Peer Review File

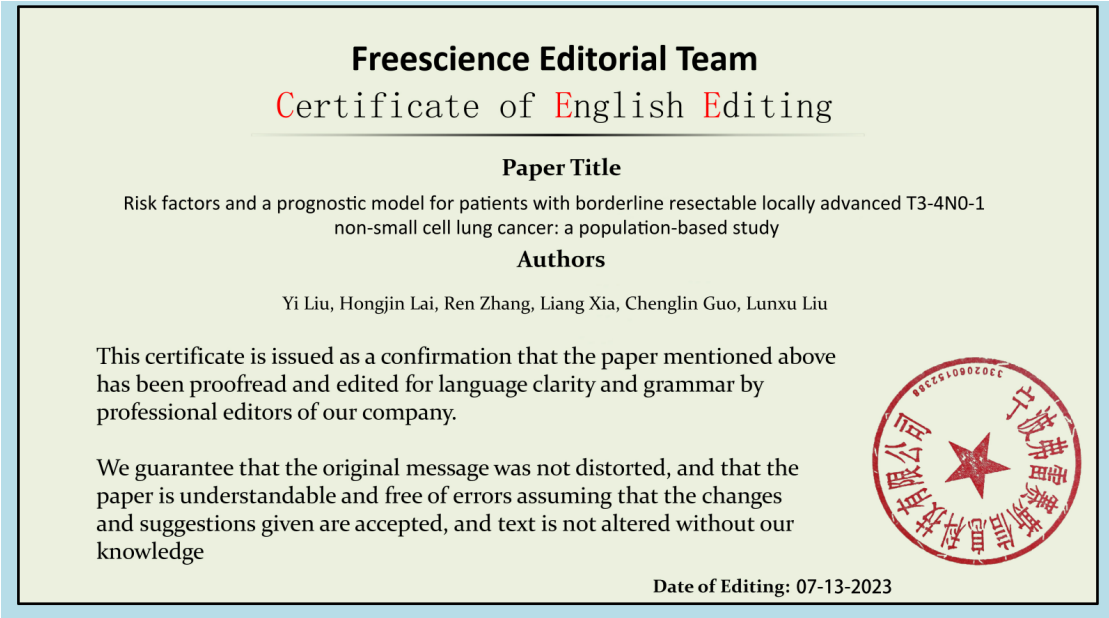
Article information: https://dx.doi.org/10.21037/tcr-23-519

Reviewer A

Introduction

Comment 1: There are several errors in grammar and diction that should be reviewed.

Reply 1: We really thank you for your thoroughly review our manuscript. We deeply apologize for the grammar and wording errors, and we have sent the manuscript to a professional language editing service for comprehensive grammar correction and improvement, ensuring the readability of the manuscript.



Freescience Editorial Team

Certificate of English Editing

Paper Title

Risk factors and a prognostic model for patients with borderline resectable locally advanced T3-4N0-1 non-small cell lung cancer: a population-based study

Authors

Yi Liu, Hongjin Lai, Ren Zhang, Liang Xia, Chenglin Guo, Lunxu Liu

This certificate is issued as a confirmation that the paper mentioned above has been proofread and edited for language clarity and grammar by professional editors of our company.

We guarantee that the original message was not distorted, and that the paper is understandable and free of errors assuming that the changes and suggestions given are accepted, and text is not altered without our knowledge

Date of Editing: 07-13-2023

Comment 2: Borderline resectable, locally advanced NSCLC encompasses more than just the T3-T4N0-N1 patients that are examined in this analysis. Are there data to suggest that there is heterogeneity in prognosis amongst these T3-4N0-1 patients that would not be predicted by TNM staging?

