Peer Review File

Article Information: Available at http://dx.doi.org/10.21037/tlcr-20-832

Reviewer A

This study is very interesting because it focuses on an unmeet clinical need, which is represented by the need of discovering a risk scoring system able to select patients who could benefit more from immune checkpoint inhibitors in non small-cell lung cancer. I've read your manuscript and these are my suggestions:

Q• In the results section of the abstract, it would be appropriate to describe with one sentence that the DNA sequencing on tumor tissue was performed in 41 patients.

A: We have briefly described the results of DNA sequencing in the abstract (page 4, line 9-13).

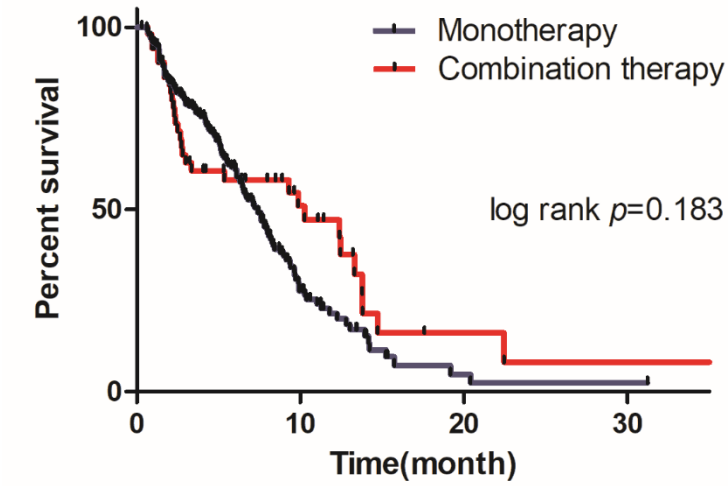
Q• The authors describe that, of all enrolled patients, 54 patients were treated with PD-1 inhibitor monotherapy, while the others received combo treatment (CT and ICIs). Did the authors evaluate the difference in terms of response between patients treated with immunotherapy and combined treatment? In my opinion, although the majority of patients received combo treatment, could be of interest to evaluate this point and, possibly, to correlate the different response with clinical features such as ECOG PS, ALC and lung/pleura metastasis on multivariate analysis.

A: We made a subgroup analysis evaluating the difference of response between NSCLC treated with monotherapy and combination therapy. Response evaluations including ORR, treatment efficacy (DCB/NDB), PFS and LEM score showed no significant difference between two groups (Table and figure were as follows).

We have done multivariate analysis (Table 2) evaluating the relationship between different response and clinical features, and the LEM scoring system was based on this multivariate analysis.

Response Evaluation	Monotherapy	Combination therapy	P value
LEM score			
Good	21 (38.9%)	90 (44.1%)	0.739
Intermediate	23 (42.6%)	83 (40.7%)	
Poor	10 (18.5%)	31 (15.2%)	
Best Response			
CR+PR	18(33.3%)	82(40.2%)	0.357
SD+PD	36(66.7%)	122(59.8%)	
Efficacy			
DCB	26(48.1%)	106(52.0%)	0.618
NDB	28(51.9%)	98(48.0%)	

Good: LEM score of 0-1; Intermediate: LEM score of 2-3; Poor: LEM score of 4-6
CR: Complete response, PR: Partial response, SD: Stable disease, PD: Progression disease;
DCB: Durable clinical benefit; NDB: Durable clinical benefit



Q• Finally, in the discussion, it would be useful to better describe the other risk scores, available in literature, in particular the LIPI score by Mezquita et al.

Awaiting the results on the association between LEM score and OS, this scoring system, possibly integrated with others, could represent a step forward in the selection of NSCLC patients at baseline.

A: We have added more informative description of the other risk scores in the part of **Discussion**, particularly LIPI score (page 18, line 13-17/ page 19, line 1-15).

Reviewer B

The manuscript entitled "Novel Risk Scoring System for Immune Checkpoint Inhibitors Treatment in Non-Small Cell Lung Cancer" by Li et al. aims to identify a risk scoring system to predict response to immune checkpoint inhibitors (ICIs).

