Characterizing tumor shrinkage as a measure of clinical benefit for immune checkpoint inhibitors

Thomas Kelleher, Junliang Cai, Nicholas AJ Botwood, Dominic F Labriola

ABSTRACT

Background We explored whether the effectiveness of immune checkpoint inhibitors (ICIs) can be characterized by incorporating a composite of duration of response (DOR) to complement traditional Response Evaluation Criteria in Solid Tumors (RECIST) criteria for objective response rate (ORR) in an intent-to-treat (ITT) population. Furthermore, the correlation of this novel endpoint, characterized by the restricted mean time in response (RMTR), with overall survival (OS) will be examined.

Methods We analyzed ORR alone or in combination with DOR (RMTR) in available phase I, II, and III trials evaluating nivolumab monotherapy or in combination with ipilimumab across solid tumor types. ORR was evaluated per RECIST V.1.1. DOR was estimated using individual patient data in ITT populations regardless of RECIST response, with non-responders imputed as zero. Associations between ORR alone or RMTR and OS were evaluated in the ITT population. DOR curves were generated using the Kaplan-Meier product limit method, and 6-month RMTR, a measure of response durability, was derived from the area under the curves. For ORR and RMTR in the ITT population, the strength of association with OS was analyzed using Pearson correlation coefficients (r).

Results Nivolumab treatment was associated with longer response durations than active control in responder and ITT populations. Similarly, ORR and RMTR were both significantly correlated with OS (ORR vs OS: r=0.684, p=0.02; RMTR vs OS: r=0.695, p=0.018).

Conclusions Combining ORR and DOR (RMTR) to objectively characterize tumor shrinkage in an ITT patient population is a novel approach that appears to correlate well with OS in patients treated with nivolumab monotherapy or in combination with ipilimumab. This endpoint may provide a more complete characterization of tumor shrinkage to incorporate into the design of future ICI clinical trials. However, confirmation of this approach will require further research.

BACKGROUND

Immune checkpoint inhibitors (ICIs) have transformed the cancer treatment paradigm.[1 2] ICIs can display different response patterns from molecularly targeted and cytotoxic agents.[2–4] To address potential differences in response patterns, it may be important to select clinical trial designs, endpoints, and statistical methods that are appropriate for ICIs.[1] While overall survival (OS) is the gold standard for defining clinical benefit, the required long follow-up may delay approval of effective treatments for patients with urgent unmet needs.[2 5 6] Therefore, there is interest in identifying alternative endpoints that would help predict ICI response and survival benefits earlier, potentially expediting drug approval and patient access.[7] Additionally, methods to more effectively capture and characterize tumor shrinkage as a suitable indicator of clinical outcome are needed.[8 9]

Objective response rate (ORR) is sometimes accepted as an endpoint supporting accelerated drug approvals.[1 10] Reduction in tumor size is captured by ORR and considered by the US Food and Drug Administration (FDA) to be a predictor of clinical benefit, potentially allowing for earlier drug approvals than OS.[11] Between March 2011 and August 2017, the FDA approved 25 ICI indications in advanced solid cancers, 15 (60%) of which were approved using ORR as the primary endpoint.[1] ORR, however, has limitations associated with the categorical nature of Response Evaluation Criteria in Solid Tumors (RECIST) and does not completely characterize the duration (DOR) or depth (DepOR) of a response. While ORR is a direct measure of antitumor activity that can also be used to track changes in tumor burden status over time and provide a comprehensive assessment of disease status,[12] ORR alone may not fully characterize ICI responses and benefits.[13] DepOR is a more quantitative measure of tumor burden over time.[14] Additionally, DepOR only measures the change in the sum of target lesions, but not non-target lesions, from baseline.[15] The clinical significance and adequacy of ORR as a surrogate for marketing approval (accelerated and regular) are dependent on other factors, including the DOR, DepOR, and complete response (CR) rates.[16 17]

The FDA recommends DOR as a key factor when considering the adequacy of ORR to support approval.[10 16] As relationships between endpoints may vary based on tumor type and ICI treatment,[7] DOR can be tailored for milestone analyses appropriate to the natural history of the tumor type and ICI efficacy. DOR is described using median estimated rates with associated 95% CIs that are compared at the trial level using Kaplan-Meier (KM) plots.[18] For patient-level representations, spider plots display individual tumor changes over time relative to baseline burden.[19] Traditionally, these data representations of DOR have been based on the subset of responding patients, defined post randomization, and are thus prone to analysis-by-responder bias.[20] New approaches to visually and statistically describe DOR in all randomized patients (ITT) are needed to avoid this bias.

Although DOR alone has not been validated as an endpoint supporting drug approval, several ICIs received accelerated approval based on its combination with ORR.[21–24] A phase III trial of patients with melanoma receiving an oncolytic virus showed that durable responses (≥6 months) were associated with clinical benefits, including significant OS improvements and clinically meaningful quality of life improvements.[25] However, associations between ORR and DOR with OS for immunotherapy have not been established, and DOR neither as an independent measure of clinical benefit nor in combination with ORR has been fully evaluated.

We assessed DOR with nivolumab monotherapy or in combination with ipilimumab in responder and ITT populations using data from phase I, II, and III studies across solid tumors. To explore effective ways to visually represent and statistically analyze DOR, we evaluated the relationship between the combination of ORR and DOR (characterized by the restricted mean time in response (RMTR)) or ORR alone with OS in the ITT population.

We propose a novel approach combining ORR and DOR (RMTR) to objectively and visually characterize tumor shrinkage in the ITT population, which our results suggest correlates with OS and helps capture clinical responses to immunotherapy.

METHODS

We retrospectively analyzed available patient-level data from completed, large, phase III, active-control, registrational trials conducted by Bristol Myers Squibb (BMS) of nivolumab as monotherapy or in combination with ipilimumab using chemotherapy as a comparator for approved indications in non-small-cell lung cancer (NSCLC), melanoma, renal cell carcinoma (RCC), and squamous cell carcinoma of the head and neck (SCCHN). For further illustration, the DOR analyses also included phase I and II, single-arm, BMS trials from NSCLC (CheckMate 012) and melanoma (CheckMate 069) that were supportive for regulatory submissions and contained the necessary endpoints. Analyses of potential differences in DOR were performed in ITT and responder populations using data from select NSCLC and melanoma trials. We evaluated associations of OS with ORR alone and in combination with DOR, measured by the RMTR, using individual patient data from randomized phase III trials across NSCLC, melanoma, RCC, and SCCHN.

Definitions

Tumor responses were evaluated per RECIST V.1.1 by local investigators and/or blinded, independent, central review (BICR) (as appropriate for each study). DOR was the time between first CR or partial response (PR) and first documented tumor progression or death due to any cause.[26 27] Patients starting subsequent anticancer therapy without prior reported progression were censored for DOR at the last evaluable tumor assessment prior to initiating subsequent therapy. ORR was the proportion of patients who achieved a CR or PR. OS was the time from randomization to death from any cause; patients alive at data cut-off were censored at the last known alive date. The responder population only included patients with a CR or PR. The ITT analysis population included patients regardless of RECIST response (ie, CR, PR, stable disease, progressive disease, non-evaluable disease, or missing response), and non-responders were imputed with a value of zero.

