleted (see page 16, line 290-298).

Changes in the text: The text has been deleted.

Comment 3: Again discussion from lines 285 to 290 should be in methods as to what the inclusion and exclusions were and not in discussion

Reply 3: We agree your comments. The content from 285 to 290 has been removed from discussion section and incorporated into method section (see page 9, line 155-159).

Changes in the text: Vital status and cancer-specific death, which were primary outcomes ofin this study, were also extracted. For radiotherapy and chemotherapy, patients were divided into two groups - "Yes" and "No/Unknown" in the SEER. Combining "No" and "Unknown" as a group would probably reduce the effect of the predictor and lead to confusion in the clinical practice. Radiotherapy and chemotherapy were not extracted in this study.

Comment 4: Discussion should be rewritten with these points in mind and referenced appropriately.

Reply 4: Thank you for your valuable comment. The section of discussion has been rewritten according to your suggestion (See page 15-21, line 278-395). All changes are present in the main text and some of examples of changes are listed as follow.

Changes in the text: "The key strengths in the present study are the following: first, the SEER database provided a total of 11,362 MSGC patients. Compared with a study by Li *et al*, the sample size in the present study was larger (4,218 vs. 14,753). This cohort, which to the best of our knowledge is the largest one used thus far, is large enough to ensure the reliability and effectiveness of the nomogram."

"M stage, which reflects distant tumor metastasis, was one of the most important predictors for cancer survival (34). In the present study, the M stage showed a significant association with OS but not CSS (P>0.05). The important role of distant metastasis in MGSC has been well-recognized (35, 36). Considering the clinical significance, the M stage was finally included."

“In the 2017 World Health Organization classification, there were over 20 different histological subtypes of MSGC, with specific features and outcomes.”