



Endpoint surrogacy in oncological randomized controlled trials with immunotherapies: a systematic review of trial-level and arm-level meta-analyses

Jianrong Zhang^{1,2#}, Wenhua Liang^{3#}, Hengrui Liang³, Xiaofei Wang², Jianxing He³

¹Brown School at Washington University in St. Louis, St. Louis, USA; ²Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, USA; ³Department of Thoracic Surgery and Oncology, First Affiliated Hospital of Guangzhou Medical University, Guangzhou Institute of Respiratory Disease & China State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou 510120, China

Contributions: (I) Conception and design: J Zhang, X Wang, W Liang; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: J Zhang; (V) Data analysis and interpretation: J Zhang, X Wang, W Liang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Prof. Xiaofei Wang, PhD. Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, USA. Email: xiaofei.wang@duke.edu; Prof. Jianxing He, MD, PhD, FACS, FRCS. Department of Thoracic Surgery and Oncology, First Affiliated Hospital of Guangzhou Medical University, Guangzhou Institute of Respiratory Disease & China State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou 510120, China. Email: drjianxing.he@gmail.com.

Abstract: Few cancer drugs or their indications achieved survival benefit in subsequent trials during postmarket period after approval based on surrogate endpoints. This causes a concern of using surrogate endpoints instead of overall survival (OS) as the primary endpoint for trial design, implementation and regulation approval. We conducted a systematic review to summarize the findings from published meta-analyses which have evaluated endpoint surrogacy for OS in oncological randomized controlled trials (RCTs) with immunotherapies. After searching articles indexed in PubMed prior to 24 February 2019, we identified a total of 11 meta-analyses for advanced multiple tumors, non-small cell lung cancer (NSCLC), urothelial carcinoma, renal cell carcinoma, melanoma; most (91%; 10/11) focused on immune checkpoint inhibitors. Although the evaluation criteria adopted by these meta-analyses for validating endpoint surrogacy were not consistent (ranging from $R^2 \geq 0.60$ to $R^2 \geq 0.80$), the results were consistent. Few studies show an association between OS and progression-free survival (PFS)/objective response rate (ORR) that met the lowest evaluation criteria ($R^2 \geq 0.60$), based on treatment effects (8%; 2/26 indications) or absolute results from experimental arm (0%; 0/11 indications). However, the association between OS and 1-year survival rate met the lowest criteria based on both the trial-level results (4/4 indications) and the arm-level results (5/5 indications). In lieu of this finding, we are supportive of an alternative endpoint, e.g., 1-year survival rate, rather than the more conventional choices PFS and ORR, as promising surrogate endpoint for OS in immunotherapy RCTs. We encourage further investigation on endpoint surrogacy based on the same or different settings, especially an assessment on survival rate at milestone time (e.g., 1-year), which has been demonstrated valuable for predicting OS in meta-analyses.

Keywords: Surrogate endpoints; randomized controlled trials (RCTs); cancer; immunotherapies

Submitted Dec 12, 2018. Accepted for publication Apr 08, 2019.

doi: 10.21037/atm.2019.04.72

View this article at: <http://dx.doi.org/10.21037/atm.2019.04.72>

Introduction

The regulatory approval of new drugs or new indications for treating diseases is ideally based on robust evidence of efficacy and safety from randomized controlled trials (RCTs). In oncology, the treatment effect of efficacy outcomes is essential for this approval process, as it reflects how clinically beneficial the new therapeutic strategy is for improving patients' outcome compared to the best alternative treatment. For example, pembrolizumab, an immunotherapy against programmed death 1 (PD-1), was proved to be clinically beneficial as a first-line treatment for advanced non-small cell lung cancer (NSCLC) compared to platinum-based chemotherapy in a phase III RCT in 2016 (1). In this trial, the primary endpoint was progression-free survival (PFS), and the secondary endpoints included overall survival (OS), objective response rate (ORR) and safety; all of these outcomes were measured with better results in pembrolizumab instead of chemotherapy (1). In the same year the United States Food and Drug Administration (FDA) approved this immunotherapy as a first-line treatment for NSCLC (2).

On the perspectives of both the FDA and the European Medicines Agency (EMA), OS is "the most reliable" and "the most persuasive" outcome among all efficacy endpoints in oncological RCTs (3,4). However, even though it has been used in RCTs as the traditional golden standard to evaluate treatment effect (5), its limitations are obvious. Compared with other efficacy endpoints, OS often requires a larger sample size, longer follow-up and higher cost. It also makes difficult to answer whether the new drug is more efficacious than the best alternative treatment during post-progression period, because after tumor progression subsequent therapies will be introduced instead of using the same originally assigned treatments (6,7). These drawbacks could be more evident when the treatment effect of the comparison treatments is not strong, or the nature of the disease history is lengthy (6).

Given the above limitations, other efficacy endpoints like PFS and ORR in RCTs have been chosen as surrogate endpoints for drug approval. These surrogate endpoints become important to demonstrate treatment effect earlier with less sample size and cost relative to OS. Considering these advantages, the FDA created a new pathway called "Breakthrough Therapy Designation", supported by the Safety and Innovation Act of 2012. The pathway expedites the approval process and the access to the medicine if the medicine "may demonstrate substantial improvement over

existing therapies on one or more clinically significant endpoints" (3). From January 2012 through December 2017, a total of 25 drugs for cancer diseases were approved by the FDA based on the Breakthrough Therapy Designation; among them, 96% (24/25) only focused on surrogate endpoints as primary endpoint in the trials (8).