The authors retrospectively analyzed 258 patients affected by advanced NSCLC treated with ICIs in a different line of therapy. To obtain a risk score Li et al. take into account clinical information and routine laboratory tests collected at the baseline of ICIs treatment. Moreover, the authors conducted in a small portion of the study population (41 patients) DNA sequencing in tumor tissue and match blood samples.

General comments:

Q:- In terms of form and language, English is quite poor. Additionally, some sentences have to be rephrased, since the clue is not completely clear and understandable.

A: This manuscript have been polished by a native English-speaking expert and some unclear sentences were rephrased as well.

Q: - Some citations i.e. 47-49 may be moved from Discussion to Introduction.

A: We made a brief introduction of these studies (Mezquita et al. and Martini et al) in the part of Introduction (page 6, line 8-13) in addition to the Discussion part.

Q:- Please add and discuss paper by Mazzaschi et al. https://doi.org/10.1016/j.lungcan.2020.07.028

A: We have added and discussed the study of Mazzaschi et al in the part of discussion (page 19, line 1-4).

Q: - Can you kindly better explain the exclusion criteria, in particular "Excluded for failure to reach evaluation of therapeutic efficacy", Fig.S1.

A: We excluded the patients **(1)** who received initial PD-1 immunotherapy at out hospital but was lost-to-follow-up thereafter (Test set, n=23; Validation set, n=87) and **(2)** those who do not receive sufficient sessions for evaluation at the time of study (Test

set, n=37; Validation set, n=42). We have added the exclusion criteria in the **part 2.1. Patient selection** (page 7, line 12-14) as well as revised **Figure S1** accordingly.

Q: - Please add data about the overall population, independently from single factor and/or LEM score e.g. PFS, OS, TTF, and ORR.

A: We have added data including best response and PFS in **Table 1**.

Most patients did not meet the endpoint of overall survival at the time of study. In the future, we would add data about OS, and we are also awaiting the results on the association between LEM score and OS.

Q: - Please spell some abbreviations mentioned in the abstract i.e ALC.

A: We have made modification as advised.

Q: - Please correct PD-(L)1 on page 3, line 21, and page 4, line 2. ICSs on page 4, line 13, and Illumia on page 6, line 18

A: We have changed PD-(L)1 to PD-1 (page 5, line 5/7), ICSs to ICIs (page 6, line 1), and Illumia to Illumina (page 9, line 10).

Q: - On page 8, lines 13-15, the authors claimed that they have adjusted the results for clinically-relevant factors such as age, smoking, prior therapy, etc. Nonetheless, considering the importance of lymphocyte count in the proposed LEM score and the potential influence of steroids treatment on this parameter, I suggest them to clarify this point. How many of these patients have received corticosteroids within 30 days before starting immunotherapy? And again…are there any correlations between steroid therapy and ALC values?

A: There was no long-term corticosteroid users included in our study, and steroids were only used before chemotherapy (or ICIs combined with chemotherapy). The ALC levels in the LEM score were tested initially at the baseline and were prior to steroid administration. Therefore, we could not determine the correlation between steroid administration and baseline ALC levels.

Reviewer C

The authors are searching for a predictive score of efficacy of ICI in metastatic NSCLC. This is an important question as currently, the only valuable parameter is PDL1 with major limitations in terms of accuracy.
Despite the interest of the initiative, there are a lot of methodological problems with the present study:

Q: - More than 200 patients were excluded. Patients that cannot reach the first evaluation for progressive disease as well as those stopping treatment for adverse reaction should be considered as treatment failure and included in the analyses. Otherwise, this is a bias selection.

