Cancer Therapy: Clinical

Evaluation of Overall Response Rate and Progression-Free Survival as Potential Surrogate Endpoints for Overall Survival in Immunotherapy Trials

Sirisha L. Mushti, Flora Mulkey, and Rajeshwari Sridhara

Abstract

Purpose: With the approval of immunotherapies for a variety of indications, methods to assess treatment benefit addressing the response patterns observed are important. We evaluated RECIST criteria–based overall response rate (ORR) and progression-free survival (PFS) as potential surrogate endpoints of overall survival (OS), and explored a modified definition of PFS by altering the threshold percentage determining disease progression to assess the association with survival benefit in immunotherapy trials.

Experimental Design: Thirteen randomized, multicenter, active-control trials containing immunotherapeutic agents submitted to the FDA were analyzed. Associations between treatment effects of ORR, PFS, modified PFS, and OS were evaluated at individual and trial levels. Patient-level responder analysis was performed for PFS and OS.

Results: The coefficient of determination (R^2) measured the strength of associations, where values near 1 imply surrogacy and values close to 0 suggest no association. At the trial level, the association between hazard ratios (HR) of PFS and OS was $R^2 = 0.1303$, and between the odds ratio (OR) of ORR and HR of OS was $R^2 = 0.1277$. At the individual level, the Spearman rank correlation coefficient between PFS and OS was 0.61. Trial-level associations between modified PFS and OS ranged between 0.07 and 0.1, and individual-level correlations were approximately 0.6. HRs of PFS and OS for responders versus nonresponders were 0.129 [95% confidence interval (CI), 0.11–0.15] and 0.118 (95% CI, 0.11–0.13), respectively.

Conclusions: Although responders exhibited longer survival and PFS than nonresponders, the trial-level and individual-level associations were weak between PFS/ORR and OS. Modifications to PFS did not improve associations. *Clin Cancer Res; 24(10); 2268–75. ©2018 AACR.*

See related commentary by Korn and Freidlin, p. 2239

Introduction

Immunotherapies, which utilize and enhance the distinctive powers of the immune system to fight cancer, are proving to be the most promising cancer treatments today. Although every cancer type is unique, immunotherapies have demonstrated various levels of clinical benefit across different cancer types. As a class, immunotherapies encompass several different types of treatments, such as checkpoint inhibitors and cancer therapeutic vaccines. Each of these treatments differ in their mechanism of action, and immune checkpoint inhibitors in particular represent a strong future for cancer treatment.

One of the first immunotherapies, ipilimumab, approved by the FDA in 2011, was the first checkpoint inhibitor to block the cytotoxic T-lymphocyte–associated antigen 4 (CTLA-4) pathway and was shown to extend survival among patients with advanced melanoma. Pembrolizumab and nivolumab, both checkpoint inhibitors that block the interaction between PD-1 and its ligands

PD-L1 and PD-L2, received accelerated approval (1, 2) from the FDA in the last quarter of 2014 based on overall response rate (ORR). Similarly, atezolizumab was approved in 2016 under accelerated approval regulation (3) based on ORR for both urothelial carcinoma and non–small cell lung cancer (NSCLC). In addition, avelumab and durvalumab, both PD-L1 agents, received accelerated approval in 2017 based on ORR for the treatment of urothelial carcinoma. Avelumab is also indicated for Merkel cell carcinoma. The ultimate goal of these agents is to prolong the life span of patients with cancer; however, some issues arise in considering the traditional efficacy endpoints for evaluating these drugs.

In evaluating efficacy, overall survival (OS) provides one of the most direct measures of true clinical benefit. OS, defined as the time from random assignment to the date of death from any cause, is a precise, objectively measured, and easy-to-interpret clinical outcome, and has been considered the most reliable and clinically meaningful endpoint for evaluating drug efficacy in oncology trials. However, due to the long-term survival advantages observed in patients treated with immunotherapies, OS may not be a realistic primary endpoint for future trials. In addition, OS may be confounded by subsequent treatments, such as a crossover to immunotherapy after disease progression. It is due to these limitations of the OS endpoint and the intent to provide patients with access to promising agents as early as possible that surrogate endpoints such as ORR and progression-free survival (PFS) have been useful in evaluating new treatments. These endpoints are

Office of Biostatistics, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland.

Corresponding Author: Sirisha L. Mushti, U.S. Food and Drug Administration, 10903, New Hampshire Avenue, Building-21, Room-3526, Silver Spring, MD 20993. Phone: 240-402-0244; E-mail: Sirisha.Mushti@fda.hhs.gov

AACR

Translational Relevance

In recent times, many immunotherapies, in particular checkpoint inhibitors, have demonstrated superior overall survival (OS) compared with existing standard of care. Future clinical trials (CT) may take much longer to complete in order to demonstrate incremental improvement in survival, and OS may not be a feasible endpoint for the evaluation of new drug products seeking marketing approval. This research evaluated if intermediate endpoints such as overall response rate, progression-free survival (PFS), or modified PFS can replace OS as a surrogate endpoint in determining the treatment effect of immunotherapies in randomized CTs. The retrospective analyses of past CTs suggest that at the individual level, a patient whose tumor responds or whose time to disease progression is longer lives longer. However, the analyses also suggest that in a randomized CT where an improvement in tumor response or improvement in PFS over existing standard of care is seen, the chances of demonstrating an improvement in OS is small.

primarily defined by changes in tumor burden and are assessable earlier, require shorter study durations and follow-up time, and may not need the larger sample sizes required in studies with OS as the primary endpoint. In addition, in single-arm trials, ORR is the only endpoint that can be meaningfully interpreted.

