European Journal of Cancer 106 (2019) 1–11

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Review

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

Run-Cong Nie [a,1], Fo-Ping Chen [b,1], Shu-Qiang Yuan [a,1], Ying-Shan Luo [b,1], Shi Chen [c], Yong-Ming Chen [a], Xiao-Jiang Chen [a], Ying-Bo Chen [a], Yuan-Fang Li [a,*,1], Zhi-Wei Zhou [a,**,1]

[a] *Department of Gastric Surgery, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China*
[b] *Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou, China*
[c] *Department of Gastric Surgery, The 6th Affiliated Hospital, Sun Yat-sen University, Guangzhou, China*

Received 12 September 2018; received in revised form 9 October 2018; accepted 10 October 2018
Available online 16 November 2018

KEYWORDS
PD-1;
PD-L1;
Surrogate end-point;
Overall survival;
RECIST criteria;
Progression-free survival

Abstract ***Background:*** We aimed to assess whether the Response Evaluation Criteria in Solid Tumors (RECIST) criteria–based objective response rate (ORR), disease control rate (DCR) and progression-free survival (PFS) could be valid surrogate end-points for overall survival (OS) in anti–programmed death-1 (PD-1)/programmed death ligand 1 (PD-L1) trials.
Methods: We systematically reviewed phase 2 and phase 3 trials of anti–PD-1/PD-L1 drug trials of advanced or recurrent solid tumours that reported OS and at least one of the RECIST criteria–based end-points. We used Spearman rank correlation to evaluate the strength of the association between these end-points and OS and a linear regression model, weighted by the sample size, to assess the association between the treatment effect on these end-points and OS. We also performed sensitivity analyses and a leave-one-out cross-validation approach to evaluate the robustness of our findings.

* *Corresponding author:* Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, 651 E Dongfeng Road, Guangzhou, Guangdong, 510060, China.
** *Corresponding author:* Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, 651 E Dongfeng Road, Guangzhou, Guangdong, 510060, China.
E-mail addresses: liyuanf@sysucc.org.cn (Y.-F. Li), zhouzhw@sysucc.org.cn (Z.-W. Zhou).
[1] These authors contributed equally to this study.

https://doi.org/10.1016/j.ejca.2018.10.011
0959-8049/© 2018 Elsevier Ltd. All rights reserved.

Results: Forty-three qualifying trails comprising 15,088 patients were eligible. PFS showed good correlation with OS (squared Spearman rank correlation coefficient [r_s^2] = 0.54; P < 0.001), while ORR and DCR illustrated moderate association with OS (r_s^2 = 0.29 and 0.28, respectively; both P < 0.001). The correlation was moderate between the treatment effects on PFS and OS (coefficient of determination [R^2] = 0.37, P < 0.001) and poor among ORR, DCR and OS (R^2 = 0.10 and 0.08, respectively); these were confirmed by sensitivity analyses (all R^2 < 0.75) and the leave-one-out cross-validation approach.
Conclusions: No RECIST criteria–based end-points could be a valid surrogate for OS. At present, we proposed to set OS as the primary end-point in anti–PD-1/PD-L1 drug trials of advanced or recurrent solid tumours.
© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Programmed death-1 (PD-1) is a coreceptor expressed predominantly by T cells that plays a critical role in physiologic regulation of the immune system [1–3]. The binding of PD-1 to its ligands, programmed death ligand 1 (PD-L1) and PD-L2, is vital for T-cell inhibition and exhaustion, which protects the human body against autoimmune diseases and assists tumour cells to escape from immunosurveillance mechanisms. Overexpression of PD-L1 on tumour cells and PD-1 on tumour-infiltrating lymphocytes has been shown to suppress anticancer efficacy of T cell, thus accelerating the progression and metastasis of the tumours [1]. Therefore, many researchers are trying to use PD-1 inhibitors to block the PD-1 pathway to use and enhance the distinctive powers of the immune system to fight against cancer. Notably, recent clinical trials have shown that anti–PD-1 agents have profound effects on solid tumour regression [4–7]; this spur anti–PD-1 to be an attractive target in cancer therapy. Currently, the Food and Drug Administration (FDA) has approved two anti–PD-1 drugs (nivolumab and pembrolizumab) and three anti–PD-L1 drugs (avelumab, atezolizumab and durvalumab) for the treatment of 11 advanced cancers; the European Medicines Agency and medicine agencies of other countries have also approved these drugs for various advanced solid tumours. Based on the favourable tumour response of PD-1 blockade in tumours with mismatch repair deficiency (dMMR) [8], the FDA even approved the PD-1 blockade across tumour types for patients with dMMR. As oncology researchers continue to develop the new therapeutic strategies of immunotherapy for advanced cancers, many clinical trials are underway to better clarify the role of PD-1 blockade in variety of tumour types; researchers also investigated the effective biomarkers, such as PD-L1 expression, Epstein–Barr virus, circulating tumour DNA and tumour mutation burden [9,10], to name but a few, to predict the response during anti–PD-1 blockade treatment.

However, there are some crucial issues under dispute when conducting anti–PD-1/anti–PD-L1 drug trials: what is the optimal end-point and how to evaluate the tumour response in these trials. In conventional clinical trials, overall survival (OS) is usually considered as the gold standard end-point because it is simple and reliable to measure and has a straightforward interpretation and clinical practicability. However, OS requires a larger sample size and prolonged follow-up duration to detect statistically significant differences, and it may be diluted by non-cancer deaths and subsequent therapies after progression. Therefore, several meta-analyses had tried to seek potential surrogacies, such as the Response Evaluation Criteria in Solid Tumors (RECIST) criteria–based progression-free survival (PFS), for OS to shorten the follow-up duration and reduce the cost of drug development [11]. Hence, some researchers set PFS and even objective response rate (ORR) as the primary end-points in anti–PD-1/PD-L1 drug trials [12–18]. Notably, the majority of the immunotherapy trials used the RECIST criteria to assess the therapeutic effects; the RECIST criteria were aimed to evaluate the activity and efficacy of new cancer therapeutics with uniformly and objectively measurable and reproducible end-points in a reliable manner [19], which was based on experience with cytotoxic agents, while it has been observed that the mechanisms of action of immunotherapeutic agents are markedly different; there exists delayed antitumour activity and increased tumour volume from immune infiltration during treatment with anti–PD-1/PD-L1 drugs [20]. Therefore, it is uncertain whether the RECIST criteria–based end-points can sufficiently reflect the antitumour effect of these drugs.

