Peer Review File

Article Information: https://dx.doi.org/10.21037/tlcr-22-642

Reviewer A

Authors analyze the impact of LIPI score as prognostic factors in III stage NSCLC. The paper is interesting and clearly written.

Reply: Thank you very much for your time and efforts reviewing out manuscript. All your comments had been very helpful. We tried to correct our manuscript based on your comments, and tried to meet the Reviewers' standards. In the revised version, we tried to focus on the N2 subgroup, which is indeed very heterogeneous and requiring much more academic attentions. We attempted to predict prognosis of the N2 subgroup by using a new combination score including parameters such as single/multi station, bulky/non-bulky, and LIPI index to reflect anatomic distribution, tumor burden and immuno-oncological backgrounds. If you allow us, we would like to change the title of the manuscript to *"The association between clinical parameters and resectability in Stage III non-small cell lung cancer, and a combination of N2 lymph node burden and LIPI score as a potential biomarker"*

We hope these efforts elevated the quality of our manuscript and added to the novelty of the paper.

However, there are certain major issues: see above responses and following:

1. First of all the authors didn't provide information which edition of TNM's classification they had applied in their work. Please add it

Reply: We appreciate your constructive comments. In the revised version, we added the version of TNM edition in the methods section

Changes in the text: Method – *Study population*; Page 7, line 13-14

"For the tumor staging, the 8th edition of TNM staging system was applied"

2. Inclusion and exclusion criteria were defined no clearly. Why patients with large call carcinoma were excluded from the study?

Reply: We appreciate your important comment, and agree that this needs more explanation. There were three cases of large cell carcinoma in our stage III NSCLC population. We did not include patients with large cell carcinoma for the analyses, because there is no consensus established regarding the treatment (treatment based on adenocarcinoma vs squamous cell types), and we thought that including these patients may affect the survival analyses. This is the same reason for exclusion of other cell types.

Changes in the text: Method – *Study population;* Page 7, line 2-14

"Among the 1,304 patients diagnosed with lung cancer between June 2008 to December 2020,

patients with clinical stage III (IIIA, IIIB, and IIIC) or pathological stage III (IIIA and IIIB) NSCLC were included. Regarding pathologic type, patients diagnosed with adenocarcinoma, squamous cell carcinoma, adenosquamous carcinoma, and NSCLC not otherwise specified (NOS) were included for analyses. The patients diagnosed with small cell lung cancer, large cell carcinoma, and others (neuroendocrine carcinoma, sarcomatoid carcinoma, pleomorphic carcinoma, mucoepidermoid carcinoma, bronchioloalveolar carcinoma, and adenoid cystic carcinoma) were excluded. For the tumor staging, the 8th edition of TNM staging system was applied. (Figure 1)."

3. The statistical analysis of cox proportional hazard is valid only if the variables entered into the model do not have significant correlation with one another. If some variables are found to have a correlation, then only one of those parameters can be used in the modelling to avoid getting biased results. This information should be provided to the readers and it signifies transparent reporting. Please address.

Reply: We appreciate your critical comment. Because TNM stages and overall stages overlap and show correlations between them, we have shown separate multivariate analyses. We set two models: Model 1 for stage IIIA-IIIC and Model 2 for TN staging.

Pulmonary functions parameters such as FEV1, FVC and DLCO could also have an interrelation and create possible bias, but none of them was included in the multivariate analyses due to statistical insignificance in the univariate analyses

We admit that we need to be more clear about how we entered the variables and avoid the potential bias, and further stated it in the Methods section

Changes in the text:

Methods – *Statistical analysis*; Page 9, line 17-20

"For multivariate analyses, we did not enter two or more parameters which could have correlations between each other and create potential bias. Staging parameters such as IIIA-IIIC and TN stage were not entered in the multivariate analyses together, and separate multivariate analyses models were made."

