Original Article



# Assessing surrogacy using restricted mean survival time ratio for overall survival in non-small cell lung cancer immunotherapy studies

**Herbert Pang[1,2*], Guangyu Yang[3], James C. Ho[4], Tiffany H. Leung[4], Qian Shi[5], Chen Hu[6], Thomas E. Stinchcombe[7], Xiaofei Wang[2]**

[1]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; [2]Department of Biostatistics and Bioinformatics, Duke University of Medicine, Durham, NC, USA; [3]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA; [4]Department of Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; [5]Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA; [6]Johns Hopkins University School of Medicine, Baltimore, MD, USA; [7]Duke Cancer Institute, Division of Medical Oncology, Durham, NC, USA

*Correspondence to:* Xiaofei Wang, PhD. Duke University Medical Center 2424 Erwin Road, Suite 1102 Hock Plaza Box 2721 Durham, NC 27710, USA. Email: xiaofei.wang@duke.edu.

**Background: Background:** Proportional hazards (PH) assumption is often violated in cancer immunotherapy studies. Restricted mean survival time (RMST) ratio is a valid metric to quantify the size of treatment effect when non-proportional hazard (NPH) is present. This study investigated the use of RMST ratio and hazard ratio (HR) in studying progression-free survival (PFS) as a surrogate endpoint for overall survival (OS) in non-small cell lung cancer immunotherapy trials.

**Methods:** Trial level data were collected from 14 phase III trials published between 2012 and 2018. A weighted least-square regression (WLSR) was performed to evaluate the trial-level surrogacy. Surrogacy was evaluated via the association between RMST ratios for PFS and OS and between HRs for PFS and OS.

**Results:** Using data extracted from published articles, low to moderate correlation (0.49) between PFS and OS was observed for HR while low correlation (0.35) was observed for RMST ratio. When trials violating PH in PFS were included, more consistent correlations for both HR (0.43) and RMST ratio (0.44) were observed.

**Conclusions:** In summary, the strength of PFS surrogacy for OS depends on whether HR or RMST ratio are chosen. RMST ratio and additional sensitivity analysis should be considered in addition to HR.

**Keywords:** Immunotherapy; lung cancer; progression-free survival (PFS); restricted mean survival time (RMST); surrogate endpoint

Submitted Aug 05, 2021. Accepted for publication Jan 31, 2022.
doi: 10.21037/cco-21-110
**View this article at:** https://dx.doi.org/10.21037/cco-21-110

## Introduction

Delayed treatment and long-term survival effects are well documented in immune check point inhibitor (ICI) trials, in which survival is used as treatment measure (1,2). With the delayed treatment effect, the proportional hazards (PH) assumption is violated and standard log-rank test is less powerful (3). A related question is whether non-proportional hazards (NPH) would impact how progression-free survival (PFS) is used as a surrogate endpoint for overall survival (OS). OS is regarded as the gold standard endpoint for evaluating the effect of new therapy in cancer. However,

* Current affiliation: Genentech, Inc., South San Francisco, CA, USA.

compared to PFS, it requires a large number of patients and longer follow-up time to accumulate the number of events required for adequate statistical power. As more cancer therapies are available, many patients are receiving therapies after the completion of study therapy, which may confound the assessment of OS. In addition, with the complexity and high cost of developing and demonstrating clinical efficacy and tolerable toxicity, the approval for cancer drugs has been slow. This has created a call for alternatives to detect signals based on surrogate endpoints for making decisions earlier and thus a potentially faster approval (4). Because of the above, there is a great need for cancer research to identify and validate surrogate endpoints for cancer clinical trials that can accurately predict the treatment effect. With the availability of a number of randomized ICI trials in advanced non-small cell lung cancer (NSCLC) and given that the PH assumption is commonly violated, it is the optimal time to investigate whether this assumption would influence the value of PFS as surrogate endpoint for OS when treatment effect is quantified by hazard ratio (HR) versus restricted mean survival time (RMST) ratio. RMST, corresponds to the area under the Kaplan-Meier curve up to a chosen time (5).

