Peer Review File

Article Information: https://dx.doi.org/10.21037/tlcr-23-866

Reviewer A

There are several issues with this manuscript. It is truly a pity as clearly this is the result of a significant amount of work by the authors.

The main issues which in my opinion make the piece unsuitable for publication are:

1. I struggle to see any clinical relevance here. Patients with resected NSCLC die from recurrence, either locoregional recurrence unsuitable for further locoregional therapies, or from metastatic dissemination – bone is only one of the possible sites, and the single metastatic site clearly associated with much worse prognosis in the clinic is the brain. I don't see how a prognostic tool for determining high risk of relapse with bone metastases could be useful.

Reply 1: Thank you for your insightful feedback. While bone metastasis is just one of the potential sites of metastasis in early-stage postoperative NSCLC patients, we acknowledge the significance of exploring this area and developing prognostic tools in this context. The rationale for this is outlined below.

Practical value: Bone metastases, although just one of the possible sites of metastasis, are common and impactful in NSCLC patients. Bone metastases can lead to significant morbidity, including bone pain, fractures, and spinal cord compression, impacting the quality of life and overall prognosis of patients. Identifying high-risk patients allows for early interventions and personalized treatment plans, potentially improving outcomes. As such, our prognostic tool helps to guide tailored treatment and monitoring strategies.

Theoretical/cognitive value: It is important to recognize that research in oncology is a continuum, and studying different aspects of metastatic spread, including bone metastases, contributes to the broader understanding of cancer progression and treatment outcomes. Insights gained from studying bone metastases can complement existing knowledge on brain metastases and other metastatic sites, leading to advancements in patient care and management.

2. The therapeutic landscape of early NSCLC has moved on from upfront surgery for all, which is the setting the authors found their work on. Significant proportions of patients will receive

neoadjuvant or perioperative chemoimmunotherapy, and there is no reference to this recent development and its relevance in the text.

Reply 2: In response to the reviewer's comment regarding the evolving therapeutic landscape of early NSCLC, we appreciate the opportunity to address this important aspect in our work.

We acknowledge the limitation of our study, which retrospectively included resectable stage I-II NSCLC patients from our hospital between 2015 and 2021. During this period, the standard treatment for resectable stage II patients in our country was surgery followed by adjuvant chemotherapy with platinum-based doublet. It was until 2023 that postoperative atezolizumab adjuvant therapy and the combination of platinum-based chemotherapy with nivolumab neoadjuvant therapy were introduced as first-line treatment options for selected patients. Therefore, in the risk factor analysis, we only included adjuvant radiotherapy or adjuvant chemotherapy as treatment factors, which did not show significance in the multivariable Cox regression analysis.

Given the increasing utilization of neoadjuvant and perioperative chemoimmunotherapy in clinical practice, we will mention our limitation regarding this framework and incorporate references to these treatment modalities and discuss their implications for patient outcomes and prognosis in our revised manuscript to provide a more comprehensive and up-to-date analysis that aligns with current clinical practices and guidelines. Thank you again for highlighting this valuable aspect.

Revise:

Adding to discussion (Discussion part, Page 10-11, Line 245-259):

In recent years, ICIs have been approved for use in the neoadjuvant and adjuvant settings based on event-free survival and disease-free survival benefit. For example, the results of PEARLS[1] and IMpower 010[2] trial indicated that platinum-based chemotherapy followed by adjuvant atezolizumab or pembrolizumab could significantly extend disease-free survival among resected stage II–IIIA NSCLC patients. According to the result of CheckMate-816[3], the efficacy of neoadjuvant platinum-based chemotherapy with nivolumab was impressive in preventing distant metastasis compared to chemotherapy group among resectable early-stage NSCLC (median event-free survival durations = 31.6 months versus 20.8; HR = 0.63; 97.38% CI 0.43–

0.91; $P = 0.005$). The metastatic diagram of early-stage resectable NSCLC would experience a great shift as the popularization of immunochemotherapy in clinic. Although our study only identified high-risk patients for bone metastasis after surgery restricted to traditional treatment regimens, the result of which can serve as a reference for future decision on clinical treatment strategies: high-risk patients predicted by this model may be prioritized for treatment with neoadjuvant/adjuvant chemoimmunotherapy to maximumly lower the possibility of postoperative bone metastases.