Reply 2: We really admire the professionalism of the reviewer. T3-4N0-1 represents a subset of borderline resectable, locally advanced NSCLC, and we have used more rigorous descriptions in our revised draft **[e.g page 1, line 3 and page 2, line 32]**. Previous study has reported several non-TNM staging risk factors, such as gender, age and histological type were associated with the prognosis of patients diagnosed with T3-4N0-1 NSCLC. However, the small sample size may introduce bias in the results (1). Furthermore, the efficacy of systematic treatment strategy such as neoadjuvant immunotherapy and induction chemotherapy in improving prognosis remains controversial (2-4). This is also one of the research questions that this study aims to investigate.

Comment 3: I think the controversy in optimal treatment of these patient has more

to do with the timing of systemic therapy than whether they receive it at all. Therefore, it would be useful to discuss how using a predictive nomogram could inform the decision for whether systemic therapy is given completely prior to surgery.

Reply 3: Thank you for your valuable comment. The timing of systemic therapy is indeed a controversial issue. According to the NCCN 3.2023 edition NSCLC guidelines, adjuvant systemic therapy has been proven beneficial for operable NSCLC patients diagnosed with T3-4N0-1. However, the impact of neoadjuvant systemic therapy on the prognosis of these patients remains controversial (5, 6). We have included a discussion on the timing of systemic therapy in the introduction section **[page 4, line 65-82]**. Our study aims to identify potential beneficiaries of neoadjuvant chemotherapy through the use of a nomogram and subsequent analysis, thereby advancing personalized precision medicine.

Methods

Comment 4: What is the rationale for only looking at patients from 2010-2015?

Reply 4: Thanks for your question. This study obtained data from patients only from the years 2010 to 2015, as certain important variables, such as systemic treatment, have not been publicly released in the latest data of the SEER database. Complete information data is only available up to the year 2015, and data from 2016 to 2023 have not been fully disclosed. In order to ensure the integrity of the cohort data, this study chose to focus on the 2010-2015 cohort for analysis.

Comment 5: For the exclusion criteria, clarify what you mean by incomplete clinical information. Were all patients with any missing data excluded?

Reply 5: Thank you for your valuable comment. During the cohort selection and exclusion process, we excluded cases with important variable records labeled as "unknown". The missing information of these variables would hinder subsequent analysis, including race, primary site, laterality, diagnostic confirmation, surgical procedure, lymph node dissection, grade classification, and systemic treatment information. We have added a corresponding description in the revised manuscript under the section of inclusion and exclusion criteria **[page 5, line 109-110]**.

Comment 6: Why was completeness of resection not included as a covariate in the Cox analysis? Your cohort does not limit it to only patients that had an R0 resection and complete resection would likely have important implications for prognosis.

Reply 6: Thank you for your valuable comment. R0 resection is indeed a significant prognostic factor. Unfortunately, it is regrettable to note that the SEER database lacks data pertaining to this particular variable, which represents one of the primary limitations of the SEER database. The potential bias resulting from the absence of information regarding R0 resection has been reported as a major limitation of this study in the discussion section **[page 11, line 290]**.