## Methods

Study level information such as study population, stage, histology, pre-treated (yes/no), immunotherapy and non-immunotherapy arms information, and primary endpoint, were extracted. The outcomes of interest to this study include OS and PFS. OS is defined as the date of randomization to the date of death due to all causes and subjects were censored at last follow-up. PFS is defined as the time from randomization to progression or death. Patients alive who had not experienced progression were censored at the last disease assessment. Surrogate measure of PFS on OS based on HR versus RMST ratio is of primary interest. In order to perform the RMST analysis, survival times were reconstructed from Kaplan-Meier curves for each treatment arm in published paper using the method of Guyot *et al.* (6). Software named "Digitizelt" (http://www.digitizeit.de/) was be used to detect the time, censoring, and survival probability. The rationale for choosing RMST is because ratios obtained from models can also be difficult to interpret when the modeling assumption is violated and RMST is model-free (5). We studied the association between PFS and OS at trial-level. A weighted least-square regression (WLSR) for log $RMST_{OS}$ and log $RMST_{PFS}$ was performed to evaluate the trial-level surrogacy, with weights equal to the sample size of the trial. Similarly, log $HR_{OS}$ and log $HR_{PFS}$ was also assessed and WLSR was fitted. To explore the impact of NPH, we also compared the "all trials" analysis with one that excludes the trials that violate NPH test for OS. The WLS R were calculated for both HR and RMST ratio for both groupings of trials. R 4.0.2 (Vienna, Austria) was the software used for statistical analysis.

## Results

A literature search was conducted using PubMed to identify phase II and phase III immunotherapy lung cancer studies published between January 2012 and October 2018. After examining the 247 initially found articles, 171 articles were excluded since they were not original articles. Among the 76 articles, another 62 articles were excluded with the reasons of exclusion being provided in Figure S1. In the end, 14 articles were eligible for further analysis (Figure S1). *Table 1* summarizes the HR and RMST ratios for PFS and OS. For PFS HR, 8 out of 16 trials showed strong PH violation with NPH test P value ≤0.01 and 4 out of 16 trials showed PH violation with NPH test P value ≤0.05. For the OS HR, 2 trials had strong PH violation with NPH test P value ≤0.01 with 1 trial had PH violation with NPH test P value ≤0.05. The study level variables, including treatment arm, control arm, histology, stage, pre-treated (yes/no), primary endpoint and primary population, are given in Table S1.

*Figure 1* illustrates the weighted least square regression line and R between OS and PFS for RMST ratio and HR with the color representing different trials and the size of the dots representing the numbers of participants in each trial. The WLS R between OS and PFS for HR and RMST ratio are 0.49 and 0.35, respectively (*Figure 1A*,*1B*). We also investigated the WLS R between OS and PFS by removing three studies with OS curves that violated NPH assumptions. In the updated curves, the WLS R between OS and PFS for RMST ratio and HR are 0.44 and 0.43, respectively (*Figure 1C*,*1D*).

NPH issue in cancer ICI trials is well-known and RMST has been proposed to deal with this issue (7). In this study, we found that PFS have low to moderate correlation with OS using HRs. However, the correlation is low between PFS and OS when RMST ratio is used. When only trials that violate the PH assumption in PFS curves were included, the results using HRs and RMST