Reference:

1. Paz-Ares, L. et al. Pembrolizumab (pembro) versus placebo for early-stage non-small cell lung cancer (NSCLC) following complete resection and adjuvant chemotherapy (chemo) when indicated: randomized, triple-blind, phase III EORTC-1416-LCG/ETOP 8-15 – PEARLS/KEYNOTE-091 study [abstract VP3-2022]. Ann. Oncol. 33, 451–453 (2022).

2. Felip, E. et al. Adjuvant atezolizumab after adjuvant chemotherapy in resected stage IB-IIIA non-small-cell lung cancer (IMpower 010): a randomised, multicentre, open-label, phase 3 trial. Lancet 398, 1344–1357 (2021).

3. Forde PM, Spicer J, Lu S, et al. Neoadjuvant Nivolumab plus Chemotherapy in Resectable Lung Cancer. N Engl J Med. 2022;386(21):1973-1985. doi:10.1056/NEJMoa2202170

Reviewer B

In this paper, Dr. Zhou et al. analyzed risk factors for bone metastases after early-stage lung cancer surgery. The analysis using monograms was novel and very interesting. The analysis is very detailed and has immediate clinical relevance. Unfortunately, however, the results obtained in the analysis are not specific to bone metastases, as there are still clinicopathologic factors that are generally associated with a high risk of tumor recurrence. In order for this study to be accepted for publication in this journal, the following points need to be clearly stated.

1. According to some guidelines regarding NSCLC, stage II and above generally require adjuvant therapy after surgery. Although there is no clear definition of early-stage lung cancer, it seems that only stage I is considered early-stage cancer, while stage II and above are generally considered advanced cancer. This paper states that early-stage NSCLC was examined, but it

needs to clearly state why stage II was included.

Reply 1: Thank you for your thorough attention to detail. We acknowledge that some Stage II NSCLC patients (mainly IIB) indeed require postoperative adjuvant therapy. However, referring to the data released by the National Lung Screening Trial (NLST) in 2023[1], only slight difference lies among the incidence of (IA: 7%; IB: 11%; IIA: 14%; IIB: 15%) after standard treatment for stage I-II NSCLC patients. Regarding the post-treatment relapse and metastasis rate, we think it's rationale to regard stage II as early-stage NSCLC patient.

Reference:

1. Potter AL, Costantino CL, Suliman RA, et al. Recurrence After Complete Resection for Non-Small Cell Lung Cancer in the National Lung Screening Trial. Ann Thorac Surg. 2023;116(4):684-692. doi:10.1016/j.athoracsur.2023.06.004

2. The presence or absence of a solid component should be discussed only in lung adenocarcinoma. All squamous cell carcinomas are solid. It seems strange to discuss the presence or absence of a solid component in a cohort with a mixture of adenocarcinoma and squamous cell carcinoma.

Reply 2: Thanks for your professional feedback. To convenient analysis, we consider radiological characteristics and histological characteristics as two distinct dimensions of NSCLC characteristics without overlapping. Thus, the proportion of solid components is at the radiological level, while squamous cell carcinoma is at histological characteristic. As a result, most squamous cell carcinoma patients are classified in the solid component (present) group based on radiological features, while adenocarcinoma patients may be in the absent group or present group, which does not affect the subsequent modeling analysis.

3. Tumor invasion diameter, lymphatic permeation, vascular invasion, pleural invasion, and pulmonary metastases are pathologic factors that should be included in the analysis.

Reply 3: Thanks for your careful considerations. Tumor invasion diameter, lymphatic permeation, vascular invasion, pleural invasion, and pulmonary metastases are important pathological factors for tumor metastasis. Still, we believe that these indicators are already

included in the postoperative pathological TNM staging, of which N1 stage was further divided into N1a single lymph node station involvement and N1b multiple lymph node station involvement manually, considering the lymphatic permeation. Therefore, there is no need to redundantly include the above indicators in the analysis.

4. It is necessary to distinguish whether the first site of recurrence after surgery is bone metastasis or whether other metastases have already been found and subsequent examination revealed bone metastasis.