Based on this premise, Mushti *et al.* [21] conducted a meta-analysis to evaluate whether PFS and ORR were potential surrogate end-points for OS in immunotherapy trials. However, they only included a small part of trials (13 positive randomised trials submitted to the FDA), which indicated the obvious selection bias in their findings. Hence, a major challenge in exploring the potential surrogate end-points for OS in

immunotherapy trials is to include all the immunotherapy trials on solid tumours. Here, in this meta-analysis, we reviewed all the eligible trials to further evaluate whether PFS, ORR and disease control rate (DCR) are potential surrogate end-points for OS.

2. Methods

2.1. Search strategy and selection criteria

In June 2018, two authors (R.-C.N. and F.-P.C.) electronically searched MEDLINE (PubMed) and Embase to identify phase 2 or phase 3 trials of anti−PD-1/PD-L1 drugs for advanced or recurrent solid tumours. Keywords for the search strategy were as follows: nivolumab, pembrolizumab, avelumab, atezolizumab, durvalumab, PD-1, PD-L1, checkpoint inhibitors, phase 2 trials and phase 3 trials. We also searched the ClinicalTrials.gov and Cochrane Library databases and used hand-searching strategy to review the references of the included trials to retrieve additional studies.

The inclusion criteria were as follows: phase 2 or phase 3 trials; anti−PD-1 or anti-PD-L1 monoclonal antibody for advanced or recurrent solid tumours as the experiment arm and reporting OS and at least one surrogate end-point (PFS, ORR and DCR). Reviews, case reports and abstracts were excluded. Two authors (R.-C.N. and F.-P.C.) extracted the following characteristics for each trial: accrual period, the phase of study (phase 2 or 3), included population, line of therapy, treatment regimen, number of patients (each arm and trial), primary end-point, median follow-up time, OS results and surrogate end-points (PFS, ORR and DCR). Discrepancies in the literature search and data extraction were resolved by two independent authors (Z.-W.Z. and Y.-F.L.).

2.2. End-point definitions

OS was defined as the time from randomisation to death from any cause. PFS was defined as the time from randomisation to the first event (progressive disease or death from any cause). ORR was defined as the proportion of confirmed complete response (CR) or partial response (PR) at the best response. DCR was defined as the percentage of confirmed CR, PR or stable disease at the best response. The responses were assessed by RECIST, version 1.1, criteria [22].

2.3. Statistical analysis

Two correlation approaches were used to assess the potential surrogate end-points for OS [23,24]. First, we used the Spearman rank correlation coefficient (r_s) to evaluate the strength of association between the surrogate end-points (median PFS, ORR and DCR) and median OS of each treatment arm (arm-level correlation). Second, we assessed the correlation between the treatment effect (natural log hazard ratio [HR] or natural log odds ratio) on surrogate end-points and OS (trial-level correlation) through a linear regression model, weighted by the sample size of each comparison [24]. For trials reporting more than two arms, we downweighted the sample size, as described by A' Hern *et al.* [25].

The squared Spearman rank correlation coefficient (r_s^2) at the arm level or the coefficient of determination (R^2) at the trial level was calculated to assess variation explained by the surrogate end-points. A squared correlation value higher than 0.75 was classified as strong, that higher than 0.5 as good, that higher than 0.25 as moderate and that equal to or lower than 0.25 as poor [11]. The candidate surrogate end-points were considered acceptable if both r_s^2 and R^2 were higher than 0.75 (strong correlation). The surrogate threshold effect (STE) [26], defined as the minimum treatment effect on the surrogate necessary to predict a non-zero 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 non-zero effect on OS.

We also performed sensitivity analysis with restriction to phase 3 trials, large trials (>200 patients), trials with mature follow-up (median follow-up $\geq$ 12 months), trials without crossover or crossover $<$ 50% and trials with first-line therapy. For each meta-analysis, we used a leave-one-out cross-validation approach to evaluate the prediction accuracy of the surrogate model [27]. Each trial was left out once, and the surrogate model was built with other trials. This model was then applied to the left-out trial, and a 95% prediction interval was calculated to compare the predicted and observed treatment effect on OS. All the statistical analyses were performed using R, version 3.4.0 (http://www.r-project.org).

3. Results

After systematic literature review (Supplementary Fig. S1), we identified 43 trials (24 phase 2 trials, 1 phase 2/3 trial and 18 phase 3 trials) comprising 15,088 patients eligible for inclusion [12−18,20,28−62]. Detailed information of these trials was presented in Table 1. Notably, there were seven (16.3%) trials [12,13,28,32,41,42,51] that had three treatment arms and 17 trials that had two treatment arms; thus, there were 38 comparisons available for trial-level analysis. The median follow-up time of the included trials varied from 5 months to 38 months, and the included studies were published between May 2015 and June 2018. Sixteen trials have reported improvement in OS (upper limit of CI for HR $<$ 1.0), and thirteen trials have reported improvement in PFS.

Fig. 1 shows the degree of association between RECIST criteria−based end-points and OS at the arm

Table 1
Characteristics of the included trials.