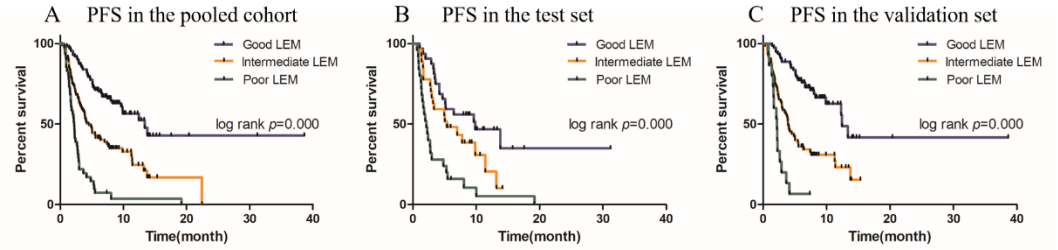
A: We did not accurately describe the exclusion criteria before and made some unnecessary misunderstanding. Here, we redefined the exclusion criteria in the **part 2.1. Patient selection** (page 7, line 12-14) as well as revised **Figure S1** accordingly.
We excluded the patients (1) who received initial PD-1 immunotherapy at out hospital but was lost-to-follow-up thereafter (Test set, n=23; Validation set, n=87) and (2) those who do not receive sufficient sessions for evaluation at the time of study (Test set, n=37; Validation set, n=42).

Q: - The primary objective of the study is not clearly stated. Which was the main judgement criteria: PFS, DCB or response? It cannot be the three at the same time. Also, from the statistical considerations we cannot extrapolate if the study has enough statistical power

A: We have added the primary endpoint of our study in the **part 2.1. Patient selection** (page 8, line 9-14). Patients were stratified as durable clinical benefit (DCB: partial or stable response lasting >6 months) and no durable benefit (NDB) groups according to published metrics. The primary endpoints were progression-free survival (PFS), and other objectives including treatment efficacy (DCB/NDB), objective response rate (ORR) and one year overall survival (OS) rate were also evaluated.

Q: - It is unclear how the authors constructed the derivation and validation groups. Apparently, they decided that patients from one hospital will be used for the derivation but without any statistical considerations. Further, the two groups are not comparable at least for PS (Chi² p <0.0001) and for type of treatment (monotherapy versus chemo-immunotherapy, Chi² p = 0.0005) but there are also discrepancies according to molecular alterations that are unknown in 11.5 and 33.9% respectively.

A: It is true that some study use split-sample or bootstrap replicates method to develop and validate a new model. However, the sample size of the current study was relatively small and underpowered, which was clarified in discussion. In addition, the aim of this study was to develop a novel prediction system rather than detect statistical significance. Difference in baseline characteristics was expected but this, on the other hand, partly support that the model could be applied in different populations.



Q: - There are no information on how was constructed the LEM score. On which basis was decided the value attributed to each parameter?

A: We conducted univariate and multivariate analyses (logistic and Cox regression analyses) to screen clinically relevant variables. Significant parameters ($p < 0.05$) including absolute lymphocyte count (ALC, **L**), Eastern Cooperative Oncology Group Performance Status (ECOG PS, **E**) and lung/pleural metastasis (**M**) were used to construct the LEM score (Table 2) .

Relative weights were based on odd ratio and hazard ratio of multivariate analyses (high HR/OR: weighted value = 3; intermediate HR/OR: weighted value = 2; low HR/OR: weighted value = 1). Weighted values were assigned to each parameter, and LEM score was the sum of weighted values of each variable (Table 3). We have added

the information on LEM score construction in **Table 3** and the part **3.3. Analyses of LEM risk scoring system.** (page 12, line 9-12)

Further, there are other problems:

Q: - table 2 is difficult to understand. Is it the results of the univariate analyses? Why adjusting variables with their own values (see section 3.2 in the text for the multivariate analysis and the term adjusted in the table 2 where the adjustement variables were "adjusted").

A: Table 2 were the results of multiple analyses (logistic and Cox regression analyses). We have made corresponding modification in the table 2 and the part **3.2. Univariate and multivariate analyses of baseline characteristics.** (page 11, line 14-15).

Q: - we do not know which were the criteria for selecting variables in the multivariate analysis (what means marginally significant?)

A: After univariate analyses, marginally significant ($p < 0.1$) factors and demographic characteristics were included for multivariate analyses. Marginally significant meant $p < 0.1$ (page 11, line 14-15).

Minor comments:

Q: - in the abstract, the term ALC must be explicited and the methodology is not adequately reported

A: We have made modification as advised and reported the methodology in detail in the abstract section. (page 3, line 10-16)

Q: - the definition of PFS and OS must be reported in the methods section

A: We have made definition of PFS and OS in the methods section. (page 8, line 10-11/ page 8, line 13-14)