Results – *Multivariate analysis on PFS and OS*; Page 12, line 12-14

"For Cox regression analysis, two models of analyses were performed due to the different sets of cancer stages (IIIA-C vs. T, N), thus all the variables entered into each model did not have significant correlation with one another."

4. How was the cut-off point for the NLR determined? Please explain it.

Reply: Thank you for the critical comment. The cut-off points for the NLR were different depending on the studies. We used derived NLR (dNLR) when calculating the LIPI score. The cut-off point for the dNLR was based on the previous study by Mezquita et al.(JAMA Oncol. 2018;4(3):351-357.), which demonstrated that dNLR greater than 3 was independently associated with OS and PFS in advanced NSCLC receiving treatment with PD-1/PD-L1

inhibitors. This information is provided in our text with the reference (7).

7. Mezquita L, Auclin E, Ferrara R, et al. Association of the Lung Immune Prognostic Index With Immune Checkpoint Inhibitor Outcomes in Patients With Advanced Non-Small Cell Lung Cancer. JAMA Oncol 2018;4:351-7.

Changes in the text: Methods – *Study design*; Page 8, line 12-18

"The LIPI score is the combination of derived neutrophil-to-lymphocyte ratio (dNLR) and LDH. dNLR is defined as absolute neutrophil count/[white blood cell concentration – absolute neutrophil count]. dNLR values greater than 3 and LDH values greater than the upper normal limit are counted as one factor. The LIPI score was used to categorize the study patients into three groups according to the number of the factors (good, 0 factors; intermediate, 1 factor; poor, 2 factors) (7). The cut-off value of LDH was defined according to the standards of each hospital."

5. I think that the data on the age, gender the type of cancer etc. should be included in the material.

Reply: This is a good point. The data on the age, gender, and the type of cancer are shown in the text (Methods – *Study design*, and Results – *Baseline characteristics*) and Table 1. We also modified Table 3 and Table 4 (cox regression analyses on PFS and OS) to include the type of cancer as a variable. There was no significant association with the cancer type and PFS in univariate analysis (Table 3 – model 1 and model 2). The cancer type showed association with OS in univariate analysis, however, did not remain significantly associated in multivariate analysis (Table 4 – model 1 and model 2).

Changes in the text: Table 3 and Table 4

Methods – *Study design*; Page 8, line 2-7

"Demographic and clinical data such as age, sex, Eastern Cooperative Oncology Group (ECOG) performance score, smoking habit, comorbidities, pulmonary function testing, and treatment modalities, pathological data such as cancer stage, pathologic type, and mutation study, and laboratory data such as complete blood count, c-reactive protein, and lactate dehydrogenase levels were collected from medical records."

Results – *Baseline characteristics*; Page 10, line 3-6

"Among the 252 patients with stage III NSCLC, the median age was 69 (IQR, 62-75) years with 191 male (75.8%) patients (Table 1). Regarding the pathologic type, 112 (44.4%) patients were diagnosed with adenocarcinomas, 125 (49.6%) with squamous cell carcinomas, 3 (1.2%) with adenosquamous carcinomas, and 12 (4.8%) with NSCLCs, NOS."

6. Most of the conclusions contain information commonly known in the treatment of NSCLC. The only the last part is about the LIPI issue and it is to general. Please correct it.

Reply: Thank you for the critical comment. Among stage III cancer, the N2 disease is a field of further researches. In the revised version, we added more detailed analysis regarding N2

disease, and attempted to make a combinatorial scoring system including bulky/nonbulky, single/multi station, and LIPI index. We have found out that this combinatorial scoring system was independently predictive of OS in the N2 subgroup. We have made additional figures and paragraph in the Results section, and further discussed in the Discussion section. Lastly, we made a brief mention in the Conclusions.