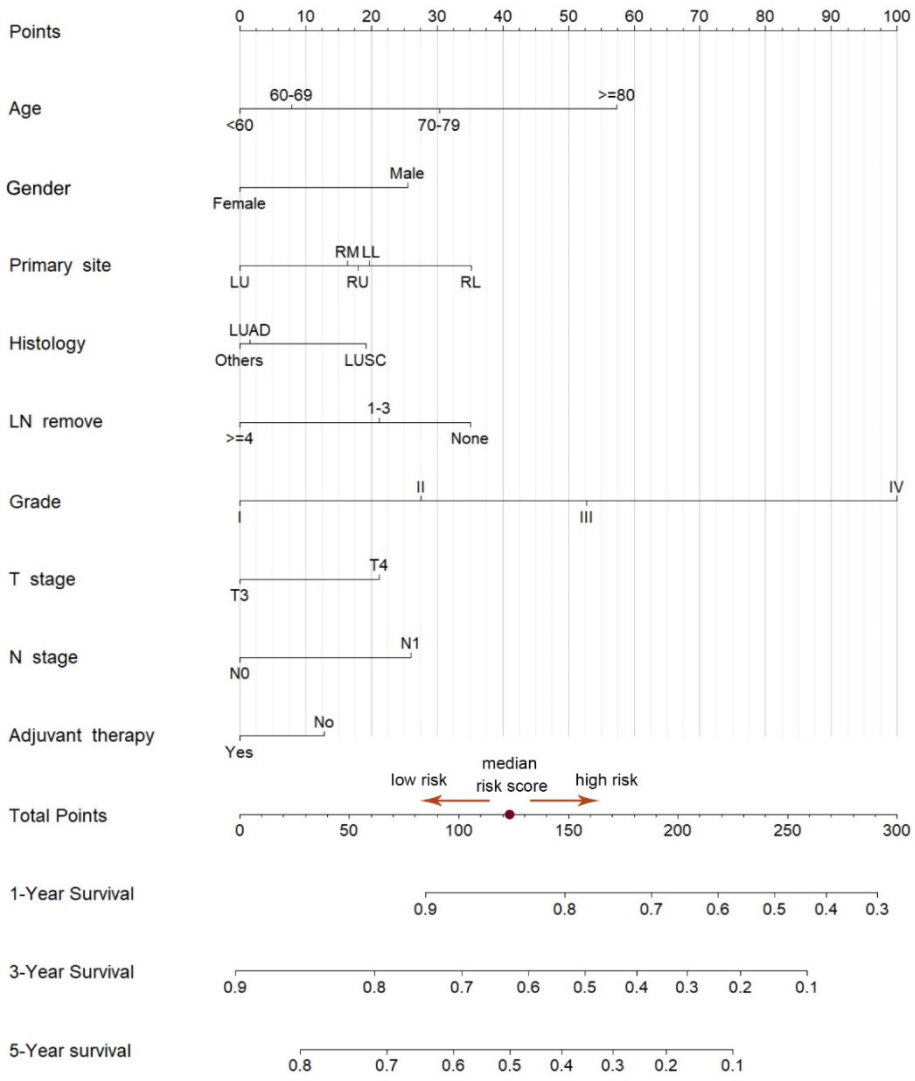
Results

Comment 7: I believe there is a typo for the median risk score on line 181.

Reply 7: Thank you for your kind reminder. We mistakenly used the risk score calculated from the original linear regression formula in the manuscript. In the revised version, we have corrected this by using the median risk score derived from the nomogram (median risk score = 122.6) **[page 9, line 212]**. It is important to note that the median risk scores from the nomogram are simply a proportional enlargement of the original risk scores. Therefore, the risk stratification groups remain unchanged, and subsequent analysis results are still reliable.

Comment 8: I would also be helpful to add a marker on the actual nomogram in figure 2 to indicate where the cut off is for high risk.

Reply 8: Thank you for your valuable comment. We have now added labels for the median risk scores in our nomogram (**Figure 3**) to provide a more intuitive differentiation of patient risk stratification.