Studies	Trial period	Type of the study	Population	Line of therapy	Experimental arm	Control arm	N	Primary end-point	Crossover	Follow-up (months)	Median OS (months)
Motzer (2015) [52]	2011–2012	Phase 2	RCC	≥2	Nivolumab	–	168	PFS	0%	Not report	18.2 vs. 25.5 vs. 24.7
Motzer (2015) [51]	2012–2014	Phase 3	RCC	2, 3	Nivolumab	Everolimus	803	OS	0%	Not report	25 vs. 19.6
Robert (2015) [56]	2013–2014	Phase 3	Melanoma	≥1	Nivolumab	Dacarbazine	418	OS	0%	16.7	NR vs. 10.8
Borghaei (2015) [31]	2013–2013	Phase 3	Non-squamous NSCLC	1–3	Nivolumab	Docetaxel	582	OS	8%	13.2	12.2 vs. 9.4
Brahmer (2015) [32]	2012–2013	Phase 3	Squamous cell NSCLC	2	Nivolumab	Docetaxel	272	OS	0%	11	9.2 vs. 6
Hamanishi (2015) [40]	2011–2014	Phase 2	Ovarian cancer	≥2	Nivolumab	–	20	ORR	0%	8	20
Fehrenbacher (2016) [36]	2013–2014	Phase 2	NSCLC	2	Atezolizumab	Docetaxel	287	OS	0%	13.0	12.6 vs. 9.7
Nishio (2016) [53]	2013–2013	Phase 2	NSCLC	2	Nivolumab	–	76	ORR	0%	16.6	17.1
Reck (2016) [55]	2014–2015	Phase 3	NSCLC	1	Pembrolizumab	Chemotherapy	305	PFS[h]	43.7%	11.2	NR for both arms
Rosenberg (2016) [57]	2014–2014	Phase 2	Urothelial carcinoma	2	Atezolizumab	–	310	ORR	0%	11.7	7.9
Weber (2016) [61]	2013–2014	Phase 2	Melanoma	≥2	Nivolumab followed by ipilimumab	Ipilimumab followed by nivolumab	140	Grade 3–5 AE	0%	19.8	NR vs. 16.9
Antonia (2016) [28]	2013–2015	Phase 1/2	SCLC	≥2	Nivolumab + ipilimumab[a]	Nivolumab	213	ORR	0%	6.6/12.0/8.67	7.7 vs. 6.0 vs. 4.4
Ferris (2016) [37]	2014–2015	Phase 3	Head and neck cancer	1	Nivolumab	ICC	361	OS	0%	5.1	7.5 vs. 5.1
Hodi (2016) [44]	2013–2014	Phase 2	Melanoma	1	Nivolumab + ipilimumab	Ipilimumab	142	ORR[c]	57%	24.5	NR for both arms
Herbst (2016) [42]	2013–2015	Phase 2/3	NSCLC	2–4	Pembrolizumab[b]	Docetaxel	1034	OS	13.1%	13.1	10.4 vs. 12.7 vs. 8.5
Kaufman (2016) [46]	2014–2015	Phase 2	Merkel cell carcinoma	≥2	Avelumab	–	88	ORR	0%	10.4	11.3
Langer (2016) [48]	2014–2016	Phase 2	NSCLC	1	Pembrolizumab + carboplatin + pemetrexed	carboplatin + pemetrexed	123	ORR	51%	10.6	NR for both arms
Hamid (2017) [41]	2012–2013	Phase 2	Melanoma	≥2	Pembrolizumab[e]	ICC	540	OS, PFS	55%	28.0	13.4 vs. 14.7 vs. 11.0
Schachter (2017) [12]	2013–2014	Phase 3	Melanoma	1	Pembrolizumab[f]	Ipilimumab	834	PFS, OS	30%	22.9	NR vs. NR vs. 16
Peters (2017) [54]	2014	Phase 2	NSCLC	≥1	Atezolizumab	–	659	ORR[g]	0%	14.6	20.1 vs. 15.5 vs. 13.2
Rittmeyer (2017) [20]	2014–2015	Phase 3	NSCLC	2–3	Atezolizumab	Docetaxel	850	OS	0%	21	13.8 vs. 9.6
Sharma (2017) [58]	2015	Phase 2	Urothelial carcinoma	≥2	Nivolumab	–	265	ORR	0%	7	8.7
Wolchok (2017) [13]	2013–2014	Phase 3	Melanoma	1	Nivolumab + ipilimumab; nivolumab	Ipilimumab	945	PFS, OS	0%	35.7 vs 38 vs 18.6	37.6 vs. NR vs. 19.9
Tawbi (2017) [59]	2015–2016	Phase 2	Soft tissue/bone sarcoma	4	Pembrolizumab	–	80	ORR	0%	17.8	12 vs. 13
Larkin (2017) [49]	2012–2014	Phase 3	Melanoma	2	Nivolumab	ICC	405	ORR, OS	23.33%	24	15.7 vs. 14.4
Toulmonde (2017) [60]	2015–2016	Phase 2	Soft tissue sarcoma	≥1	Cyclophosphamide + pembrolizumab	–	50	ORR, PFS	0%	6.8	9.2 vs. 5.6 vs. 7.1 vs. NR
Balar (2017) [29]	2014–2015	Phase 2	Urothelial carcinoma	1	Atezolizumab	–	119	ORR	0%	17.2	15.9
Bauml (2017) [30]	2014–2015	Phase 2	Head and neck cancer	≥1	Pembrolizumab	–	171	ORR	0%	7	8
Bellmunt (2017) [14]	2014–2015	Phase 3	Urothelial carcinoma	2	Pembrolizumab	ICC	542	OS, PFS	0%	14.1	10.3 vs. 7.4
Carbone (2017) [34]	2014–2015	Phase 3	NSCLC	1	Nivolumab	ICC	423	PFS[i]	58%	13.5	14.4 vs. 13.2

Hida (2017) [43]	2014–2015	Phase 2	NSCLC	2	Nivolumab	–	35	ORR	0%	Not reported	16.3
Kang (2017) [45]	2014–2016	Phase 3	Gastric cancer	≥3	Nivolumab	Placebo	493	OS	0.6%	8.8	5.26 vs. 4.14
Kudo (2017) [47]	2014–2014	Phase 2	Oesophageal squamous cell carcinoma	≥1	Nivolumab	–	65	ORR	0%	10.8	10.8
Calabro (2018) [33]	2015–2016	Phase 2	Mesothelioma	1, 2	Tremelimumab + durvalumab	–	40	ORR	0%	19.2	16.6
D'Angelo (2018) [35]	2015–2016	Phase 2	Sarcoma	≥2	Nivolumab + ipilimumab	Nivolumab	83	ORR	0%	14.2	14.3 vs. 10.7
Gandhi (2018) [15]	2016–2017	Phase 3	NSCLC	1	Chemotherapy + pembrolizumab	Chemotherapy + placebo	616	OS, PFS	50%	10.5	NR vs. 11.3
Giaccone (2018) [39]	2015–2016	Phase 2	Thymic carcinoma	≥2	Pembrolizumab	–	40	ORR	0%	20	24.9
Garassino (2018) [38]	2014–2015	Phase 2	NSCLC	≥3	Durvalumab	–	376	ORR[d]	0%	Not reported	9.9 vs. 13.3 vs. 9.3 vs. 10.9
Long (2018) [50]	2014–2017	Phase 2	Melanoma with brain metastases	≥1	Nivolumab + ipilimumab; nivolumab	–	76	ORR	0%	17	18.5 vs. NR
Zhu (2018) [62]	2016–2017	Phase 2	HCC	≥2	Pembrolizumab	–	104	ORR	0%	12.3	12.9
Socinski (2018) [16]	2015–2016	Phase 3	NSCLC	1	Atezolizumab + bevacizumab + carboplatin + paclitaxel	Bevacizumab + carboplatin + paclitaxel	692	PFS, OS	0%	15.4	19.2 vs. 14.7
Shitara (2018) [17]	2015–2016	Phase 3	Gastric or gastroesophageal junction cancer	2	Pembrolizumab	Paclitaxel	395	OS, PFS	10%	8.5	9.1 vs. 8.3
Motzer (2018) [18]	2014–2016	Phase 3	RCC	1	Nivolumab + ipilimumab	Sunitinib	847	OS, ORR, PFS	0%	25.2	NR vs. 26

NSCLC, non–small-cell lung cancer; SCLC, small cell lung cancer; RCC, renal cell carcinoma; HCC, hepatocellular carcinoma; ORR, objective response rate; OS, overall survival; NR, not reached; AE, adverse event; ICC, investigator's choice chemotherapy; EGFR, epidermal growth factor receptor; ALK, anaplastic lymphoma kinase.