DOR analyses

DOR was estimated using individual patient-level data in responder and ITT populations. DOR curves were generated using the KM product limit method for each treatment arm for the investigator-assessed and/or BICR-assessed responder population and the ITT population. Median DOR along with two-sided 95% CIs, based on log–log transformed CIs for the survivor function, were calculated. Treatment effects were estimated using the Cox-proportional hazards model to calculate HRs as has been done previously.[28–30] Statistical testing for differences in KM curves utilized the log-rank test. RMTR is the same as the area under the DOR curve, which is how it represents a robust measure of the average time in response for all randomized patients. Differences in RMTR can be used as a measure of treatment effect. Estimates of response rates at specific time points (0, 6, 12, 18, and 24 months) were calculated based on KM plots. The primary time point analyzed was RMTR through 6 months unless otherwise noted, as other time points provided similar results.

Analyses of associations between ORR and OS and the combination of ORR with DOR and OS

RMTR was calculated through 6 months using the area under DOR KM curves. Associations between OS HRs and the differences in ORR and RMTR were summarized by Pearson correlation coefficients (r) computed using the linear regression model, with r values close to 1 indicating strong associations.

RESULTS

Nine completed phase I, II, and III trials (n=3551) of nivolumab monotherapy or combination therapy with ipilimumab for NSCLC (CheckMate 012, 017, 026, 057, and 063) or melanoma (CheckMate 037, 066, 067, and 069) were selected to study DOR differences in patients treated with ICI therapy compared with chemotherapy (table 1). Associations between ORR and RMTR with OS were evaluated using individual patient data from nine randomized phase III studies (n=5323; table 1) across multiple indications (melanoma: CheckMate 037, 066, and 067; NSCLC: CheckMate 017, 026, and 057; RCC: CheckMate 025 and 214; SCCHN: CheckMate 141).

DOR analyses in patient responders from NSCLC and melanoma studies

Five completed NSCLC studies (CheckMate 012, 017, 026, 057, and 063) were included in DOR analyses of investigator-assessed and/or BICR-assessed responders (n=359).[4 31–34] Patients treated with nivolumab alone or combined with ipilimumab showed more durable responses than patients treated with available chemotherapy across all NSCLC studies (figure 1A).

DOR KM curves for all patient responders were also calculated for patients from four melanoma studies (n=585; CheckMate 037, 066, 067, and 069; figure 1B).[35–38] Similar to DOR analyses in NSCLC, nivolumab regimens produced durable responses in patients with melanoma, with prolonged DOR (>30 months) observed across all studies.

DOR analyses of the ITT population in NSCLC and melanoma studies

DOR was further analyzed in the ITT population (n=1641) from five NSCLC studies, which included all patients regardless of RECIST response with non-responders imputed as zero.[4 31–34] Nivolumab monotherapy or in combination with ipilimumab was associated with prolonged DOR in the NSCLC ITT population (figure 2A). This figure further illustrates how combining ORR with DOR can be presented visually and in a way that is more clinically informative than either endpoint alone. For example, the initial drop in the KM curve and the estimate at 0 months is the estimate of the ORR, while DOR is illustrated by the remainder of the KM curve; the visual representation tells more about the course of treatment over time (eg, altering natural history of the tumor). For example, KM curves that are flat over time represent patients with a longer DOR, while KM curves with many step-downs represent declining DOR over time. Therapeutic response can be assessed by both magnitude (ORR) and duration (DOR).[17] Combining ORR analyses with DOR can provide a comprehensive measure of tumor responses that is more clinically informative than either endpoint alone. For example, in figure 2B, the line for NIVO-037 crosses the 0-month x-axis at 26.5%, which represents the ORR (table 1), while the percentage of responders still in response at different time points (6,

12, and 30 months) is presented by the remainder of the DOR KM plot. Figure 2B further shows that nivolumab monotherapy and combination therapy with ipilimumab produced greater response rates with longer DOR (>30 months) observed across all studies versus active control.

Similar to results from NSCLC studies, DOR analyses of the ITT population (n=1548) from melanoma studies (figure 2B) showed that both nivolumab monotherapy and combination therapy with ipilimumab produced greater response rates with longer DOR (>30 months) observed across all studies versus active control.[35–38]

Combining ORR with DOR analyses in the ITT population

Estimates of RMTR were investigated in ITT populations from CheckMate 026, an NSCLC study that did not meet its primary endpoint of progression-free survival (PFS), and CheckMate 066, a melanoma study that met its primary endpoint of OS.[32] CheckMate 026 was used to demonstrate different ways that the combination of ORR with DOR can be summarized and compared in the ITT population for a negative study and to summarize RMTR estimates in the NSCLC population (figure 3A). Curves for nivolumab can be compared with chemotherapy using traditional log-rank tests (p=0.175) and treatment effects estimated with typical HRs and 95% CIs (0.98 (0.82 to 1.18)). Additionally, milestone 'survival' rates can be reported as KM estimates at specific time points (eg, 6, 12, and 18 months (20.1% vs 13.3%, 13.6% vs 7.0% and 9.2% vs 4.8%, respectively)). RMTR can also be assessed through certain time points; here, a 12-month time point was chosen to represent RMTR as the average time in response for the first year. Finally, treatment benefits can be estimated as differences in RMTR between groups and represent the difference in average time in response over the first year. At 12 months, RMTR (95% CI) was 2.4 (1.87 to 2.93) months for nivolumab and 2.2 (1.76 to 2.64) months for investigator's choice chemotherapy; a difference of 0.2 months in the first year (figure 3A).

CheckMate 066 was used to demonstrate combining ORR with DOR in the ITT population for a positive study and to summarize RMTR estimates in the melanoma population.[36] DOR curves for nivolumab versus dacarbazine were compared using traditional log-rank tests (p≤0.001), and treatment effects were estimated with a HR (95% CI) of 0.44 (0.35 to 0.55). Milestone survival rates were 38.8 versus 6.9 at 6 months, 35.1 versus 5.3 at 12 months, and 33.5 versus 5.3 at 18 months, respectively. At 12 months, the RMTR (95% CI) was 4.63 (3.87 to 5.4) months for nivolumab and 1.02 (0.61 to 1.43) months for dacarbazine, a difference of 3.6 months in the first year (figure 3B).