The most commonly used criteria to define progression and tumor response to treatment are RECIST (4, 5). However, the RECIST criteria were developed based on experience with cytotoxic agents, and it has been observed that the mechanisms of action of immunotherapeutic agents are markedly different from cytotoxic drugs. There is a need to evaluate if the RECIST criteria are appropriate to define ORR and PFS in patients receiving immunotherapies.

Atypical response patterns in patients receiving immunotherapies have led to challenges in evaluating clinical benefit using the traditional RECIST criteria. Patients may exhibit delay in response, immediate reduction to target tumor burden accompanied by presence of a new lesion, or even an initial increase in tumor burden (tumor flare or pseudo-progression) followed by a decrease in tumor burden. At the initial stage of treatment, a pseudo-progression may be observed, leading to discontinuation of treatment despite the potential of a delayed clinical response. To address the challenges observed in evaluating disease progression based on a RECIST-defined PFS endpoint, we evaluated ORR, PFS, and a modified RECIST criteria–defined PFS endpoint as potential surrogate endpoints for OS.

Several methods have been proposed in the literature by Prentice (6) and Buyse and colleagues (7–9) to validate surrogate endpoints (10–12). The associations can be characterized using a simple Spearman method when censoring is not an issue or evaluated using multitrial joint modeling methods, such as a copula bivariate survival model (13). In part 1 of our analysis, we present evaluations of RECIST criteria–based ORR and PFS as surrogate endpoints for OS using the individual-level and trial-level surrogacy described by Oba and colleagues (14) irrespective of site and histology of disease. Furthermore, using these same methods, we explore a modified RECIST criteria for PFS as a surrogate endpoint of OS in part 2.

Materials and Methods

We conducted a meta-analysis using the data from immunotherapy trials of anti–PD-1/PD-L1 agents submitted to the FDA as either initial or supplemental biological license applications for marketing approval across multiple indications between 2014 and 2016. For our analysis, we selected randomized, multicenter, active-controlled trials that were designed with head-to-head comparisons or add-on designs. Tumor assessments were made using RECIST v1.1 criteria for patients in the intent-to-treat population (as randomized) who had measurable lesions at baseline and at least one post-baseline visit.

The outcome measures considered in this analysis were OS, PFS, and ORR. OS was defined as the time since randomization to death. OS was censored at the last follow-up date for patients who were alive at the time of data cutoff. ORR was defined as the proportion of patients who achieved a complete or partial response as their best overall response based on RECIST v1.1 criteria. Patients with nonevaluable or unknown response status were considered nonresponders. PFS was defined as the time since randomization to progression or death, whichever occurred first. The censoring rules for PFS followed the censoring rules described in the FDA guidance on clinical trials endpoints (15). Disease progression (PD) was determined using criteria as described by RECIST v1.1, where a significant increase ($\geq 20\%$) in size of target lesions as compared with nadir with an absolute increase in size of at least 5 mm, development of new lesion (s), or unequivocal progression in an existing nontarget tumor was defined as progression.

In part 1 of the analysis, data and results based on ORR, PFS, and OS as submitted by the applicant were used. In part 2 of the analysis, a consistent approach was implemented using investigator-evaluated tumor assessments to define progression using the standard RECIST v1.1 criteria across all trials. The threshold percentage for defining PD due to target lesions, based on percent increase in sum of the longest diameters of all target lesions from nadir, was modified to range from the RECIST definition of 20% up to 45% in 5% increments. In this modified PFS analysis, the date of unequivocal progression of nontarget tumors and existence of new lesions were utilized as recorded in the database to indicate progression.

Surrogacy evaluations were performed at the trial and individual levels. The Spearman rank correlation coefficient was used to assess individual-level associations between PFS and OS considering the individual patient data from all the clinical trials. A patient-level responder analysis was performed to compare PFS and OS between responders and nonresponders irrespective of the treatment assignment using the pooled data set. In this analysis, the hazard ratios (HR) of PFS and OS were calculated using the Cox proportional hazards models stratified by study. Of note, this responder analysis was a nonrandomized comparison between the treatment groups among the responders.

For trial-level analyses, the associations between treatment effect based on ORR, PFS, and OS were evaluated using a weighted linear regression model with weights equal to the sample size of each randomized comparison. The treatment effects based on PFS and OS were measured by the log of HR estimated using an unstratified Cox proportional hazards model with treatment as the covariate. The odds ratio (OR) was considered when assessing the relative effect of ORR. The coefficient of determination (R^2) value computed from the weighted linear regression model was

used to measure the strength of association between PFS/ORR and OS. These associations are presented graphically using bubble plots. The trial-level association defined at the population level evaluates the ability of a surrogate endpoint to predict the effect of treatment on the true endpoint. An R^2 value close to 1 suggests a perfect surrogate endpoint, and an R^2 equal to 0 suggests no association between the surrogate endpoint and the true benefit endpoint. Because achieving $R^2 = 1$ is nearly impossible, we prospectively chose $R^2 = 0.80$ as the cutoff value to establish PFS and ORR as a validated surrogate endpoint for OS in this specific context. A surrogate endpoint is considered to be validated for use in phase III clinical trials when a strong association is demonstrated at both individual and trial levels.