**Table 1** Summary of HR and RMST ratio for PFS and OS

| Trial | PFS HR (95% CI) | PFS RMST ratio | OS HR (95% CI) | OS RMST ratio |
|---|---|---|---|---|
| Antonia *et al.*, 2017 | 0.55 (0.45, 0.67) | 1.52 (1.31, 1.77) | 0.66 (0.52, 0.84) | 1.13 (1.05, 1.21) |
| Barlesi *et al.*, 2018 | 1.05 (0.83, 1.32)** | 1.07 (0.89, 1.27) | 0.90 (0.73, 1.12) | 1.07 (0.93, 1.22) |
| Borghaei *et al.*, 2015 | 0.91 (0.76, 1.09)** | 1.23 (1.04, 1.47) | 0.75 (0.62, 0.91)** | 1.16 (1.03, 1.30) |
| Brahmer *et al.*, 2015 | 0.63 (0.48, 0.82)** | 1.65 (1.30, 2.09) | 0.59 (0.44, 0.78) | 1.43 (1.19, 1.72) |
| Carbone *et al.*, 2017 | 1.19 (0.97, 1.45)** | 0.92 (0.78, 1.09) | 1.08 (0.87, 1.35) | 0.95 (0.85, 1.06) |
| Fehrenbacher *et al.*, 2016 | 0.94 (0.73, 1.21)* | 1.07 (0.85, 1.35) | 0.73 (0.54, 0.98) | 1.13 (0.99, 1.30) |
| Gandhi *et al.*, 2018 | 0.53 (0.43, 0.64) | 1.49 (1.31, 1.71) | 0.50 (0.39, 0.65) | 1.26 (1.14, 1.38) |
| Govindan *et al.*, 2017 | 0.92 (0.78, 1.08)* | 1.07 (0.94, 1.22) | 0.91 (0.77, 1.07) ** | 1.07 (0.96, 1.19) |
| Herbst *et al.*, 2016, 10 mg | 0.79 (0.67, 0.94) ** | 1.29 (1.12, 1.50) | 0.63 (0.51, 0.78) * | 1.29 (1.15, 1.44) |
| Herbst *et al.*, 2016, 2 mg | 0.87 (0.74, 1.04)** | 1.18 (1.01, 1.36) | 0.73 (0.59, 0.89) | 1.20 (1.07, 1.34) |
| Langer *et al.*, 2016 | 0.54 (0.32, 0.92) | 1.30 (1.04, 1.63) | 0.95 (0.44, 2.01) | 1.00 (0.87, 1.16) |
| Lynch *et al.*, 2012, A | 0.68 (0.47, 0.98) | 1.23 (1.00, 1.53) | 0.86 (0.58, 1.27) | 1.12 (0.89, 1.41) |
| Lynch *et al.*, 2012, B | 0.87 (0.6, 1.25)* | 1.07 (0.84, 1.35) | 0.98 (0.67, 1.45) | 1.00 (0.78, 1.28) |
| Reck *et al.*, 2016 | 0.49 (0.36, 0.65)** | 1.53 (1.30, 1.81) | 0.61 (0.42, 0.9) | 1.17 (1.03, 1.32) |
| Rittmeyer *et al.*, 2017 | 0.94 (0.81, 1.08)** | 1.14 (0.99, 1.32) | 0.73 (0.62, 0.86) | 1.20 (1.09, 1.33) |
| Socinski *et al.*, 2018 | 0.61 (0.51, 0.73)* | 1.40 (1.25, 1.56) | 0.78 (0.64, 0.96) | 1.10 (1.00, 1.21) |

**, strong PH violation with NPH test P value ≤0.01; *, PH violation with NPH test P value ≤0.05. HR, hazard ratio; RMST, restricted mean survival time; PFS, progression-free survival; OS, overall survival; PH, proportional hazards; NPH, non-proportional hazards.

ratios results were closer. Therefore, the presence of NPH may affect the concordance between PFS and OS. Wang *et al.* (8) studied a similar topic and concluded that milestone RMST may serve well as a surrogate endpoint for OS HR in multiple cancers. In a more recent article by Kok *et al.* (9) with a different set of studies, their results suggest a slightly different conclusion that 6-month PFS could reliably estimate 12-month OS. One of the strengths of our study is that we consider surrogacy in NSCLC rather than across different cancer types which can make the results difficult to interpret. Moreover, we also used non-milestone RMST and PFS for the analysis which makes full use of the survival curve unlike the above two mentioned studies. This study is not without limitations. Ideally, it would be good to conduct surrogacy analysis at individual data level as well. However, it has been noted previously that trial-level surrogacy analysis produces decent results with sufficient number of trials (e.g., N>10) available (10).