These two examples illustrate the value of RMTR. Figure 3A provides a scenario where RMTR (at time point 0), by incorporating DOR, shows a change in direction of the treatment comparison between nivolumab and investigator's choice chemotherapy relative to response rates. While RMTR shows an average time in response of 2.4 months over the first 12 months for nivolumab, compared

Table 1 Monotherapy and combination studies of nivolumab in NSCLC and melanoma included in DOR analyses

Study #/population	Study objective	Study design and treatment regimen	ORR per RECIST	Median OS, months HR (CI)
NSCLC				
CheckMate 012/chemo-naive recurrent stage IIIb or stage IV NSCLC[34]	To assess safety and tolerability	Phase I, randomized (1:1:1), open-label, multiple-cohort study of NIVO as monotherapy or in combination with IPI NIVO (3 mg/kg Q2W) +IPI (1 mg/kg Q12W) cohort 1 (n=38) NIVO (3 mg/kg Q2W) +IPI (1 mg/kg Q6W) cohort 2 (n=39)	47% (cohort 1) 38% (cohort 2)	Not available
CheckMate 017/previously treated advanced or metastatic SQ NSCLC after one prior platinum doublet-based chemo regimen[39]	To compare OS of NIVO with docetaxel	Phase III, randomized (1:1), open-label study of NIVO (n=135) vs docetaxel (n=137) NIVO: 3 mg/kg Q2W Docetaxel: 75mg/m² Q3W	20% vs 9%	9.2 v 6.0HR=0.59 (95% CI 0.44 to 0.79)
CheckMate 026/chemo-naive stage IV or recurrent PD-L1-expressing NSCLC[32]	To compare PFS (BICR) of NIVO with IC chemo in patients with ≥5% PD-L1+ tumor expression	Phase III, open-label, randomized study of NIVO (n=271) vs IC chemo (n=270) NIVO: 3 mg/kg IV Q2W IC chemo: gemcitabine with cisplatin or carboplatin; paclitaxel with carboplatin; pemetrexed with cisplatin or carboplatin Q3W	26% vs 33%	14.4 vs 13.2 HR=1.02 (95% CI 0.80 to 1.30)
CheckMate 057/previously treated advanced or metastatic NSQ NSCLC after one prior platinum doublet-based chemo regimen[4]	To compare OS of NIVO with docetaxel	Phase III, randomized (1:1), open-label study of NIVO (n=292) vs docetaxel (n=290) NIVO: 3 mg/kg Q2W Docetaxel: 75mg/m² Q3W	19% vs 12%	12.2 vs 9.4 HR=0.73 (96% CI 0.59 to 0.89)
CheckMate 063/advanced, refractory SQ NSCLC after platinum doublet-based chemo and at least one additional systemic therapy[33]	To assess ORR (BICR) of NIVO	Single-arm, phase II study of NIVO monotherapy (n=117) NIVO: 3 mg/kg IV Q2W	15%	8.2
Melanoma				
CheckMate 037/advanced melanoma s/p anti-CTLA-4 therapy, and if *BRAF* V[600] mutation +s/p BRAF inhibitor[37]	To estimate ORR in the NIVO group and compare OS of NIVO with IC	Phase III, randomized (2:1), open-label study of NIVO (n=272) vs IC chemo (n=133) NIVO: 3 mg/kg IV Q2W IC: dacarbazine 1000 mg/m² IV Q3W or carboplatin (AUC 6) IV and paclitaxel 175 mg/m² Q3W	26.5% vs 9.0%	15.74 vs 14.39 HR=0.95 (95.54% CI 0.73 to 1.24)
CheckMate 066/previously untreated, BRAF WT unresectable or metastatic melanoma[36]	To compare the OS of NIVO with dacarbazine	Phase III, randomized (1:1), double-blind study of NIVO (n=210) vs dacarbazine (n=208) NIVO: 3 mg/kg IV Q2W Dacarbazine: 1000mg/m² Q3W	40.0% vs 13.9%	NR vs 10.8 HR=0.42 (99.79% CI 0.25 to 0.73)

Table 1 Continued

Study #/population	Study objective	Study design and treatment regimen	ORR per RECIST	Median OS, months HR (CI)
CheckMate 067/previously untreated, unresectable, or metastatic melanoma[35]	To compare PFS and OS of NIVO monotherapy with IPI monotherapy, NIVO+IPI to IPI monotherapy	Phase III, randomized (1:1:1), double-blind study of NIVO (n=316) or NIVO+IPI (n=314) vs IPI (n=315). NIVO: 3 mg/kg IV Q2W. NIVO+IPI: NIVO 1 mg/kg+IPI 3 mg/kg Q3W for four doses then NIVO 3 mg/kg Q2W. IPI: 3 mg/kg Q3W for four doses	PD-L1 positive vs negative 57.5% vs 41.3% (NIVO). 72.1% vs 54.8% (NIVO+IPI). 21.3% vs 17.8% (IPI)	NR vs 19.98 (NIVO mono vs IPI mono) – HR=0.63 (98% CI 0.48 to 0.81)
CheckMate 069/previously untreated, unresectable, or metastatic melanoma[32]	To compare ORR, as determined by investigators, of NIVO+IPI with IPI monotherapy	Phase II, randomized (2:1), double-blind study of NIVO+IPI (n=95) vs IPI (n=47). NIVO+IPI: NIVO 1 mg/kg+IPI 3 mg/kg Q3W for four doses; then NIVO 3 mg/kg Q2W. IPI: 3 mg/kg Q3W for four doses	61.1% vs 10.8% among BRAF WT patients	NR vs NR HR=0.73 (95% CI 0.39 to 1.36)
SCCHN				
CheckMate 141/recurrent or metastatic platinum-refractory SCCHN[41]	To compare OS of NIVO monotherapy with IC	Phase III, randomized (2:1), open-label study of NIVO (n=240) vs IC (n=121). NIVO: 3 mg/kg IV Q2W. IC: cetuximab, methotrexate, or docetaxel	13.3% vs 5.8%	7.5 vs 5.1 HR=0.70 (97.73% CI 0.51 to 0.96)
RCC				
CheckMate 025/advanced or metastatic clear-cell RCC previously treated with 1–2 regimens of antiangiogenic therapy[40]	To compare OS of NIVO monotherapy with everolimus	Phase III, randomized (1:1), open-label study of NIVO (n=410) vs everolimus (n=411). NIVO: 3 mg/kg IV Q2W. Everolimus: 10mg PO daily	25% vs 5%	25.0 vs 19.6 HR=0.73 (98.5% CI 0.57 to 0.93)
CheckMate 214/previously untreated, advanced, or metastatic RCC[42]	To compare OS, PFS, and ORR of NIVO+IPI with sunitinib	Phase III, randomized (1:1) study of NIVO+IPI (n=550) vs sunitinib (n=546). NIVO+IPI: NIVO 3 mg/kg+IPI 1 mg/kg IV Q3W for four doses; then NIVO 3 mg/kg Q2W. Sunitinib: 50mg PO daily for 4 weeks, then 2 weeks off	39% vs 32%	NR vs 32.9 HR=0.68 (99.8% CI 0.49 to 0.95)