Since surrogate endpoints have been used in RCTs for decades, it is important to know whether those approved drugs based on surrogate endpoints still lead to significant OS benefit in subsequent postmarketing trials. A study has been conducted to answer this question: only 14% (5/36) of drugs approved on the basis of surrogate endpoints achieved OS benefit in subsequent postmarketing trials (9). The proportion of OS benefit was also not high in another study: among 22 RCTs indicating PFS benefit, only 36% of them still achieved benefit based on mature OS in the follow-up reports (10).

The above findings reflect the concern of using surrogate endpoints as the only primary endpoint for trial implementation and drug approval. Therefore, meta-analyses that evaluated the validation of endpoint surrogacy for OS have been published consistently during the past 10 years (10-15). These meta-analyses were conducted mostly in the settings of targeted therapies or chemotherapies (10-15); surprisingly, most of them indicated a low strength of association between OS and traditional surrogate endpoints.

In the setting of immunotherapies, the association between OS and other endpoints revealed by meta-analyses has yet to be systematically summarized in a review. As we acknowledge, the therapeutic pattern of immunotherapies is different from targeted therapies and chemotherapies. Unlike chemotherapies, patients with immunotherapies (such as immune checkpoint inhibitors) could have a longer survival. Immunotherapies often have delayed treatment effects compared to chemotherapies and targeted therapies. Given these different patterns, the endpoints' surrogacy in the immunotherapy setting for patients with cancers deserves to be analyzed and summarized. This could be of great interest to clinicians, researchers, policymakers and pharmaceutical professionals, especially considering immunotherapies have showed a sufficient superiority of both efficacy and safety compared to the best alternative treatment in the new era of medicine (16-18).

Methods

We conducted the review in a compliance with the Preferred Reporting Items for Systematic Reviews and

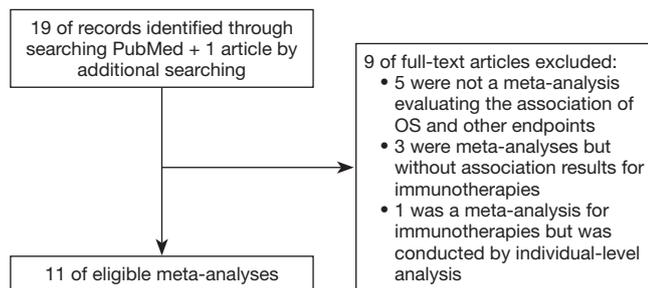


Figure 1 The PRISMA flow chart. PRISMA, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses; OS, overall survival.

Meta-Analyses (PRISMA) Statement (19). Articles were retrieved in PubMed according to the time interval from the inception date of the database to 24 February 2019. The search formula was: [Overall Survival (title/abstract)] AND (Endpoint* OR End Point*) AND (Immune OR Immunotherapy OR Immunotherapies) AND (Correlation or Linear) AND Trials. We also conducted an extra hand researching in PubMed for eligible article(s). There were no restrictions on language or publication type.

For inclusion criteria, eligible studies were meta-analyses that evaluated the association between OS and other endpoints only based on oncological RCTs, and that considered immunotherapy as the experimental treatment of each included RCT, and that used either correlation analysis or linear regression as statistical method. Since the highest level for evaluating endpoint surrogacy is trial-level (13), in this review we primarily considered trial-level association results, which were based on treatment effects [e.g., OS hazard ratio (HR) and PFS HR]; we also extracted arm-level association results, which were based on absolute results of immunotherapies within experimental arm (e.g., median OS and 12-month survival rate). As endpoint surrogacy in immunotherapies for cancer is a new topic, we did not have restrictions on the type of immunotherapies, type of non-OS endpoints, trial phase, treatment line, tumor type and tumor stage.

The following characteristics of eligible meta-analyses were extracted: author, publication year, source of included trials, retrieval time range, main inclusion criteria, statistical method (correlation analysis or linear regression), evaluation criteria for the association between endpoints, total number of included RCTs and patients, treatment phase, type of the immunotherapies (according to the inclusion criteria or the results of included RCTs), as well as trial-level and/or arm-

level association results. For the statistical method of each meta-analysis, we also identified more detailed information, including whether the method was weighted by sample size, or performed on a logarithmic scale.

Our primary interest was trial-level association based on treatment effects between OS and other endpoints, including PFS and ORR; our secondary interest was arm-level association based on absolute results within experimental arm. If the included meta-analyses used a correlation analysis with correlation coefficient (r) as measure, we calculated coefficient of determination (R^2) though taking the square of the value of r . Since there are no standard evaluation criteria for association results, we presented the exact results of included meta-analyses based on R^2 , and also made an exploratory evaluation on endpoint surrogacy, by using the lowest criteria used in our included meta-analyses.

Results

Through article searching, we found 19 articles in PubMed and one eligible meta-analysis by hand searching; among those 20 articles, five were not meta-analyses that evaluated the association of OS and other endpoints, three were meta-analyses but without association results for immunotherapies, and one was a meta-analysis for immunotherapies but it was conducted based on individual-level analysis. Therefore, a total of 11 studies were eligible according to our inclusion/exclusion criteria (*Figure 1*) (20-30). In these studies, surrogate endpoints included PFS (20,21,23-30), three-month PFS (29), 6-month PFS (20,23,29), ORR (20-25,27,28), disease control rate (DCR) (23,24), 1-year survival (26,28,30) and 2-year survival (30).

The characteristics related to study design are summarized in *Table 1*. Most included meta-analyses (91%; 10/11) were based on a systematic research through electronic databases for published articles, conference abstracts, and/or registered information; one meta-analysis (9%; 1/11) was conducted according to the convenience sample—the RCTs submitted to the FDA for approval (21).