In our study, we mainly focused on trial level surrogacy. To be comprehensive, future work should consider looking at individual level surrogacy as well. For both individual level and trial level surrogacy investigation, a bivariate model for RMST Ratio like the one proposed for HR can be considered (11). Even with surrogacy validation at both levels, we should be cautious in using surrogate endpoints to replace treatment estimation based on gold standard endpoint like OS.

## Summary and conclusion

Our results highlight the potential problems with using traditional analytics alone for surrogacy investigation in presence of NPH in cancer ICI trials. Based on the above, researchers are encouraged to consider other measures such as RMST ratio for studying surrogate endpoints and conduct additional analysis to understand the impact of trials that violate the PH assumption. As subject level surrogacy analysis can complement trial level analysis, further research on the concordance between RMST ratio and HR should involve subject level surrogacy analysis if individual patient data is available.

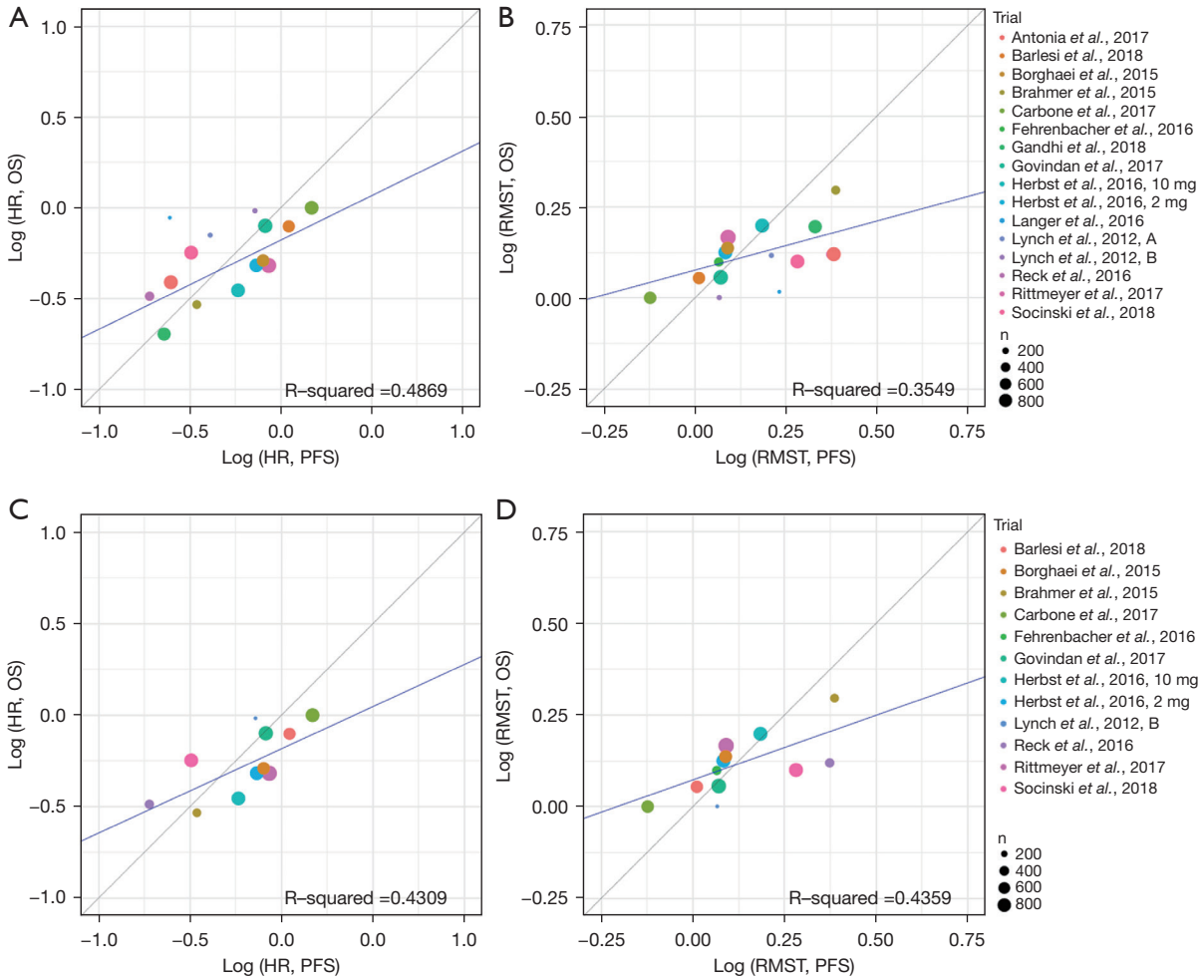


**Figure 1** Weighted least square regression line and $R^2$ between OS and PFS for RMST ratio and HR. (A) HR—all studies; (B) RMST ratio—all studies; (C) HR—studies with NPH OS removed; (D) RMST ratio—studies with NPH OS removed. HR, hazard ratio; OS, overall survival; PFS, progression-free survival; RMST, restricted mean survival time; NPH, non-proportional hazards.

## Acknowledgments

We would like to thank Liyuan Fan for her help with some of the data extraction.

