Ther Adv Med Oncol
2020, Vol. 12: 1–10
DOI: 10.1177/1758835920929583

Surrogate endpoints for overall survival in anti-programmed death-1 and anti-programmed death ligand 1 trials of advanced melanoma

Run-Cong Nie#, Shu-Qiang Yuan#, Yun Wang#, Xue-Bin Zou#, Shi Chen#, Shu-Man Li, Jin-Ling Duan, Jie Zhou, Guo-Ming Chen, Tian-Qi Luo, Zhi-Wei Zhou* and Yuan-Fang Li*

Abstract

Background: We assessed the surrogacy of objective response rate (ORR), disease control rate (DCR) and progression-free survival (PFS) for overall survival (OS) in anti-PD-1/PD-L1 trials of metastatic melanoma through a meta-analysis of randomized controlled trials (RCTs).

Methods: PubMed and EMBASE were searched for phase II/III RCTs till June 2019 investigating anti-PD-1/PD-L1 agents. Treatment effect (hazard ratio or odds ratio) on potential surrogates (ORR/DCR/PFS) and OS were collected. At trial level, we assessed the correlation between treatment effect on potential surrogates and OS, weighted by sample size, fixed and random effect models, and calculated the surrogate threshold effect (STE). Sensitivity analyses and leave-one-out cross-validation approach were performed to evaluate the robustness of our findings.

Results: We included 8 RCTs (4110 patients; 11 comparisons). We did not identify strong correlations between ORR [coefficient of determination (R^2): 0.09–0.25], DCR (0.41–0.57) and OS. However, we noted a strong correlation between PFS and OS, with R^2 of 0.82 in sample size, 0.75 in fixed effect and 0.72 in random effect model weighting, the robustness of which was further verified by leave-one-out cross-validation approach. Sensitivity analyses with restriction to trials with less than 50% crossover, phase III trials, large trials and first-line trials strengthened the correlation (0.78–0.94). The STE for PFS was 0.78.

Conclusions: PFS may be the appropriate surrogate for OS in anti-PD-1/PD-L1 trials of metastatic melanoma. A future anti-PD-1/PD-L1 trial would need less than 0.78 for PFS of the upper limit of confidence interval to predict an OS benefit.

Keywords: immune checkpoint, overall survival, PD-1, PD-L1, surrogate endpoint

Received: 18 October 2019; revised manuscript accepted: 28 April 2020.

Correspondence to:
Yuan-Fang Li
Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, No. 651 Dongfeng Eastern Road, Guangzhou, Guangdong, 510060, China
liyuanf@sysucc.org.cn

Zhi-Wei Zhou
Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, No. 651 Dongfeng Eastern Road, Guangzhou, Guangdong, 510060, China
zhouzhw@sysucc.org.cn

Run-Cong Nie
Shu-Qiang Yuan
Guo-Ming Chen
Tian-Qi Luo
Department of Gastric Surgery & Melanoma Surgical Section, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Yun Wang
Department of Hematologic Oncology, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Xue-Bin Zou
Department of Ultrasound, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Introduction

In the past 10 years, an unprecedented revolution in the treatment landscape for metastatic melanoma has yielded continuously improving survival outcomes for these patients.[1] Before 2011, metastatic melanoma was associated with devastating outcomes, with a median overall survival (OS) of approximately 9 months and 3-year OS of approximately 12%.[2] However, the identification of negative immune checkpoints [cytotoxic T-lymphocyte associated protein 4 (CTLA-4) or programmed cell death 1 (PD-1)][3,4] has tremendously changed the standards of clinical performance for therapies for this disease. Since the regulatory approval of a CTLA-4 inhibitor (ipilimumab) by the US Food and Drug Administration (FDA), 3-year OS of treated patients has increased to 30%.[2,5] Currently, other newer agents that

Shi Chen
Department of Gastric Surgery, The 6th Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

Shu-Man Li
Jin-Ling Duan
Jie Zhou
Department of Experimental Research (Cancer Institute), Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

#These authors contributed equally to this study.

*These authors are co-senior authors.

block the binding of PD-1 to its ligand, programmed death ligand 1 (PD-L1), within the cancer microenvironment have attracted more interest from oncology researchers. Notably, several clinical trials demonstrated that, compared with ipilimumab monotherapy alone, anti-PD-1/PD-L1 therapy with or without ipilimumab further improved the survival of patients with metastatic melanoma, leading to 4-year OS of 53% and 46%, respectively.[5–7]