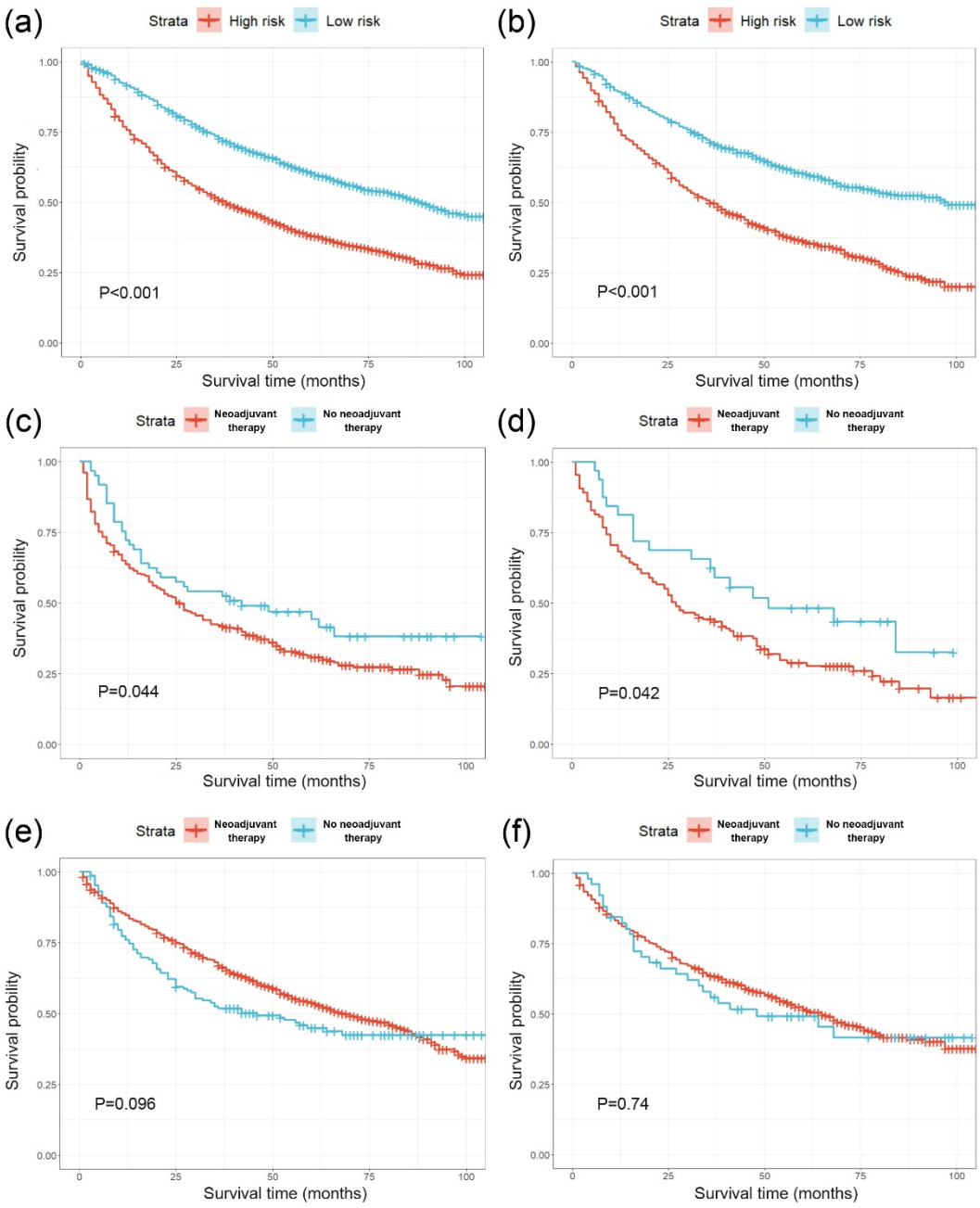


Comment 9: It looks like the lines 182-192 the call outs are to figure 4 a-f when it should be figure 5 a-f.

Reply 9: Thank you for pointing out this mistake. We have corrected this error in the revised manuscript **[page 9, line 221]**. Furthermore, we have meticulously reviewed the revised version to ensure the accuracy of our text.

Comment 10: Also, in figure 5 c-f, I think the strata are incorrectly labelled as "high risk" and "low risk" when they should be "neoadjuvant therapy" and "no neoadjuvant therapy." Otherwise, it is very unclear to me how the subgroup analysis was performed.

Reply 10: Thank you for pointing out this mistake. Indeed, the right labels should be "neoadjuvant therapy" and "no neoadjuvant therapy". We have corrected this error in the revised manuscript (**Figure 5 c-f**). Once again, we apologize for the proofreading oversight in our manuscript.



Discussion

Comment 11: There are several grammatical/diction errors that should be reviewed.