*Funding:* The research work was partially supported by NIH P01CA142538 (Wang) and NIH R01AG066883 (Wang), HMRF grant of Hong Kong 16172901 (Pang, Ho), University Postgraduate Fellowships of HKU Foundation (Leung) and Postgraduate scholarship of the University of Hong Kong (Leung).

## Footnote

*Peer Review File:* https://cco.amegroups.com/article/view/10.21037/cco-21-110/prf

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at https://cco.amegroups.com/article/view/10.21037/cco-21-110/coif). HP reports HMRF grant of Hong Kong 16172901, an NIHU01 grant from FDA, stock options from Roche, and personal fees from Genentech, outside the submitted work. TL reports University Postgraduate Fellowships of HKU Foundation. QS reports consulting/advisory role from Yiviva Inc., Boehringer Ingelheim Pharmaceuticals, Inc., Regeneron Pharmaceuticals, Inc., Hoosier Cancer Research Network (to QS), Honorarium/speaker role from Chugai Pharmaceutical Co., Ltd., stocks from Johnson & Johnson,

Amgen, and Merck & CO. (to QS), research funds from Celgene/BMS, Roche/Genentech, Janssen, Novartis. CH reports support from NCI/NIH (U10-CA180822), grant from RTOG Foundation, and consulting fees from Merck & Co. and D1Med Technology Co. TES reports receiving grants or contracts from Genentech/Roche (Institution), AstraZeneca (Institution), Takeda (Institution), Advaxis (Institution), Regeneron (Institution), and Mirati (Institution); and participation on a Data Safety Monitoring Board or Advisory Board of Takeda, AstraZeneca, Genentech/Roche, Foundation Medicine, Pfizer, EMD Serono, Novartis, Daiichi Sankyo¸ Lilly, Medtronic, Puma Biotechnology, Janssen Oncology, Regeneron, Turning Point Therapeutics, Sanofi/Aventis. XW serves as an unpaid editorial board member of *Chinese Clinical Oncology*. 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.