Reply 4: Thanks for your valuable input. While we acknowledge the importance of distinguishing the first site of recurrence after surgery, we regret to admit that we do not have the detailed information to make this distinction because of the retrospective nature of our data. We appreciate your understanding in this matter.

5. Detection of bone metastases by blood tests should be discussed. Have tumor markers or alkaline phosphatase been examined for their relationship to the detection of bone metastases?

Reply 5: Thank you for your expert insight. Alkaline phosphatase (ALP) has been clinically proven to be one of the most sensitive indicators for detecting bone metastases[1], facilitating early detection of bone metastasis in the body. However, due to the retrospective nature of our study, no exact tests results on tumor or serum markers examined for their relationship to detection of bone metastases were recorded.

Yet, the purpose of this study is to identify high-risk patients for potential postoperative bone metastasis, rather than screening for patients with bone metastases. The predictive model we have constructed utilizes four easily obtainable clinical parameters to efficiently identify high-risk patients for postoperative bone metastasis, allowing for a more targeted arrangement of the frequency of bone metastasis monitoring indicators such as alkaline phosphatase and whole-body bone scan during postoperative follow-up. Therefore, we hold the view that the association between alkaline phosphatase levels/tumor biomarkers and bone metastasis is not the focus of this study and does not require extensive discussion. However, we will add the discussion for individualized follow-up strategy with more frequent high-sensitivity imaging or serum tests for bone metastases in our manuscript (Discussion part, Page 11, Line 261-271).

Reference:

1. Knapp BJ, Devarakonda S, Govindan R. Bone metastases in non-small cell lung cancer: a narrative review. J Thorac Dis. 2022;14(5):1696-1712. doi:10.21037/jtd-21-1502

6. There is no description of what specific tests should be added to a population that has a high incidence of bone metastases.

Reply 6: Thank you for your valuable feedback. As responded in comment 5, we will add the description that "This model aids clinicians in identifying high-risk patients who may benefit from high-sensitivity tests for bone metastases, such as alkaline phosphatase tests and whole-body scans, during their personalized follow-up." in the discussion.

7. To demonstrate that the biomarker is specific for bone metastasis, it is important to compare it with cases of metastasis other than bone.

Reply 7: Thank you for your valuable feedback. Regarding your suggestion to compare the biomarker with cases of metastasis other than bone to demonstrate its specificity for bone metastasis, we would like to further elaborate on our stance. Indeed, the primary focus of our study is on bone metastasis, and the predictive factors we have identified have been shown to effectively predict the risk of bone metastasis occurrence. While comparing with other types of metastasis may potentially contribute to further validating the specificity of these predictive factors, we believe it is not the primary focus of our study. Our research aims to establish a predictive model for bone metastasis in early-stage NSCLC to identify high-risk patients in clinical practice and to rationalize screening strategies. Therefore, our emphasis lies in validating the effectiveness of these predictive factors in early bone metastasis detection to facilitate timely therapeutic interventions and improve patient outcomes. We appreciate your suggestion, but focusing on comparing with other metastasis types may stray from our main research objectives. Thank you sincerely!

Reviewer C

The authors have performed univariable and multivariable cox regression to develop and

evaluate a nomogram for bone metastasis in 2106 patients with early-stage (I-II) NSCLC. Internal validation with optimism correction on 1000 bootstrap resamples has been performed. The model showed optimism-corrected C-index of 0.82, 0.79 and 0.78 at 1-year, 3-year and 5-year timepoints.

1. The dataset includes 54 (2.6%) events. Considering the class imbalance, the model trained on such data tend to be biased towards the majority class, resulting in overfitting and leading to a poor performance on new data. How do the authors have addressed this issue in their analysis?

Reply 1: We acknowledge the reviewer's concern regarding the potential risk of overfitting due to class imbalance. However, it is important to note that our study population is limited to early-stage NSCLC patients (Stage I-II), where the risk of bone metastasis is relatively low. This results in a small number of outcome events in our dataset. Additionally, our variables include solid component and detailed N staging, making it challenging to validate externally with larger cohorts. We have described this limitation in the discussion section. We look forward to further validation of our findings in larger cohorts.