Regarding the inclusion criteria of included meta-analyses, most included meta-analyses (91%; 10/11) investigated the trials with immune checkpoint inhibitors (5 for PD-1/PD-L1 inhibitor only), except one meta-analysis (9%; 1/11) included the trials with other immunotherapies like interferon (IFN)- α -2, IL-2, autologous cytokine-induced killer (29). As to tumor type, five meta-analyses (45%; 5/11) were conducted based on advanced multiple

Table 1 Characteristics of included meta-analyses for research design

Author & year	Source of included trials	Retrieval time range	Main inclusion criteria	Statistical method	Evaluation criteria for the association between endpoints
Ritchie <i>et al.</i> 2018 (20)	Articles, conference abstracts	Jan 2000 to Jan 2017	Advanced stage; solid tumors; checkpoint inhibitor [‡] ; phase II or III	C [‡] and L [‡]	No
Mushiti <i>et al.</i> 2018 (21)	Convenience sample*	2014 to 2016	PD-1/PD-L1 inhibitor; RECIST v1.1	C ^{‡†} and L ^{‡†}	R ² : ≥0.80 indicates a validated surrogacy
Roviello <i>et al.</i> 2017 (22)	Articles	NR to Feb 2017	Checkpoint inhibitor [‡] ; phase II or III; OS or PFS as primary endpoint; each arm with ≥10 patients	L ^{‡†}	No
Kaufman <i>et al.</i> 2018 (23)	Articles, conference abstracts and information in trials registries	Jan 2005 to Mar 2017	Melanoma, NSCLC, HNSCC, RCC, or urothelial carcinoma; Checkpoint inhibitor [‡]	L ^{‡†}	R ² : ≤0.49 as low, 0.49–0.72 as medium, ≥0.72 as high
Nie <i>et al.</i> 2019 (24)	Articles, and information in trials registries	NR to Jun 2018	Advanced or recurrent stage; PD-1/PD-L1 inhibitor; phase II or III	C [‡]	R ² : ≤0.25 as poor, 0.25–0.50 as moderate, 0.50–0.75 as good, ≥0.75 as strong
Shukuya <i>et al.</i> 2016 (25)	Articles, conference abstracts	Jan 2012 to Feb 2016	NSCLC; PD-1/PD-L1 inhibitor as monotherapy	C [‡]	No
Zhao <i>et al.</i> 2018 (26)	Articles, conference abstracts and information in trials registries	NR to May 2018	Advanced stage; NSCLC; second-line treatment	C, L ^{‡††}	R ² : >0.60 as a clinically relevant surrogacy
Ito <i>et al.</i> 2019 (27)	Articles, conference abstracts	No mention	Advanced stage; NSCLC; PD-1/PD-L1 inhibitor as monotherapy compared with chemotherapies; phase II or III; OR and HR stratified by PD-L1 expression	C [‡] and L [‡]	
Abdel-Rahman 2018 (28)	Articles, conference abstracts	NR to Aug 2017	Advanced stage; urothelial carcinoma; RCC; PD-1/PD-L1 inhibitor as monotherapy; median OS and ORR, PFS or 1-year survival rate	C	r: ≤0.19 as very weak, 0.20–0.39 as weak, 0.40–0.59 as moderate, 0.60–0.79 as strong, ≥0.80 as very strong
Bria <i>et al.</i> 2015 (29)	Articles, conference abstracts	NR to Dec 2013	Advanced stage; RCC; therapeutic strategies approved by FDA or EMA; phase III; 1st line	C and L	No
Petrelli <i>et al.</i> 2016 (30)	Articles	NR to Jul 2015	Metastatic stage; cutaneous melanoma only; checkpoint inhibitor [‡] ; phase II or III	C ^{‡†} and L ^{‡†}	r: 0.60–0.79 as strong, ≥0.80 as very strong

*, the randomized controlled trials which were submitted to the FDA for approval; †, PD-1/PD-L1/CTLA4 inhibitor; ‡, the statistical analysis was weighted by sample size; ††, the statistical analysis was performed on a logarithmic scale; ‡, multiple linear regression based on adjustment. NR, no restriction; C, correlation analysis; L, linear regression; RECIST, the Response Evaluation Criteria In Solid Tumors; OS, overall survival; PFS, progression-free survival; ORR, objective response rate; OR, odds ratio; HR, hazard ratio; FDA, the U.S. Food and Drug Administration; EMA, the European Medical Agency; TTP, time to progression; NSCLC, non-small cell lung cancer; HNSCC, head and neck squamous-cell carcinoma; RCC, renal cell carcinoma.

tumors (20-24), three of which had an additional analysis on NSCLC (20,22,23). Another six meta-analyses focused on specific advanced tumors [3 NSCLC (25-27), 1 urothelial carcinoma and renal cell carcinoma (28), 1 renal cell carcinoma alone (29), 1 melanoma (30)].

For statistical method, even though correlation analysis and/or linear regression were used, our included meta-analyses performed them not in an exactly same way: most meta-analyses (82%; 9/11) were weighted by sample size (20-27,30); among them, four meta-analyses performed the statistics on a logarithmic scale (21,22,26,30), and two conducted adjustment in linear regression model (23,26). For association results, six meta-analyses had evaluation criteria for considering validated endpoint surrogacy for OS, including $R^2 \geq 0.80$ (21), $R^2 \geq 0.75$ (24), $R^2 \geq 0.72$ (23), $R^2 \geq 0.64$ (28,30), and $R^2 \geq 0.60$ (26).