Results

We identified 13 randomized clinical trials consisting of 6,722 patient records. Of note, there were four trials with three treatment arms (two experimental arms and one control arm). For the purpose of this analysis, these three-arm trials were treated as two separate trials each, duplicating the control arm in both trials. As a result, there were 17 randomized comparisons available for analysis from the 13 clinical trials selected. These trials included both head-to-head (treatment A vs. treatment B) comparisons, where an experimental monotherapy (treatment A) is compared directly against an active control (treatment B), and add-on (treatment A vs. treatment A + B) treatment comparisons, where an experimental agent is added to a standard-of-care background.

Table 1 summarizes the trials included in the analyses (16–28). For each study, the disease indication, primary or coprimary and secondary endpoints, total number of randomized patients and corresponding randomization allocation ratio, unstratified HR of OS, PFS as defined by RECIST criteria v1.1, unstratified OR for ORR, and response rates in the experimental versus control arm are listed.

Diseases represented in these trials include melanoma (6), NSCLC (5; 1 for squamous and 4 for nonsquamous), renal cell carcinoma (1), and head and neck cancer (1). In two of these trials, patients in the control arm were allowed to cross over to the experimental arm upon documented progression. The HRs ranged between 0.42 and 0.86 for OS and 0.41 and 0.93 for PFS, and the minimum OR for ORR was 1.67.

Part 1: Responder analysis between PFS and OS

Based on the pooled analysis of 6,722 patients, the ORR was 32% in the experimental groups and 12.7% in the control group. Figure 1A and B present the Kaplan–Meier plots for OS and PFS, respectively, for the responders and nonresponders in the experimental and control arms. The responders had longer survival times compared with nonresponders [HR = 0.14; 95% confidence interval (CI), 0.12–0.16], irrespective of treatment group. Similarly, the responders had longer PFS (HR = 0.12; 95% CI, 0.11–0.14).

Part 1: Individual-level and trial-level association results

The individual-level association based on rank correlation between PFS and OS was not strong ($P = 0.61$). The trial-level association between PFS and OS is shown in Fig. 2A. In this figure, each dot in the plot represents a single trial, with the size of each dot proportional to the size of the trial and color of the dot referring to the disease indication. The line represents the regression line fitted using the HR of PFS as a predictor of OS HR on a logarithmic scale. The trial-level R^2 value computed using the weighted linear regression model was 0.1303, demonstrating a weak association between PFS and OS. The trial-level R^2 value measuring the association between ORR and OS was only 0.1277, representing a weak association between ORR and OS (Fig. 2B). The OR for ORR in one study was an outlier (13.56; Table 1), and excluding this from the analysis resulted in an R^2 value of 0.1213. This suggests that, although responders lived longer without disease progression and death, both PFS and ORR are not good surrogate endpoints for OS.

Table 1. Summary of clinical trials

Randomized comparisons	Indication	Endpoints Primary	Endpoints Secondary	Number of patients	Randomization ratio	OS HR	PFS HR	ORR OR	Response rates (trt vs. control)
1	Melanoma	PFS (IRC), OS	ORR (IRC)	359	1:1	0.86	0.598	6.1957	22.5 vs. 4.5%
2	Melanoma	PFS (IRC), OS	ORR (IRC)	360	1:1	0.771	0.498	8.6163	28.7 vs. 4.5%
3	Melanoma	OS, PFS (IRC)	ORR (IRC)	557	1:1	0.673	0.567	3.542	41.3 vs. 16.6%
4	Melanoma	OS, PFS (IRC)	ORR (IRC)	555	1:1	0.67	0.581	3.3401	39.9 vs. 16.6%
5	Melanoma	ORR (IRC), OS	PFS (IRC)	182[a]	2:1	0.864	0.736	4.9762	31.2 vs. 8.3%
6	Melanoma	OS	ORR, PFS (INV)	418	1:1	0.418	0.432	4.1149	40 vs. 13.9%
7	Melanoma	PFS (INV), OS	ORR (INV)	631	1:1	0.708	0.572	3.2949	43.7 vs. 19%
8	Melanoma	PFS (INV), OS	ORR (INV)	629	1:1	0.702	0.432	5.7838	57.6 vs. 19%
9	Melanoma	ORR (INV)	PFS (INV)	142	2:1	0.792	0.396	13.5655	55.8 vs. 8.5%
10	NSCLC[b]	OS, PFS (IRC)	ORR (IRC)	687	1:1	0.732	0.869	1.6951	25.5 vs. 16.8%
11	NSCLC[b]	OS, PFS (IRC)	ORR (IRC)	689	1:1	0.631	0.788	1.694	25.5 vs. 16.8%
12	NSCLC[c]	PFS (IRC)	OS, ORR (IRC)	305	1:1	0.614	0.499	2.032	47.6 vs. 30.9%
13	NSCLC	OS	PFS, ORR (INV)	287	1:1	0.675	0.912	1.0476	15.3 vs. 14.7%
14	NSCLC non-SQ	OS	ORR, PFS (INV)	582	1:1	0.75	0.909	1.6742	19.2 vs. 12.4%
15	NSCLC SQ	OS	ORR, PFS (INV)	272	1:1	0.59	0.63	2.6042	20 vs. 8.8%
16	RCC	OS	ORR, PFS (INV)	821	1:1	0.758	0.873	5.9323	25.1 vs. 5.4%
17	SCCHN	OS	PFS, ORR (INV)	361	2:1	0.691	0.882	2.5055	13.3 vs. 5.8%

