Peer Review File

Article information: http://dx.doi.org/10.21037/tlcr-21-199

Reviewer A

Comment 1: The work is interesting and well designed. However, the discussion and conclusions are unhelpful. The message is not clear, the discussion of the results in the clinical context is not clear. There is no intuition of added value to what is already known

Response 1: Thank you for your comment.

Reviewer B: The authors performed a meta-analysis of published data to evaluate the consistency of using PFS as a surrogate marker for OS in randomized phase II and III trials testing immune checkpoint inhibitors (ICI) for the treatment of metastatic NSCLC. The present study shows interesting results that might be appealing for the readers of this journal, nevertheless, some adjustments are necessary to make the results more consistent and sound.

Major Comments

Comment 1: In the abstract's conclusion, it is stated that PFS is an adequate surrogate for OS in studies testing ICI in second line, whilst the results shown point in the opposite direction.

Response 1: Thank you for your comment. I should apologize for this mistake. Our current findings suggested that PFS could be a potential alternative endpoint for OS in trials with immunotherapy as first–line setting, but PFS should be cautiously interpreted without OS data for trials with immunotherapy as second or above line treatment. We have revised the relevant sentences in the new version of manuscript.

The following text has been revised:

Abstract

"PFS could be a potential alternative endpoint for OS in trials with immunotherapy as first–line setting, but PFS should be cautiously interpreted without OS data for trials with immunotherapy as second or above line treatment."

Discussion

"PFS is a potential surrogate endpoint for OS in trials of advanced NSCLC treated with anti–PD-(L)1–based combination therapy or trials with first–line immunotherapy. However, it should be cautiously interpreted in the absence of OS for trials of anti–PD-(L)1 inhibitor–based monotherapy or trials with immunotherapy as second or later–line treatment, which might provide valuable

clues to guide the development of improved regulatory endpoints for immunotherapy in advanced NSCLC."

Comment 2: The authors decided to pool up together studies testing ICI used both as first line treatment as well as in second or further lines. This strategy, by itself, might impact the ability to demonstrate any correlation between PFS and OS, since, most studies evaluating ICI for NSCLC in second or further lines of treatment did not observed any difference in PFS among treatment groups. Besides, patients included in studies testing ICI in second line have gone through a biological selection and do not reflect the population treated in first line.

Response 2: Thank you for your good comment. We do agree with your opinion that pooling up together studies testing ICI used both as first line treatment as well as in second or further lines would impact the ability to demonstrate correlation between PFS and OS. To clarify the real correlation between PFS and OS, we have conducted a series of subgroup analysis: (1) trial phase (III vs II); (2) histological type (NSCLC vs non-squamous NSCLC only vs squamous NSCLC only); (3) specific anti–PD-(L)1 drug; (4) treatment group (combination therapy vs monotherapy); (5) biomarker selection (yes vs no); and (6) lines of treatment (first-line vs second or above lines vs others); (7) primary endpoint (OS vs non-OS); (8) follow-up duration (< 24 months vs ≥ 24 months). For each subgroup analysis, we calculated the P value for interaction ($P_{interaction}$) by using a meta-regression model. We also noticed the differences on eligible patients between trials testing ICI in second-line treatment and those testing ICI in first-line setting. We performed the key subgroup analysis based on the lines of treatment (first-line vs second or above lines vs others). The rHR was 0.91 (95% CI, 0.84 to 0.99) for trials with immunotherapy as first–line setting, and 1.17 (95% CI, 1.06 to 1.29) for trials with immunotherapy as second or above line treatment ($P_{interaction} < 0.01$).

Comment 3: Similarly, the authors grouped together, in their preliminary analysis, studies testing anti-PD1/PD-L1 drugs in monotherapy, in combination with chemotherapy and in combination with an anti-CTLA4 drug. The selection criteria were quite different among these trials, especially in those testing ICI as monotherapy, in which patients should bear tumors with some degree of PD-L1 expression (usually high expression). I believe it would be better to perform these analysis separately.