[a] Experimental arm included two cohorts: 1 mg/kg of nivolumab plus 3 mg/kg of ipilimumab (61 patients) and 3 mg/kg of nivolumab plus 1 mg/kg of ipilimumab (54 patients).

[b] Experimental arm included two cohorts: 2 mg/kg of pembrolizumab (344 patients) and 10 mg/kg of pembrolizumab (346 patients).

[c] ORR for BRAF V600 wild type.

[d] The median OS had not been reached in cohort 3 (EGFR−/ALK− with at least 90% of tumour cells with PD-L1 expression) of this study. Thus, we included cohort 1 (EGFR+/ALK + NSCLC with at least 25%, or less than 25%, of tumour cells with PD-L1 expression) and cohort 2 (EGFR−/ALK− with at least 25%, or less than 25%, of tumour cells with PD-L1 expression) of this study.

[e] Experimental arm included two cohorts: 2 mg/kg of pembrolizumab (180 patients) and 10 mg/kg of pembrolizumab (181 patients).

[f] Experimental arm included two cohorts: 1.2 g/kg of pembrolizumab every 3 weeks in first-line (142 patients), second-line (271 patients) and third-line (254 patients) therapy.

[g] Experimental arm included three cohorts: 10 mg/kg of atezolizumab every 2 weeks (279 patients) and 3 weeks (277 patients).

[h] PFS for patients with a PD-L1 expression level of ≥5%.

[i] PFS for patients with a PD-L1 expression level of ≥50%.

level. Through Spearman rank analysis, we observed good correlation between PFS and OS ($r_s^2 = 0.54$, $P < 0.001$) but moderate correlation between ORR ($r_s^2 = 0.29$, $P < 0.001$) or DCR ($r_s^2 = 0.28$, $P < 0.001$) and OS (Fig. 1 and Table 2). Meanwhile, at the trial level, we observed moderate correlation between the treatment effect on PFS and OS ($R^2 = 0.37$, $P < 0.001$; Fig. 2A and Table 2) and poor correlation between ORR or DCR and OS ($R^2 = 0.10$, $P = 0.053$ between ORR and OS; $R^2 = 0.08$, $P = 0.147$ between ORR and OS; Table 2). As estimated from the linear regression model ($\log [HR_{OS}] = -0.09 + 0.44 \times \log [HR_{PFS}]$), every 1 reduction of risk for PFS only provides 0.44 reduction for OS, indicating that reduction in risk of tumour progress as evaluated by the RECIST criteria is moderate associated with reduced death.

Given that there may have potential heterogeneity, owing to the variety of tumour types and crossover effects, in our study, we then performed subgroup analysis and sensitivity analysis to verify the robustness of our findings. First, we performed subgroup analysis according to the tumour types; good correlation between PFS and OS in melanoma ($R^2 = 0.61$, $P = 0.002$; STE = 0.83; Fig. 2B) and moderate correlation in non–small-cell lung cancer (NSCLC) ($R^2 = 0.44$, $P = 0.018$; STE = 1.01; Fig. 2C) and other cancer types ($R^2 = 0.29$, $P = 0.058$; STE = 1.06; Fig. 2D) were observed. Then, we performed sensitivity analyses in 14 trials without a crossover rate and 21 trials with a crossover rate less than 50% respectively to eliminate the effect of crossover; no strong correlations between PFS and OS were noted with restriction to trials without crossover ($R^2 = 0.31$) or the trials with <50% crossover ($R^2 = 0.41$) (Supplementary Table S1).

Table 2
Correlation analysis between surrogate end-points and OS at the arm level and trial level.

Surrogate end-point	Arm level			Trial level		
	No. of treatment arms	r_s^2	P value	No. of trials	R^2	P value
PFS	27	0.54	<0.001	37[a]	0.37	<0.001
ORR	24	0.29	<0.001	36[b]	0.10	0.053
DCR	19	0.28	<0.001	28[b]	0.08	0.147

OS, overall survival; PFS, progression-free survival; ORR, objective response rate; DCR, disease control rate; r_s^2, the squared Spearman rank correlation coefficient; R^2, coefficient of determination.
[a] Natural log hazard ratio in the analysis.
[b] Natural log odds ratio in the analysis.

Finally, we performed other sensitivity analyses with restriction to phase 3 trials, large trials or trials with mature follow-up; all did not identify strong correlations between PFS and OS (all $R^2 < 0.75$, Supplementary Table S1).

Additionally, we performed a leave-one-out cross-validation approach to assess the accuracy of PFS in predicting OS. For PFS, the HRs for OS fell within the 95% prediction interval in eight of 13 for melanoma (Fig. 3A), in seven of 12 for NSCLC (Fig. 3B) and in eight of 13 for other cancer types (Fig. 3C), indicating that the treatment effect on PFS is not predictive for that of OS.

4. Discussion

In randomised controlled trials of solid tumours, alternative end-points, such as PFS and disease-free survival,