Reply 11: We really thank you for your thoroughly review our manuscript. We deeply apologize for the grammar and wording errors, and we have sent the manuscript to a professional language editing service for comprehensive grammar correction and improvement, ensuring the readability of the manuscript.

Comment 12: It would be informative here to very briefly highlight how nomograms have been used in other cancer types and what advantages they have over TNM staging. For example, are they better able to predict which patients may

benefit from specific therapies?
Reply 12: Thank you for your valuable comment. In the discussion section, we have added the application of nomogram in other types of cancer and their advantages over TNM staging (7-9). Additionally, we have presented references to confirm the application of nomograms in predicting the potential benefits of specific therapies for patients (10, 11) **[page 9, line 228-232]**.

Comment 13: Clarify what you mean by "less surgical damage of adenocarcinoma" on line 222.
Reply 13: Thank you for your valuable comment. "Less surgical damage of adenocarcinoma" indicates that adenocarcinoma of the lung is mostly peripheral. Surgical procedures for peripheral lung cancer are typically conducted away from the main bronchus and adjacent major blood vessels, resulting in reduced surgical damage. We have revised the statement "less surgical damage of adenocarcinoma" in the revised manuscript to "less surgical damage in peripheral lung cancer" to make it more comprehensible **[page 10, line 256-257]**.

Comment 14: Regarding your discussion around your findings on neoadjuvant therapy, another way people think of neoadjuvant is as a test of tumor biology. Since you only include patients that received surgery in your analysis it's hard to say whether neoadjuvant could be a useful tool for improving patient selection for surgery.
Reply 14: Thank you for your valuable comment. In addition to therapeutic purposes, neoadjuvant therapy is also reported to utilize as a test of tumor biology (12). Indeed, due to the inclusion of only surgically treated patients in our study, we were unable to assess whether neoadjuvant therapy could serve as a useful tool for improving patient surgical selection. We have addressed this in the discussion section. On the other hand, the conclusion we can draw is that the combination of neoadjuvant therapy can improve OS in patients with resectable T3-4N1 NSCLC who undergo surgery, and this conclusion is reliable **[page 11, line 269-273]**.

Comment 15: Given these patients were from 2010-2015, immunotherapy likely was not being used for the patients that received neoadjuvant. Could you comment as to whether you think your findings for T3-4N0 patients not benefiting from induction therapy would hold up if their regimen included checkpoint inhibitors?
Reply 15: Thank you for your valuable comment. Neoadjuvant immunotherapy is now being used in various types of cancer. However, since our data did not include information on immunotherapy, the potential benefits of neoadjuvant immunotherapy for T3-4N0-1 patients still require further discussion. In the revised manuscript, we have addressed this issue in the discussion section by incorporating the latest research findings:
"Neoadjuvant immunotherapy has been employed in various types of cancer (13, 14). Since our data did not include information on immunotherapy, the potential benefits of neoadjuvant immunotherapy in T3-4N0 patients still require further discussion. A

recent meta-analysis has reported that neoadjuvant chemoradiotherapy achieves a higher rate of major pathological response (MPR) in patients diagnosed with stage II/III NSCLC compared to neoadjuvant chemotherapy alone. Furthermore, MPR has been found to be associated with improved overall survival (OS) (HR 80.0, 72.0-88.0, P<0.001) (13). This suggests that neoadjuvant immunotherapy seems to be beneficial for patients diagnosed with both T3-4N0 and T3-4N1 NSCLC. It is worth noting that factors influencing the achievement of major pathological response (MPR) should be taken into consideration before administering neoadjuvant immunotherapy (15, 16)." **[page 11, line 278-287]**

Furthermore, since our study cohort only received neoadjuvant chemotherapy, we have revised the term "neoadjuvant therapy" to the more accurate term "neoadjuvant chemotherapy" in the revised manuscript to avoid ambiguity.

