



Differences in treatment effect size between progression-free survival and overall survival in anti-PD-1/PD-L1 inhibitors-based trials in advanced NSCLC: a systematic review and meta-analysis

Zhirui Zhou^{1#}, Shengxiang Ren^{2#}, Lingxiao Chen³, Caicun Zhou², Tao Jiang^{2,4}

¹Radiation Oncology Center, Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai, China; ²Department of Medical Oncology, Shanghai Pulmonary Hospital & Thoracic Cancer Institute, Tongji University School of Medicine, Shanghai, China; ³Institute of Bone and Joint Research, Kolling Institute, Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia; ⁴Department of Pulmonary Medicine, Shanghai Respiratory Research Institute, Zhongshan Hospital Affiliated to Fudan University, Shanghai, China

Contributions: (I) Conception and design: T Jiang, S Ren, C Zhou; (II) Administrative support: T Jiang, S Ren, C Zhou; (III) Provision of study materials or patients: T Jiang, S Ren, C Zhou; (IV) Collection and assembly of data: T Jiang, Z Zhou, L Chen; (V) Data analysis and interpretation: T Jiang, Z Zhou, L Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Tao Jiang, PhD. Department of Medical Oncology, Shanghai Pulmonary Hospital & Thoracic Cancer Institute, Tongji University School of Medicine, No. 507, Zheng Min Road, Shanghai 200433, China; Department of Pulmonary Medicine, Shanghai Respiratory Research Institute, Zhongshan Hospital Affiliated to Fudan University, No. 180, Fenglin Road, Shanghai 200032, China. Email: tonyjiangdr@163.com.

Background: To investigate the differences in treatment effect sizes between progression-free survival (PFS) and overall survival (OS) in advanced non-small cell lung cancer (NSCLC) treated with programmed cell death 1 (PD-1) and its ligand (PD-L1) blockade-based treatments.

Methods: The differences in treatment effect sizes between PFS and OS were assessed by using a ratio of hazard ratio (rHR): the HR for PFS to that for OS. A random effects meta-analysis across trials was conducted to generate the combined rHR. We also evaluated the feasibility of adopting PFS as the surrogate of OS by using Spearman correlation coefficient (*R*) between $\log\text{HR}_{\text{PFS}}$ and $\log\text{HR}_{\text{OS}}$.

Results: A total of 27 randomized controlled trials (RCTs) with 15,590 patients were included. Treatment effect sizes were comparable, on average, for OS than for PFS (pooled rHR, 0.98; 95% CI, 0.91 to 1.08). Subgroup analysis revealed that treatment effect sizes were greater for OS than for PFS for trials with immunotherapy as second or above line treatment (rHR, 1.17; 95% CI, 1.06 to 1.29), while the differences were greater for PFS than for OS for trials with immunotherapy as first-line setting (rHR, 0.91; 95% CI, 0.84 to 0.99; $P_{\text{interaction}} < 0.01$). The coefficient of determination was 40% and *R* was 0.63 between $\log\text{HR}_{\text{PFS}}$ and $\log\text{HR}_{\text{OS}}$. Subgroup analysis showed that coefficient of determination and *R* were 62% and 0.79 in trials with immunotherapy as first-line setting, 22% and 0.47 in trials with immunotherapy as second or above line treatment, respectively.

Discussion: Treatment effect sizes between PFS and OS were roughly consistent in trials with different anti-PD-(L)1 inhibitor-based therapies. 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.

Keywords: Programmed cell death 1 (PD-1); programmed cell death ligand 1 (PD-L1); lung cancer; survival; end point

Submitted Mar 11, 2021. Accepted for publication May 13, 2021.

doi: 10.21037/tlcr-21-199

View this article at: <http://dx.doi.org/10.21037/tlcr-21-199>

Introduction

Immune checkpoint inhibitors (ICIs) targeting programmed cell death protein 1 (PD-1) and its ligand (PD-L1) significantly improve the prognosis and have shifted the treatment paradigm in patients with advanced non-small cell lung cancer (NSCLC) (1,2). Unlike platinum-based chemotherapy and molecular targeted therapy, anti-PD-(L)1 inhibitors-based treatments exhibit unconventional response patterns including hyperprogression (3,4), pseudoprogression (5), long-lasting response and response after treatment discontinuation (6,7), which could result in several unique types of Kaplan-Meier curves such as “long tail” of overall survival (OS) curves and “initial overlap or crossover” of progression-free survival (PFS) curves (8,9). As a result, primary endpoint determination would also encounter challenges in the era of immunotherapy clinical trials.

Generally, OS is the standard primary endpoint for phase III clinical trials in solid tumors (10,11). However, it would need long-term follow-up and can be influenced by crossover or the effects of post-progression treatments. Therefore, PFS is often leveraged as an acceptable surrogate endpoint for OS, especially in the era of targeted therapy, though the reliability of PFS remains controversial (12-14). Recently, several phase III trials investigating the efficacy of anti-PD-(L)1 inhibitor plus chemotherapy adopted PFS as single primary endpoint (15-17). Moreover, given the rapid development of novel ICIs and growing need for patients to access innovative treatments (18,19), regulatory authorities are more likely to favor PFS for making decisions regarding drug approval and licensing (20-23). However, in advanced NSCLC, whether PFS is a reliable surrogate endpoint for OS in clinical trial design of anti-PD-(L)1 inhibitor-based therapeutic strategies remains undetermined.

In this study, we performed a systematic review and meta-analysis of published randomized controlled trials (RCTs) and aimed to investigate the differences in treatment effect sizes between PFS and OS, and the surrogacy of PFS for OS in patients with advanced NSCLC treated with anti-PD-(L)1 inhibitor-based treatments. We present the following article in accordance with the PRISMA reporting checklist (available at <http://dx.doi.org/10.21037/tlcr-21-199>).

Methods

Online search strategy

We performed a systematic online search for RCTs that

compared anti-PD-(L)1 inhibitors-based therapies with standard treatments in advanced NSCLC via PubMed, the Cochrane Library, and EMBASE in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) reporting guideline. Relevant conferences abstracts and presentations including the American Society of Clinical Oncology, the American Association for Cancer Research, the European Society for Medical Oncology and the World Lung Cancer Conference were also searched until June 1, 2020. We also manually screened the references of each eligible study until no additional articles could be added. The main keywords used for online search were “lung neoplasm”, “lung tumor”, “lung cancer”, “checkpoint inhibitor”, “programmed cell death 1”, “programmed cell death ligand 1”, “PD-1”, “PD-L1” and “randomized”. The full online search strategies were listed in [Supplemental Material](#).