Abbreviations: INV, investigator; IRC, independent review committee; non-SQ, nonsquamous; RCC, renal cell carcinoma; SCCHN, squamous cell carcinoma of head and neck; SQ, squamous.
[a]Only 182 patients who had at least 6-month follow-up were considered in the computation of OS, PFS HRs, and ORR OR.
[b]Patient population: NSCLC with PD-L1 positive at 1%.
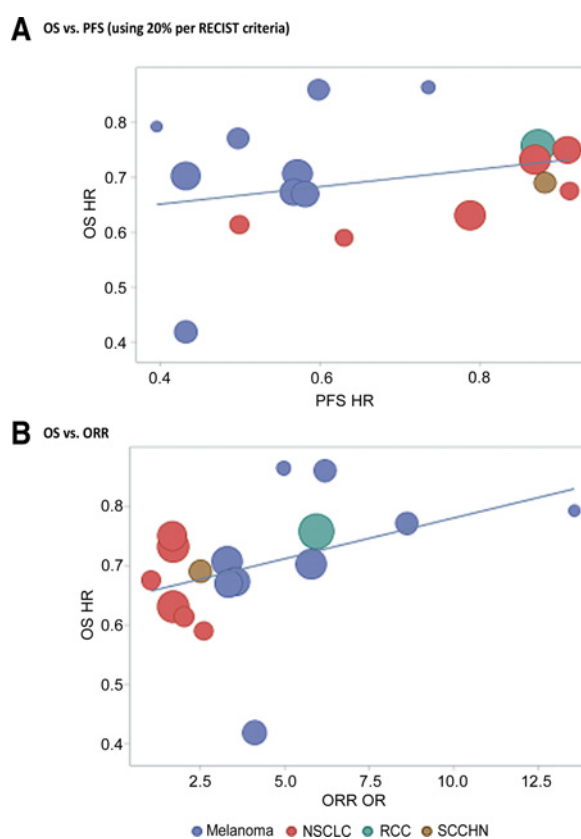[c]Patient population: NSCLC with PD-L1 positive at 50%.
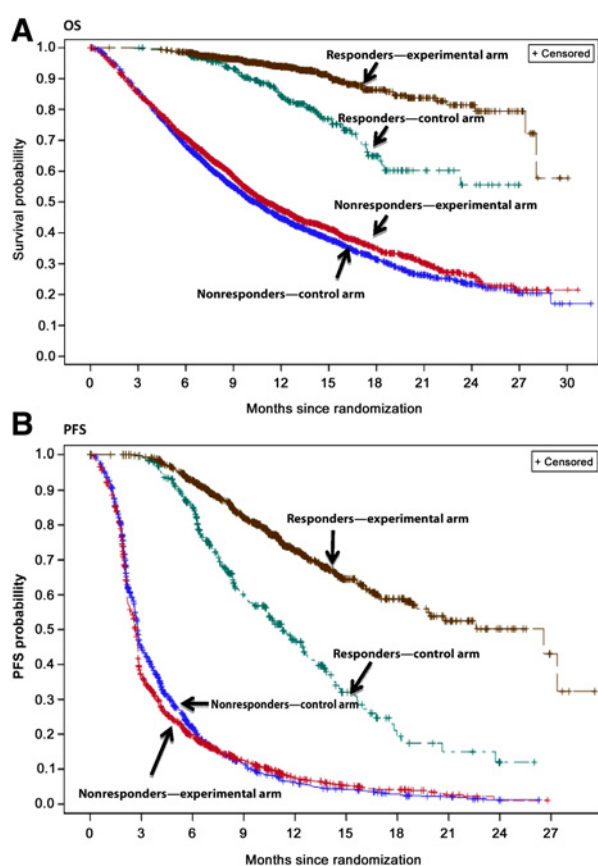
Figure 1.
Patient-level responder analysis results for OS and PFS.

Figure 2.
Trial-level association results.

Part 2: Modified PFS analysis

Using data from the 13 randomized trials, the typical disease progression assessment based on the standard RECIST v1.1 criteria in the treatment and control arms is summarized in Table 2. The primary component of tumor progression was further characterized based on target, nontarget, new lesion, or progression based on new and nontarget lesions together at the same assessment. A majority of the progression events (52%) were due to a significant increase ($\geq$20% per RECIST) in target tumor lesions.

By altering the threshold for progression from a 20% increase in sum of longest diameters in the target lesion from nadir to an increase of 25%, 30%, 35%, 40%, and 45%, we explored whether the higher burden threshold was more relevant in evaluating immunotherapeutic agents and as such, potentially better associated with OS. Table 3 presents the number of PFS events recalculated using the standard 20% cutoff as well as each modified threshold for target lesions. The first row in Table 3 contains the total PFS events from the pooled data. Beneath, results are further broken down into PFS events by study.