The rapid advances in anti-PD-1/PD-L1 therapy for melanoma has spurred researchers and physicians to explore more effective therapies to further extend the clinical benefit; however, a critical issue that is still under investigation is what is the optimal endpoint and how should tumour response be evaluated in anti-PD-1/PD-L1 trials for metastatic melanoma. In conventional randomized clinical trials of melanoma, OS is considered the gold standard for the endpoint because it is simple to measure, easy to interpret, and unbiased. However, of the use of OS requires prolonged follow-up durations and larger sample sizes to detect statistically significant differences, consideration of the effect of subsequent therapies after progression that might prolong survival, and the risk of noncancer deaths. Therefore, reliable endpoints that could be used as surrogates for OS in metastatic melanoma could shorten the follow-up period and reduce the cost of drug development. A previous meta-analysis reported that progression-free survival (PFS) could be considered a valid surrogate for OS in dacarbazine-controlled randomized trials of metastatic melanoma.[8] Nonetheless, in the era of immunotherapy, PD-1/PD-L1 inhibitors rather than dacarbazine are assigned as the control arm, and the mechanisms of action of anti-PD-1/PD-L1 agents are markedly distinct from those of cytotoxic agents; there is delayed antitumour activity,[9] pseudoprogression[10] and hyperprogressive disease[11] during anti-PD-1/PD-L1 therapy. Therefore, it is still uncertain whether PFS or other Response Evaluation Criteria in Solid Tumors (RECIST) criteria-defined endpoints [including objective response rate (ORR) and disease control rate (DCR)] can sufficiently reflect the antitumour effect of these drugs in melanoma.

Based on this premise, we performed this meta-analysis to assess the correlation between PFS, ORR, DCR and OS in trials of anti-PD-1/PD-L1 drugs for metastatic melanoma.

Methods

Search strategy and selection criteria

In June 2019, we systematically searched the Medline (PubMed), Embase, ClinicalTrials.gov and Cochrane Library databases. We also manually searched the references of the included trials and abstracts of two conference proceedings [the 2019 American Society of Clinical Oncology (ASCO) annual meeting and the European Society for Medical Oncology (ESMO) 2018 congress] to retrieve additional studies. We searched for the following concepts and linked them together with the AND operator: 'nivolumab', 'pembrolizumab', 'avelumab', 'atezolizumab', 'durvalumab', 'PD-1', 'PD-L1', 'checkpoint inhibitors', 'melanoma' and 'randomized controlled trial' (Box 1, Supplemental materials).

We included phase II or phase III trials of unresectable, advanced or recurrent melanoma that used PD-1/PD-L1 inhibitors in the experimental arm and any therapy in the control arm. We required trials to report the hazard ratios (HRs) for OS and PFS and/or odds ratios (ORs) for ORR and DCR. We excluded reviews, abstracts, case reports, studies that were not published as full-text articles and studies with cohorts of less than 50 patients. Two authors (RCN and SQY) extracted the following characteristics for each trial: population, study phase, experiment arm, control arm, number of patients, primary endpoint, crossover, follow-up period, OS results and surrogate endpoints (PFS, ORR and DCR). Discrepancies in the literature search and data extraction were resolved by two senior authors (ZWZ and YFL).

Endpoint definitions

OS was defined as the time from randomization to death from any cause. PFS was defined as the time from randomization to progressive disease or death from any cause. ORR was defined as the proportion of confirmed complete response (CR) or partial response (PR) at the point of best overall response. DCR was defined as the percentage of confirmed CR, PR or stable disease at the point of best overall response.

Statistical analysis

We assessed the correlation between the treatment effect (HR or OR) among the surrogate endpoints (PFS, ORR, and DCR) and OS using

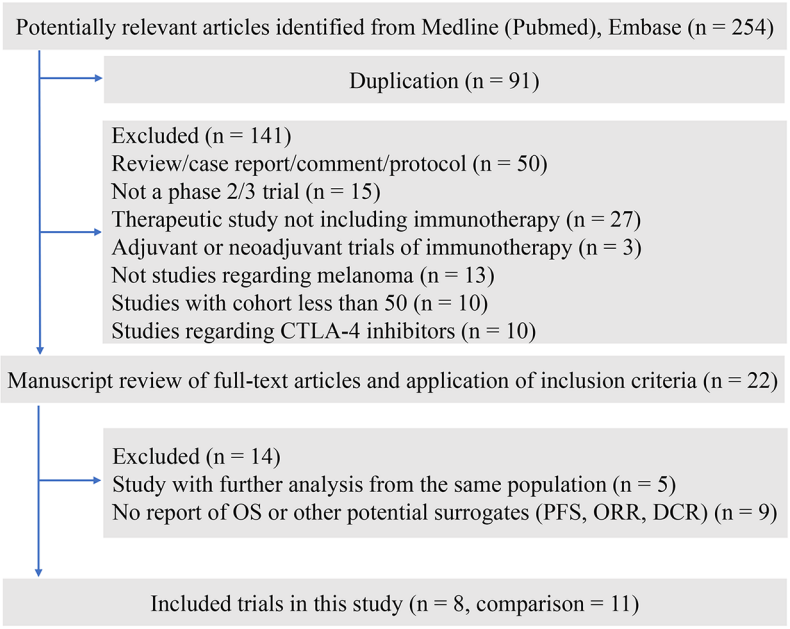

Figure 1. Study flow diagram of the included studies in this meta-analysis.
DCR, disease control rate; ORR, objective response rate; OS, overall survival; PFS, progression-free survival.

a linear regression model.[12] To interpret the differences between studies with respect to study size and precision of HR estimates, we weighted the analysis proportionally to the study sample size or to the precision of the observed treatment effects. Hence, we used a fixed effect model and a random effect model as the weighting strategies.[13] While the fixed effect meta-analysis is based on the presumption that a common treatment effect exists among every trial and uses the estimated inverse variance as weights, the random effect meta-analysis permits treatment effect discrepancy from trial to trial and merges the potential among-trial variation of effects into the weights. Overall, we applied three weighting strategies (sample size, fixed effect and random effect). According to A'Hern *et al.*,[14] we downweighted the sample size if trials reported more than two treatment arms.