Inclusion and excluded criteria

We included RCTs that compared the efficacy of anti-PD-(L)1 as monotherapy or in combination with standard treatment in patients with advanced NSCLC. Retrospective studies, cost effectiveness analyses, quality of life studies, single-arm phase I or II trials, randomized studies did not report the efficacy of anti-PD-(L)1 as monotherapy or in combination with standard treatment in patients with advanced NSCLC or that compared anti-PD-(L)1 treatment with other immunotherapies or included cases less than 100 were excluded. When duplicate publications for the same study occurred, we included only the most recent and complete publications or the ones supporting the approval by the US Food and Drug Administration (FDA). Three authors (TJ, ZRZ and SXR) independently conducted the selection of all included studies. Discrepancies were settled by discussion and consensus.

Data extraction

Three authors (TJ, ZRZ and SXR) independently carried out the data extraction based on PRISMA statement by using a standardized data collection form. The following items were extracted from each eligible study: trial name, published year, national clinical trials (NCT) identification number, trial phase, total patients, histological type, anti-PD-(L)1 drug, biomarker status, lines of treatment, treatment group, primary endpoint, and follow-up duration. They also independently extracted the hazard

ratios (HRs) and the related 95% confidence intervals (CIs) for PFS and OS. Where available, we included the most updated survival data.

Quality assessment

The methodologic quality for each included study was assessed according to the Cochrane Collaboration handbook (24) based on the original publication or its update and the supplemental materials. The adequacy of the following aspects was evaluated: random sequence generation, allocation concealment, blinding of participants, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other potential threats to validity. The risk of bias of each aspect was classified as low, high, or unknown. Discrepancies in

data extraction and quality assessment were resolved by discussion to achieve consensus among all investigators.

Statistical analysis

We evaluated the differences in treatment effect size between OS and PFS by using a two-step process as previously described (25,26). Briefly, we compared differences in outcome measures by using a ratio of HRs (rHR, defined as the ratio of the HR for PFS to the HR for OS) within each trial. An rHR <1.0 suggested that the treatment effect size was greater for PFS than for OS while rHR >1.0 indicated that the treatment effect size was greater for OS than for PFS. Due to the dependency between PFS and OS outcomes, the standard error (SE) of the log rHR was calculated as follows:

$$SE(\log rHR) = \sqrt{Var(\log HR_{PFS}) + Var(\log HR_{OS}) - \rho \times 2 \times SE(\log HR_{PFS}) \times SE(\log HR_{OS})} \quad [1]$$

Since individual patient-level data is unavailable, we could not assess the dependency between PFS and OS outcomes. Thus, we calculated the SE(log rHR) under the assumption of no dependency ($\rho=0$). Considering the heterogeneous group of patient populations, the pooled rHR across the studies was obtained by combining the rHR of each study using a random-effects model. Heterogeneity among trials was evaluated with the I^2 statistic and the Cochran Q χ^2 test. Assumption of homogeneity was considered invalid for values of $P < 0.10$. In a sensitivity analysis, we surveyed the robustness of the estimation of the pooled rHR for arbitrary correlations between the $\log HR_{PFS}$ and $\log HR_{OS}$, ranging from no dependency ($\rho=0$) to complete dependency ($\rho=1.00$) (26).

We then conducted the following subgroup analyses: (I) trial phase (III vs. II); (II) histological type (NSCLC vs. non-squamous NSCLC only vs. squamous NSCLC only); (III) specific anti-PD-(L)1 drug; (IV) treatment group (combination therapy vs. monotherapy); (V) biomarker selection (yes vs. no); and (VI) lines of treatment (first-line vs. second or above lines vs. others); (VII) primary endpoint (OS vs. non-OS); (VIII) follow-up duration (<24 vs. ≥ 24 months). For each subgroup analysis, we calculated the P value for interaction ($P_{interaction}$) by using a meta-regression model.

We utilized a least squares-linear regression model weighted by within-trial variance and determined the coefficient of determination (the proportion of variance in treatment effect on OS that is predictable from the

treatment effect on PFS) (27). The correlation between HRs of PFS and OS was estimated using the Spearman correlation coefficient (R) between $\log HR_{PFS}$ and $\log HR_{OS}$. Then we conducted surrogate end point validation by using a regression of $\log HR_{PFS}$ and $\log HR_{OS}$. Two-sided $P < 0.05$ was considered as statistical significance. All data were analyzed by using IBM SPSS statistical software version 22.0 and R v3.6.3 (<http://www.r-project.org>; the R Foundation for Statistical Computing, Vienna, Austria).

Results

Studies identification and characteristics

As shown in *Figure 1*, 439 publications and 12 conferences abstracts and presentations were retrieved via the initial online search, and 58 duplicates were excluded. After reviewing the titles and abstracts, 175 studies were excluded and 218 relevant publications remained for eligibility. After a detailed review, 191 studies were removed because they were reviews/comments, incomplete reports, irrelevant articles and studies without usable data. Finally, 27 RCTs involving 15,590 patients with NSCLC were included in the meta-analysis (<https://cdn.amegroups.cn/static/public/tlcr-21-199-1.pdf>).

In total, two trials (7%) were phase II trials, one (4%) was phase II/III trial and the remaining 24 (89%) were phase III trials (<https://cdn.amegroups.cn/static/public/tlcr-21-199-1.pdf>). 14 trials were done in patients with advanced