2. The calibration plots show that the probabilities are limited to a narrow range of >0.85. the authors have mentioned: “The calibration curve showed the model was well calibrated.”. How do the authors describe the model behavior for patients with low-risk?

Reply 2: Thank you for highlighting the calibration plots depicting probabilities largely concentrated within a narrow range exceeding 0.85. This observation is primarily attributed to our study cohort, which predominantly consists of early-stage (I-II) NSCLC patients. As these patients inherently exhibit a lower risk of metastasis, it's unsurprising that the calibration curve reflects probabilities skewed towards higher values, resulting in a range from 0.8 to 1. However, despite this concentration of probabilities in the higher range, it's essential to note that the calibration curve closely parallels the line of perfect calibration, particularly near a slope of 1. This alignment indicates that our model effectively calibrates predicted probabilities with observed outcomes across the spectrum of risk. In our manuscript, we have emphasized the significance of this alignment and asserted that our model is well calibrated. Nonetheless, we acknowledge the need for cautious interpretation, particularly for patients with lower risk scores.

While the probabilities may appear compressed in the higher range, the alignment with the ideal calibration line underscores the reliability of our model's predictions across varying risk levels. Thank you sincerely!

3. The calibration curves also should be corrected for optimism.

Reply 3: Thank you for your attention to our study and for your suggestion regarding the calibration curves. We fully appreciate your concern regarding the potential optimism in the calibration curves. Due to the limited sample size in our study, we were unable to perform sample splitting or external validation to correct for overfitting. Nevertheless, we will acknowledge this limitation cautiously in the discussion section and welcome future studies with larger sample sizes to further validate the performance of our model (Discussion part, Page 11, Line 278-280).

4. The authors have considered Events Per Predictor>=10 for sample size calculation. This is a rough estimation. Recent guidelines suggest more robust methods to calculate the sample size for development and validation of prediction models (DOI: 10.1136/bmj.m441). I suggest to calculate the sample size based on the four proposed equations in this paper and add that as a supplementary material.

Reply 4: In response to the reviewer's suggestion, we have employed a new sample size calculation method. Using the “pmsampsize” function from the pmsampsize package (DOI: 10.1136/bmj.m441)[1], we used the following parameters, including a Cox-Snell R-squared of 0.05, 10 candidate predictor parameters, an overall event rate of 0.0256, a prediction timepoint of 1 year, and an anticipated average follow-up time of 1 year. This ensured adequate statistical power to detect effects of interest. The final calculated total sample size required was 1750, with 45 outcome events, which barely met our sample size requirements (See following Figure).

```
> library(pmsampsize)
> pmsampsize(
+   type = "s", #"s" specifies sample size calculation for a prediction model with a survival outcome
+   csrsquared = 0.05, #the expected value of the Cox-Snell R-squared of the new model
+   parameters = 10, #the number of candidate predictor parameters for potential inclusion
+   rate = 0.0256,#the overall event rate in the population of interest
+   timepoint = 1, #the timepoint of interest for prediction
+   meanfup = 1, #anticipated average follow-up time
+   seed = 123456#set random-number seed
+ )
NB: Assuming 0.05 acceptable difference in apparent & adjusted R-squared
NB: Assuming 0.05 margin of error in estimation of overall risk at time point = 1
NB: Events per Predictor Parameter (EPP) assumes overall event rate = 0.0256

             Samp_size Shrinkage Parameter CS_Rsq Max_Rsq Nag_Rsq  EPP
Criteria 1        1750     0.900        10   0.05   0.212   0.235 4.48
Criteria 2         913     0.825        10   0.05   0.212   0.235 2.34
Criteria 3 *      1750     0.900        10   0.05   0.212   0.235 4.48
Final SS          1750     0.900        10   0.05   0.212   0.235 4.48

 Minimum sample size required for new model development based on user inputs = 1750,
 corresponding to 1750 person-time** of follow-up, with 45 outcome events
 assuming an overall event rate = 0.0256 and therefore an EPP = 4.48

 * 95% CI for overall risk = (0.018, 0.033), for true value of 0.025 and sample size n = 1750
 **where time is in the units mean follow-up time was specified in
```

Reference:

1. Riley, R. D. et al. Calculating the sample size required for developing a clinical prediction model. Bmj 368, m441, doi:10.1136/bmj.m441 (2020).