We calculated the weighted coefficient of determination (R^2) to quantify the variation explained by the surrogate endpoints. We considered the correlation between OS and surrogate endpoints to be strong if the R^2 exceeded 0.75.[15,16] The surrogate threshold effect (STE),[16] defined as the minimum treatment effect on the surrogate necessary to predict a nonzero effect on the OS, was calculated. For future trials, the upper limit of the confidence interval (CI) for the estimated surrogate treatment effect should fall below the STE to predict a nonzero effect on OS. The STE in this study were performed through sample size weighting strategy.

Since the estimated treatment effect of OS can be influenced by crossover design, sample size and other potential factors, we performed several sensitivity analyses that restricted the analyses to trials with crossover $<50\%$, phase III trials, large trials (comparisons with >300 patients) trials, and trials with first-line therapy. For each meta-analysis, we applied an internal validation through leave-one-out analysis to evaluate the prediction accuracy of the surrogate model.[17] Each trial was left out once, and the surrogate model was built with other trials. This model was then re-applied to the left-out trial, and a 95% prediction interval was calculated to compare the predicted and observed treatment effect on OS. All statistical analyses were performed using R version 3.6.0 (http://www.r-project.org).

Results

After systematically screening 254 relevant articles (Figure 1), we identified eight trials (three phase II trials and five phase III trials) comprising 4110 subjects that were eligible for inclusion.[5–7,18–22] Table 1 shows detailed information from the

Table 1. Characteristics of the included trials.

Studies	Population	Study phase	Experimental arm	Control arm	*n*	Primary endpoint	Crossover	Median follow-up (months)
Hodi *et al.*,[6] CheckMate 069	Histologically confirmed, unresectable stage III or IV metastatic melanoma	II	Nivolumab plus ipilimumab	Ipilimumab	142	ORR[a]	57%	24.5 months
Hamid *et al.*,[18] KEYNOTE 002	Advanced melanoma with progression after two or more ipilimumab doses, previous BRAF or MEK inhibitor or both, if $BRAF^{V600}$ mutant-positive	II	Pembrolizumab 2 mg/kg; Pembrolizumab 10 mg/kg	ICC	540	OS, PFS	55%	28.0 months
Schachter *et al.*,[7] KEYNOTE 006	Ipilimumab-naive unresectable or advanced melanoma; <1 prior therapy;	III	Pembrolizumab every 2 weeks; Pembrolizumab every 3 weeks	Ipilimumab	834	PFS, OS	30%	22.9 months
Ascierto *et al.*,[19] CheckMate 066	Unresectable previously untreated stage III or IV melanoma, without a BRAF mutation	III	Nivolumab	Dacarbazine	418	OS	0%	38.4 months for nivolumab, and 38.5 months for dacarbazine
Hodi *et al.*,[5] CheckMate 067	Untreated, unresectable stage III or IV melanoma, known $BRAF^{V600}$ mutation status	III	Nivolumab plus ipilimumab; Nivolumab	Ipilimumab	945	PFS, OS	0%	46.9 months for nivolumab plus ipilimumab, 36.0 months for nivolumab, and 18.6 months for ipilimumab
Larkin *et al.*,[20] CheckMate 037	Unresectable stage IIIC or IV metastatic melanoma	III	Nivolumab	ICC	405	ORR, OS	23.33%	24 months
Long *et al.*,[22] KEYNOTE 022	Untreated, unresectable stage III or IV melanoma, known $BRAF^{V600}$ mutation status	III	Pembrolizumab plus epacadostat	Pembrolizumab plus placebo	706	PFS, OS	0%	12·4 months
Ascierto *et al.*,[21] KEYNOTE 252	Unresectable stage III or metastatic stage IV melanoma	II	Dabrafenib, trametinib plus pembrolizumab	Dabrafenib, trametinib plus placebo	120	PFS	0%	9.6 months

ICC, investigator's choice-chemotherapy; NR, not reached; ORR, objective response rate; OS, overall survival; PFS, progression-free survival.
[a]ORR for $BRAF^{V600}$ wild type.

included trials. The median follow-up duration of the included trials varied from 9.6 months to 46.9 months. We noted that three trials had three treatment arms.[5,7,18] In order to avoid the over-fitting correlation, we excluded the comparisons of pembrolizumab 10 mg/kg *versus* pembrolizumab 2 mg/kg in the KEYNOTE 002 trial, pembrolizumab every 3 weeks *versus* pembrolizumab every 2 weeks in the KEYNOTE 006 trial, and nivolumab plus ipilimumab *versus* nivolumab in the CheckMate 067 trial. Therefore, all the trials included 11 comparisons for quantitative analysis. Six comparisons reported improvement in OS (upper limit of CI for HR < 1.0), and eight comparisons reported improvement in PFS.