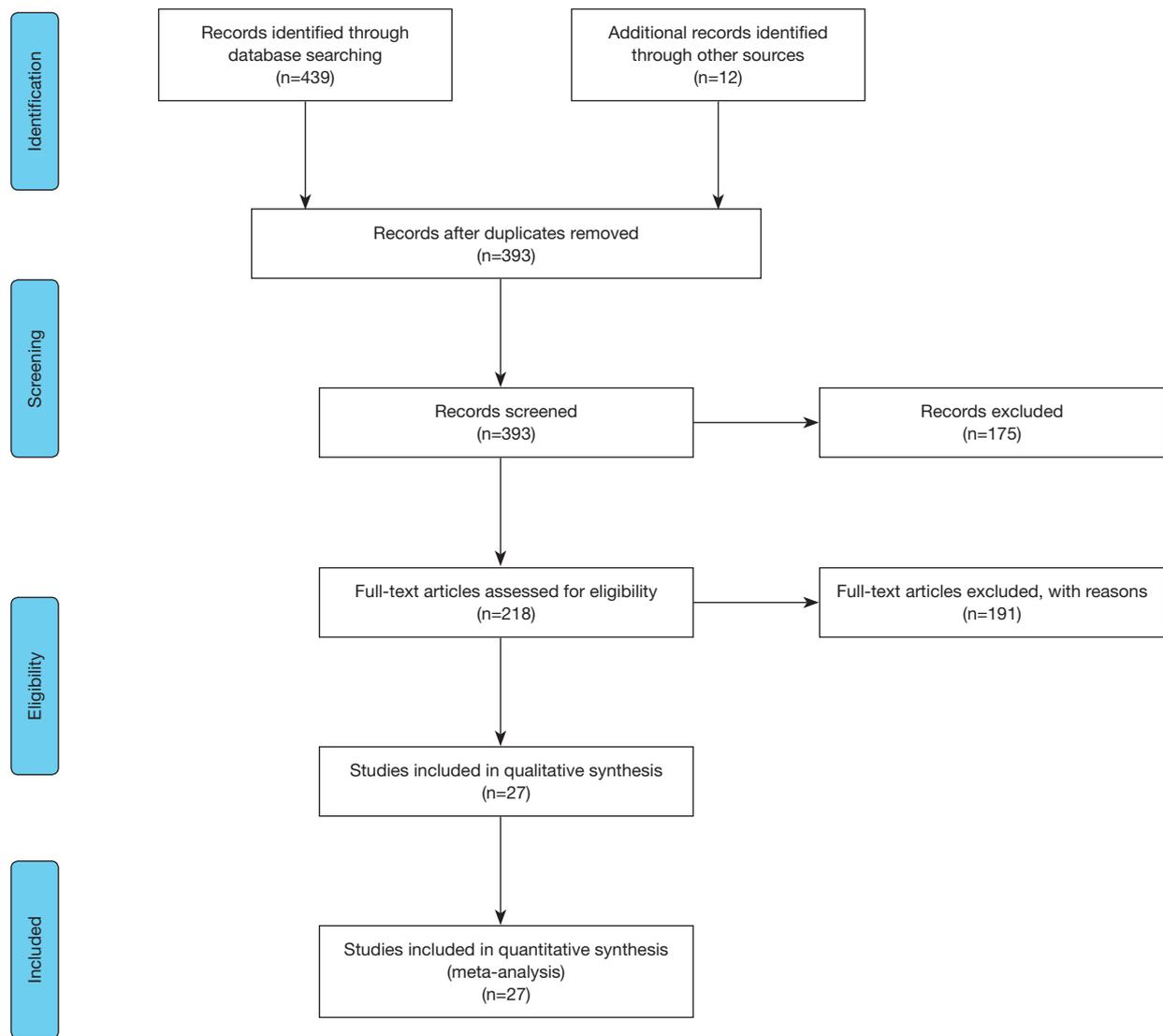


Figure 1 Study selection based on the Preferred Reporting Items for Systematic Reviews and Meta-analyses reporting guideline.

NSCLC, nine trials in non-squamous NSCLC, and four in squamous NSCLC. The included studies covered 16 trials with anti-PD-1 inhibitors (including six with nivolumab, seven with pembrolizumab, one with camrelizumab and two with sintilimab) and 11 trials with anti-PD-L1 inhibitors (including one with avelumab and seven with atezolizumab, and three with durvalumab) compared with control groups receiving standard treatments. 11 trials preset the biomarker including PD-L1 expression and tumor mutation burden. 18 studies were conducted with the immunotherapy as first-line setting, seven with the immunotherapy as second or above line treatment, and two with the immunotherapy as

third or above line treatment. Thirteen studies tested anti-PD-(L)1 inhibitor as a single agent and 14 as combination therapy. The primary end points were OS in 10 studies, PFS in seven studies, both OS and PFS as co-primary end points in nine studies, and objective response rate (ORR) in one study. All included studies were well-designed with well-defined main outcomes. Data from four trials (EMPOWER-Lung-1, IMpower132, KEYNOTE-033, NA2020, MYSTIC, ORIENT-12 and CheckMate 9LA) were extracted from conference presentations. The assessment of methodologic quality for each included study is provided in [Table S1](#).

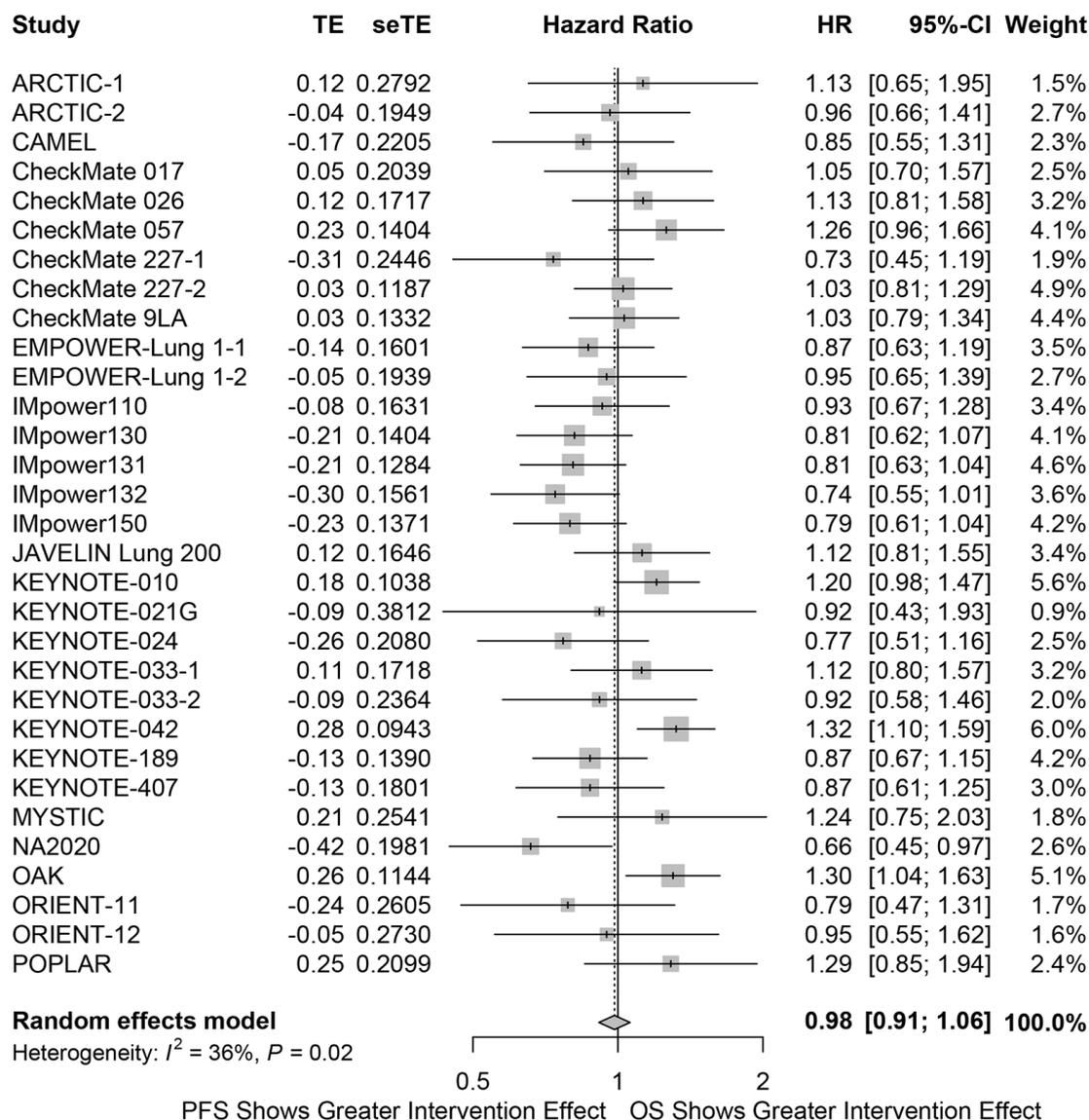


Figure 2 Ratios of hazard ratios (rHRs) between OS and PFS for 27 trials. An rHR <1.0 suggests greater treatment effect size for PFS than for OS.