AUC, area under the curve; BICR, blinded independent central review; chemo, chemotherapy; CTLA-4, cytotoxic T lymphocyte antigen-4; DOR, duration of response; IC, investigator's choice; IPI, ipilimumab; IV, intravenous; MONO, monotherapy; NIVO, nivolumab; NR, not reached; NSCLC, non-small-cell lung cancer; NSQ, non-squamous; ORR, objective response rate; OS, overall survival; PD-L1, programmed death ligand 1; PFS, progression-free survival; PO, take by mouth; QXW, every X weeks; RCC, renal cell carcinoma; RECIST, Response Evaluation Criteria in Solid Tumors; SCCHN, squamous cell carcinoma of the head and neck; s/p, status post; SQ, squamous; WT, wild type.
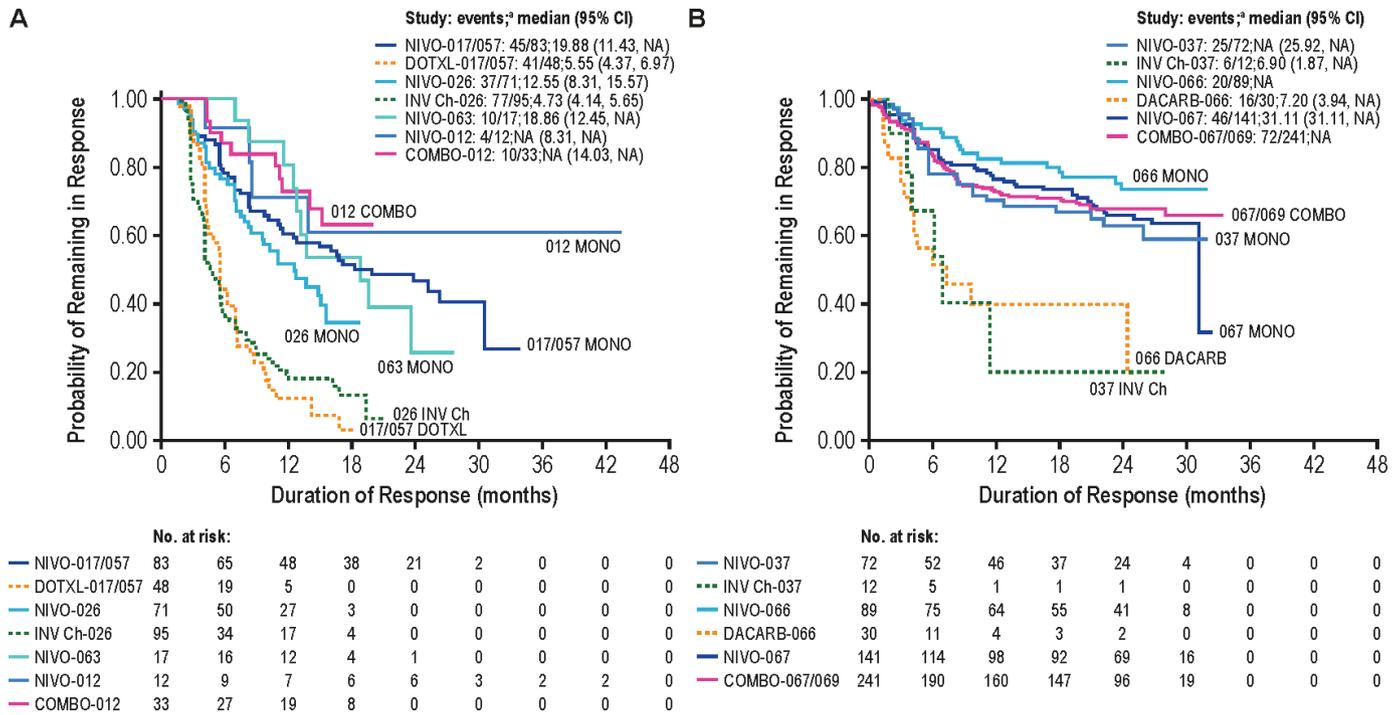
Figure 1 Kaplan-Meier plots of DOR in all responders in (A) NSCLC and (B) melanoma[b] studies. COMBO, combination NIVO+IPI therapy; DACARB, dacarbazine; DOR, duration of response; DOTXL, docetaxel; INV Ch, investigator's choice chemotherapy; IPI, ipilimumab; MONO, NIVO monotherapy; NA, not applicable; NIVO, nivolumab; NSCLC, non-small-cell lung cancer. [a]An event stands for non-responder or initial responder subsequently having lost response (eg, disease progression). [b]Each pairwise comparison from study 067 was included.
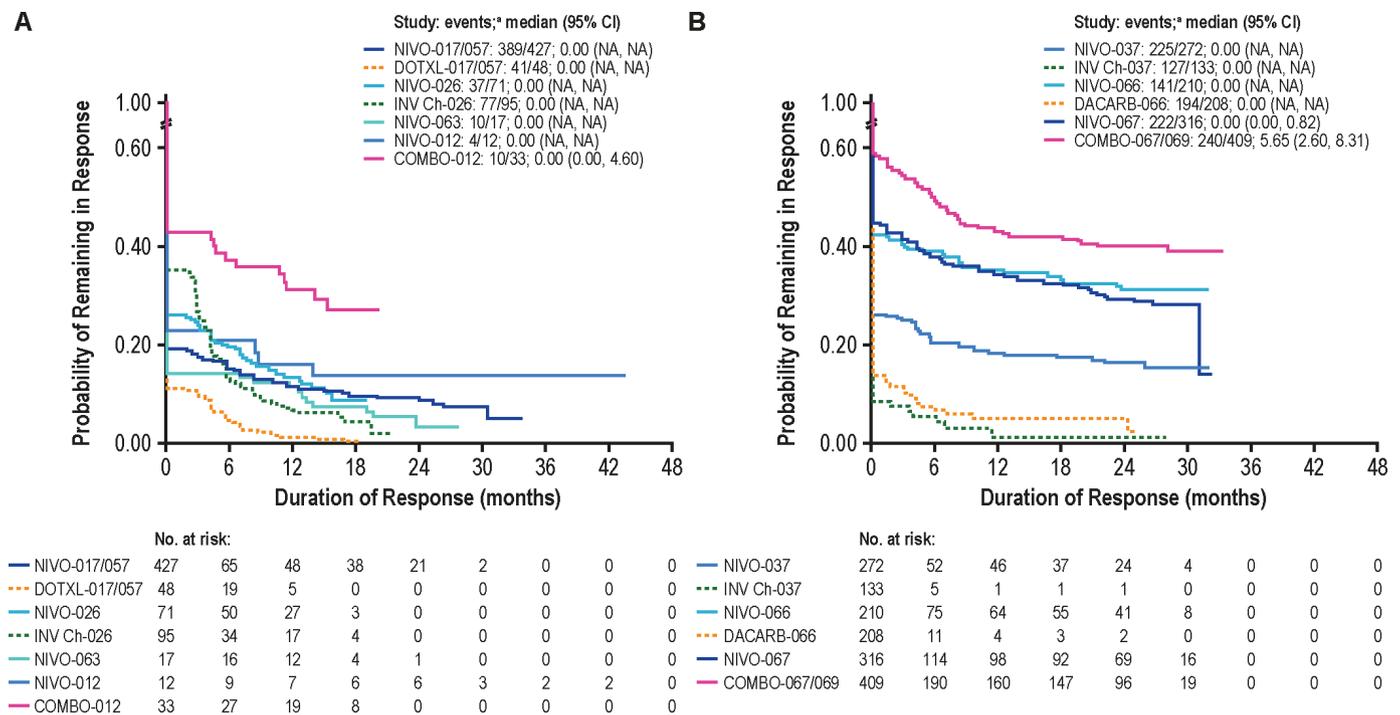
Figure 2 Kaplan-Meier plots of DOR analyses in the ITT population for all randomized patients with (A) NSCLC and (B) melanoma[b]. COMBO, combination NIVO+IPI therapy; DACARB, dacarbazine; DOR, duration of response; DOTXL, docetaxel; INV Ch, investigator's choice chemotherapy; IPI, ipilimumab; ITT, intent-to-treat; NA, not applicable; NIVO, nivolumab; NSCLC, non-small-cell lung cancer. [a]An event stands for non-responder or initial responder subsequently having lost response (eg, disease progression). [b]Study 069 did not contain a NIVO monotherapy arm.
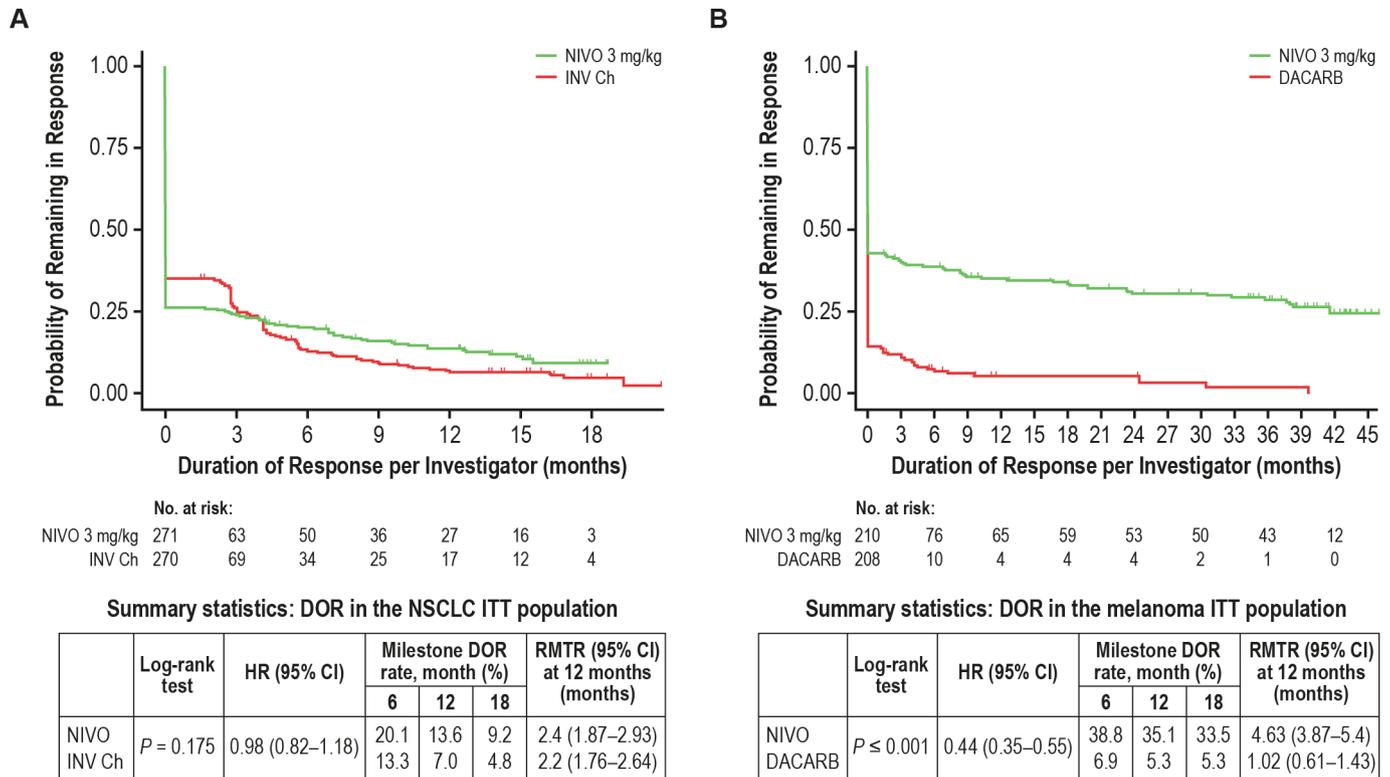
Figure 3 (A) DOR for the NSCLC ITT population treated with nivolumab (green curve) compared with investigator's choice chemotherapy (red curve) in the CheckMate 026 trial. (B) DOR for the melanoma ITT population treated with nivolumab (green curve) or dacarbazine (red curve) in the CheckMate 066 trial. DACARB, dacarbazine; DOR, duration of response; INV Ch, investigator's choice chemotherapy; ITT, intent-to-treat; NIVO, nivolumab; NSCLC, non-small-cell lung cancer; RMTR, restricted mean time in response.