We first derived the degree of association between potential endpoints and OS through three

weighting strategies. As shown in Table 2, we observed that the correlations between ORR (sample size: 0.25, 95% CI –0.01 to 0.99; fixed effect: 0.10, –0.09 to 0.88; random effect: 0.09, –0.10 to 0.86; Supplemental Figure S1A), DCR (sample size: 0.57, 0.11–0.99; fixed effect: 0.44, 0.03–0.99; random effect: 0.42, 0.02–0.99; Supplemental Figure S1B) and OS were not strong enough to support the robust surrogacy of DCR or ORR for OS. Thus, we then focused on the potential surrogacy of PFS for OS and plotted the HRs for PFS and OS (Figure 2A). Deducing the correlation coefficient by weighting for sample size, we noted a strong correlation between PFS and OS (0.82, 0.41–0.99; Table 2). While presuming no difference between therapy type and treatment effect on PFS and OS (fixed effect model) slightly weakened the degree of association (0.75, 0.30–0.99), allowing for different therapy types to have a differential effect on PFS and OS (random effect model) weakened the association (0.72, 0.25–0.99).

$HR_{OS} = 0.215 + 0.845 \times HR_{PFS}$, where HR_{PFS} represents the HR for PFS and HR_{OS} represents the predicted HR for OS. This model indicates that every 1% PFS risk reduction due to anti-PD-1/PD-L1 treatment can induce 0.845% risk reduction of OS. We then calculated the STE of 0.78, indicating that a future anti-PD-1/PD-L1 trial would need less than 0.78 for PFS of the upper limit of the confidence interval to predict an OS benefit (Figure 2A).

Table 2. Correlation analysis between surrogate endpoints and OS.

Surrogate endpoint	Weighted coefficient of determination, R^2 (95% CI)	*p* value
DCR		
Sample size	0.57 (0.11–0.99)	0.007
Fixed effect	0.44 (0.03–0.99)	0.024
Random effect	0.42 (0.02–0.99)	0.031
ORR		
Sample size	0.25 (–0.01 to 0.99)	0.118
Fixed effect	0.10 (–0.09 to 0.88)	0.338
Random effect	0.09 (–0.10 to 0.86)	0.360
PFS		
Sample size	0.82 (0.41–0.99)	<0.001
Fixed effect	0.75 (0.30–0.99)	<0.001
Random effect	0.72 (0.25–0.99)	<0.001

CI, confidence interval; DCR, disease control rate; ORR, objective response rate; OS, overall survival; PFS, progression-free survival.

There may have been potential heterogeneity due to the crossover effects and sample size; in our study, we noted two outliers in the plot of the HRs for PFS and OS (Figure 2A). We observed that these two outliers were mainly from two comparisons of the studies of Checkmate 069[6]

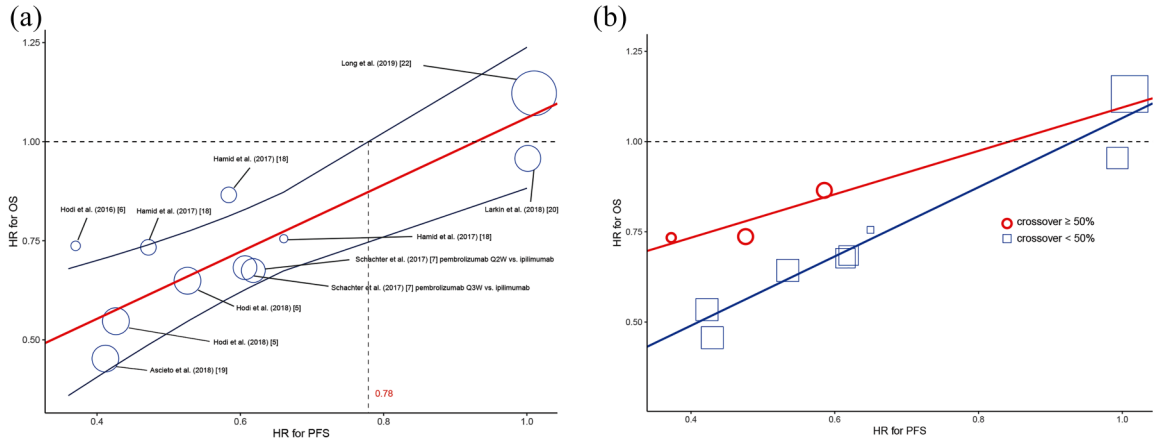

Figure 2. Correlation between treatment effects on overall survival and progression-free survival. Each trial is represented by a circle, with the size of the circle being proportional to the sample size. (A) The blue line represents the 95% prediction limit of the regression line (red line). Model equation: $HR_{OS} = 0.215 + 0.845 \times HR_{PFS}$, $R2_{sample\ size} = 0.82$ with $p < 0.001$, STE = 0.78; (B) Crossover <50% (blue hollow rectangle; $R2_{sample\ size} = 0.94$ with $p < 0.001$) *versus* ⩾50% (red hollow circle; $R2_{sample\ size} = 0.76$ with $p = 0.329$).
HR, hazard ratio; OS, overall survival; PFS, progression-free survival; STE, surrogate threshold effect.