Reviewer B

Major concerns

Comment 1: Given the changes between the 7th and 8th lung cancer TNM classification and staging system, it would be beneficial if authors could please define T3 and T4 for the readers. Given the 8th edition was implemented after the data collection period (2010-2015), it is important to clarify whether CT imaging or surgical/pathological findings were used to define T3 and T4 when restaging the patients based on the 8th edition.

Reply 1: Thank you for your valuable comment. We have attached a **supplementary file 1** in the revised manuscript to provide a detailed explanation of the implementation criteria for TNM re-staging. Furthermore, all patients in the cohort underwent T staging based on pathological diagnosis **[page 5, line 104-105]**.

Supplementary file 1. The TNM re-stage criteria based on the 8th edition of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual

T stage

T3

Any of the following characteristics:

1) Size: >5 cm but <7 cm.
2) Local invasion: direct invasion of chest wall (including superior sulcus tumors), parietal pleura (PL3), phrenic nerve, or parietal pericardium.

T4

Any of the following characteristics:

1) Size>7 cm.
2) Airway location: invasion of the carina or trachea.
3) Local invasion: diaphragm, mediastinum, heart, great vessels, recurrent laryngeal nerve, esophagus or vertebral body.

N stage (Not corrected; same as the 7th edition)

N0

No regional lymph nodes involvement.

N1

Involvement of ipsilateral peribronchial and/or ipsilateral hilar lymph nodes (includes direct extension to intrapulmonary nodes).

Comment 2: Could the authors please explain how the diagnosis of N1 disease was made? Was it determined through EBUS TBNA or surgical lymph node resection? If some patients did not undergo lymph node resection, that raises concerns about the possibility of missing true pathological diagnoses and metastasized diseases. It is worth crosschecking the patients who did not have lymph node resection with the type of surgery they received. If those patients also underwent segmentectomy or wedge resection, it is likely that they had limited lung function (i.e. low FEV1) and were sicker, which could impact prognosis.

Reply 2: Thank you for your valuable comment. According to the description in the SEER database "SEER Manual Section V: Stage at Diagnosis", the N staging is determined based on a combination of clinical diagnosis and surgical pathology. The clinical N staging provides a reference for the extent of lymph node dissection during surgery, while the final determination of N staging is based on pathological examination following lymph node removal. Therefore, the possibility of false negatives is minimal. Additionally, we crosschecked the surgical procedures for patients who did not undergo lymph node dissection and found that only one patient out of a total of 5,054 underwent segmental resection or wedge resection. While such patients undergoing segmental resection or wedge resection may experience disease progression and impact prognosis, this single case does not appear to affect the overall results of the large sample size.

Comment 3: The analysis does not appear to include immunotherapy or tyrosine kinase inhibitors. It is important to note that these treatment options may not have

been widely available during the data collection period. Given the availability of the new oncological treatment options and their incorporation into NCCN guidelines for the same targeted patient population as described in this manuscript, the clinic utility of the findings reported here may be limited.

Reply 3: Thank you for your valuable comment. Immunotherapy or tyrosine kinase inhibitors (TKIs) in combination with chemotherapy have been recommended as new neoadjuvant/adjuvant treatment options for patients with T3-4N0-1 NSCLC, according to the latest NCCN guideline (17). This is an exciting breakthrough. Our study did not consider the use of immunotherapy or TKIs, which is a limitation that we acknowledge in the limitation section of the article. However, according the latest NCCN guideline, chemotherapy remains the cornerstone of systemic treatment. As a large-scale study, the conclusions of our research still serve as a reference for the clinical application of neoadjuvant chemotherapy. Additionally, we have added a discussion section that explores the application of immunotherapy, making it more relevant to clinical practice **[page 11, line 278-287]**.

Comment 4: It is advisable for the authors to report the overall p value for each independent variable in their univariable and multivariable analyses (i.e. Table 2), particularly since they selected variables based on these p-values.

Reply 4: Thank you for your valuable comment. In the revised manuscript, we have reported the p-values for each independent variable in both the univariate logistic regression and multivariate logistic regression analyses, making the variable selection logic clearer **[page 8, line 187-197]**.