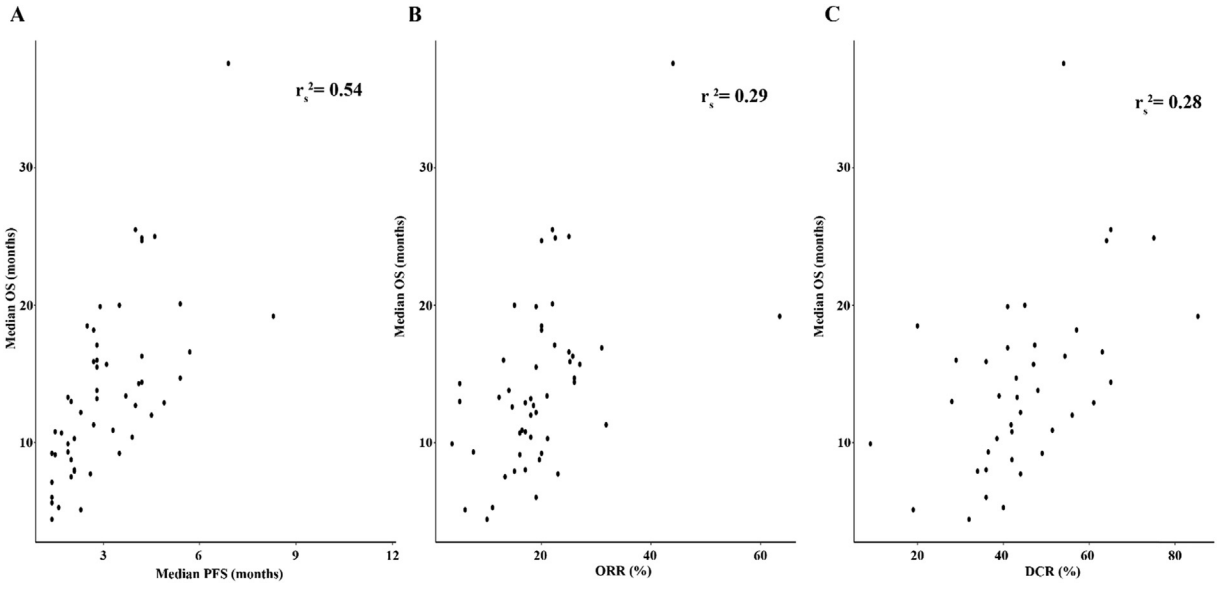

Fig. 1. Correlation between surrogate end-points and OS. (A) Median OS and median PFS ($r_s^2 = 0.54$, $P < 0.001$); (B) median OS and ORR (%) ($r_s^2 = 0.29$, $P < 0.001$) and (C) median OS and DCR (%) ($r_s^2 = 0.28$, $P < 0.001$). OS, overall survival; PFS, progression-free survival; ORR, objective response rate; DCR, disease control rate; r_s^2, the squared Spearman rank correlation coefficient.

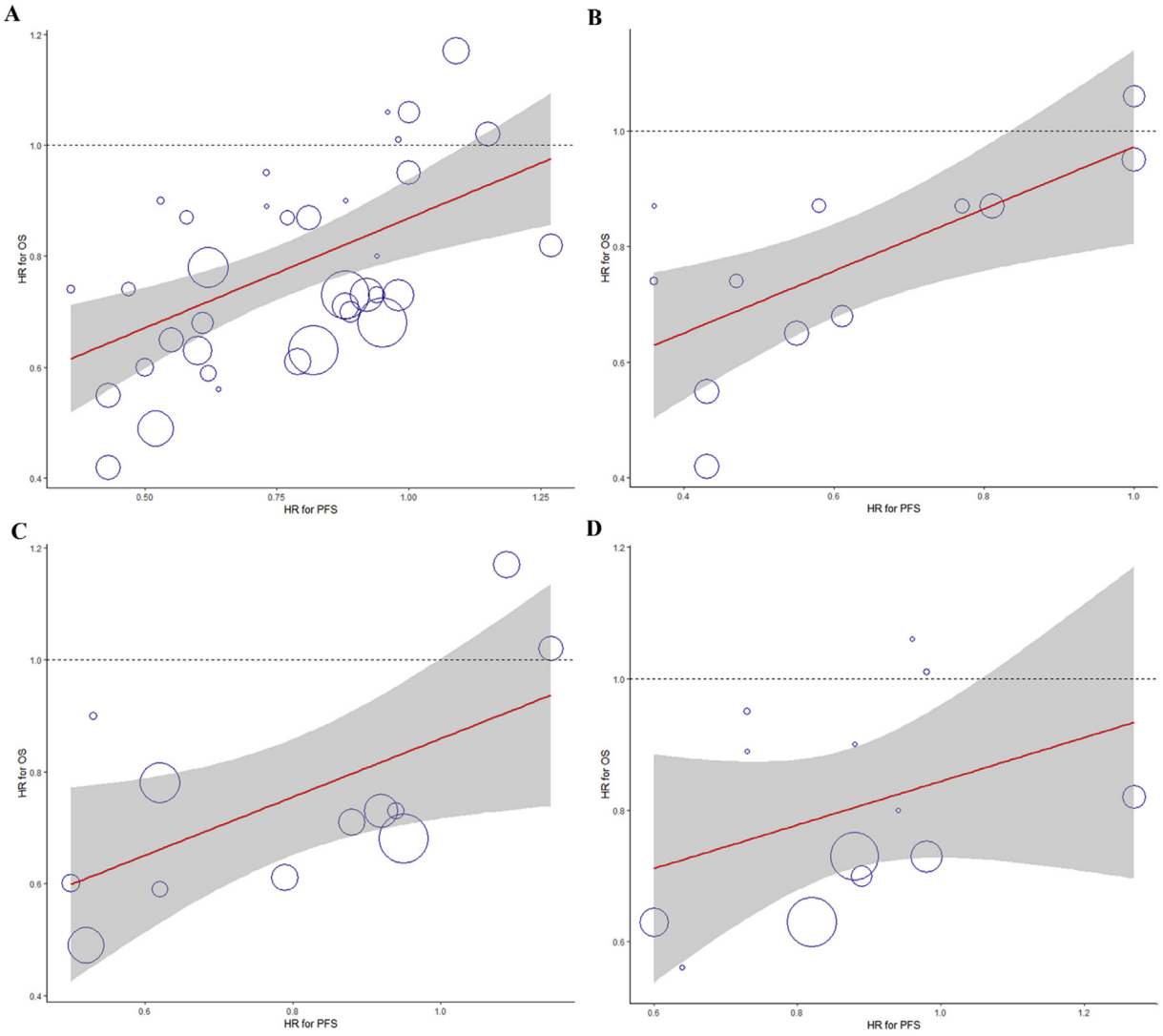

Fig. 2. Correlation between treatment effects on OS and PFS. Each trial is represented by a circle, with the size of the circle being proportional to the sample size. The gray area represents the 95% prediction limit of the regression line (red line). (A) Total tumour type (R^2 = 0.37, P < 0.001; STE = 1.07); (B) melanoma (R^2 = 0.61, P = 0.002; STE = 0.84); (C) NSCLC (R^2 = 0.44, P = 0.018; STE = 1.00) and (D) other cancer type (R^2 = 0.290, P = 0.058; STE = 1.06). OS, overall survival; PFS, progression-free survival; NSCLC, non–small-cell lung cancer; R^2, coefficient of determination; STE, surrogate threshold effect. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