Table 3. Sensitivity analysis of the correlation between PFS and OS.

	Weighted coefficient of determination, R^2 (95% CI)	STE
Total population[5–7,18–22]		0.78
Sample size	0.82 (0.41–0.99)	
Fixed effect	0.75 (0.30–0.99)	
Random effect	0.72 (0.25–0.99)	
Trials with <50% crossover[5,7,19–22]		0.82
Sample size	0.94 (0.60–0.99)	
Fixed effect	0.94 (0.58–0.99)	
Random effect	0.94 (0.58–0.99)	
Phase III trials[5,7,19,20,22]		0.79
Sample size	0.95 (0.64–0.99)	
Fixed effect	0.94 (0.63–0.99)	
Random effect	0.94 (0.63–0.99)	
Comparisons with >300 patients[5,7,18–20,22]		0.78
Sample size	0.86 (0.43–0.99)	
Fixed effect	0.78 (0.29–0.99)	
Random effect	0.78 (0.29–0.99)	
Trials on first-line treatment[5–7,19,21,22]		0.76
Sample size	0.91 (0.51–0.99)	
Fixed effect	0.90 (0.49–0.99)	
Random effect	0.83 (0.34–0.99)	

CI, confidence interval; OS, overall survival; PFS, progression-free survival; STE, surrogate threshold effect.
The STE in this study were performed through sample size weighting strategy.

and KEYNOTE 002.[18] Notably, these two studies had the similar feature: phase II designs with obvious crossover (55% and 57%, respectively), thus resulting in the discordant change between PFS and OS. Hence, we performed several sensitivity analyses (Table 3) and noted that restriction of the analysis to eight comparisons with crossover rates less than 50% (0.94–0.94; Figure 2B, Supplemental Figure S2A) demonstrated a perfect correlation between treatment effect on PFS and OS; the three comparisons with crossover rate >50% indicated a weakened correlation between treatment effect on PFS and OS ($R2 = 0.76$ for sample size weighting, $p = 0.329$; Figure 2B). Then, we included phase III trials; the degree of association between PFS and OS was based on seven comparisons of five trials, excluding the four comparisons from the CheckMate 069,[6] KEYNOTE 002[18] and KEYNOTE 252[21] studies, and included data from 3308 subjects. An extreme strong correlation (0.94 to 0.95) between PFS and OS was noted upon restriction to phase III trials (Table 3; Supplemental Figure S2B). In addition, we also performed other sensitivity analyses that restricted analyses to large trials and trials on first-line treatment; all these analyses exhibited strong to very strong correlations (0.78–0.91) between PFS and OS (Table 3; Supplemental Figure S2C–D).

Finally, we performed a leave-one-out cross validation approach to assess the accuracy of PFS in predicting OS. We noted that the observed HR for OS fell between the limits of the 95% prediction intervals in all the 11 comparisons, indicating that the treatment effect on PFS is a valid predictor of OS (Figure 3).

Discussion

In the present study, we found that the correlations between DCR/ORR and OS were not strong, indicating that the treatment effect on these two endpoints was not predictive of OS. Notably, we found a strong correlation between PFS and OS (0.72–0.82), irrespective of the applied weighting strategies. Sensitivity analyses that were restricted to the trials with less than 50% crossover, phase III trials and first-line trials further yielded stronger or even nearly perfect correlations (0.83–0.94) between PFS and OS; the leave-one-out cross-validation approach also confirmed that the effects observed on PFS were adequate to predict the treatment effect on OS. Therefore, we propose the use of PFS as the surrogate endpoint for OS in anti-PD-1/PD-L1 trials of metastatic melanoma.

The treatment landscape of metastatic melanoma has dramatically transitioned from cytotoxic agents to targeted drugs and now to anti-PD-1/PD-L1 agents,[23] and such changes have translated into enormous survival benefits for melanoma patients with metastatic disease. Recently, the update of survival data from the CheckMate 067