Treatment effect size differences between PFS and OS

Twenty trials (74%) found consistent results between OS and PFS, 13 showing statistical significance for both, and seven showing both as non-statistical significance. A statistically significant benefit was shown for PFS but not for OS in five trials and OS but not for PFS in four trials. Treatment effect sizes were similar, on average, for OS than for PFS (pooled rHR, 0.98; 95% CI, 0.91 to 1.06), with

significant heterogeneity across trials ($I^2=36\%$, $P=0.02$; *Figure 2*). This result remains consistent on sensitivity analyses by testing different levels of dependency (range, 0.00–1.00) between OS and PFS (*Figure S1*).

Subgroup analyses (*Figure 3*) showed that the treatment effect sizes did not differ for OS than for PFS in phase III (rHR, 0.98; 95% CI, 0.91 to 1.05) and phase II trials (rHR, 1.19; 95% CI, 0.83 to 1.71; $P_{\text{interaction}}=0.30$). Intriguingly, the

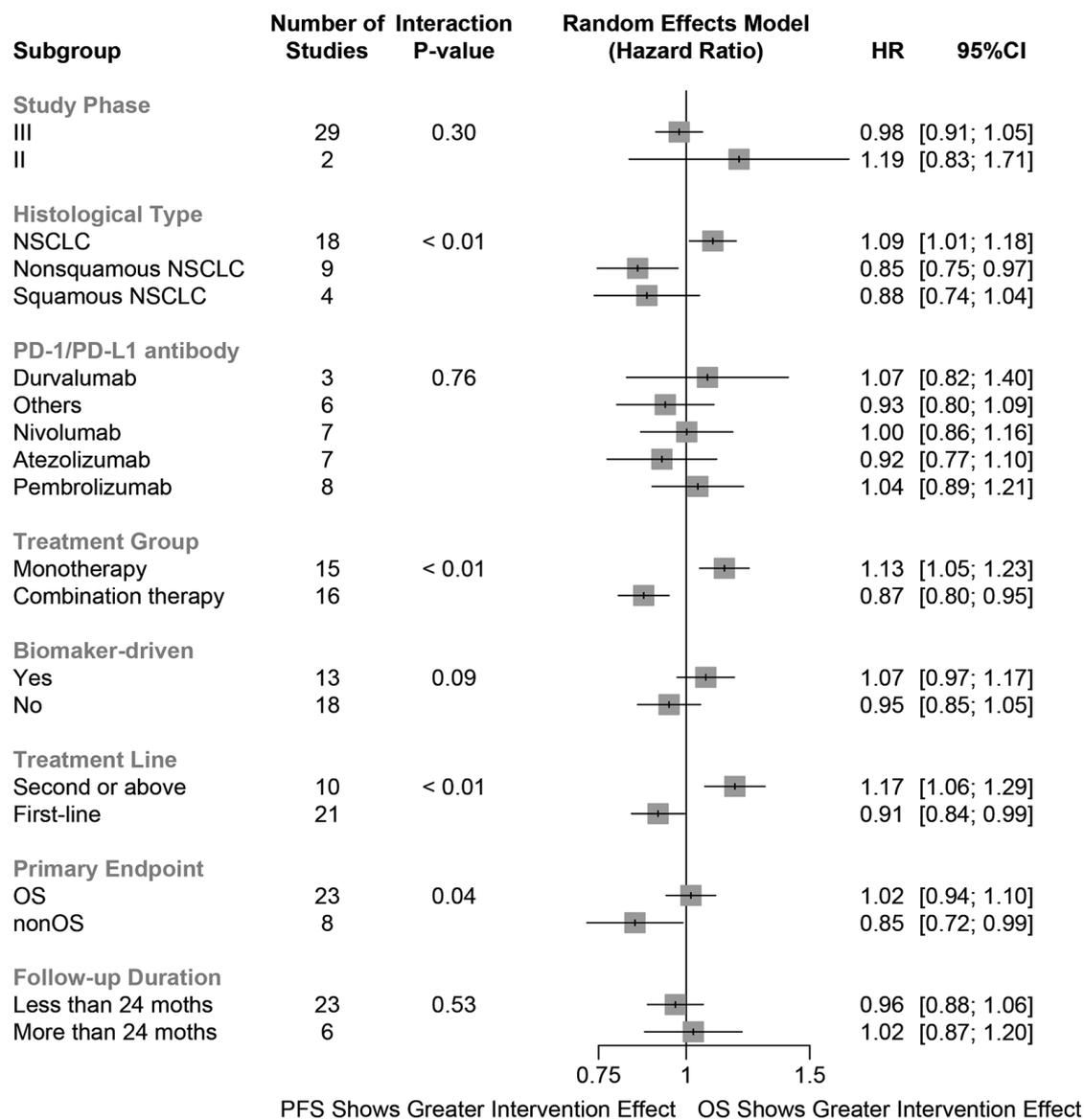


Figure 3 Subgroup analyses of ratios of hazard ratios between OS and PFS according to trial phase (III vs. II), histological type (NSCLC vs. non-squamous NSCLC vs. squamous NSCLC), specific anti-PD-(L)1 drug, treatment group (combination therapy vs. monotherapy), biomarker selection (yes vs. no), lines of treatment (first-line vs. second or above lines vs. others), primary endpoint (OS vs. non-OS), follow-up duration (<24 vs. ≥24 months).

treatment effect sizes were 9% greater for OS than for PFS in NSCLC (including both squamous and non-squamous NSCLC; rHR, 1.09; 95% CI, 1.01 to 1.18) instead of squamous NSCLC only (rHR, 0.85; 95% CI, 0.75 to 0.97), and non-squamous NSCLC only (rHR, 0.88; 95% CI, 0.74 to 1.04). The interaction test among histological types was statistically significant ($P_{interaction} < 0.01$). On stratification by

specific drug, the treatment effect sizes were comparable for OS than for PFS and interaction test was not statistically significant ($P_{interaction} = 0.76$). Notably, the differences in treatment effect sizes were greater for OS than for PFS for trials of anti-PD-(L)1 monotherapy (rHR, 1.13; 95% CI, 1.05 to 1.23) while it was greater for PFS than for OS for trials of combination therapy (rHR, 0.87; 95% CI, 0.80 to