Due to the nature of this modification, the impact on assessment of PD and the number of PFS events decreased as the threshold values increased. The PFS events calculated using a higher threshold value were a subset of the PFS events calculated using a lower threshold value, because the criteria for PD due to target tumors were defined to meet a "minimum" percentage increase in the sum of target lesions. Thus, the PD defined using the 45% cutoff will also be included in the PD based on 20%, 25%, 30%, 35%, and 40%. The differences that were observed in number of PFS events by varying the cutoff values were due solely to the target tumor response. For example, when applying the 20% cutoff criteria (PFS_20), almost 52% of the progressions were based on target tumors, as shown in Table 2. When the cutoff value was increased to 25% (PFS_25), only 50% of progressions were due to increase in target tumor burden. For $n = 128$ patients, the target lesion increase in tumor size was $\geq$20% but <25%; thus, they were omitted when computing PFS_25. It is important to note that many of these trials allowed for treatment beyond initial RECIST-defined progression; thus, there is the potential for a patient whose initial value was between the 20% and 25% cut points to appear as a progression at a later date if his or her tumor burden increased to that higher threshold value.

Table 2. Reason for progression (using RECIST criteria of 20% for target tumors) in each treatment group

PFS events (20% cutoff)	Experimental	Control	Total
Deaths	437	361	798
Progression type[a]	2,328	1,715	4,043
Target tumor	1,194 (51%)	902 (53%)	2,096 (52%)
Nontarget tumor	235 (10%)	190 (11%)	425 (11%)
New lesion	676 (29%)	436 (25%)	1,112 (28%)
New lesion and nontarget	223 (10%)	187 (11%)	410 (10%)

[a]Follow-up postprogression was different in the treatment and control arms.

Mushti et al.

Table 3. PFS events using the modified RECIST criteria

| | Number of PFS events | | | | | |
N (%)	PFS_20 (20% cutoff)	PFS_25 (25% cutoff)	PFS_30 (30% cutoff)	PFS_35 (35% cutoff)	PFS_40 (40% cutoff)	PFS_45 (45% cutoff)
Overall	5,680 (72%)	5,529 (71%)	5,408 (69%)	5,303 (68%)	5,222 (67%)	5,147 (66%)
Study 1	309 (86%)	301 (84%)	293 (82%)	290 (81%)	288 (80%)	284 (79%)
Study 2	292 (81%)	284 (79%)	276 (77%)	272 (76%)	268 (74%)	265 (74%)
Study 3	376 (68%)	365 (66%)	356 (64%)	347 (62%)	339 (61%)	335 (60%)
Study 4	384 (69%)	375 (68%)	365 (66%)	358 (65%)	350 (63%)	346 (62%)
Study 5	112 (62%)	108 (59%)	105 (58%)	103 (57%)	101 (55%)	101 (55%)
Study 6	281 (67%)	276 (66%)	272 (65%)	266 (64%)	263 (63%)	259 (62%)
Study 7	414 (66%)	404 (64%)	397 (63%)	386 (61%)	384 (61%)	377 (60%)
Study 8	388 (62%)	380 (60%)	373 (59%)	363 (58%)	358 (57%)	352 (56%)
Study 9	70 (49%)	68 (48%)	67 (47%)	67 (47%)	67 (47%)	65 (46%)
Study 10	508 (74%)	495 (72%)	484 (71%)	476 (69%)	467 (68%)	462 (67%)
Study 11	505 (73%)	493 (72%)	481 (70%)	470 (68%)	457 (66%)	452 (66%)
Study 12	192 (63%)	188 (62%)	187 (61%)	185 (61%)	184 (60%)	182 (60%)
Study 13	241 (84%)	233 (81%)	230 (80%)	224 (78%)	220 (77%)	218 (76%)
Study 14	478 (82%)	463 (80%)	451 (77%)	445 (76%)	440 (76%)	430 (74%)
Study 15	229 (84%)	225 (83%)	223 (82%)	219 (81%)	218 (80%)	216 (79%)
Study 16	604 (74%)	582 (71%)	565 (69%)	555 (68%)	544 (66%)	534 (65%)
Study 17	297 (82%)	289 (80%)	283 (78%)	277 (77%)	274 (76%)	269 (75%)

Part 2: Individual-level association results between PFS and OS

The rank correlations between the observed values of PFS at different thresholds for progression and survival times of each patient were calculated to assess the individual-level associations and are presented in Fig. 3. A correlation value close to 0.6 suggests that the associations between each of the modified PFS endpoints and OS were not strong at the individual level, and the consistent correlation value suggests that the different thresholds had no impact on the associations. It is unclear what impact, if any, the nontarget and new lesion progressions played in contributing to the lack of association between PFS and OS; however, >45% of progressions were based on these types of response.

Part 2: Trial-level association results between PFS and OS

The trial-level association between OS and the traditional PFS endpoint, derived using investigator-evaluated tumor assessments to define progression based on standard RECIST criteria, was $R^2 = 0.0559$. Figure 3 shows the trial-level associations between each of the modified PFS and OS. The R^2 values are low, ranging between 0.07 and 0.1 for each of the different cutoff values considered. Similar to the results observed with the RECIST-defined PFS, the modified PFS did not result in an improved surrogate endpoint for OS.