The association results of the 11 included meta-analyses are presented in *Table 2*. In the five meta-analyses based on multiple tumors, the median number of included trials and patients was 20 and 10,300, respectively; in another six meta-analyses based on specific tumors, the median number of included trials and patients was 9.5 and 2,926.5, respectively.

Association of trial-level treatment effects between endpoints

Among all 26 indications of treatment effects between OS and traditional surrogate endpoints (PFS and ORR), only 8% (2/26) of them met the lowest evaluation criteria ($R^2 \geq 0.60$) for endpoint surrogacy in all included studies. The R^2 of OS HR and PFS HR ranged from 0.13 to 0.50 (20,21,23,24), and that of OS HR and ORR OR ranged from <0.01 to 0.52 (20,21-24), in meta-analyses based on multiple cancers. In meta-analyses for NSCLC, R^2 of OS HR and PFS HR ranged from 0.40 to 0.85 (20,26,27), and that of OS HR and ORR OR ranged from <0.01 to 0.57 (20,22,27). In a meta-analysis for advanced renal cell carcinoma, the R^2 of OS HR and PFS HR was 0.40 (29). Apart from these, the R^2 of OS HR and DCR OR ranged from 0.08 to 0.96 in meta-analyses based on multiple cancers (23,24).

Three meta-analyses evaluated the surrogacy of survival rate for OS. Two were using the ratio of the 1-year survival rates between treatment arms, with the R^2 value of 0.78 and 0.64, respectively, between the ratio and the OS HR in trials with PD-1/PD-L1 inhibitor for advanced NSCLC and renal cell carcinoma (26,28). The other one was using the difference of the 1-year survival rates between treatment arms; the R^2 between the difference and the OS HR was 0.74

and 0.85, respectively, in trials with immune checkpoint inhibitors and ipilimumab only (CTLA4 inhibitor) for metastatic melanoma (30). In the same study, the difference of 2-year survival rates was also investigated for its association with OS HR: for immune checkpoint inhibitors, the R^2 was 0.69; for ipilimumab only, the R^2 was 0.25 (30).

Association of absolute results within experimental arm

Among all 11 indications between median OS and the outcomes of traditional surrogate endpoints (median PFS and ORR), none of them met the lowest evaluation criteria ($R^2 \geq 0.60$) for endpoint surrogacy in all included studies. The R^2 of median OS with median PFS, ORR, 1-year survival rate and 2-year survival rate ranged from <0.01 to 0.54 (24-27,28,30), 0.01 to 0.29 (24,25,28), 0.65 to 0.97 (26,28,30), 0.62 to 0.66 (30), respectively.

The R^2 of 12-month OS rate with six-month PFS rate ranged from 0.48 to 0.71 (20,29), and that with ORR ranged from <0.01 to 0.44 (20). The R^2 of 18-month OS rate with six-month PFS rate ranged from 0.58 to 0.95 (23), and the R^2 of nine-month OS rate with three-month PFS rate was 0.63 (29).

Discussion

Through summarizing the characteristics and results of previously published meta-analyses, this systematic review provides an overview of the association between overall survival and other endpoints in oncological RCTs with immunotherapies. These meta-analyses focused on advanced multiple tumors, NSCLC, urothelial carcinoma, renal cell carcinoma, and melanoma; most of them (91%; 10/11) investigated the trials with immune checkpoint inhibitors, including PD-1, PD-L1, or CTLA4 inhibitors. Even though their evaluation criteria for validating endpoint surrogacy were not consistent (ranging from $R^2 \geq 0.60$ to $R^2 \geq 0.80$), we found that their results were consistent. Overall, most the trial-level association results (92%; 24/26) between OS and traditional surrogate endpoints (PFS, ORR) did not meet the lowest evaluation criteria in included meta-analyses ($R^2 \geq 0.60$), exempt from some results for NSCLC. According to arm-level association results, none of all 11 indications met the lowest evaluation criteria ($R^2 \geq 0.60$) for endpoint surrogacy. However, the results between OS and 1-year survival met the lowest criteria in all indications regardless of trial-level results (4/4) or arm-level results (5/5), indicating its promising value to predict overall survival.

Table 2 Characteristics of included meta-analyses for association between endpoints