with an RMTR of only 2.2 months for investigator's choice chemotherapy, response rates were ~25% and ~33% for nivolumab versus investigator's choice chemotherapy, respectively, as seen from at time point 0 in the figure. In contrast to this first example, figure 3B provides an example of more consistent treatment comparisons when analyzing ORR and RMTR. Response rates of ~40% and ~14% were observed for nivolumab and dacarbazine, respectively, while RMTRs were 4.63 and 1.02 months over the first 12 months for nivolumab and dacarbazine, respectively.

Associations between OS and RMTR and between ORR and OS in the ITT population across different tumor types

A total of 5323 patients in 9 phase III randomized studies across tumor types (NSCLC, melanoma, RCC, and SCCHN) were evaluated for OS, ORR, and RMTR.[4 32 35–42] Table 1 summarizes the objectives, study designs, treatment regimens, and previously reported ORR and OS results from these studies.

The relationships between treatment-related differences in RMTR and ORR, separately, with OS were assessed in the ITT population (figure 4). A significant correlation between 6-month treatment differences in ORR and HRs for OS was observed ($r=0.684$; p=0.02; figure 4A). Similarly, a comparable significant correlation was observed between OS HRs and RMTR treatment differences at the 6-month cut-off ($r=0.695$; p=0.018;

figure 4B). Correlations between RMTRs and HRs for OS at other cutoffs (3, 12, 18, and 24 months) were similar (figure 5).

DISCUSSION

These results demonstrate how tumor shrinkage can be characterized using DOR or a combination of DOR and ORR (RMTR) in the ITT populations of patients treated with ICIs across multiple tumor types. Combining ORR analyses with DOR can provide a comprehensive measure of tumor responses that cannot be obtained from ORR analyses alone. Nivolumab monotherapy or combination therapy with ipilimumab appears to be associated with a longer DOR versus active control in both responder and ITT populations. To our knowledge, this is the first report of an association between RMTR and OS in patients from an ITT population treated with ICIs. The novel endpoint RMTR may influence clinical trial design and complement traditional endpoints such as OS to capture responses to immunotherapy.