are increasingly serving the surrogate for OS to reduce the sample size, follow-up duration and cost of trials; nonetheless, it is necessary to ensure that these end-points are valid surrogates for OS. In the immunotherapy trials, the valid surrogate end-point for OS is underdetermined; the value of conventional RECIST criteria–based end-points, such as PFS and ORR, as surrogate end-points requires further investigation because the mechanisms of action of immunotherapeutic agents are markedly different from cytotoxic agents. Notably, in the recent meta-analysis, Mushti *et al.* did not observe strong associations between PFS/ORR and OS [21]; their findings provided results from an individual patient-level analysis restricted to trials submitted for regulatory approval; thus, we expand the analyses to a broader range of trials, including negative studies, to further evaluate the value of RECIST criteria–based ORR, DCR and PFS as surrogate end-points for OS in this study. We noted a good correlation between PFS (r_s^2: 0.54) and OS and moderate correlations between ORR (r_s^2: 0.29) or DCR (r_s^2: 0.28) and OS. Additionally, in terms of treatment effects, we failed to observe strong correlations between RECIST criteria–based end-points and OS (R^2 ranging from 0.01 to 0.69); the leave-one-out cross-validation approach also confirmed that the effects observed on PFS were not adequate to predict the treatment effect on OS. Therefore, our analyses could further confirm that PFS, ORR and DCR could not be regarded as valid surrogate end-points for OS.

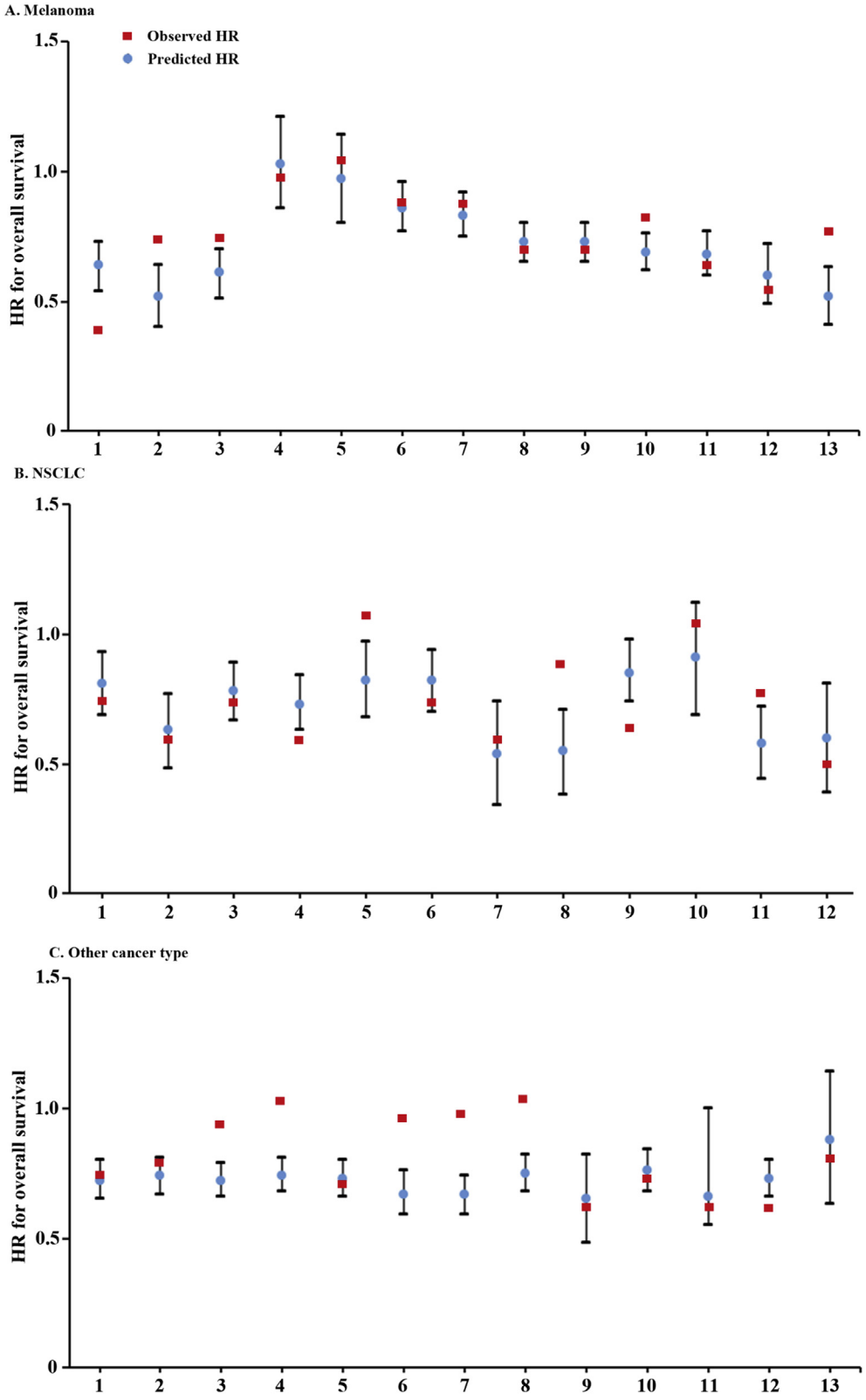

Fig. 3. Leave-one-out cross-validation analysis of the prediction of OS by treatment effect on PFS. (A) Melanoma; (B) NSCLC and (C) other cancer type. Predicted HRs for OS are calculated from the observed HR on PFS of that particular trial and the surrogate model built on all other trials. Observed HRs are shown for OS. All values are shown with 95% prediction intervals. OS, overall survival; PFS, progression-free survival; NSCLC, non–small-cell lung cancer; HR, hazard ratio.

In the most recent decade, the therapeutic blockade of the PD-1 signalling pathway has led to one of the greatest advances in the management strategies of cancer treatment, especially for advanced or recurrent tumours. Several clinical trials had identified favourable therapeutic effects of anti–PD-1/PD-L1 drugs for various malignancies, including NSCLC [15,16,20,31,32,42,48,55], melanoma [12,13,49,56,63], urothelial carcinoma [14,29] and other kinds of cancers [18,28,46]. It is well accepted that OS is the standard end-point for clinical trials; however, several phase 3 anti–PD-1/PD-L1 drug trials used PFS and ORR as the primary end-point [34,55] or coprimary end-point [12–18]. In the past 4 years, the FDA had approved pembrolizumab and nivolumab for advanced melanoma, atezolizumab for urothelial carcinoma and avelumab for Merkel cell cancer based on ORR benefit from these drugs, through accelerated and breakthrough filing pathways. This is suspicious because the correlation between ORR and OS was weak through our observation and Mushti's meta-analysis [21]. To evaluate the robustness of our findings, we performed sensitivity analyses with restriction to phase 3 trials, large trials (>200 patients), trials with mature follow-up (median follow-up ≥12 months), trials without crossover or crossover <50% and trials with first-line therapy; none of the sensitivity analyses could identified a strong correlation between the RECIST criteria–based end-points and OS. To the end, we used a leave-one-out cross-validation approach in each meta-analysis to evaluate the accuracy of the surrogate end-point in predicting OS. Disappointingly, the power of RECIST-based PFS in predicting OS is low. Therefore, in future immunotherapy trials, it is better to choose OS as the primary end-point until optimal valid surrogate end-point was identified. With the increment of anti–PD-1/PD-L1 drug trials of solid tumours, developing of better method to assess therapeutic effect of immunotherapy is still urgently needed.