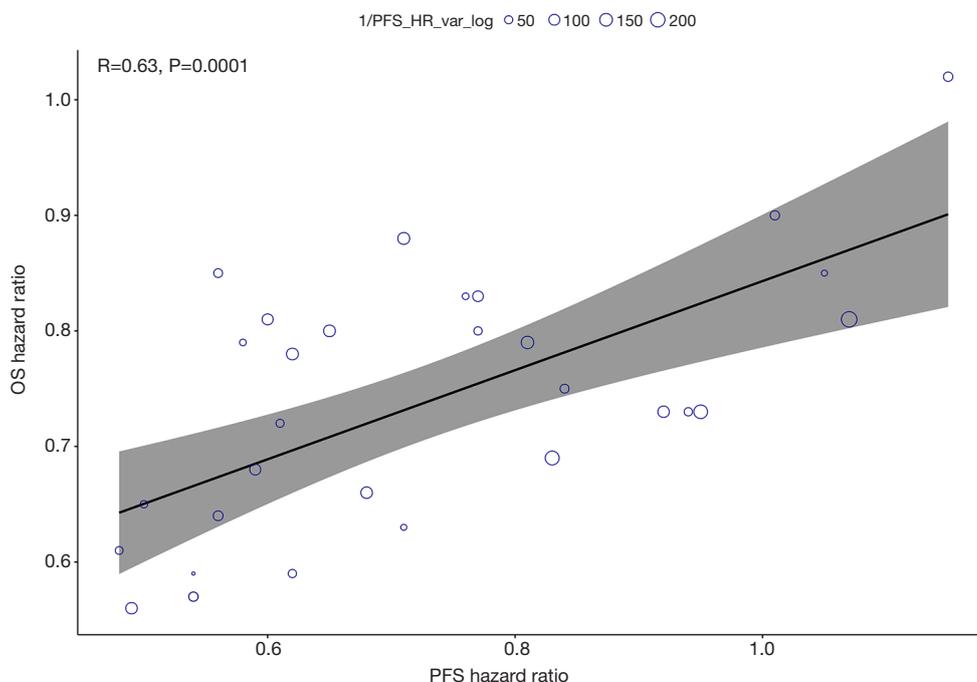


Figure 4 Correlation between OS and PFS. Each circle represents a trial. The size of the circle is proportional to the inverse of the variance. The lines represent 95% prediction intervals for a trial with a median weight.

0.95). The interaction test result was statistically significant ($P_{\text{interaction}} < 0.01$). We also found clear differences by treatment lines. 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 ($P_{\text{interaction}} < 0.01$). The differences in treatment effect sizes were also observed by primary endpoints. The rHR was 1.02 (95% CI, 0.94 to 1.10) for trials with OS as primary endpoint, and 0.85 (95% CI, 0.72 to 0.99) for trials with other primary endpoints ($P_{\text{interaction}} = 0.04$). However, we did not find any significant differences by biomarker status ($P_{\text{interaction}} = 0.09$) or follow-up duration ($P_{\text{interaction}} = 0.53$).

Surrogacy metrics

The Spearman correlation coefficient (R) between $\log\text{HR}_{\text{OS}}$ and $\log\text{HR}_{\text{PFS}}$ was 0.63 ($P = 0.0001$), and the co-efficient of determination in the weighted linear regression model was 0.40 (Figure 4); thus, 40% of the variability in the OS effect could be explained by the variability in the PFS effect. Of note, the correlation between $\log\text{HR}_{\text{OS}}$ and $\log\text{HR}_{\text{PFS}}$ appeared to be stronger for anti-PD-(L)1 monotherapy group ($R = 0.75$; $P = 0.0012$; Figure S2A) than anti-PD-(L)1

based combination therapy ($R = 0.68$; $P = 0.0038$; Figure S2B). The co-efficient of determination was 0.56 in monotherapy group and 0.46 in combination treatment group, respectively. Subgroup analyses also showed that trials with first-line immunotherapy ($R = 0.79$; $P < 0.0001$; Figure S3A) had significantly higher correlation between $\log\text{HR}_{\text{OS}}$ and $\log\text{HR}_{\text{PFS}}$ than trials with immunotherapy as second or later-line treatment ($R = 0.47$; $P = 0.1700$; Figure S3B).

Discussion

In this systematic review and meta-analysis of 27 RCTs with 15,590 patients, treatment effect sizes were comparable, on average, between PFS and OS. Although more than one half of all included trials showed consistent statistical significance for both PFS and OS, a moderate correlation was observed between HRs of PFS and OS at trial level. Subgroup analyses showed that the differences in treatment effect sizes were greater for PFS than for OS in trials with first-line immunotherapy, and a significant correlation was observed between HRs of PFS and OS. Conversely, the differences in treatment effect sizes were greater for OS than for PFS in trials with immunotherapy as second

or later-line treatment, while a limited correlation was observed between HRs of PFS and OS.

We observed that the agreements between PFS and OS outcomes in 74% of the included trials, which were inconsistent with a previous publication that both qualitative and quantitative differences in treatment effects were found between PFS and OS in immunotherapy trials (26). In this recent meta-epidemiological study in trials of US FDA-approved oncology immunotherapy drugs except anti-PD-(L)1 inhibitor, treatment effect sizes were greater for PFS than for OS, which is consistent with surrogacy metrics (26). As a result, caution must be taken when interpreting PFS in the absence of OS. In contrast, the current study only included anti-PD-(L)1 inhibitor trial and found consistent treatment effect sizes between PFS and OS when evaluating surrogacy metrics including R and co-efficient of determination in the weighted linear regression model. Although R showed significant correlation, co-efficient of determination was relatively low, with only 40% of the variability in the OS effect explained by the variability in the PFS effect. Similarly, a recent meta-analysis included trials in solid tumors with PD-1 inhibitors and found that there was no significant correlation between OS and PFS in terms of medians and gains in medians, even though their HRs were significantly correlated (14), indicating that PFS could not be a very ideal surrogate for OS when investigating PD-1 inhibitors-based treatments in the overall population.

Due to the obvious heterogeneity of combined rHR in whole group, we then conducted subgroup analyses. The results revealed no significant difference regarding study phase, specific drugs, biomarker status, or follow-up duration. Notably, the differences in treatment effect sizes were greater for OS than PFS for monotherapy trials and trials with first-line immunotherapy. Moreover, a highly significant correlation between HRs of PFS and OS was observed in monotherapy group and first-line setting group, suggesting a high feasibility of utilizing PFS as surrogate for OS in this clinical scenario. On the contrary, treatment effect sizes were greater for PFS than OS in combination therapy trials and trials with immunotherapy as second or later-line treatment. Only a moderate correlation between HRs of PFS and OS was found, which were consistent with previous findings (28-30), indicating that PFS might not be suitable as an independent surrogate for OS in anti-PD-(L)1 inhibitors-based combination treatment trials or trials with immunotherapy as second or later-line treatment. Therefore, OS or OS combined with PFS should be still regarded as the standard primary endpoints in these clinical

trials design.