Discussion

Intermediate endpoints based on tumor measurement, such as ORR and PFS, have been routinely used to evaluate new therapies in oncology trials and have served as the endpoints likely to predict clinical benefit for the consideration of accelerated approval of drugs. In recent immunotherapy studies, such as in nonsquamous NSCLC (23), despite a weak-to-moderate PFS treatment effect, significant improvement in the OS treatment effect was observed. In cases such as this, promising products may not have been approved if PFS was the primary endpoint of the study evaluating the product. Given the differing mechanisms of action of immunotherapy products compared with standard chemotherapy agents, the question arises if the definition of disease response and progression using standard RECIST v1.1 criteria are appropriate for evaluating this new and promising

class of products. In trying to answer this question, we first evaluated if ORR and PFS as currently defined using RECIST v1.1 were reasonable surrogate endpoints for OS, and second, in order to improve the degree of surrogacy, we used modified criteria for progression by imposing more stringent criteria.

We investigated if there were strong associations between PFS and OS, and ORR and OS based on 13 trials with 17 randomized comparisons. Based on the responder analysis, it was observed that responders were associated with longer survival and PFS durations compared with nonresponders. The analyses showed that the surrogacy evaluations at the trial level resulted in weak associations between the RECIST criteria–based endpoints of PFS and ORR with respect to OS.

Taking into account that there may be delayed response preceded by actual disease progression, modifying the burden level at which we determine progression could potentially address this issue. Even with higher burden levels, the surrogacy evaluations both at the trial and individual levels resulted in weak associations between the modified PFS and OS, similar to the RECIST-defined PFS. Thus, using a more stringent criterion to define PFS did not result in a better association between PFS and OS. Of note, the endpoints based on tumor response may not always capture the drug's effect on OS and hence may result in weak associations. Demonstrating stronger associations between the surrogate endpoints (ORR and PFS) and OS at the individual level is a necessary but not sufficient condition to conclude that ORR and PFS can replace OS in assessing the clinical benefit in immuno-oncology trials.

The challenges noted in using RECIST to evaluate immunotherapies have not gone unnoticed in the oncology community. There have been multiple attempts to modify the RECIST algorithm, from the Immune-Related Response Criteria (irRC 2009; ref. 29) to immune-related RECIST (irRECIST 2013–2014; refs. 30, 31), and finally the most current Immune RECIST (iRECIST 2017; ref. 32) guidelines. Many developers have also defined their own set of modifications (33). Most of these efforts have centered on how to include new lesions either in the tumor burden totals or as an element of progression, whether to use unidimensional or bidimensional measurements of target lesions, and how to proceed after conventional progression is
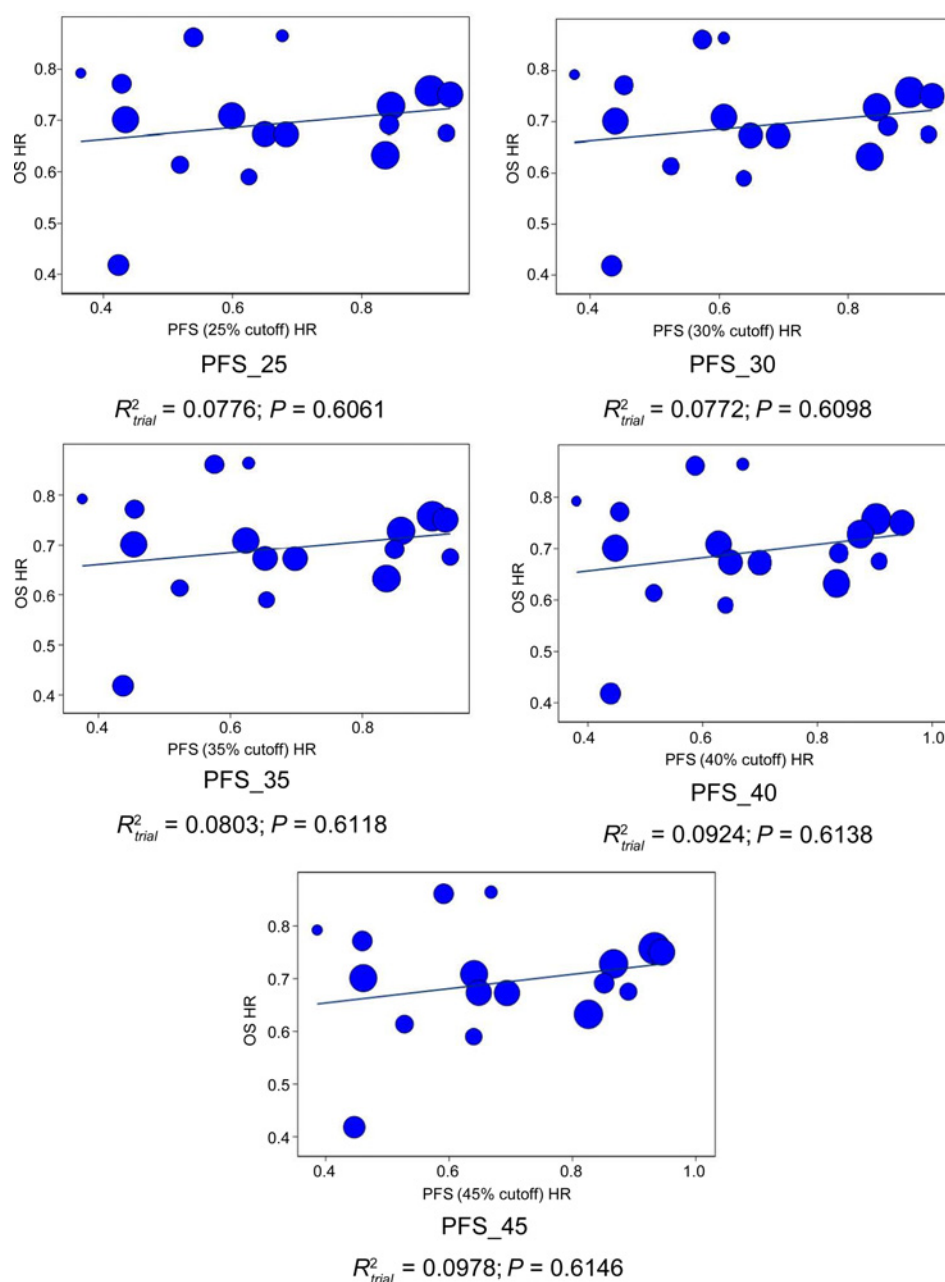
Figure 3.
Individual- and trial-level association between OS and modified PFS.