It is well noted that, in contrast to chemotherapeutic agents, anti–PD-1/PD-L1 drugs exert the distinct antitumour effect through blocking the PD-1 pathway to activate the cytotoxic T-cell response; this can cause increased immune cell infiltration around a metastatic focus. Therefore, anti–PD-1/PD-L1 drugs may induce an initial increase in the tumour volume, and the delayed antitumour activity can, thereafter, produce late shrinkage of the tumour volume [20]; this may cause pseudoprogression and premature cessation of treatment. The weak correlations between ORR, DCR and OS in our analyses also suggested the pseudoprogression in the immunotherapy trials. In addition, Beaver *et al.* [64] explored the effect of continuation of treatment beyond RECIST-based progression in patients with unresectable or metastatic melanoma; they found that 95 patients (14%) in the treatment beyond progression cohort could have 30% or more decreased tumour burden and longer OS time, indicating that selected progression patients might still benefit from treatment with anti–PD-1 monoclonal antibody. Hence, optimisation for the RECIST criteria is needed for anti–PD-1/PD-L1 drug trials, and the immune-related response criteria (irRC), which incorporates the measurement of new lesions into the total tumour burden and describes additional patterns of the tumour response that may occur after the initial tumour expansion, may satisfy the demand [65,66]. However, because only four immunotherapy trials have reported irRC-based outcomes [47,52,57,59], we failed to assess the value of irRC-based end-points as surrogates for OS. In March 2017, the RECIST Working Group had published a consensus guideline, iRECIST, to provide a standard approach to assess the tumour response in immunotherapy trials; nonetheless, they also pointed out that the robustness of the iRECIST consensus guidelines required further validation through the initial collection of the immunotherapeutic warehouse [67]. Therefore, in future immunotherapy trials, iRECIST-based end-points are proposed to be reported.

Notably, some of the eligible trials allowed crossover at the time of disease progression on control therapy; this may result in positive results of PFS but negative results of OS. Flaherty *et al.* [68] demonstrated that the inclusion of trials with crossover could weaken the correlation between PFS and OS in advanced melanoma. Therefore, eliminating the potential effect of crossover is very vital to clarify the correlation between PFS and OS in immunotherapy trials. Thus, we performed the sensitivity analysis with restriction to trials without the crossover rate and trials with the crossover rate less than 50%. Nevertheless, the correlations between PFS and OS remained moderate, indicating that PFS is still not a robust surrogate for OS for trials without crossover.

There are few notable limitations of our study. Foremost, the immunotherapy was mainly delivered as multiple-line therapy in these trials; thus, heterogeneity may exist in our study. Hence, we performed sensitivity analyses with restriction to trials in first-line therapy; this confirmed that the correlation between PFS and OS was moderate ($R^2 = 0.45$). Next, many of the eligible trials have small sample sizes and short follow-up durations, which may result in fairly wide CIs of HR for treatment effects and thus confuse our results of correlation coefficients. Nonetheless, in sensitivity analyses with restriction to large-scale trials and trials with a long follow-up duration, the correlations between PFS and OS were still moderate. Finally, because only four trials have reported irRC-based outcomes, it is difficult to evaluate whether these end-points were valid surrogates for OS. Further studies were proposed to include enough trials to assess the value of the irRC-based or iRECIST-based outcomes.

In conclusion, the correlations between RECIST criteria–based end-points and OS were not strong; RECIST criteria–based end-points were not valid surrogates for OS in anti–PD-1/PD-L1 drug trials of advanced or recurrent solid tumours. At present, OS should be set as the primary end-point in immunotherapy trials.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Conflict of interest statement

None declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2018.10.011.

References

[1] Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. Annu Rev Immunol 2008;26:677–704.
[2] Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Canc 2012;12:252–64.
[3] Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science 2018;359:1350–5.
[4] Brahmer JR, Drake CG, Wollner I, et al. Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. J Clin Oncol 2010;28:3167–75.
[5] Garon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. N Engl J Med 2015;372:2018–28.
[6] Robert C, Ribas A, Wolchok JD, et al. Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab-refractory advanced melanoma: a randomised dose-comparison cohort of a phase 1 trial. Lancet 2014;384:1109–17.
[7] Wolchok JD, Kluger H, Callahan MK, et al. Nivolumab plus ipilimumab in advanced melanoma. N Engl J Med 2013;369:122–33.
[8] Le DT, Uram JN, Wang H, et al. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med 2015;372:2509–20.
[9] Kim ST, Cristescu R, Bass AJ, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. Nat Med 2018;24:1449–58.
[10] McDermott DF, Huseni MA, Atkins MB, et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. Nat Med 2018;24:749–57.
[11] Mauguen A, Pignon J-P, 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.
[12] 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;390:1853–62.
[13] Wolchok JD, Chiarion-Sileni V, Gonzalez R, et al. Overall survival with combined nivolumab and ipilimumab in advanced melanoma. N Engl J Med 2017;377:1345–56.
[14] Bellmunt J, de Wit R, Vaughn DJ, et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. N Engl J Med 2017;376:1015–26.
[15] Gandhi L, Rodriguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N Engl J Med 2018;378:2078–92.
[16] Socinski MA, Jotte RM, Cappuzzo F, et al. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. N Engl J Med 2018:2288–301.
[17] Shitara K, Ozguroglu M, Bang YJ, et al. Pembrolizumab versus paclitaxel for previously treated, advanced gastric or gastro-oesophageal junction cancer (KEYNOTE-061): a randomised, open-label, controlled, phase 3 trial. Lancet 2018:123–33.
[18] Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. N Engl J Med 2018;378:1277–90.
[19] Litiere S, Collette S, de Vries EG, et al. RECIST - learning from the past to build the future. Nat Rev Clin Oncol 2017;14:187–92.
[20] 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.
[21] Mushti SL, Mulkey F, Sridhara R. Evaluation of overall response rate and progression-free survival as potential surrogate endpoints for overall survival in immunotherapy trials. Clin Canc Res 2018;24:2268–75.
[22] Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Canc 2009;45:228–47.
[23] Buyse M, Molenberghs G, Burzykowski T, et al. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 2000;1:49–67.
[24] Buyse M, Burzykowski T, Michiels S, Carroll K. Individual- and trial-level surrogacy in colorectal cancer. Stat Methods Med Res 2008;17:467–75.
[25] A'Hern RP, Ebbs SR, Baum MB. Does chemotherapy improve survival in advanced breast cancer? A statistical overview. Br J Canc 1988;57:615–8.
[26] Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharmaceut Stat 2006;5:173–86.
[27] Julious SA, Campbell MJ, Walters SJ. Predicting where future means will lie based on the results of the current trial. Contemp Clin Trials 2007;28:352–7.
[28] Antonia SJ, López-Martin JA, Bendell J, et al. Nivolumab alone and nivolumab plus ipilimumab in recurrent small-cell lung cancer (CheckMate 032): a multicentre, open-label, phase 1/2 trial. Lancet Oncol 2016;17:883–95.
[29] Balar AV, Galsky MD, Rosenberg JE, et al. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. Lancet 2017;389:67–76.
[30] Bauml J, Seiwert TY, Pfister DG, et al. Pembrolizumab for platinum- and cetuximab-refractory head and neck cancer: results from a single-arm, phase II study. J Clin Oncol 2017;35:1542–9.
[31] 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.
[32] 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.
[33] Calabro L, Morra A, Giannarelli D, et al. Tremelimumab combined with durvalumab in patients with mesothelioma (NIBIT-