Author & year	Total trials/ patients (N)	Phase	Treatment (Inclusion criteria or results)	Association based on trial-level treatment effects, R ²	Association based on absolute results within experimental arm, R ²			
Multiple cancers								
Ritchie <i>et al.</i> 2018 (20)	20*/10,828	II, III	Exp: checkpoint inhibitor [§] ; Con: targeted therapy, chemotherapy or placebo	PFS HR & OS HR: 0.18 ORR OR & OS HR: 0.32	6-month PFS & 12-month OS rates: 0.55 ^{††} ORR & 12-month OS rate: 0.01			
		III	Exp: checkpoint inhibitor [§]	PFS HR & OS HR: 0.22 ORR OR & OS HR: 0.52	6-month PFS & 12-month OS rates: 0.52 ORR & 12-month OS rate: <0.01			
		II, III	Exp: PD-1/PD-L1 inhibitor	PFS HR & OS HR: 0.50 ORR OR & OS HR: 0.32	6-month PFS & 12-month OS rates: 0.48 ORR & 12-month OS rate: 0.18			
Mushti <i>et al.</i> 2018 (21)	13 [#] /6,722	NR	Exp: PD-1/PD-L1 inhibitor	PFS HR & OS HR: 0.13 ORR OR & OS HR: 0.13	–			
Roviello <i>et al.</i> 2017 (22)	17 [§] /8,994	II, III	Exp: checkpoint inhibitor [§] ; Con: chemotherapy or placebo	ORR OR & OS HR: 0.47	–			
			Exp: PD-1/PD-L1 inhibitor	ORR OR & OS HR: 0.18	–			
			Exp: CTLA4 inhibitor	ORR OR & OS HR: <0.01	–			
Kaufman <i>et al.</i> 2018 (23)	27 [†] /10,300	NR	Exp: checkpoint inhibitor [§] and/or other regimens	PFS HR & OS HR: 0.37 6-month PFS ratio & OS HR: 0.33 ORR OR & OS HR: 0.07 DCR OR & OS HR: 0.27	6-month PFS & 18-month OS rates: 0.62 – – –			
			Exp: checkpoint inhibitor [§] only	PFS HR & OS HR: 0.45 6-month PFS ratio & OS HR: 0.91 ^{††} ORR OR & OS HR: 0.27 DCR OR & OS HR: 0.96 ^{††}	– – – –			
			Exp: PD-1/PD-L1 inhibitor and/or other regimens	–	6-month PFS & 18-month OS rates: 0.70			
			Exp: PD-1/PD-L1 inhibitor only	–	6-month PFS & 18-month OS rates: 0.58			
			Exp: CTLA4 inhibitor and/or other regimens	–	6-month PFS & 18-month OS rates: 0.95 ^{††}			
			Nie <i>et al.</i> 2019 (24)	43 [†] /15,088	II, III	Exp: PD-1/PD-L1 inhibitor	PFS HR & OS HR: 0.37 ORR OR & OS HR: 0.10 DCR OR & OS HR: 0.08	Median PFS & Median OS: 0.54 ORR & median OS: 0.29 DCR & median OS: 0.28

Table 2 (continued)

Table 2 (continued)

Author & year	Total trials/ patients (N)	Phase	Treatment (Inclusion criteria or results)	Association based on trial-level treatment effects, R ²	Association based on absolute results within experimental arm, R ²
NSCLC					
Ritchie <i>et al.</i> 2018 (20)	8/NG	II, III	Exp: checkpoint inhibitor [§] ; Con: targeted therapy, chemotherapy or placebo	PFS HR & OS HR: 0.40 ORR OR & OS HR: 0.46	6-month PFS & 12-month OS rates: 0.58 ORR & 12-month OS rate: 0.44
Roviello <i>et al.</i> 2017 (22)	7/3,369	II, III	Exp: checkpoint inhibitor [§] ; Con: chemotherapy or placebo	ORR OR & OS HR: <0.01	–
Kaufman <i>et al.</i> 2018 (24)	NG	NR	Exp: checkpoint inhibitor [§] and/or other regimens	–	6-month PFS & 18-month OS rates: 0.81 ^{††}
			Exp: PD-1/PD-L1 inhibitor and/or other regimens	–	6-month PFS & 18-month OS rates: 0.77 ^{††}
			Exp: PD-1/PD-L1 inhibitor only	–	6-month PFS & 18-month OS rates: 0.67
Shukuya <i>et al.</i> 2016 (25)	10/2,379	I-III	Exp: PD-1/PD-L1 inhibitor monotherapy	–	Median PFS & median OS: 0.22
				–	ORR & median OS: 0.20
Zhao <i>et al.</i> 2018 (26)	5/3,024	NR	Exp: PD-1/PD-L1 inhibitor	PFS HR & OS HR: 0.84 ^{††}	Median PFS & median OS: 0.18
				1-year survival ratio & OS HR: 0.78 ^{††}	1-year survival rate & median OS: 0.97 ^{††}
Ito <i>et al.</i> 2019 (27)	7/3,752	II, III	Exp: PD-1/PD-L1 inhibitor monotherapy; Con: chemotherapy	PFS HR & OS HR: 0.85 ORR OR & OS HR: 0.57	– –
			Exp: PD-1/PD-L1 inhibitor monotherapy; Con: chemotherapy (subgroup with PD-L1 high expression)	PFS HR & OS HR: 0.59 ORR OR & OS HR: 0.25	– –
Advanced urothelial carcinoma					
Abdel-Rahman 2018 (28)	9/1,699	NR	Exp: PD-1/PD-L1 inhibitor monotherapy	–	Median PFS & median OS: <0.01
				–	ORR & median OS: 0.01
				–	1-year survival rate & median OS: 0.65 ^{††}

Table 2 (continued)

Table 2 (continued)

Author & year	Total trials/ patients (N)	Phase	Treatment (Inclusion criteria or results)	Association based on trial-level treatment effects, R ²	Association based on absolute results within experimental arm, R ²
Advanced renal cell carcinoma					
Abdel-Rahman 2018 (28)	4/1,093	NR	Exp: PD-1/PD-L1 inhibitor monotherapy	1-year survival ratio & OS HR: 0.64	Median PFS & median OS: 0.16 ORR & median OS: 0.16 1-year survival rate & median OS: 0.89 ^{††}
Bria <i>et al.</i> 2015 (29)	9/2,829	III	Exp ^{**} : NR (non-checkpoint inhibitor); Con ^{**} : NR	PFS HR & OS HR: 0.40	6-month PFS & 12-month OS rates: 0.71 ^{††} 3-month PFS & 9-month OS rates: 0.63 ^{††}
Metastatic melanoma					
Petrelli <i>et al.</i> 2016 (30)	13/3,373	II, III	Exp: checkpoint inhibitor [§] Exp: ipilimumab (CTLA4 inhibitor)	Delta 1-year survival & OS HR: 0.74 Delta 2-year survival & OS HR: 0.69 Delta 1-year survival & OS HR: 0.85 Delta 2-year survival & OS HR: 0.25	Median PFS & median OS: 0.20 1-year survival rate & median OS: 0.86 ^{††} 2-year survival rate & median OS: 0.62 Median PFS & median OS: 0.26 1-year survival rate & median OS: 0.86 2-year survival rate & median OS: 0.66