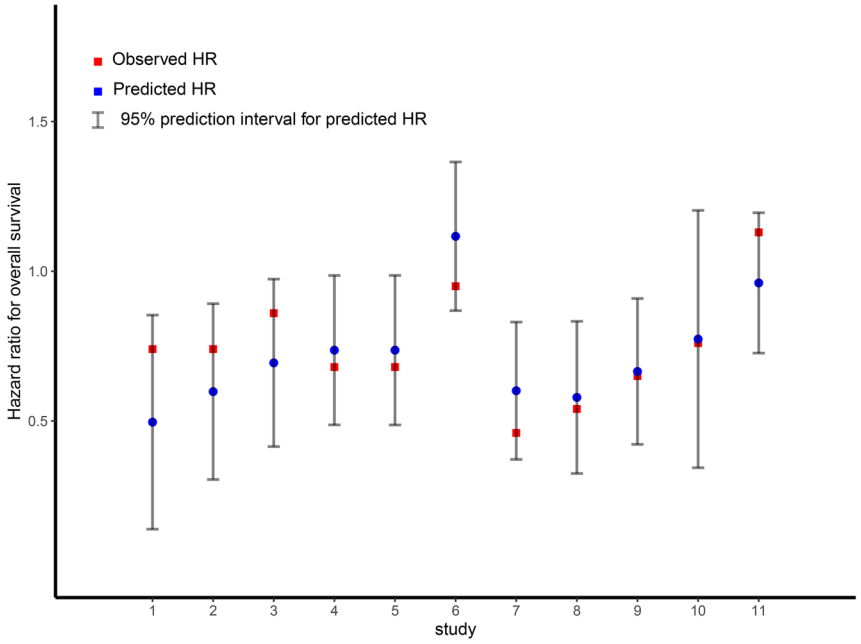

Figure 3. Leave-one-out cross-validation analysis of the prediction of OS by treatment effect on PFS: observed HR for OS for left-out trial *versus* predicted HR for OS and 95% prediction interval for predicted HR for OS. To assess model accuracy, a leave-one-out cross-validation strategy was used: each unit of analysis was left out once, and the linear model was then constructed from scratch using the remaining data.[17] This model was then re-applied to the left-out study in order to compare the predicted and observed treatment effect on OS. Based on the linear regression models, a 95% prediction interval was calculated to compare the predicted and observed treatment effect on OS.
HR, hazard ratio; OS, overall survival; PFS, progression-free survival.

trial reported a 4-year OS rate of 53% in the nivolumab plus ipilimumab group, which is an extravagant expectation for both clinicians and patients 10 years ago. The researchers are now evaluating the potential role of combination regimens, such as PD-1/PD-L1 inhibitors in combination with innate immune stimulants[24] or molecularly targeted agents (ClinicalTrials.gov identifiers: NCT02130466, NCT02967692, and NCT02908672), to enhance the therapeutic effect and minimize the risk of toxicities associated with combination therapy. It is well recognized that OS is the standard endpoint for clinical trials; however, several trials have set ORR[25–27] or PFS[5–7,18,20] as the primary or coprimary endpoints in anti-PD-1/PD-L1 trials of metastatic melanoma before these endpoints were validated as surrogates for OS. A meta-analysis by Mushti[28] reported that the associations between PFS/ORR and OS were too weak to support these RECIST-defined endpoints as surrogates for OS in anti-PD-1/PD-L1 trials of solid tumours. Nonetheless, their analysis was based on 13 positive trials approved by the FDA, which indicated a selection biases in their findings. In addition, the correlation between RECIST-defined endpoints and OS in the melanoma subpopulation was not reported. Our previous study noted a good correlation between PFS and OS in anti-PD-1/PD-L1 trials in metastatic melanoma.[29] In the present analysis, we applied more rigorous criteria using three weighting strategies to address this urgent issue, and our findings further validated that correlations between DCR/ORR and OS were not strong. Surprisingly, we identified a strong correlation between PFS and OS, which was further verified through extensive sensitivity analyses and leave-one-out cross-validation. We believe that the robust correlation between PFS and OS in anti-PD-1/PD-L1 therapy of melanoma is mainly attributable to the fact that melanoma is an aggressive tumour and that the subsequent lines of therapy are limited if patients develop progressive disease after anti-PD-1/PD-L1 therapy. Therefore, we propose that in future anti-PD-1/PD-L1 trials for metastatic melanoma, PFS is considered for use as the surrogate endpoint for OS.

STE is an alternative measure for surrogate endpoint validation.[16] Using a surrogate endpoint

with an STE closer to 1, it would be easier to predict an OS benefit. In the present study, we found that the STE was 0.78 for PFS. In addition, we noted that six of eight comparisons that reported PFS with an upper limit CI for HR < 0.78 reported improvement in OS, and all the three comparisons that reported PFS with an upper limit CI for HR $\geqslant 0.78$ failed to report improvement in OS; thus the accuracy rate of an STE of 0.78 was 81.8% (9/11). Therefore, an anti-PD-1/PD-L1 trial in metastatic melanoma producing a hazard reduction of at least 22% (upper 95% CI of HR < 0.78) for disease progression or death, could expect to promise a statistically significant reduction in OS.