documented, that is, confirmation of PD or continuation until confirmation can be achieved. Creating a consensus on how to evaluate objective response and progression in immunotherapy trials in a consistent and effective manner is needed.

There are a few limitations noted in this analysis. First, the evaluations were based solely on the anti–PD-1/PD-L1 agents submitted to the FDA for approval, and hence all trials considered were positive and no negative trials were available to assess their impact on surrogacy. Second, in some studies, the appearance of new lesions in the patients treated with immunotherapeutic agents were no longer considered as unequivocal progression, and hence the therapy was continued for a longer duration and patients were followed past initial progression. However, the

control arms were not followed for similar durations, resulting in a difference in follow-up durations between the treatment and control arms. On the other hand, given the issues in assessing progression using standard RECIST criteria, an extended duration of immunotherapy treatments has facilitated the exploration of the treatment effects of these agents. Adherence to this new model of data collection will give an opportunity to make potential changes to RECIST, which was not possible before due to the nonavailability of assessments.

We also acknowledge that we have pooled all diseases together due to the small number of studies within each disease. The degree of association between PFS and OS may vary among different diseases. In the exploratory analysis, when assessed separately,

associations between PFS and OS in patients with melanoma were weaker than associations in patients with NSCLC (at PFS_45 $R^2 =$ 0.08 in melanoma studies vs. $R^2 = 0.59$ in NSCLC studies), although sample sizes among specific disease types are small for these methods and both values are still below any that would indicate surrogacy-level association (34). This hypothesis of differences in degree of association among the diseases needs further exploration with the accumulation of more data and experience.

The lack of association could be due to a subgroup of patients benefitting from the therapy. However, at this time, a well-defined subgroup of patients based on molecular biomarkers or clinical features who are more likely to respond has not been identified. In addition, the experience and length of follow-up in these studies are limited, as some of the trial level results included in this study are based on interim analyses. Extended follow-up with enhanced experience in multiple products can provide further exploration of intermediate endpoints as surrogate endpoints in the future.

Given the uncertainty that surrounds the use of PFS as an endpoint to assess survival benefit (35, 36) and the weak associations of PFS, ORR, and modified PFS endpoints with OS, as shown in this analysis, there is a need to further explore novel intermediate endpoints (37) that would better serve as a surrogate endpoint for OS in assessing the clinical benefit of immunotherapeutic agents and accelerate the drug approval process.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: S.L. Mushti, F. Mulkey, R. Sridhara
Development of methodology: S.L. Mushti, R. Sridhara
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.L. Mushti, R. Sridhara
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.L. Mushti, F. Mulkey, R. Sridhara
Writing, review, and/or revision of the manuscript: S.L. Mushti, F. Mulkey, R. Sridhara
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.L. Mushti, F. Mulkey, R. Sridhara
Study supervision: S.L. Mushti, R. Sridhara