*, 9 NSCLC, 4 melanoma, 2 small cell lung cancer, 1 renal cell carcinoma, 1 head and neck cancer, 1 prostate cancer, 1 gastric cancer, 1 malignant mesothelioma; #, 6 melanoma, 5 NSCLC, 1 renal cell carcinoma, 1 head and neck cancer; §, 7 NSCLC, 4 melanoma, 3 small cell lung cancer, 1 renal cell carcinoma, 1 head and neck cancer, 1 prostate cancer; †, 14 melanoma, 9 NSCLC, 2 urothelial carcinoma, 1 renal cell carcinoma, 1 Head and neck squamous-cell carcinoma; ‡, 14 NSCLC, 8 melanoma (1 with brain metastases), 4 urothelial carcinoma, 3 renal cell carcinoma, 3 sarcoma (1 soft tissue/bone sarcoma, 1 soft tissue sarcoma, 1 sarcoma), 2 head and neck cancer, 1 mesothelioma, 1 thymic carcinoma, 1 hepatic cell carcinoma, 1 gastric cancer, 1 gastric or gastroesophageal junction cancer, 1 esophageal squamous cell carcinoma, 1 ovarian cancer, 1 small cell lung cancer, 1 Merkel cell carcinoma; §, PD-1/PD-L1/CTLA4 inhibitor; **, comparison of the four included trials for evaluating the trial-level association of treatment effects between PFS and OS: (Study 1) Interferon (IFN)- α -2 + 13-cis retinoic acid vs. IFN- α -2; (Study 2) IFN- α + IL-2 + 5-fluorouracil vs. IFN- α ; (Study 3) autologous cytokine-induced killer vs. IL-2 + IFN- α -2a; (Study 4) IFN- α + intravenous IL-2 vs. IFN- α + subcutaneous IL-2; ††, the results were accepted by the authors for considering endpoint surrogacy based on their evaluation criteria. NR, no restriction; Exp, experimental arm; Con, control arm; NG, not given; HR, hazard ratio; OR, odds ratio; DCR, disease control rate; Delta, difference.

One-year survival is a type of millstone survival, which was well introduced at the Brookings Conference on Clinical Cancer Research in Washington, DC (November 2013). Actually, millstone survival (including 1-year survival) is an intermediate endpoint of OS based on a cross-sectional assessment at a pre-specified time point during ongoing trials, in order to provide trialists with a first

glimpse of treatment efficacy and safety, especially in trials with long term survival and delayed treatment effects (31). Dislike chemotherapies, patients with immunotherapies (like immune checkpoint inhibitors) could have a longer survival; also, immunotherapies could present delayed treatment effects compared to chemotherapies and targeted therapies. Therefore, theoretically millstone survival could

be a great intermediate endpoint in oncological RCTs with immunotherapies, which is well supported by the results in our included meta-analyses. Nowadays, one acceptable recommendation for evaluating the strength of the association between endpoints is the criteria proposed by the German Institute of Quality and Efficiency in Health Care (IQWiG) guidelines (13-15,32). According to the criteria, if we score the strength as low, medium and high correlations when the R^2 value is lower than or equal to 0.49, between 0.49 and 0.72, as well as equal to or higher than 0.72, the value of 1-year survival is also convincing: most indications demonstrated a high correlation between OS and 1-year survival, not just based on trial-level treatment effects (75%; 3/4), but also based on absolute results within experimental arm (80%; 4/5).

Despite of 1-year survival, the value of other non-traditional endpoints needs more evidence in future investigations. Those non-traditional endpoints include 2-year survival, 6-month PFS, 3-month PFS, DCR, and a novel endpoint in an individual-level meta-analysis we found during our article searching, called the intermediate response endpoint (IME). It is defined as “having no nontarget lesion progression, no new lesion appearance, and reaching a target lesion response determined by baseline tumor burden, tumor reduction depth, and tumor change dynamics within 1 year after randomization”; the R^2 between OS HR and IME OR was 0.68 (33).

With respect to traditional surrogate endpoints, their associations with OS could be low. If we still use the IQWiG criteria, a low correlation between OS and PFS can be found in eight (67%) of 12 indications based on treatment effects, and in six (86%) of seven indications based on absolute results within experimental arm. The strength of the association between OS and another traditional surrogate endpoint—ORR—is also low in most indications: 86% (12/14) based on the treatment effects, and 100% (4/4) based on absolute results within experimental arm. In fact, low association between OS and traditional surrogate endpoints is not a rare issue in other treatments, including targeted therapies and chemotherapies. As we acknowledge, Sherrill *et al.* conducted the first related systematic review of meta-analyses in 2010, focusing on solid tumors (mostly advanced or metastatic) with disease-progression endpoints, which included PFS, ORR, time to progression (TTP), disease-free survival (DFS), or even-free survival (EFS) (11). Among 45 indications from 22 included meta-analyses, which had the results based on treatment effects, more than half (57.7%; 26/45) were considered as a low correlation ($R^2 \leq 0.49$). Another

four systematic reviews of meta-analyses were conducted subsequently, also focusing on the association between OS and surrogate endpoints in oncological RCTs. A low trial-level correlation can be found in 62.5% (10/16), 51.6% (32/62), 56.2% (73/130), 38% (34/89) of indications from included meta-analyses, respectively (11-15). Furthermore, the systematic review of meta-analyses conducted by Haslam *et al.* (15) also provided a subset of results based on immunotherapies: none of the four meta-analyses focusing on immunotherapies indicated high correlation; among these four meta-analyses, three (75%) indicated low correlation, based on the IQWiG criteria.