Response 3: Thanks. Although we performed several subgroup analyses based on the different features, distinct selection criteria among different included

publications would compromise the finale results of this meta-analysis. We have added this unavoidable factor as one of the significant limitations in the Discussion part of our revised manuscript.
The following text has been added:
"Forth, selection criteria were quite different among these included trials, especially in those testing ICI as monotherapy. Although we performed several subgroup analyses based on the different features, this difference would still compromise the findings of this meta-analysis. Future investigation should focus on a specific group of populations. Lastly, we only included trials with anti-PD-(L)1 inhibitor in advanced NSCLC. Therefore, it is not suitable to generalize the current findings to other ICIs and other tumor types."

Comment 4: Some studies included in this meta-analysis do not have mature OS data, how have the authors integrated them in their analysis?
Response 4: Thanks. Although these trials did not report the mature OS data, they reported the up-to-date HR values (until June 1, 2020 according to the online search strategy of study protocol). We just integrated the HRs of eligible studies.

Comment 5: It does not seem appropriate to include two subpopulations of the same trial, except they represent comparisons between different treatment arms, in the same meta-analysis. Henceforth, CHECKMATE 227-1, KEYNOTE-033-2 and EMPOWER-Lung1-2 should be excluded from the general analysis. KEYNOTE-033-2 and EMPOWER-Lung1-2 could be pooled up together with other similar trials in a meta-analysis evaluating ICI as monotherapy in first line, though.
Response 5: Thanks for your suggestion. Although CHECKMATE 227-1, KEYNOTE-033-2 and EMPOWER-Lung1-2 had the same control group to the CHECKMATE 227-2, KEYNOTE-033-1 and EMPOWER-Lung1-1, respectively, they had totally different experimental groups. Thus, we only integrated the relevant HRs of each arm.

Comment 6: The proportionality of the death or disease progression probability is not maintained along all the extent of the curve in trials testing ICIs, with a tail developing after some follow-up time. Besides, sometimes, mainly in second line, differences between PFS curves only develop after 3 to 6 months. How do these characteristics might affect the correlation between OS and PFS in these trials?
Response 6: Thanks for your comment. Your concern is very significant. In fact, this is also the conundrum of immunotherapy. Researchers have tried to develop

several methods or tools to accurately assess the proportionality of the death or disease progression probability in trials testing ICIs. To date, we did not observe better results or alternatives than the survival curves together with HRs. Definitely, these characteristics could affect the correlation between OS and PFS in these trials, but currently we do not have a better way to avoid this influence.

Minor comments

Comment 7: Minor typing/writing review.

Response 7: Thanks. The manuscript has been revised with new line and page numbers in the text, some grammar and spelling errors have also been corrected. Furthermore, the current version of manuscript has been revised by an English speaker, Prof. Jun Zhang, from Department of Cancer Biology, University of Kansas Cancer Center, University of Kansas Medical Center, Kansas City, USA.

Comment 8: The authors inform they excluded irrelevant studies. What was the definition of irrelevant?

Response 8: Thanks. Irrelevant studies included publications did not report the efficacy of anti–PD-(L)1 as monotherapy or in combination with standard treatment in patients with advanced NSCLC. In the methodological section, the exclusion criteria, we made a statement.

Comment 9: Paragraphs describing treatment effect by subgroup should be separated in order to better organize the ideas and make the text more approachable for the reader.

Response 9: Thanks. To avoid the manuscript appears too long and difficult to read and understand, we showed the subgroup results in a paragraph. But we separately discussed the meaningful subgroup results in the Discussion part.

Comment 10: The authors state they observed a greater effect of ICI on OS for trials that included all histologies when compared with those that included only squamous or non-squamous histologies. Could they elaborate on the reasons for these results? They might be the result of an enrichment for trials testing ICI in second line in the first subgroup.

Response 10: Thanks for your good comment. In fact, we also have no idea that treatment effect of ICI on OS for trials that included all histologies was greater than those that included only squamous or non-squamous histologies. Maybe the different trial design, control arm, and included populations together with distinct biological selection (e.g. PD-L1 expression, TMB, etc.) were the potential

reasons.

Comment 11: The authors did not include the CCTG BR34 study in their list of eligible trials.

Response 11: Thanks. The deadline of online search is June 1, 2020. Although the ASCO.com reported the abstract of CCTG BR34 study, we did not find more details of this study at that time. Thus, we did not include the CCTG BR34 study in the list of eligible trials.