References

1. Chuk MK, Chang JT, Theoret MR, Sampene E, He K, Weis SL, et al. FDA approval summary: accelerated approval of pembrolizumab for second-line treatment of metastatic melanoma. Clin Cancer Res 2017; 23:5666–70.
2. Hazarika M, Chuk MK, Theoret MR, Mushti S, He K, Weis SL, et al. FDA approval summary: nivolumab for treatment of unresectable or metastatic melanoma following progression on ipilimumab. Clin Cancer Res 2017; 23:3484–8.
3. Ning YM, Suzman D, Maher EV, Zhang L, Tang S, Ricks T, et al. FDA approval summary: atezolizumab for the treatment of patients with progressive advanced urothelial carcinoma after platinumcontaining chemotherapy. Oncologist 2017;22:743–9.
4. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent DJ, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228–47.
5. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors (RECIST guidelines). J Natl Cancer Inst 2000;92:205–16.
6. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989;8:431–40.
7. Buyse M, Burzykowski T, Carroll K, Michiels S, Sargent DJ, Miller LL, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. J Clin Oncol 2007;25:5218–24.
8. Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. Biometrics 1998;54:1014–29.
9. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 2000;1:49–67.
10. Bruzzi P, Del Mastro L, Sormani MP, Bastholt L, Danova M, Focan C, et al. Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients. J Clin Oncol 2005;23:5117–25.
11. Miksad RA, Zietemann V, Gothe R, Schwarzer R, Conrads-Frank A, Schnell-Inderst P, et al. Progression-free survival as a surrogate endpoint in advanced breast cancer. Int J Technol Assess Health Care 2008; 24:371–83.
12. Imai H, Mori K, Wakuda K, Ono A, Akamatsu H, Shukuya T, et al. Progression-free survival, post-progression survival, and tumor response as surrogate markers for overall survival in patients with extensive small cell lung cancer Ann Thorac Med 2015;10:61–6.
13. Chen Z. A flexible copula model for bivariate survival data [dissertation]. Rochester (NY): University of Rochester; 2014.
14. Oba K, Paoletti X, Alberts S, Bang YJ, Benedetti J, et al. Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. J Natl Cancer Inst 2013;105: 1600–7
15. Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. Silver Spring (MD): U.S. Food and Drug Administration; 2007. Available from: https://www.fda.gov/downloads/drugsGuidanceComplianceRegulatoyInformation/Guidance/UCM071590.pdf.
16. Ribas A, Puzanov I, Dummer R, Schadendorf D, Hamid O, Robert C, et al. Pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory melanoma (KEYNOTE-002): a randomised, controlled, phase 2 trial. Lancet Oncol 2015;16:908–18
17. Robert C, Schachter J, Long GV, Arance A, Grob JJ, Mortier L, et al. Pembrolizumab versus ipilimumab in advanced melanoma. N Engl J Med 2015;372:2521–32.
18. Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, 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.
19. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med 2016;375:1823–33
20. Brahmer J, Reckamp KL, Baas P, Crino L, Eberhardt WE, Poddubskaya E, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. N Engl J Med 2015;373:123–35.
21. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. N Engl J Med 2015;373:1803–13.
22. Weber JS, D'Angelo SP, Minor D, Hodi FS, Gutzmer R, Neyns B, et al. Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. Lancet Oncol 2015; 16:375–84.
23. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N Eng J Med 2015;373:1627–39.
24. Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, et al. Nivolumab in previously untreated melanoma without BRAF mutation. N Engl J Med 2015;372:320–30.
25. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. N Engl J Med 2015;373:23–34.

26. Postow MA, Chesney J, Pavlick AC, Robert C, Grossmann K, McDermott D, et al. Nivolumab and ipilimumab versus ipilimumab in untreated melanoma. N Engl J Med 2015;372:2006–17.

27. Ferris RL, Blumenschein G Jr, Fayette J, Guigay J, Colevas AD, Licitra L, et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. N Engl J Med 2016;375:1856–67.

28. Fehrenbacher L, Spira A, Ballinger M, Kowanetz M, Vansteenkiste J, Mazieres J, 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.

29. Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbé C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. Clin Cancer Res 2009;15:7412–20.

30. Nishino M, Giobbie-Hurder A, Gargano M, Suda M, Ramaiya NH, Hodi FS. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. Clin Cancer Res 2013;19:3936–43.

31. Bohnsack O, Ludajic K, Hoos A. Adaptation of the immune-related response criteria: irRECIST. Ann Oncol 2014;25:iv361–72.

32. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol 2017;18:e143–52

33. Ballinger M, Lyons B, Winslow N, Fine G, Chen DS. Non-classical response patterns, immune-modified RECIST and immune-modified PFS: potential optimization of radiographic endpoints for immno-therapies [workshop presentation]. In: Proceedings of the FDA-AACR Immuno-Oncology Drug Development Workshop; 2016 Oct 13–14; Washington, DC. Available from: http://www.aacr.org/advocacypo licy/governmentaffairs/documents/I-O%20Day2%20slidedeck-safe. pdf.

34. Shi Q, Flowers CR, Hiddemann W, Marcus R, Herold M, Hagenbeek A, et al. Thirty-month complete response as a surrogate end point in first-line follicular lymphoma therapy: an individual patient-level analysis of multiple randomized trials. J Clin Oncol 2017; 35:552–60.

35. Fiteni F, Westeel V, Pivot X, Borg C, Vernerey D, Bonnetain F. Endpoints in cancer clinical trials. J Visc Surg 2014;151:17–22.

36. Chakravarty A, Sridhara R. Use of progression-free survival as a surrogate marker in oncology trials: some regulatory issues. Stat Methods Med Res 2008;17:515.

37. Sridhara R. Challenges in interpreting results based on traditional endpoints. In: Proceedings of the FDA-AACR Immuno-Oncology Drug Development Workshop; 2016 Oct 13–14; Washington, DC. Available from: www.aacr.org/AdvocacyPolicy/GovernmentAffairs/ Documents/I-O%20Day2%20SlideDeck-safe.pdf.

Clinical Cancer Research

Evaluation of Overall Response Rate and Progression-Free Survival as Potential Surrogate Endpoints for Overall Survival in Immunotherapy Trials

Sirisha L. Mushti, Flora Mulkey and Rajeshwari Sridhara

Updated version	Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-17-1902

Cited articles	This article cites 32 articles, 7 of which you can access for free at: http://clincancerres.aacrjournals.org/content/24/10/2268.full#ref-list-1
Citing articles	This article has been cited by 7 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/24/10/2268.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://clincancerres.aacrjournals.org/content/24/10/2268. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.