Given the rapid progress and development of novel immunotherapies and unique patterns of response and progression, there is an urgent need to explore additional alternative endpoints to accurately depict early activity, and optimize combinatorial strategies. A recent meta-analysis investigated milestone rates in 25 trials and found milestone rates (e.g., OS at 12 months) could be used as a potential early endpoint to measure or describe drug activity in immunotherapy trials in lung cancer (23). However, milestone rate could not account for the totality of the survival curve and the effect of censoring before the milestone time point. More recently, Ritchie *et al.* reported that the 6-month PFS rate correlated moderate strongly with the 12-month OS rate while ORR correlated poorly with 6-month PFS rate or 12-month OS rate in the phase II and III trials of checkpoint inhibitors in advanced solid tumors, indicating that 6-month PFS rate should be recommended in place of ORR as an end point in future phase 2 checkpoint-inhibitor trials (31). Other attempts that leveraged immune RECIST and immune-modified PFS as the surrogate endpoints for survival benefit showed promising effect but validation with large datasets and iterative modifications is still needed (32,33). Considering surrogate endpoints that would enable earlier assessments of treatment effects would be useful, continued investigation and optimization of study endpoints and their surrogates in cancer immunotherapy is strongly warranted.

Several limitations should be acknowledged. First, in the absence of individual patients' data, we cannot evaluate the within-trial differences in treatment effect size between OS and PFS. We thereby consider a conservative strategy by assessing the variance of the logHR without dependence. In fact, the variability of rHR remained unchanged as shown in the sensitivity analysis (Figure S1). Second, some subgroup analyses (e.g., histological types and specific anti-PD-1/PD-L1 inhibitors) should be cautiously interpreted due to the small number of eligible studies. Third, several more recent trials only reported the initial results on HRs of PFS or OS without details, hence we did not include them in this meta-analysis. 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.

Conclusions

In conclusion, the current evidence suggests that treatment effect sizes between PFS and OS might be roughly consistent in trials with different immunotherapeutic strategies in advanced NSCLC. 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.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (grant 81871865 to Dr. C Zhou, grant 81972167 to Dr. S Ren, grant 82003231 to Dr. Z Zhou). The Backbone Program of Shanghai Pulmonary Hospital (grant FKGG1802 to Dr. S Ren), Shanghai Pujiang Talent Plan (grant 2019PJD048 to Dr. S Ren), Shanghai Key disciplines of Respiratory (grant 2017ZZ02012 to Dr. C Zhou), Shanghai Major Diseases Multidisciplinary Cooperation Diagnosis and Treatment Construction Project, and the Shanghai Sailing Program (No. 20YF1407500).

Footnote

Reporting Checklist: The authors have completed the PRISMA reporting checklist. Available at <http://dx.doi.org/10.21037/tlcr-21-199>

Peer Review File: Available at <http://dx.doi.org/10.21037/tlcr-21-199>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tlcr-21-199>). CZ serves as an Editor-in-Chief of *Translational Lung Cancer Research* from Aug 2020 to Jul 2022. TJ serves as an unpaid editorial board member of *Translational Lung Cancer Research* from Feb 2018 to Jan

2021. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature* 2018;553:446-54.
2. Hanna NH, Schneider BJ, Temin S, et al. Therapy for Stage IV Non-Small-Cell Lung Cancer Without Driver Alterations: ASCO and OH (CCO) Joint Guideline Update. *J Clin Oncol* 2020;38:1608-32.
3. Champiat S, Ferrara R, Massard C, et al. Hyperprogressive disease: recognizing a novel pattern to improve patient management. *Nat Rev Clin Oncol* 2018;15:748-62.
4. Fuentes-Antrás J, Provencio M, Díaz-Rubio E. Hyperprogression as a distinct outcome after immunotherapy. *Cancer Treat Rev* 2018;70:16-21.
5. Guitoli G, Baldessari C, Bertolini F, et al. Are we ready to describe response or progression to immunotherapy in lung cancer? *Crit Rev Oncol Hematol* 2019;138:112-9.
6. Herbst RS, Garon EB, Kim DW, et al. Long-Term Outcomes and Retreatment Among Patients With Previously Treated, Programmed Death-Ligand 1 Positive, Advanced Non-Small-Cell Lung Cancer in the KEYNOTE-010 Study. *J Clin Oncol* 2020;38:1580-90.
7. Swami U, Monga V, Bossler AD, et al. Durable Clinical Benefit in Patients with Advanced Cutaneous Melanoma after Discontinuation of Anti-PD-1 Therapies Due to Immune-Related Adverse Events. *J Oncol* 2019;2019:1856594.

8. Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* 2013;39:1-10.
9. Hegde PS, Chen DS. Top 10 Challenges in Cancer Immunotherapy. *Immunity* 2020;52:17-35.
10. Mauguen A, Pignon JP, Burdett S, et al. Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: a re-analysis of meta-analyses of individual patients' data. *Lancet Oncol* 2013;14:619-26.
11. Shi Q, Schmitz N, Ou FS, et al. Progression-Free Survival as a Surrogate End Point for Overall Survival in First-Line Diffuse Large B-Cell Lymphoma: An Individual Patient-Level Analysis of Multiple Randomized Trials (SEAL). *J Clin Oncol* 2018;36:2593-602.
12. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst* 2009;101:1642-9.
13. Foster NR, Renfro LA, Schild SE, et al. Multitrial Evaluation of Progression-Free Survival as a Surrogate End Point for Overall Survival in First-Line Extensive-Stage Small-Cell Lung Cancer. *J Thorac Oncol* 2015;10:1099-106.
14. Gyawali B, Hey SP, Kesselheim AS. A Comparison of Response Patterns for Progression-Free Survival and Overall Survival Following Treatment for Cancer With PD-1 Inhibitors: A Meta-analysis of Correlation and Differences in Effect Sizes. *JAMA Netw Open* 2018;1:e180416.
15. Zhou C, Chen G, Huang Y, et al. A Randomized Phase 3 Study of Camrelizumab plus Chemotherapy as 1st Line Therapy for Advanced/Metastatic Non-Squamous Non-Small Cell Lung Cancer. *Journal of Thoracic Oncology* 2019;14:S215-S6.
16. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* 2016;375:1823-33.
17. Carbone DP, Reck M, Paz-Ares L, et al. First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. *N Engl J Med* 2017;376:2415-26.
18. Hoos A. Development of immuno-oncology drugs - from CTLA4 to PD1 to the next generations. *Nat Rev Drug Discov* 2016;15:235-47.
19. Rothschilds AM, Wittrup KD. What, Why, Where, and When: Bringing Timing to Immuno-Oncology. *Trends Immunol* 2019;40:12-21.
20. Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? *J Clin Oncol* 2012;30:1030-3.
21. Kim C, Prasad V. Cancer Drugs Approved on the Basis of a Surrogate End Point and Subsequent Overall Survival: An Analysis of 5 Years of US Food and Drug Administration Approvals. *JAMA Intern Med* 2015;175:1992-4.
22. Zhao F. Surrogate End Points and Their Validation in Oncology Clinical Trials. *J Clin Oncol* 2016;34:1436-7.
23. Blumenthal GM, Zhang L, Zhang H, et al. Milestone Analyses of Immune Checkpoint Inhibitors, Targeted Therapy, and Conventional Therapy in Metastatic Non-Small Cell Lung Cancer Trials: A Meta-analysis. *JAMA Oncol* 2017;3:e171029.
24. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0)*. The Cochrane Collaboration 2011. Available online: <https://handbook-5-1.cochrane.org/>
25. Sterne JA, Juni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;21:1513-24.
26. Tan A, Porcher R, Crequit P, et al. Differences in Treatment Effect Size Between Overall Survival and Progression-Free Survival in Immunotherapy Trials: A Meta-Epidemiologic Study of Trials With Results Posted at ClinicalTrials.gov. *J Clin Oncol* 2017;35:1686-94.
27. Ciani O, Buyse M, Garside R, et al. Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer. *J Clin Epidemiol* 2015;68:833-42.
28. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med* 2015;373:123-35.
29. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* 2015;373:1627-39.
30. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet* 2017;389:255-65.
31. Ritchie G, Gasper H, Man J, et al. Defining the Most Appropriate Primary End Point in Phase 2 Trials of Immune Checkpoint Inhibitors for Advanced Solid Cancers: A Systematic Review and Meta-analysis. *JAMA Oncol* 2018;4:522-8.
32. Seymour L, Bogaerts J, Perrone A, et al. iRECIST:

- guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol* 2017;18:e143-52.
33. Hodi FS, Ballinger M, Lyons B, et al. Immune-Modified Response Evaluation Criteria In Solid Tumors

(imRECIST): Refining Guidelines to Assess the Clinical Benefit of Cancer Immunotherapy. *J Clin Oncol* 2018;36:850-8.

Cite this article as: Zhou Z, Ren S, Chen L, Zhou C, Jiang T. Differences in treatment effect size between progression-free survival and overall survival in anti-PD-1/PD-L1 inhibitors-based trials in advanced NSCLC: a systematic review and meta-analysis. *Transl Lung Cancer Res* 2021;10(6):2562-2572. doi: 10.21037/tlcr-21-199

Search strategies

Search date was from the inception through March 1st, 2020.

PubMed search strategy

((“Nivolumab”[Substance] OR “Nivolumab”[All Fields] OR “Opdivo”[All Fields] OR “BMS-936558”[All Fields]) OR (“Pembrolizumab”[Substance] OR “Pembrolizumab”[All Fields] OR “lambrolizumab”[All Fields] OR “keytruda”[All Fields] OR “MK-3475”[All Fields]) OR (“Atezolizumab”[Substance] OR “Atezolizumab”[All Fields] OR “Tecentriq”[All Fields] OR “MPDL3280A”[All Fields]) OR (“Durvalumab”[Substance] OR “Durvalumab”[All Fields]) OR (“Avelumab”[Substance] OR “Avelumab”[All Fields]) OR (“Camrelizumab”[All Fields] OR “SHR-1210”[All Fields]) OR (“Sintilimab”[All Fields] OR “IBI308”[All Fields]) OR (“Toripalimab”[All Fields] OR “JS001”[All Fields]) OR (“PD-1”[All Fields] OR “PD-L1”[All Fields])) AND (“lung neoplasms”[Mesh] OR “lung cancer”[All Fields] OR “lung tumor”[Title/Abstract]) AND (“randomized”[All Fields])

EMBASE search strategy

(‘Nivolumab’ OR ‘Opdivo’ OR ‘BMS-936558’ OR ‘Nivo’ OR ‘Pembrolizumab’ OR ‘lambrolizumab’ OR ‘keytruda’ OR ‘SCH 900475’ OR ‘MK-3475’ OR ‘Atezolizumab’ OR ‘MSB0010718C’ OR ‘Tecentriq’ OR ‘RO5541267’ OR ‘RG7446’ OR ‘MPDL3280A’ OR ‘Durvalumab’ OR ‘MEDI-4736’ OR ‘MEDI4736’ OR ‘Avelumab’ OR ‘Camrelizumab’ OR ‘SHR-1210’ OR ‘Sintilimab’ OR ‘IBI308’ OR ‘Toripalimab’ OR ‘JS001’ OR ‘checkpoint inhibit’ OR ‘PD-1’ OR ‘PD-L1’) AND (‘lung neoplasms’ OR ‘lung carcinoma’ OR ‘lung cancer’ OR ‘lung tum’) AND (‘randomized’)

Cochrane database search strategy

(“Nivolumab” OR “Pembrolizumab” OR “Atezolizumab” OR “Avelumab” OR “Durvalumab” OR “Camrelizumab” OR “Sintilimab” OR “Toripalimab” OR “PD-1” OR “PD-L1”) AND (“lung neoplasms” OR “lung cancer” OR “lung tumor”) AND (“randomized”)

Table S1 The methodologic quality assessment for each included study

Trial ID	Random sequence generation	Allocation concealment	Blinding of participants	Blinding of outcome assessment	Incomplete outcome data	Selective outcome reporting	Other potential threats to validity
ARCTIC	Low	Low	Low	Low	Low	Low	Low
CAMEL	Low	Low	Low	Low	Low	Low	Low
CheckMate 017	Low	Low	Low	Low	Low	Low	Low
CheckMate 026	Low	Low	Low	Low	Low	Low	Low
CheckMate 057	Low	Low	Low	Low	Low	Low	Low
CheckMate 227	Low	Low	Low	Low	Low	Low	Low
CheckMate 9LA	Low	Low	Low	Low	Low	Low	Low
EMPOWER-Lung 1	Low	Low	Low	Low	Low	Low	Unknown
IMpower110	Low	Low	Low	Low	Low	Low	Low
IMpower130	Low	Low	Low	Low	Low	Low	Low
IMpower131	Low	Low	Low	Low	Low	Low	Low
IMpower132	Low	Low	Low	Low	Low	Low	Low
IMpower150	Low	Low	Low	Low	Low	Low	Low
JAVELIN Lung 200	Low	Low	Low	Low	Low	Low	Low
KEYNOTE-010	Low	Low	Low	Low	Low	Low	Low
KEYNOTE-021G	Low	Low	Low	Low	Low	Low	Low
KEYNOTE-024	Low	Low	Low	Low	Low	Low	Low
KEYNOTE-033	Low	Unknown	Low	Low	Low	Low	Unknown
KEYNOTE-042	Low	Low	Low	Low	Low	Low	Low
KEYNOTE-189	Low	Low	Low	Low	Low	Low	Low
KEYNOTE-407	Low	Low	Low	Low	Low	Low	Low
MYSTIC	Low	Unknown	Low	Low	Low	Low	Unknown
NA2020	Low	Unknown	Low	Low	Low	Low	Unknown
OAK	Low	Low	Low	Low	Low	Low	Low
ORIENT-11	Low	Low	Low	Low	Low	Low	Low
ORIENT-12	Low	Unknown	Low	Low	Low	Low	Unknown
POPLAR	Low	Low	Low	Low	Low	Low	Low