## References

1. Hoos A. Evolution of end points for cancer immunotherapy trials. Ann Oncol 2012;23 Suppl 8:viii47-52.
2. Chen TT. Statistical issues and challenges in immuno-oncology. J Immunother Cancer 2013;1:18.
3. Fine GD. Consequences of delayed treatment effects on analysis of time-to-event endpoints. Drug Inf J 2007;41:535-9.
4. Lassere MN, Johnson KR, Boers M, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. J Rheumatol 2007;34:607-15.
5. Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. J Clin Oncol 2014;32:2380-5.
6. Guyot P, Ades AE, Ouwens MJ, et al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 2012;12:9.
7. Alexander BM, Schoenfeld JD, Trippa L. Hazards of Hazard Ratios - Deviations from Model Assumptions in Immunotherapy. N Engl J Med 2018;378:1158-9.
8. Wang ZX, Wu HX, Xie L, et al. Correlation of Milestone Restricted Mean Survival Time Ratio With Overall Survival Hazard Ratio in Randomized Clinical Trials of Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. JAMA Netw Open 2019;2:e193433.
9. Kok PS, Cho D, Yoon WH, et al. Validation of Progression-Free Survival Rate at 6 Months and Objective Response for Estimating Overall Survival in Immune Checkpoint Inhibitor Trials: A Systematic Review and Meta-analysis. JAMA Netw Open 2020;3:e2011809.
10. Renfro LA, Shi Q, Xue Y, et al. Center-Within-Trial Versus Trial-Level Evaluation of Surrogate Endpoints. Comput Stat Data Anal 2014;78:1-20.
11. Buyse M, Burzykowski T, Michiels S, et al. Individual- and trial-level surrogacy in colorectal cancer. Stat Methods Med Res 2008;17:467-75.

**Cite this article as:** Pang H, Yang G, Ho JC, Leung TH, Shi Q, Hu C, Stinchcombe TE, Wang X. Assessing surrogacy using restricted mean survival time ratio for overall survival in non-small cell lung cancer immunotherapy studies. Chin Clin Oncol 2022;11(1):7. doi: 10.21037/cco-21-110