RMTR is a robust, clinically interpretable measure of the average time in response that can be applied to an ITT patient population and used as an alternative measure of treatment effect.[43] As an endpoint, RMTR may be an earlier efficacy readout and may capture unique aspects of ICI treatment benefits compared with traditional
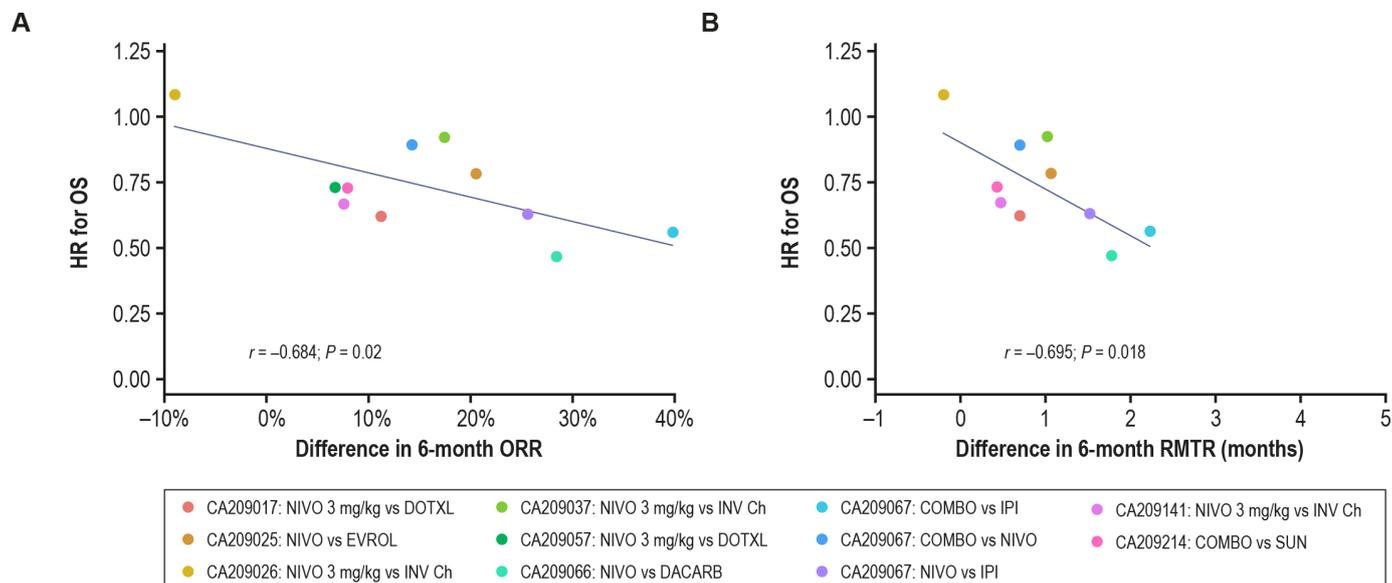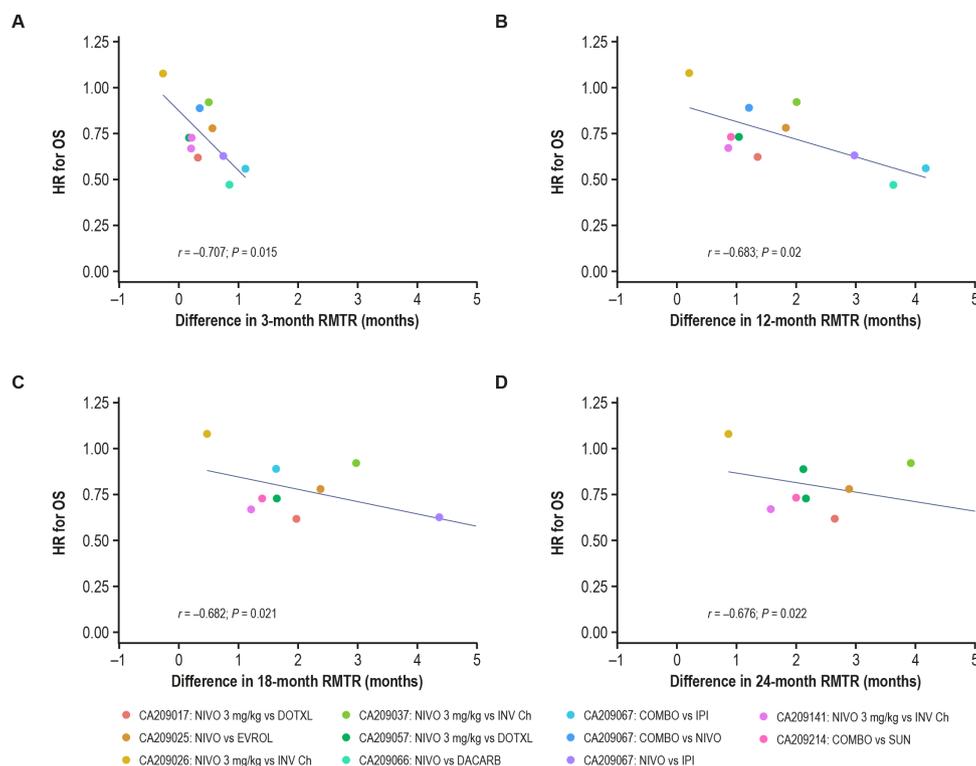
Figure 4 OS HR relationships between treatment arms at 6 months. Blue lines indicate a line of best fit from a linear regression model. (A) Relationship between the differences in ORR treatment differences and OS HRs. Each data point corresponds to an individual trial. (B) Correlation of the RMTR differences for combined ORR and DOR (RMTR) with the OS HRs. COMBO, combination NIVO+IPI therapy; DACARB, dacarbazine; DOR, duration of response; DOTXL, docetaxel; EVROL, everolimus; INV Ch, investigator's choice chemotherapy; IPI, ipilimumab; NIVO, nivolumab; ORR, objective response rate; OS, overall survival; RMTR, restricted mean time in response; SUN, sunitinib.

endpoints. This may impact patient care, as earlier treatment decisions could be made about continuing or switching treatments. Response durability is often measured by comparing DOR between treatment arms among responders.[28] Longer maintenance of response associates with lack of disease progression; a reasonable

Figure 5 OS HR relationships between treatment arms at various time points. Blue lines indicate a line of the best fit from a linear regression model. Correlation of the RMTR differences for combined ORR and DOR (RMTR) with the OS HRs at (A) 3 months, (B) 12 months, (C) 18 months, and (D) 24 months. COMBO, combination NIVO+IPI therapy; DACARB, dacarbazine; DOR, duration of response; DOTXL, docetaxel; EVROL, everolimus; INV Ch, investigator's choice chemotherapy; IPI, ipilimumab; NIVO, nivolumab; OS, overall survival; RMTR, restricted mean time in response; SUN, sunitinib.

indication that OS should be extended as well. Thus, a novel composite of ORR and DOR could provide more comprehensive information on clinical responses with ICIs than ORR alone.

Advantages of tumor-based endpoints, such as DOR and ORR, compared with time-to-event endpoints, such as OS and PFS, are that the former can be directly attributed to the therapy and might provide an earlier readout of drug activity.[13 44] In our study, RMTR and OS were correlated at 6, 12, 18, and 24 months, with similar levels of significance observed at all time points. These preliminary findings suggest that a 6-month cut-off is a reasonable duration of follow-up for DOR/RMTR analyses of ICIs.

ICIs can pose unique challenges for clinical trial design in part related to their mechanisms of action, including heterogeneous mechanisms of primary and acquired resistance, potentially delayed onset of immunologic responses, and optimization of dosing and duration.[45] ORR and response duration are important for evaluating efficacy of oncology therapeutics. A measure characterizing both of these responses, such as RMTR, is likely to be a reliable predictor of long-term clinical benefits.