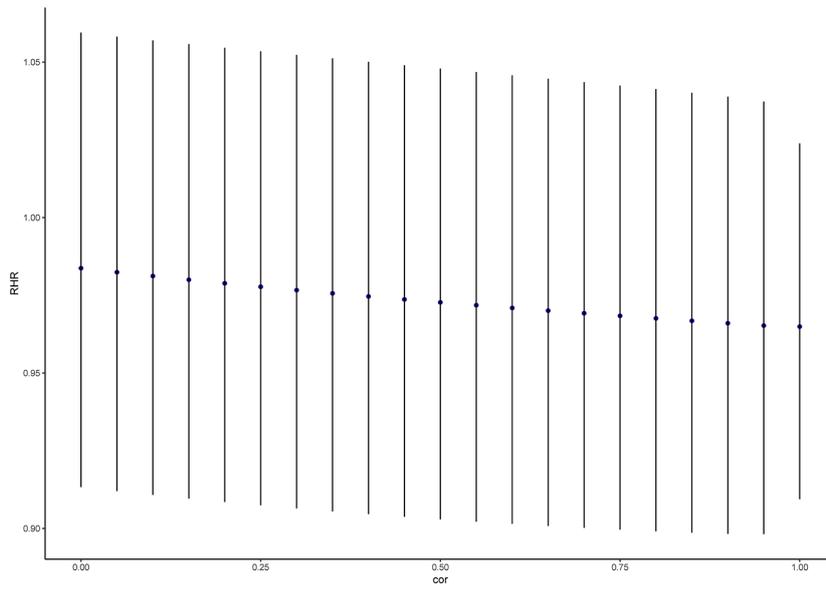


Figure S1 Effect of dependency between OS and PFS. A ratio of hazard ratio (rHR) <1.0 indicates greater treatment effect size for PFS than for OS. A ρ of 0 indicates no dependency and a ρ of 1.0 indicates complete dependency. With increasing ρ , the variability of within-trial rHRs decreased, which resulted in increased between-trials heterogeneity.

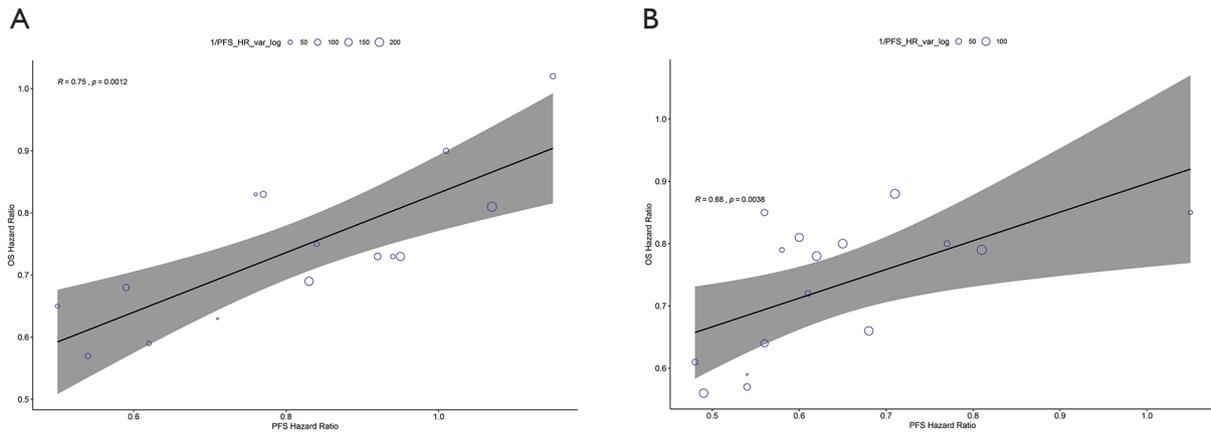


Figure S2 Correlation between OS and PFS according to the treatment patterns. (A) Monotherapy group; (B) combination therapy group. Each circle represents a trial. The size of the circle is proportional to the inverse of the variance. The lines represent 95% prediction intervals for a trial with a median weight.

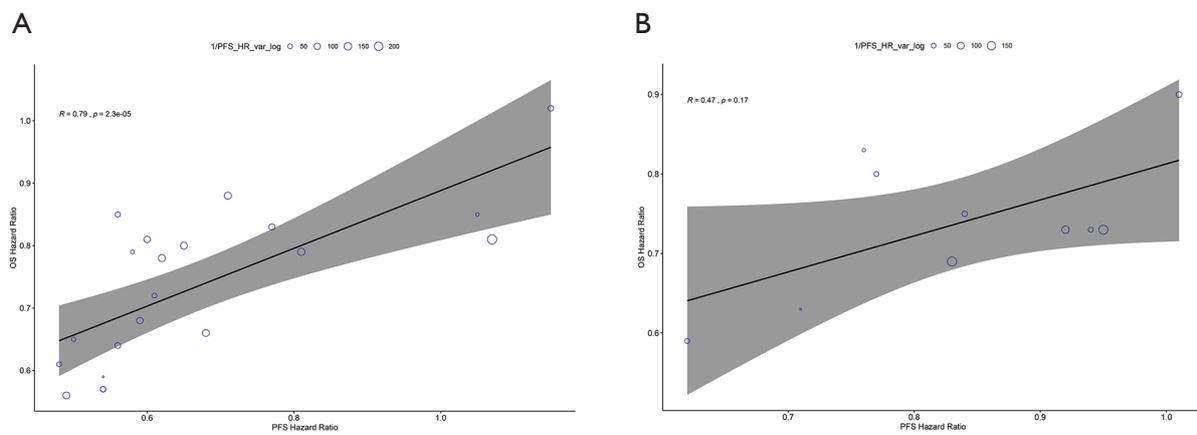


Figure S3 Subgroup analysis of correlation between PFS and OS according to lines of treatment. (A) First-line treatment; (B) second or above-line treatment. Each circle represents a trial. The size of the circle is proportional to the inverse of the variance. The lines represent 95% prediction intervals for a trial with a median weight.