Anti-PD-1/PD-L1 agents exert an antitumour effect by activating effector T cells, resulting in T cells circulating throughout the body that can identify cognate antigens presented by cancer cells.[30] Thus, patients who receive anti-PD-1/PD-L1 therapy might develop immune-related response patterns, wherein they initially experience transitory tumour swelling that meets conventional response criteria for progression but is later followed by decreased tumour burden. Beaver and colleagues reported that 14% of patients with metastatic melanoma who continued PD-1 inhibitor treatment beyond RECIST-defined progression experienced delayed tumour decrease and prolonged overall survival,[10] indicating the existence of pseudoprogression and a rationale for continuing PD-1 inhibitor treatment for these patients. To distinguish pseudoprogression from truly progressive disease, oncologists modified the conventional RECIST criteria and developed new response criteria, including immune-related response criteria using bidimensional measurements (irRC),[31] the revised irRC using unidimensional measurements based on the original RECIST (referred to as irRECIST),[32] and now the immune RECIST (iRECIST).[33] However, we should bear in mind that the iRECIST requires further validation and that the overall incidence of pseudoprogression in melanoma is low, ranging from 8% to 14%.[10,30,34] Therefore, we propose that the RECIST-defined PFS could be set as the primary or coprimary endpoint in anti-PD-1/PD-L1 trials of metastatic melanoma. In addition, we should also set the iRECIST-defined endpoints as the secondary endpoints in future investigations.

Notably, crossover from the control arm to a highly active experimental therapy at the time of progression may result in positive results for PFS but negative results for OS. We observed that half of included trials reported a 23–57% crossover rate, and the sensitivity analysis showed that the inclusion of trials with crossover weakened the correlation between PFS and OS; however, we also recognized that only three comparisons had the crossover rate $>50\%$, which would reduce the power to make this conclusion. Therefore, with the emergence of anti-PD-1/PD-L1 trials that allow crossover design, the correlation between PFS and OS needs further investigation.

Our study has several notable limitations. First, since the included trials were derived from multiple-line therapy, heterogeneity may exist in our analysis. Hence, we performed a sensitivity analysis that restricted the analysis to trials with first-line therapy and confirmed the very strong correlation between PFS and OS (0.83–0.91). Next, some eligible trials had a small sample size, short follow-up times and phase II designs, accounting for fairly wide confidence intervals of the HR for treatment effects. Thus, the surrogacy of PFS for OS is still not trustworthy in trials with small sample sizes or short follow-up durations, or phase II trials. In addition, the evaluation of PFS in the included studies might be either based on an independent review committee or investigator, which can bias our conclusions. Lastly, our study was performed at the trial level instead of at the individual level.

Conclusion

PFS may be the appropriate surrogate for OS in anti-PD-1/PD-L1 trials of metastatic melanoma. A future similar anti-PD-1/PD-L1 trial would need less than 0.78 for PFS of the upper limit of the confidence interval to predict an OS benefit.

Author contributions

Study design: Run-Cong Nie, Shu-Qiang Yuan and Yun Wang; Manuscript writing and revision: Run-Cong Nie, Yuan-Fang Li and Ying-Bo Chen; Literature retrieval: Run-Cong Nie and Shu-Qiang Yuan; Discretion of eligibility: Shi Chen and Shu-Man Li; Data extraction: Run-Cong Nie, Shu-Qiang Yuan, Jin-Ling Duan, Jie Zhou and Guo-Ming Chen; Statistical analysis: Run-Cong Nie, Tian-Qi Luo and Zhi-Wei Zhou

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that there is no conflict of interest.

ORCID iD

Run-Cong Nie https://orcid.org/0000-0003-3595-6304

Supplemental material

Supplemental material for this article is available online.