We provide preliminary evidence that RMTR appears to be comparable to ORR alone in its association with survival outcomes across multiple tumor types. However, prospective studies and analyses comparing DOR and RMTR with other endpoints (eg, PFS, durable response rate[25]) are needed to learn about their potential utility in clinical trials and how these novel endpoints compare with traditional RECIST measures of tumor response. Higher-level evidence is required to extensively establish surrogacy for OS.

In recent years, several approved immunotherapies have demonstrated durable responses,[28 46 47] and accelerated approvals have been granted based on ORR.[10] However, strong associations between ORR and OS for ICIs have yet to be established. Our identification of a significant association between ORR and OS is novel, and differs from recent meta-analyses, which reported only weak associations in studies of both immunotherapies and targeted therapies.[7 44 48–50] Weak correlations in these analyses were likely related to multiple factors, including a crossover study design, patients continuing on additional therapies, long postprogression survival, study heterogeneity (involving sample sizes, disease types and settings, lines of therapy and follow-up durations), and antitumor mechanisms of action associated with ICIs that may contribute to delayed responses or pseudoprogression, which may lead to premature treatment cessation.[7 44 48 49]

An FDA-sponsored meta-analysis of 14 randomized, active-controlled trials of targeted and standard therapies including 12 567 patients with advanced NSCLC found no association between ORR and OS (R^2=0.09; 95% CI 0 to 0.33).[44] The authors suggested that ORR alone may not be the optimal endpoint for the expedited approval of cytotoxic therapies or immunotherapies.[44] Another meta-analysis of 13 randomized, multicentre, active-controlled

trials of immunotherapies submitted for FDA approval found that the association between ORR and OS was weak (R^2=0.1277).[48] Unlike the data presented here, which used differences in response rates to assess the relative effects of ORR, this previous study used ORs. An inherent challenge with ORs is the potential exaggeration of small differences, a significant issue when response rates are low. In contrast, our approach utilized differences in ORR that have a simpler interpretation based on absolute numbers rather than ratios. Unlike the previously reported study-level meta-analyses, individual patient-level data in this study were analyzed, providing more precision to the association analyses between ORR and DOR with OS. Another possible reason for different results could be that both meta-analyses only included trials submitted for regulatory approval which were positive and thus subject to selection bias, whereas our study included both positive and negative trials.[44 48]

Advantages and limitations of DOR in the ITT population

DOR analyses have traditionally been based on the subset of responding patients, defined post randomization, and thus subject to analysis-by-responder bias, which may cause overestimation of treatment effects.[28] However, assessing DOR in the ITT population could reduce the potential bias caused by exclusion of non-responder patients. The value of the ITT approach versus a responder subgroup analysis is that it could increase accuracy of predicting outcomes. Here, we show that DOR and DOR in combination with ORR (RMTR) can be analyzed in the ITT population to help define clinically meaningful responses with ICIs.

Our results from trials of nivolumab monotherapy and nivolumab in combination with ipilimumab across multiple tumor types demonstrated that treatment differences for both ORR and RMTR in the ITT population appeared to be significantly associated with OS. Importantly, DOR together with ORR could potentially be a reliable endpoint for characterizing tumor shrinkage in an overall ITT patient population that also associates with survival benefit. DOR combined with ORR can be characterized in multiple ways, including KM curves, RMTR, medians and milestone analyzes. Furthermore, visual representations like KM curves may allow prescribers to easily observe both response rates and duration.

Our novel approach using a composite endpoint of ORR and DOR in the ITT population may be particularly important in the immunotherapy era, in part because it may account for some of the nuances of responses that occur with ICIs. Our analysis of the area under the DOR curves to estimate RMTR in the ITT population is also an innovative approach for characterizing responses with ICIs.

Our study is limited by potential bias associated with retrospective analyses.[51] However, we have provided data that establish a rationale for prospective studies of associations between DOR/RMTR and OS. RMTR in the ITT population as a potential endpoint has limitations.

First, HR estimates are not readily interpretable because of 'non-proportional hazards'. When the proportional hazard assumption is violated, it limits the interpretation of HR, as it is not clear with which estimate the HR is associated.[52] Second, RMTR estimates require a time-point cut-off and benefits associated with an extended duration may be missed. In addition to response bias, DOR assessment can be influenced by varied frequencies of follow-up and tumor re-evaluation.[53] Larger prospective studies are needed to determine whether the combined ORR and DOR strategy may be a reliable and earlier indicator of survival benefit.

Another study limitation was that in our pooled analysis of patients receiving either first-line or subsequent lines of nivolumab, the potential impact of treatment line on our results is uncertain. In addition, our study did not assess responses by tumor type; however, there are visual differences between the NSCLC and melanoma graphs (figures 1, 2 and 5). The potential impact of treatment line and tumor type on RMTR should be studied further.

Overall, our results suggest that the association between OS and the combination of higher ORR with prolonged DOR is consistent with that of ORR alone. DOR in conjunction with ORR has the potential to both characterize tumor shrinkage and potentially predict OS benefit. This endpoint combination may be useful in early-stage clinical development to characterize a drug candidate's potential for survival benefit early on during a clinical study or to accelerate approval in late-stage ICI development.

CONCLUSIONS

ORR and RMTR appear to be associated with survival in patients treated with nivolumab monotherapy or with nivolumab plus ipilimumab combination therapy across multiple tumor types. Combining ORR and DOR to evaluate ICI therapy provides a visual, less biased and more informative analysis of the ITT population. Results showed similar associations of this combined endpoint to OS as ORR alone. The novel combination of ORR and DOR (RMTR) is a potentially useful and reliable endpoint for characterizing tumor shrinkage and clinical benefit, that is associated with OS. Our findings provide a strong rationale for further studies needed to establish the potential clinical utility of this endpoint. Further exploration of DOR and RMTR endpoints is warranted, particularly with new therapeutic modalities that have novel mechanisms of action or that involve novel drug combinations involving immunotherapy.

REFERENCES

1. Ferrara R, Pilotto S, Caccese M, *et al*. Do immune checkpoint inhibitors need new studies methodology? *J Thorac Dis* 2018;10:S1564–80.
2. Anagnostou V, Yarchoan M, Hansen AR, *et al*. Immuno-oncology trial endpoints: capturing clinically meaningful activity. *Clin Cancer Res* 2017;23:4959–69.
3. Hodi FS, O'Day SJ, McDermott DF, *et al*. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* 2010;363:711–23.
4. 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.
5. Roever L. Endpoints in clinical trials: advantages and limitations. *Evidence Based Medicine and Practice* 2016;1:e111.
6. Saad ED, Buyse M. Statistical controversies in clinical research: end points other than overall survival are vital for regulatory approval of anticancer agents. *Ann Oncol* 2016;27:373–8.
7. Kaufman HL, Schwartz LH, William WN, *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.
8. He X, Zhang Y, Ma Y, *et al*. Optimal tumor shrinkage predicts long-term outcome in advanced nonsmall cell lung cancer (NSCLC) treated with target therapy: result from 3 clinical trials of advanced NSCLC by 1 institution. *Medicine* 2016;95:e4176.
9. Cremolini C, Loupakis F, Antoniotti C, *et al*. Early tumor shrinkage and depth of response predict long-term outcome in metastatic colorectal cancer patients treated with first-line chemotherapy plus bevacizumab: results from phase III tribe trial by the Gruppo Oncologico del Nord Ovest. *Ann Oncol* 2015;26:1188–94.
10. Johnson JR, Ning Y-M, Farrell A, *et al*. Accelerated approval of oncology products: the food and drug administration experience. *J Natl Cancer Inst* 2011;103:636–44.
11. US Food and Drug Administration. Accelerated approval, 2018. Available: https://www.fda.gov/forpatients/approvals/fast/ucm405447.htm
12. Aykan NF, Özatlı T. Objective response rate assessment in oncology: current situation and future expectations. *World J Clin Oncol* 2020;11:53–73.