Comment 5: The author performed multiple univariable analyses (Table 2), which poses a risk of inflating type I error. To mitigate the risk, I recommend that the authors consider employing methods such as Bonferroni correction, Holm-Bonferroni correction or false discovery rate control.

Reply 5: Thank you for your valuable comment. We applied the Bonferroni correction method to adjust for multiple hypothesis testing in the results of the univariate analysis, using a significance threshold of $P < 0.004$. P-values ranging from 0.004 to 0.05 were considered as suggestive evidence. Finally, we combined statistical significance with clinical importance to complete the variable selection process for the multivariate model **[page 7, line 166-169]**.

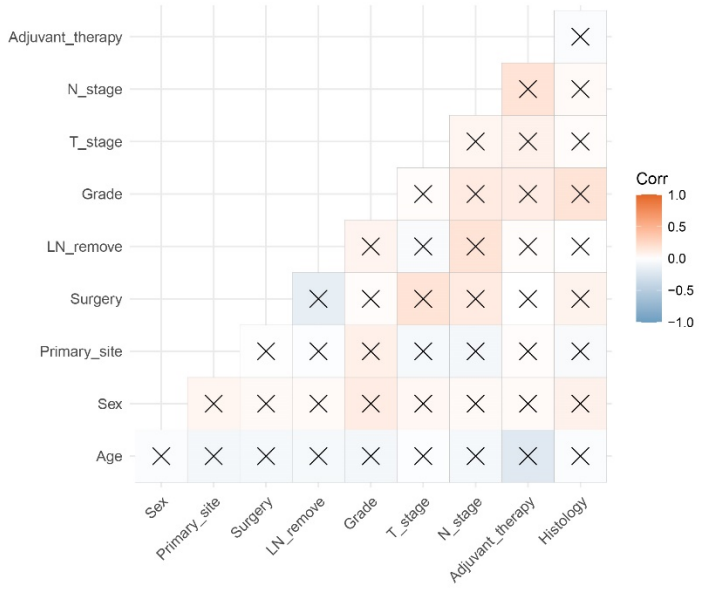
Comment 6: The authors used univariable analysis to screen for variables that enter multivariable analysis, and then used that to select variables for their predictive model. However, it is worth noting that some articles argue against using this approach to screen for predictive factors, as it may raise questions about the statistical soundness.

Reply 6: Thank you for your valuable comment. Univariate analysis is used to screen variables for inclusion in multivariate analysis, and then multivariate analysis is employed to select variables for the predictive model. This is a classic approach for constructing predictive models. However, some scholars have raised concerns about

statistical and rational issues associated with this method, particularly regarding variable correlation and multicollinearity (18).

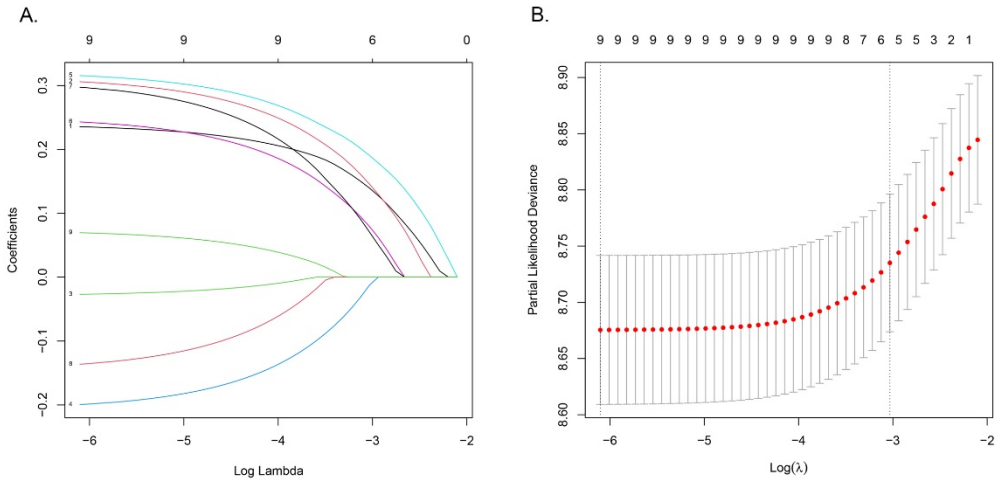
To address these issues, we conducted additional analyses, including correlation analysis and LASSO regression analysis. The correlation analysis results indicated no significant correlation among the selected variables (**Supplementary Figure 1**). The LASSO regression results further confirmed that all nine variables were important, and there was no multicollinearity bias present (**Supplementary Figure 2**) **[page 8, line 197-200]**.