MESO-1): an open-label, non-randomised, phase 2 study. Lancet Respir Med 2018;6:451–60.
[34] 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.
[35] D'Angelo SP, Mahoney MR, Van Tine BA, et al. Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials. Lancet Oncol 2018;19:416–26.
[36] Fehrenbacher L, Spira A, Ballinger M, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. Lancet 2016;387:1837–46.
[37] Ferris RL, Blumenschein Jr G, Fayette J, et al. Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. N Engl J Med 2016;375:1856–67.
[38] Garassino MC, Cho BC, Kim JH, et al. Durvalumab as third-line or later treatment for advanced non-small-cell lung cancer (ATLANTIC): an open-label, single-arm, phase 2 study. Lancet Oncol 2018;19:521–36.
[39] Giaccone G, Kim C, Thompson J, et al. Pembrolizumab in patients with thymic carcinoma: a single-arm, single-centre, phase 2 study. Lancet Oncol 2018;19:347–55.
[40] Hamanishi J, Mandai M, Ikeda T, et al. Safety and antitumor activity of Anti-PD-1 antibody, nivolumab, in patients with platinum-resistant ovarian cancer. J Clin Oncol 2015;33:4015–22.
[41] 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 Canc 2017;86:37–45.
[42] Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 2016;387:1540–50.
[43] Hida T, Nishio M, Nogami N, et al. Efficacy and safety of nivolumab in Japanese patients with advanced or recurrent squamous non-small cell lung cancer. Cancer Sci 2017;108:1000–6.
[44] 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;17:1558–68.
[45] Kang YK, Boku N, Satoh T, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2017;390:2461–71.
[46] Kaufman HL, Russell J, Hamid O, et al. Avelumab in patients with chemotherapy-refractory metastatic Merkel cell carcinoma: a multicentre, single-group, open-label, phase 2 trial. Lancet Oncol 2016;17:1374–85.
[47] Kudo T, Hamamoto Y, Kato K, et al. Nivolumab treatment for oesophageal squamous-cell carcinoma: an open-label, multicentre, phase 2 trial. Lancet Oncol 2017;18:631–9.
[48] Langer CJ, Gadgeel SM, Borghaei H, et al. Carboplatin and pemetrexed with or without pembrolizumab for advanced, non-squamous non-small-cell lung cancer: a randomised, phase 2 cohort of the open-label KEYNOTE-021 study. Lancet Oncol 2016;17:1497–508.
[49] 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;36:383–90.
[50] Long GV, Atkinson V, Lo S, et al. Combination nivolumab and ipilimumab or nivolumab alone in melanoma brain metastases: a multicentre randomised phase 2 study. Lancet Oncol 2018;19:672–81.
[51] Motzer RJ, Escudier B, McDermott DF, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. N Engl J Med 2015;373:1803–13.
[52] Motzer RJ, Rini BI, McDermott DF, et al. Nivolumab for metastatic renal cell carcinoma: results of a randomized phase II trial. J Clin Oncol 2015;33:1430–7.
[53] Nishio M, Hida T, Atagi S, et al. Multicentre phase II study of nivolumab in Japanese patients with advanced or recurrent non-squamous non-small cell lung cancer. ESMO Open 2016;1,e000108.
[54] Peters S, Gettinger S, Johnson ML, et al. Phase II trial of atezolizumab as first-line or subsequent therapy for patients with programmed death-ligand 1-selected advanced non-small-cell lung cancer (BIRCH). J Clin Oncol 2017;35:2781–9.
[55] 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.
[56] Robert C, Long GV, Brady B, et al. Nivolumab in previously untreated melanoma without BRAF mutation. N Engl J Med 2015;372:320–30.
[57] Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet 2016;387:1909–20.
[58] Sharma P, Retz M, Siefker-Radtke A, et al. Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial. Lancet Oncol 2017;18:312–22.
[59] Tawbi HA, Burgess M, Bolejack V, et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. Lancet Oncol 2017;18:1493–501.
[60] Toulmonde M, Penel N, Adam J, et al. Use of PD-1 targeting, macrophage infiltration, and Ido pathway activation in sarcomas a phase 2 clinical trial. JAMA Oncology 2018;4:93–7.
[61] Weber JS, Gibney G, Sullivan RJ, et al. Sequential administration of nivolumab and ipilimumab with a planned switch in patients with advanced melanoma (CheckMate 064): an open-label, randomised, phase 2 trial. Lancet Oncol 2016;17:943–55.
[62] Zhu AX, Finn RS, Edeline J, et al. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. Lancet Oncol 2018:940–52.
[63] Hazarika M, Chuk MK, Theoret MR, et al. U.S. FDA approval summary: nivolumab for treatment of unresectable or metastatic melanoma following progression on ipilimumab. Clin Canc Res 2017;23:3484–8.
[64] 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;19:229–39.
[65] 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 Canc Res 2009;15:7412–20.
[66] 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;34:1510–7.
[67] 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.
[68] 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.