For such a low correlation of the trial-level association between OS and traditional surrogate endpoints (PFS and ORR), several considerations have been well explained by the five previous systematic reviews of meta-analyses (11-15) as well as our 11 included meta-analyses (20-30). Their considerations involve the issues of study design, such as the availability of subsequent treatment(s) after tumor progression (especially based on a cross-over study design), sample size, time of follow-up, and the time interval as well as the accuracy of tumor assessment. Further, their considerations involve the issues of clinical or biological situations, including cancer disease and treatment type. For example, in colorectal cancer and extensive small cell lung cancer, the association between OS and PFS tended to be strong; however, such a strong association was hard to be found in other cancer diseases (11-15). Also, the association could be different due to different treatments; interestingly, we found that our study based on the setting of immunotherapies has a higher proportion of indications with a low correlation between OS and traditional surrogate endpoints, compared to the previous five systematic reviews mostly based on the settings of targeted therapies and chemotherapies (11-15), according to the IQWiG criteria. This indicates that using traditional surrogate endpoints to predict patients' final survival lacks of sufficient evidence in the setting of immunotherapies (34).

Our review has some limitations that should be discussed. First, surrogate endpoints may present different levels of association with OS in different settings (e.g., treatment regimen, treatment line, cancer type, cancer stage); for example, in our review some of trial-level association results between OS and PFS indicated high correlation for NSCLC but not for multiple tumors or other specific cancer types. Therefore, meta-analyses by using RCTs for multiple tumors might attenuate the strength of association between OS and other endpoints for a specific cancer type. More

investigations for a specific cancer type are recommended. Second, our included meta-analyses did not consider the narrow time frame to observe OS improvements. This means that during the time when those meta-analyses investigated the strength of association between surrogate endpoints and OS, the results (HR) of OS in their included RCTs may be not mature. Furthermore, if those RCTs did not implement OS as primary endpoint (surrogate endpoint was primary endpoint and OS was secondary endpoint), not enough statistical power may exist for the results of OS so that the OS results may be not statistically significant and not as consistent as the results of surrogate endpoints (which could be significant). The above two conditions lead to a potential bias in meta-analysis. As our acknowledgement, none of our included meta-analyses mentioned that the included results of OS were based on enough period of follow-up for a mature estimation.

Third, our review was conducted only based on the progress of immunotherapies at the present time point; also, the results from included meta-analyses are limited to specific cancer types (advanced multiple tumors, NSCLC, urothelial carcinoma, renal cell carcinoma, melanoma) with insufficient number of RCTs (especially phase III RCTs with mature OS after a long period of follow-up). Since immunotherapies are still a new treatment that many trials are ongoing or being developed, more meta-analyses for validating endpoint surrogacy for OS in trials with these cancer types and other cancer types should be encouraged, including current surrogate endpoints or new endpoints, especially when mature OS results in those trials are published. Fourth, as there are no commonly acceptable criteria for evaluating the validation of surrogate endpoints for OS in oncological RCTs, in this review we described how our included meta-analyses evaluated their results (criteria ranging from $R^2 \geq 0.60$ to $R^2 \geq 0.80$), and used the lowest criteria ($R^2 \geq 0.60$) from included meta-analyses to evaluate all the results from these meta-analyses. Argument may be made that such criteria we used is not strict, but we are surprised that few results of OS and traditional surrogate endpoints (PFS and ORR) met the criteria.

Conclusions

In conclusion, due to few evidences to support a strong association between OS and traditional surrogate endpoints like PFS and ORR, we think that OS should be re-considered as the primary endpoint in RCTs with immunotherapies, in order to provide convincing

evidence to support clinical benefit from new cancer drugs/indications. However, given the benefit of surrogate endpoints, as well as the feature that surrogate endpoints may present different levels of association with OS in different settings (e.g., treatment regimen, treatment line, cancer type, cancer stage), we encourage more investigations on surrogate endpoints based on the same or different settings, especially an assessment on survival rate at milestone time (e.g., 1-year), which has been demonstrated valuable for predicting OS in meta-analyses.

Acknowledgments

We sincerely thank Ms. Carolyn Smith (Senior Tutor, The Writing Center, Washington University in St. Louis) for providing suggestions on manuscript writing.

Footnote

Conflicts of Interest: This review was presented at the SCT's 40th Annual Meeting on May 21, 2019, at New Orleans, USA.

References

1. Reck M, Rodríguez-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.
2. Pai-Scherf L, Blumenthal GM, Li H, et al. FDA Approval Summary: Pembrolizumab for Treatment of Metastatic Non-Small Cell Lung Cancer: First-Line Therapy and Beyond. *Oncologist* 2017;22:1392-9.
3. US Food and Drug Administration. Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Available online: <http://www.fda.gov/downloads/Drugs/Guidances/ucm071590.pdf>
4. European Medicines Agency. Guideline on the evaluation of anticancer medicinal products in man. Available online: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/11/WC500238764.pdf
5. Johnson JR, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol* 2003;21:1404-11.
6. Dancey JE, Dodd LE, Ford R, et al. Recommendations for the assessment of progression in randomised cancer treatment trials. *Eur J Cancer* 2009;45:281-9.
7. Zhang J. Application and Approval of Cancer Drugs in China: Acceleration Should be Kept in Progress. *AME*