Supplementary figure 1. Variable correlation analysis.



Supplementary figure 1. The correlation analysis plot demonstrates that there is no strong correlation between any of the variables included in this study. The symbol "X" represents a lack of significant correlation.

Supplementary figure 2. Least absolute shrinkage and selection operator (LASSO) regression detects multicollinearity of variables.



Supplementary figure 2. (A) The coefficient distribution plot and (B) the error parameter plot. LASSO regression results suggest that all nine variables we selected are important, and the final model does not suffer from multicollinearity issues.

Minor concerns

Comment 1: In the first paragraph of the introduction, the authors mentioned "the 8th edition of national comprehensive cancer network guideline". The current version of the NCCN guideline is version 3.2023. It is unclear if the authors were referring to the 8th edition of TNM classification by AJCC?

Reply 1: Thank you for your valuable comment. We apologize for the error in the original manuscript. The guidelines mentioned in the first paragraph of the introduction should refer to the 3.2023 version of the NCCN NSCLC Guidelines. The corresponding statement has been corrected in the revised manuscript **[page 4, line 63]**.

Comment 2: In the first paragraph of introduction, in line 63, the authors mentioned "complete (or near-complete) pathological response. It would be helpful if the authors could provide a definition of "pathological response" in the manuscript as clinical practice typically relies on radiological response for such evaluations.

Reply 2: Thank you for your valuable comment. The pathological response mentioned in the manuscript was indeed assessed based on radiological response. In the revised manuscript, we have added the definition of "complete pathological response" as follows: "complete pathological response was defined as the complete radiologic disappearance of all measurable or assessable disease" **[page 4, line 72-74].**

Comment 3: In the first paragraph of introduction, in line 64, "neoadjuvant therapy could effectively degrade the NTM stage of selected patients…". Did the authors mean "downgrade" instead of "degrade"?

Reply 3: Thank you for your valuable comment. Yes, the intended expression is "downgrade" instead of "degrade". The corresponding statement has been corrected in the revised manuscript **[page 4, line 70]**.

Comment 4: The authors excluded 1,007 subjected because of incomplete clinical information as described in methods. Please explain what 'incomplete clinical information' means.

Reply 4: Thank you for your valuable comment. During the cohort selection and exclusion process, we excluded cases with important variable records labeled as "unknown". The missing information of these variables would hinder subsequent analysis, including race, primary site, laterality, diagnostic confirmation, surgical procedure, lymph node dissection, grade classification, and systemic treatment information. We have added a corresponding description in the revised manuscript under the section of inclusion and exclusion criteria **[page 5, line 109-110]**.

Reviewer C

Comment 1: P7, line 154-155; Totally, 305 (8.62%) patients received neoadjuvant therapy and 1,678 (47.43%) patients received adjuvant therapy. It is better to describe the contents of the treatment (e.g., chemotherapy, radiotherapy).

Reply 1: Thank you for your valuable comment. In this study, the neoadjuvant/adjuvant therapy involved was exclusively chemotherapy. We have made the necessary corrections to all relevant descriptions in the revised manuscript **[e.g. page 2, line 40 and 44]**.