Changes in the text:

Methods – *Study design*; Page 8, line 22 ~ Page 9, line 1-7

"The subgroup analysis was done to describe N2 disease more specifically. The baseline characteristics and survival outcomes of clinical N2 subgroup were analyzed using new scoring system. The parameters included in the scoring system were lymph node station (single or multi), lymph node volume (non-bulky or bulky), and LIPI score (0, 1, or 2). Single lymph node station was scored as 0, and multi-station was scored as 1. Non-bulky lymph node was scored as 0, while bulky lymph node was scored as 1. The final score was counted as the sum of lymph node station, lymph node volume, and LIPI scores, which ranges from 0 to 4 (Supplementary figure 1)."

Results – *N2 subgroup analysis*; Page 13, line 21-23 ~ Page 14, line 1-16

"The baseline characteristics were analyzed among clinical N2 subgroup patients (n = 93, Supplementary Table 1). The median age was 68 (IQR, 61-75) years, and 71 (76.3%) patients were male. Regarding lymph node status, 40 (43%) were single station, while 35 (37.6%) were multi-station. Twelve (12.9%) patients showed bulky mediastinal lymph node, and 63 (67.7%) patients had non-bulky non-bulky N2 nodes. Among the N2 subgroup, 23 (24.7%) patients received surgery as the first-line treatment, and same number of patients were treated with chemotherapy and CCRT.

Kaplan-Meier curves for PFS and OS comparison according to the combinatorial scoring system which is the sum of lymph node (LN) station (single/multi-station), LN volume (bulky/non-bulky), and LIPI score parameters showed statistically significant difference ($p = 0.002$, and $p = 0.003$, respectively, Supplementary Figure 2 and Supplementary Figure 3).

In cox regression analysis on PFS, however, the combinatorial scores (LN station + LN volume + LIPI score) did not show statistical significance in univariate analysis ($p = 0.057$). The score was significantly associated with OS in both univariate and multivariate analysis (Supplementary Table 2, $p = 0.018$, and $p = 0.016$, respectively). The risk of mortality increased significantly as the combinatorial score increased. When compared to the score 0 group (reference), score 2 group showed HR of 9.498 (95% CI 2.200-41.011, $p = 0.003$) and score 3 group showed HR of 20.083 (95% CI 2.454-164.356, $p = 0.005$), respectively."

Discussion; Page 17, line 7-21

"Making the decision of treatment modalities in Stage IIIA-N2 disease is more challenging. As N2 involvement exhibits heterogeneous disease entity, treatment often requires bi- or trimodalities, and the clinical role of complete resection requires further investigation (17). In guidelines on decision of treatment options for Stage III N2 disease, lymph node extent

(single station or multi-station, single zone or multizone) and lymph node volume (non-bulky or bulky) are included as criteria (18). Stage III N2 disease is heterogeneous in terms of tumor burden, anatomical distribution of tumor cells, and treatment modalities. We believe that multiple factors should be considered when predicting outcomes of this heterogenous patients group. In the subgroup analysis of patients with N2 diseases, we attempted to make new scoring system. Lymph node station and volume parameters reflect the anatomical distribution and tumor burden of the N2 disease. LIPI score focuses on the onco-immunological background, as this biomarker has shown association with clinical outcomes in NSCLC patients undergoing immunotherapy (7). As was shown in our results, this multi-parameter scoring system may predict survival outcomes of the N2 disease. However, larger N2 populations are necessary for validation."

Conclusion; Page 18, line 9-10

"In clinical N2 subgroup, LN status combined with LIPI score may have predictive value for OS."

Tables; Supplementary Table 1, Supplementary Table 2

Figures; Supplementary Figure 1, Supplementary Figure 2, Supplementary Figure 3

Reviewer B

The manuscript entitled "The association between clinical parameters and resectability in stage III non-small cell lung cancer and LIPI score as a potential biomarker" describes the factors related to the resectability among the stage III non-small cell lung cancer (NSCLC) patients. The authors also investigated the role of LIPI score as a predictive marker for overall survival (OS) and progression-free survival (PFS).

The heterogeneity in stage III NSCLC patients is mainly due to the N status, which had already been demonstrated in the past. Therefore, the N2 status should be described more in detail when discussing the resectability, in other word, single station N2 or multiple station N2.