5. The variable selection (forward, backward and etc.) and its criterion (AIC, BIC, P and etc.) hasn't been mentioned in the analysis section. Table 2 shows that all predictors have remained in multivariable analysis. However, the nomogram only includes four predictors which had significant P-value in multivariable analysis.

Reply 4: Thank you for your insightful comments regarding the variable selection process. We acknowledge that the details of variable selection methods and criteria were not explicitly mentioned in the analysis section. To clarify, we initially conducted univariable Cox regression analysis to perform preliminary variable screening. Subsequently, significant variables from the univariate analysis, along with clinically relevant predictors, were collectively included in the final multivariable model. Finally, for the nomogram construction, we only included the predictors with significant p-values in the final multivariable analysis to ensure the clinical relevance and statistical significance of the model (Methods part, Page 5-6, Line 119-123). Thank you sincerely!

6. It is suggested to add decision curve analysis.

Reply 6: We appreciate the attention to detail and the valuable input provided. However, we found that decision curve analysis did not yield satisfactory results in our study. This may be attributed to the relatively low occurrence of bone metastasis in early-stage lung cancer cases. Considering this limitation, we believe that incorporating decision curve analysis may not add significant value to our study and could potentially divert focus from our main results. Thank you sincerely!

7. How do the authors explain the remarkably extended confidence interval of some of the predictors?

Reply 7: Thank you for raising this point. We believe the remarkably extended confidence intervals observed for some predictors can be primarily attributed to the inclusion of patients with low metastatic risk in stages I-II, as previously mentioned. The inherently low risk of bone metastasis in this patient subset results in a relatively small number of outcome cases, leading to larger confidence intervals. Thank you sincerely!

8. Adjuvant radiotherapy has been performed only for 8 patients. This can not be a potential predictor during variable selection.

Reply 8: Thank you for bringing up this point. We appreciate your attention to detail. Following your suggestion, we have not only removed the variable regarding adjuvant radiotherapy but have also recalculated the multivariable analysis accordingly (Table 2).

9. English language should be improved throughout the manuscript. Below few corrections:

• Line 42: proves valuable -> proves to be

• Metastases is plural. Are the authors referring to more than one bone metastasis?

• Rephrase this sentence: “The potential for postoperative metastasis may become more pronounced within the expanding early-stage NSCLC population”

• Line 71: discern -> identify

• Line 80: “Previous works have made an effort”-> efforts

• Line 105: radiotherapy is repeated twice instead of chemotherapy

• “Baseline clinical and demographic characteristics were summarized in Table 1.”: were -> are

Reply 9: Thank you for your feedback on the manuscript. We will work on improving the English language throughout the document and make the necessary corrections. All the suggestions mentioned below have been corrected.

10. Line 40: “confirming the reliability of the model” should be removed. The method does not assess the reliability of the model.

Reply 10: Thank you for your feedback. This sentence has been removed according to your professional advice.

11. “Prediction model” and “Predictive model” has been used interchangeably in the manuscript. Please, keep the lexical consistency.

Reply 11: Thank you for your feedback. All the “prediction model” have been changed into “predictive model” in the manuscript.

12. Line 155-156: “Pathological stage I-II NSCLC patients who received lobectomy were included for analysis” is method, not result.

Reply 12: Thank you for your feedback. The sentence was deleted in the part of result. And this inclusion criteria was mentioned in the method part.

13. Instead of naming the “surv_cutpoint” function, please briefly describe the logic behind it.

Reply 13: We appreciate the reviewer's attention to detail. The "surv_cutpoint" functio n determines the optimal cut-off for one or multiple continuous variables simultaneousl y using maximally selected rank statistics. This method is outcome-oriented, providing a cut-off value corresponding to the most significant relationship with the outcome. Fo r further information, please refer to: https://www.rdocumentation.org/packages/survminer /versions/0.4.9/topics/surv_cutpoint.Thank you sincerely!

14. Table 1: “Ever smoker”: does it mean former?

Reply 14: Thank you for your valuable feedback on our manuscript. We acknowledge and

agree with your suggestion to modify "Ever smoker" to "Former smoker" in the baseline table.