13 McCoach CE, Blumenthal GM, Zhang L, *et al*. Exploratory analysis of the association of depth of response and survival in patients with metastatic non-small-cell lung cancer treated with a targeted therapy or immunotherapy. *Ann Oncol* 2017;28:2707–14.

14 Lee C-K, Kim S-S, Park S, *et al*. Depth of response is a significant predictor for long-term outcome in advanced gastric cancer patients treated with trastuzumab. *Oncotarget* 2017;8:31169–79.

15 Morgensztern D, Ko A, O'Brien M, *et al*. Association between depth of response and survival in patients with advanced-stage non-small cell lung cancer treated with first-line chemotherapy. *Cancer* 2019;125:2394–9.

16 US Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics: guidance for industry, 2018. Available: https://www.fda.gov/media/71195/download

17 Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist* 2008;13 Suppl 2:19–21.

18 Suvannasankha A, Fausel C, Juliar BE, *et al*. Final report of toxicity and efficacy of a phase II study of oral cyclophosphamide, thalidomide, and prednisone for patients with relapsed or refractory multiple myeloma: a Hoosier Oncology Group trial, HEM01-21. *Oncologist* 2007;12:99–106.

19 Chia PL, Gedye C, Boutros PC, *et al*. Current and evolving methods to visualize biological data in cancer research. *J Natl Cancer Inst* 2016;108:djw031.

20 Garnett SA, Martin M, Jerusalem G, *et al*. Comparing duration of response and duration of clinical benefit between fulvestrant treatment groups in the confirm trial: application of new methodology. *Breast Cancer Res Treat* 2013;138:149–55.

21 EMD Serono Inc. BAVENCIO® (avelumab) [package insert] Rockland, MA, 2019. Available: https://www.emdserono.com/content/dam/web/corporate/non-images/country-specifics/us/pi/bavencio-pi.pdf

22 AstraZeneca. IMFINZI® (durvalumab) [package insert] Wilmington, DE, 2020. Available: https://www.azpicentral.com/imfinzi/imfinzi.pdf

23 Bristol Myers Squibb. YERVOY® (ipilimumab) [package insert] Princeton, NJ, 2020. Available: http://packageinserts.bms.com/pi/pi_yervoy.pdf

24 Bristol Myers Squibb. OPDIVO® (nivolumab) [package insert], 2020. Available: https://packageinserts.bms.com/pi/pi_opdivo.pdf

25 Kaufman HL, Andtbacka RHI, Collichio FA, *et al*. Durable response rate as an endpoint in cancer immunotherapy: insights from oncolytic virus clinical trials. *J Immunother Cancer* 2017;5:72.

26 FDA. Clinical trial endpoints for the approval of cancer drugs guidance for industry, 2018. Available: https://www.fda.gov/media/71195/download

27 Huang B, Tian L, Talukder E, *et al*. Evaluating treatment effect based on duration of response for a comparative oncology study. *JAMA Oncol* 2018;4:874–6.

28 Korn EL, Othus M, Chen T, *et al*. Assessing treatment efficacy in the subset of responders in a randomized clinical trial. *Ann Oncol* 2017;28:1640–7.

29 Badar T, Szabo A, Wadleigh M, *et al*. Real-World outcomes of adult B-cell acute lymphocytic leukemia patients treated with Inotuzumab Ozogamicin. *Clin Lymphoma Myeloma Leuk* 2020;20:556–60.

30 Cheminant M, Dreyling M, Hermine O. Recent advances in the first-line treatment of mantle cell lymphoma. *Ann Lymphoma* 2020;4:2.

31 Vokes EE, Ready N, Felip E, *et al*. Nivolumab versus docetaxel in previously treated advanced non-small-cell lung cancer (CheckMate 017 and CheckMate 057): 3-year update and outcomes in patients with liver metastases. *Ann Oncol* 2018;29:959–65.

32 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.

33 Rizvi NA, Mazières J, Planchard D, *et al*. Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial. *Lancet Oncol* 2015;16:257–65.

34 Hellmann MD, Rizvi NA, Goldman JW, *et al*. Nivolumab plus ipilimumab as first-line treatment for advanced non-small-cell lung cancer (CheckMate 012): results of an open-label, phase 1, multicohort study. *Lancet Oncol* 2017;18:31–41.

35 Larkin J, Chiarion-Sileni V, Gonzalez R, *et al*. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med* 2015;373:23–34.

36 Robert C, Long GV, Brady B, *et al*. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med* 2015;372:320–30.

37 Weber JS, D'Angelo SP, Minor D, *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.

38 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.

39 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.

40 Motzer RJ, Escudier B, McDermott DF, *et al*. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med* 2015;373:1803–13.

41 Ferris RL, Blumenschein G, Fayette J, *et al*. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. *N Engl J Med* 2016;375:1856–67.

42 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.

43 Huang B, Kuan P-F. Comparison of the restricted mean survival time with the hazard ratio in superiority trials with a time-to-event end point. *Pharm Stat* 2018;17:202–13.

44 Blumenthal GM, Karuri SW, Zhang H, *et al*. Overall response rate, progression-free survival, and overall survival with targeted and standard therapies in advanced non-small-cell lung cancer: US food and drug administration trial-level and patient-level analyses. *J Clin Oncol* 2015;33:1008–14.

45 Baik CS, Rubin EH, Forde PM, *et al*. Immuno-oncology clinical trial design: limitations, challenges, and opportunities. *Clin Cancer Res* 2017;23:4992–5002.

46 Chuk MK, Chang JT, Theoret MR, *et al*. FDA approval summary: accelerated approval of pembrolizumab for second-line treatment of metastatic melanoma. *Clin Cancer Res* 2017;23:5666–70.

47 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 Cancer Res* 2017;23:3484–8.

48 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.

49 Nie R-C, Chen F-P, Yuan S-Q, *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.

50 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.

51 Norvell DC. Study types and bias-Don't judge a study by the abstract's conclusion alone. *Evid Based Spine Care J* 2010;1:7–10.

52 Roychoudhury S, Anderson KM, Ye J. Robust design and analysis of clinical trials with non-proportional hazards: a straw man guidance from a cross-pharma Working group. *arXiv*:1908.07112 [stat.AP].

53 Eisenhauer EA, Therasse P, Bogaerts J, *et al*. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47.