Although the results showed that the unresectable group had more patients with T4 and N2/N3 background, worse lung functions, and smoking habit, this observation is not novel.

The resectable stage III NSCLC showed better PFS and OS than the patients with unresectable tumor, which is the conclusion of the study, is also quite ordinary. And there are lots of confounding factors between resectability and prognosis.

The relationship between LIPI score and PFS/OS is not novel.

I think that the authors should try to combine the background factors and the patients' immune status or LIPI score together to create other prognostic factors.

Reply: Thank you for your time and efforts reviewing this article. All your comments had been

very critical and constructive. We fully agree that we need to show something more, especially for the N2 subgroup.

Among stage III cancer, N2 subgroup is very heterogeneous and no consensus had been reached regarding the treatment approach and how we predict prognosis. Furthermore, multiple factors should be considered when making the decision to either completely resect or undergo other non-surgical treatment.

We have took your comment regarding N2 stage very gravely, and further underwent additional data acquirement about 93 N2 patients, most importantly single/multi station, bulky/non-bulky. We used these additional data for both predicting prognosis and resectablity in patients with N2 disease. For predicting prognosis, we made combinatorial scoring system including previously shown LIPI index, bulky/non-bulky N2 disease, single/multi station N2 involvement.

***New scoring system**

The parameters included in the new scoring system were lymph node station (single or multi), lymph node volume (non-bulky or bulky), and LIPI score (0, 1, or 2). Single lymph node station was scored as 0, and multi-station was scored as 1. Non-bulky lymph node was scored as 0, while bulky lymph node was scored as 1. The final score was counted as the sum of lymph node station, lymph node volume, and LIPI score, which ranges from 0 to 4 (Supplementary Figure 1). We believe that the two additional parameters reflect the anatomical distribution and tumor burden of the N2 disease, while LIPI index more focuses on the onco-immunological background. We further made this mention in the Discussion section.

***Additional results in the revised version**

Log rank tests and Kaplan-Meier curves showed that PFS and OS according to the lymph node (LN) station + LN volume + LIPI score were significantly different ($p = 0.002$, and $p = 0.003$, respectively, Supplementary figure 2 and figure 3). In cox regression analysis on PFS, however, LN station + LN volume + LIPI score did not show statistical significance in univariate analysis ($p = 0.057$). The score was significantly associated with OS in both univariate and multivariate analysis (Supplementary Table 2, $p = 0.018$, and $p = 0.016$, respectively). The risk for mortality significantly increased as the combinatorial score increased. We think these results on predicting outcomes in N2 subgroups are important and further described in the Results section as in text and the separate tables.

***Resectability**

However, regarding resectability, no significant association was shown. Logistic regression analysis was also performed for unresectability with LN station + LN volume + LIPI score, however, it did not show statistical significance in univariate analysis ($p = 0.961$). In addition, when we underwent logistic regression separately for LN station status, and LN volume separately, these parameters did not show significant association with resectability. We assume that many other confounding factors exist, so the combinatorial score alone cannot predict

resectability.

***Change in title**

We agree that this new scoring system needs more validation in the larger population, but we believe that it reflects heterogeneous feature of the N2 disease. If you allow us, we would like to change the title of the manuscript to *"The association between clinical parameters and resectability in Stage III non-small cell lung cancer, and a combination of N2 lymph node burden and LIPI score as a potential biomarker"*

We hope these additional analyses had been novel enough and meet your standard. Thank you very much.

Changes in the text:

Methods – *Study design*; Page 8, line 22 ~ Page 9, line 1-7

Results – *N2 subgroup analysis*; Page 13, line 21-23 ~ Page 14, line 1-16

Discussion; Page 17, line 7-21

Conclusion; Page 18, line 9-10

Tables; Supplementary Table 1, Supplementary Table 2

Figures; Supplementary Figure 1, Supplementary Figure 2, Supplementary Figure 3