References

1. Schadendorf D, van Akkooi ACJ, Berking C, *et al.* Melanoma. *Lancet* 2018; 392: 971–984.
2. Robert C, Thomas L, Bondarenko I, *et al.* Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med* 2011; 364: 2517–2526.
3. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* 2012; 12: 252–264.
4. Ribas A and Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Science* 2018; 359: 1350–1355.
5. Hodi FS, Chiarion-Sileni V, Gonzalez R, *et al.* Nivolumab plus ipilimumab or nivolumab alone versus ipilimumab alone in advanced melanoma (CheckMate 067): 4-year outcomes of a multicentre, randomised, phase 3 trial. *Lancet Oncol* 2018; 19: 1480–1492.
6. Hodi FS, Chesney J, Pavlick AC, *et al.* Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *Lancet Oncol* 2016; 17: 1558–1568.
7. Schachter J, Ribas A, Long GV, *et al.* Pembrolizumab versus ipilimumab for advanced melanoma: final overall survival results of a multicentre, randomised, open-label phase 3 study (KEYNOTE-006). *Lancet* 2017; 390: 1853–1862.
8. Flaherty KT, Hennig M, Lee SJ, *et al.* Surrogate endpoints for overall survival in metastatic melanoma: a meta-analysis of randomised controlled trials. *Lancet Oncol* 2014; 15: 297–304.
9. Chen DS and Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature* 2017; 541: 321–330.
10. Beaver JA, Hazarika M, Mulkey F, *et al.* Patients with melanoma treated with an anti-PD-1 antibody beyond RECIST progression: a US food and drug administration pooled analysis. *Lancet Oncol* 2018; 19: 229–239.
11. Ferrara R, Mezquita L, Texier M, *et al.* Hyperprogressive disease in patients with advanced non-small cell lung cancer treated with PD-1/PD-L1 inhibitors or with single-agent chemotherapy. *JAMA Oncol* 2018; 4: 1543–1552.
12. Buyse M, Burzykowski T, Michiels S, *et al.* Individual- and trial-level surrogacy in colorectal cancer. *Stat Methods Med Res* 2008; 17: 467–475.
13. DerSimonian R and Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177–188.
14. A'Hern RP, Ebbs SR and Baum MB. Does chemotherapy improve survival in advanced breast cancer? A statistical overview. *Br J Cancer* 1988; 57: 615–618.
15. Michiels S, Le Maitre A, Buyse M, *et al.* Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analyses of individual patient data. *Lancet Oncol* 2009; 10: 341–350.
16. Burzykowski T and Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat* 2006; 5: 173–186.
17. Julious SA, Campbell MJ and Walters SJ. Predicting where future means will lie based on the results of the current trial. *Contemp Clin Trials* 2007; 28: 352–357.
18. Hamid O, Puzanov I, Dummer R, *et al.* Final analysis of a randomised trial comparing pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory advanced melanoma. *Eur J Cancer* 2017; 86: 37–45.
19. Ascierto PA, Long GV, Robert C, *et al.* Survival outcomes in patients with previously untreated BRAF wild-type advanced melanoma treated with nivolumab therapy: three-year follow-up of a randomized phase 3 trial. *JAMA Oncol* 2019; 5: 187–194.
20. Larkin J, Minor D, D'Angelo S, *et al.* Overall survival in patients with advanced melanoma who received nivolumab versus investigator's choice chemotherapy in CheckMate 037: a randomized, controlled, open-label phase III trial. *J Clin Oncol* 2018; 36: 383–390.
21. Ascierto PA, Ferrucci PF, Fisher R, *et al.* Dabrafenib, trametinib and pembrolizumab or placebo in BRAF-mutant melanoma. *Nat Med* 2019; 25: 941–946.
22. Long GV, Dummer R, Hamid O, *et al.* Epacadostat plus pembrolizumab versus placebo

plus pembrolizumab in patients with unresectable or metastatic melanoma (ECHO-301/KEYNOTE-252): a phase 3, randomised, double-blind study. *Lancet Oncol* 2019; 20: 1083–1097.

23. Luke JJ, Flaherty KT, Ribas A, *et al*. Targeted agents and immunotherapies: optimizing outcomes in melanoma. *Nat Rev Clin Oncol* 2017; 14: 463–482.
24. Ribas A, Medina T, Kummar S, *et al*. SD-101 in combination with pembrolizumab in advanced melanoma: results of a phase Ib, multicenter study. *Cancer Discov* 2018; 8: 1250–1257.
25. Postow MA, Chesney J, Pavlick AC, *et al*. Nivolumab and ipilimumab versus ipilimumab in untreated melanoma. *N Engl J Med* 2015; 372: 2006–2017.
26. Chesney J, Puzanov I, Collichio F, *et al*. Randomized, open-label phase II study evaluating the efficacy and safety of talimogene laherparepvec in combination with ipilimumab versus ipilimumab alone in patients with advanced, unresectable melanoma. *J Clin Oncol* 2018; 36: 1658–1667.
27. Tawbi HA, Forsyth PA, Algazi A, *et al*. Combined nivolumab and ipilimumab in melanoma metastatic to the brain. *N Engl J Med* 2018; 379: 722–730.
28. Mushti SL, Mulkey F and Sridhara R. Evaluation of overall response rate and progression-free survival as potential surrogate endpoints for overall survival in immunotherapy trials. *Clin Cancer Res* 2018; 24: 2268–2275.
29. Nie RC, Chen FP, Yuan SQ, *et al*. Evaluation of objective response, disease control and progression-free survival as surrogate end-points for overall survival in anti-programmed death-1 and anti-programmed death ligand 1 trials. *Eur J Cancer* 2019; 106: 1–11.
30. Chiou VL and Burotto M. Pseudoprogression and immune-related response in solid tumors. *J Clin Oncol* 2015; 33: 3541–3543.
31. Wolchok JD, Hoos A, O'Day S, *et al*. Guidelines for the evaluation of immune therapy activity in solid tumors: Immune-related response criteria. *Clin Cancer Res* 2009; 15: 7412–7420.
32. Nishino M, Giobbie-Hurder A, Gargano M, *et al*. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. *Clin Cancer Res* 2013; 19: 3936–3943.
33. Seymour L, Bogaerts J, Perrone A, *et al*. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol* 2017; 18: e143–e152.
34. Hodi FS, Hwu WJ, Kefford R, *et al*. Evaluation of immune-related response criteria and RECIST v1.1 in patients with advanced melanoma treated with pembrolizumab. *J Clin Oncol* 2016; 34: 1510–1517.