- Med J 2018;3:57.
8. Puthumana J, Wallach JD, Ross JS. Clinical Trial Evidence Supporting FDA Approval of Drugs Granted Breakthrough Therapy Designation. *JAMA* 2018;320:301-3.
 9. Kim C, Prasad V. Cancer Drugs Approved on the Basis of a Surrogate End Point and Subsequent Overall Survival: An Analysis of 5 Years of US Food and Drug Administration Approvals. *JAMA Intern Med* 2015;175:1992-4.
 10. Wayant C, Vassar M. A comparison of matched interim analysis publications and final analysis publications oncology clinical trials. *Ann Oncol* 2018. [Epub ahead of print].
 11. Sherrill B, Kaye JA, Sandin R, et al. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *Onco Targets Ther* 2012;5:287-96.
 12. Davis S, Tappenden P, Cantrell A. A Review of Studies Examining the Relationship between Progression-Free Survival and Overall Survival in Advanced or Metastatic Cancer [Internet]. London: National Institute for Health and Care Excellence (NICE);2012 Aug. Available online: <http://www.ncbi.nlm.nih.gov/books/NBK425826/>
 13. Prasad V, Kim C, Burotto M, et al. The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses. *JAMA Intern Med* 2015;175:1389-98.
 14. Savina M, Gourgou S, Italiano A, et al. Meta-analyses evaluating surrogate endpoints for overall survival in cancer randomized trials: A critical review. *Crit Rev Oncol Hematol* 2018;123:21-41.
 15. Haslam A, Hey SP, Gill J, et al. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate end-points and overall survival in oncology. *Eur J Cancer* 2019;106:196-211.
 16. Emens LA, Ascierto PA, Darcy PK, et al. Cancer immunotherapy: Opportunities and challenges in the rapidly evolving clinical landscape. *Eur J Cancer* 2017;81:116-29.
 17. Farkona S, Diamandis EP, Blasutig IM. Cancer immunotherapy: the beginning of the end of cancer? *BMC Med* 2016;14:73.
 18. Whiteside TL, Demaria S, Rodriguez-Ruiz ME, et al. Emerging Opportunities and Challenges in Cancer Immunotherapy. *Clin Cancer Res* 2016;22:1845-55.
 19. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
 20. Ritchie G, Gasper H, Man J, et al. Defining the Most Appropriate Primary End Point in Phase 2 Trials of Immune Checkpoint Inhibitors for Advanced Solid Cancers: A Systematic Review and Meta-analysis. *JAMA Oncol* 2018;4:522-8.
 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 Cancer Res* 2018;24:2268-75.
 22. Roviello G, Andre F, Venturini S, et al. Response rate as a potential surrogate for survival and efficacy in patients treated with novel immune checkpoint inhibitors: A meta-regression of randomised prospective studies. *Eur J Cancer* 2017;86:257-65.
 23. Kaufman HL, Schwartz LH, William WN Jr, et al. Evaluation of classical clinical endpoints as surrogates for overall survival in patients treated with immune checkpoint blockers: a systematic review and meta-analysis. *J Cancer Res Clin Oncol* 2018;144:2245-61.
 24. 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.
 25. Shukuya T, Mori K, Amann JM, et al. Relationship between Overall Survival and Response or Progression-Free Survival in Advanced Non-Small Cell Lung Cancer Patients Treated with Anti-PD-1/PD-L1 Antibodies. *J Thorac Oncol* 2016;11:1927-39.
 26. Zhao S, Zhang Z, Zhang Y, et al. Progression-free survival and 1-year milestone survival as surrogates for overall survival in previously treated advanced non-small cell lung cancer. *Int J Cancer* 2019;144:2854-66.
 27. Ito K, Miura S, Sakaguchi T, et al. The impact of high PD-L1 expression on the surrogate endpoints and clinical outcomes of anti-PD-1/PD-L1 antibodies in non-small cell lung cancer. *Lung Cancer* 2019;128:113-9.
 28. Abdel-Rahman O. Surrogate end points for overall survival in trials of PD-(L)1 inhibitors for urinary cancers: a systematic review. *Immunotherapy* 2018;10:139-48.
 29. Bria E, Massari F, Maines F, et al. Progression-free survival as primary endpoint in randomized clinical trials of targeted agents for advanced renal cell carcinoma. Correlation with overall survival, benchmarking and power analysis. *Crit Rev Oncol Hematol* 2015;93:50-9.
 30. Petrelli F, Coinu A, Cabiddu M, et al. Early analysis of surrogate endpoints for metastatic melanoma in immune checkpoint inhibitor trials. *Medicine (Baltimore)* 2016;95:e3997.
 31. Chen TT. Milestone Survival: A Potential Intermediate

- Endpoint for Immune Checkpoint Inhibitors. *J Natl Cancer Inst* 2015. doi: 10.1093/jnci/djv156.
32. Institute for Quality and Efficiency in Health Care (IQWiG). Validity of surrogate endpoints in oncology: executive summary. Available online: http://www.iqwig.de/download/A10-05_Executive_Summary_v1-1_Surrogate_endpoints_in_oncology.pdf
33. Gao X, Zhang L, Sridhara R. Exploration of a Novel Intermediate Response Endpoint in Immunotherapy Clinical Studies. *Clin Cancer Res* 2018;24:2262-7.
34. Roviello G. To surrogate or not surrogate: an ancient dilemma without a happy ending. *J Hosp Manag Health Policy* 2019;3:6.

Cite this article as: Zhang J, Liang W, Liang H, Wang X, He J. Endpoint surrogacy in oncological randomized controlled trials with immunotherapies: a systematic review of trial-level and arm-level meta-analyses. *Ann Transl Med* 2019;7(11):244. doi: 10.21037/atm.2019.04.72