## Supplementary

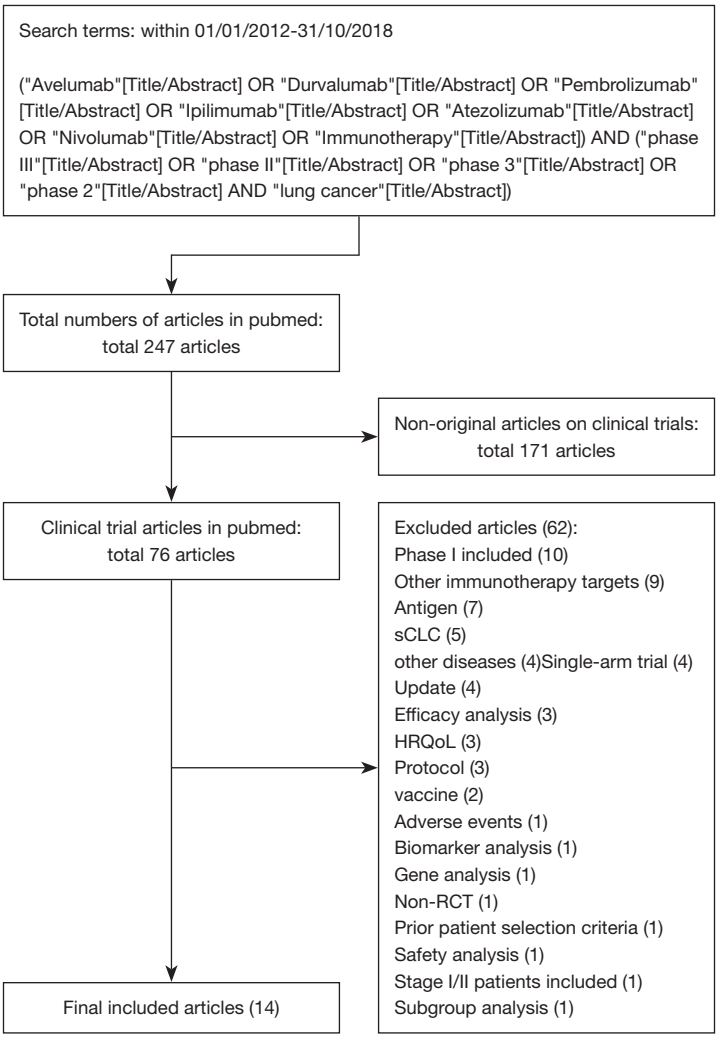


**Figure S1** Study selection flow chart. SCLC, small cell lung cancer; HRQoL, health-related quality of life; RCT, randomised controlled trial.


https://dx.doi.org/10.21037/cco-21-110

**Table S1** Summary of 14 published randomized IM trials (16 treatment comparisons) in stage III/IV NSCLC

| Trial | Treatment | Control | Histology | Stage | Pre treated | Primary endpoints | Primary population |
|---|---|---|---|---|---|---|---|
| Antonia *et al.*, 2017 | durvalumab | placebo | NSCLC | 3 | Y | OS + PFS | All comers |
| Barlesi *et al.*, 2018 | avelumab | docetaxel | NSCLC | 3b/4/r | Y | OS | PD-L1≥1% |
| Borghaei *et al.*, 2015 | nivolumab | docetaxel | Nonsquamous | 3b/4/r | Y | OS | All comers |
| Brahmer *et al.*, 2015 | nivolumab | docetaxel | Squamous | 3b/4 | Y | OS | All comers |
| Carbone *et al.*, 2017 | nivolumab | plat-based chemo | NSCLC | 4/r | N | PFS | PD-L1≥5% |
| Fehrenbacher *et al.*, 2016 | atezolizumab | docetaxel | NSCLC | 3b/4/r | Y | OS | All comers |
| Gandhi *et al.*, 2018 | Pembrolizumab + plat-based chemo | Placebo + plat-based chemo | Nonsquamous | 4 | N | OS + PFS | All comers |
| Govindan *et al.*, 2017 | Ipilimumab + CP | Placebo + CP | Squamous | 4/r | N | OS | All comers |
| Herbst *et al.*, 2016, 10 mg | pembrolizumab 10 mg/kg | docetaxel | NSCLC | 3b/4/r | Y | OS + PFS | PD-L1≥1% |
| Herbst *et al.*, 2016, 2 mg | pembrolizumab 2 mg/kg | docetaxel | NSCLC | 3b/4/r | Y | OS + PFS | PD-L1≥1% |
| Langer *et al.*, 2016 | Pembrolizumab + car + peme | Car + peme | Nonsquamous | 3b/4 | N | OR | All comers |
| Lynch *et al.*, 2012, A | phased ipilimumab + CP | Placebo + CP | NSCLC | 3b/4 | N | irPFS | All comers |
| Lynch *et al.*, 2012, B | concurrent ipilimumab + CP | Placebo + CP | NSCLC | 3b/4 | N | irPFS | All comers |
| Reck *et al.*, 2016 | pembrolizumab | plat-based chemo | NSCLC | 4 | N | PFS | PD-L1≥50% |
| Rittmeyer *et al.*, 2017 | atezolizumab | docetaxel | NSCLC | 3b/4 | Y | OS | All comers |
| Socinski *et al.*, 2018 | Atezolizumab + CP | Bevacizumab + CP | Nonsquamous | 4/r | Y + N | OS + PFS | All comers |

NSCLC, non-small cell lung cancer; Pre, Previously; OS, overall survival; PFS, progression-free survival; Ep, end point; Popu, population; Pub, publish; pac, paclitaxel; car, carboplatin; peme, pemetrexed; plat, platinum; chemo, chemotherapy; nons, nonsquamous; CP, carboplatin + paclitaxel.


https://dx.doi.org/10.21037/cco-21-110