

Immuno-oncology Trial Endpoints: Capturing Clinically Meaningful Activity

Valsamo Anagnostou¹, Mark Yarchoan¹, Aaron R. Hansen², Hao Wang¹, Franco Verde¹, Elad Sharon³, Deborah Collyar⁴, Laura Q.M. Chow⁵, and Patrick M. Forde¹



Abstract

Immuno-oncology (I-O) has required a shift in the established paradigm of toxicity and response assessment in clinical research. The design and interpretation of cancer clinical trials has been primarily driven by conventional toxicity and efficacy patterns observed with chemotherapy and targeted agents, which are insufficient to fully inform clinical trial design and guide therapeutic decisions in I-O. Responses to immune-targeted agents follow nonlinear dose-response and dose-toxicity kinetics mandating the development of novel response evaluation criteria. Biomarker-driven surrogate endpoints may better capture the mechanism of action and biological response to I-O agents and could be incorporated prospectively in early-phase I-O clinical trials.

While overall survival remains the gold standard for evaluation of clinical efficacy of I-O agents in late-phase clinical trials, exploration of potential novel surrogate endpoints such as objective response rate and milestone survival is to be encouraged. Patient-reported outcomes should also be assessed to help redefine endpoints for I-O clinical trials and drive more efficient drug development. This paper discusses endpoints used in I-O trials to date and potential optimal endpoints for future early- and late-phase clinical development of I-O therapies. *Clin Cancer Res*; 23(17): 4959–69. ©2017 AACR.

See all articles in this *CCR Focus* section, "Clinical Trial Design Considerations in the Immuno-oncology Era."

Introduction

In the last decade, rapid advances in our understanding of the human immune system have led to a new paradigm of treating cancer with agents that modulate the immune system described as immuno-oncology (I-O; ref. 1, 2). I-O, particularly the novel class of immune-checkpoint inhibitors (ICI), has transformed cancer therapeutics with notable clinical benefit observed in a diverse array of solid tumors, including melanoma (3, 4), non-small cell lung cancer (NSCLC; refs. 5–8), head and neck cancer (9), renal cancer (RCC; ref. 10), bladder cancer (11), Hodgkin lymphoma (12), and mismatch repair-deficient colon cancer (13). These ICI agents target immune regulatory pathways, thereby "releasing the brakes" and allowing the immune system to eliminate cancer cells (14). Several ICIs are now FDA approved for the treatment of a variety of cancers, including ipilimumab, an inhibitor of

cytotoxic T-lymphocyte-associated protein 4 (CTLA4), and nivolumab, pembrolizumab, and atezolizumab, which inhibit programmed cell death protein 1 (PD-1) or its ligand, PD-L1. However, many additional I-O agents in clinical development target a number of different immune modulatory pathways (for review, see ref. 15).

Such agents demonstrate fundamentally different tumor response kinetics from cytotoxic cancer therapies and warrant a reconsideration of conventional efficacy endpoints. Response patterns with immunotherapy differ from those of cytotoxic agents. Traditional clinical trial endpoints, like overall response rate (ORR) and progression-free survival (PFS), may be limited in their ability to predict long-term survival of patients treated with I-O agents. Conventional response evaluation criteria in solid tumors version 1.1 (RECIST vs1.1) may underestimate the benefit from I-O agents (16, 17). Immune-related response criteria (irRC) have been proposed to characterize patterns of response (18) and novel clinical endpoints are used in an attempt to account for unconventional tumor kinetics; however, these are not universally validated. In this article, we discuss traditional efficacy endpoints that have been used in oncology clinical trials to date, assess endpoints currently used in I-O trials, provide an overview of the challenges with efficacy endpoints in early- and late-stage I-O trials, and offer recommendations for the future clinical investigation of these therapies.

¹The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, Maryland. ²Division of Medical Oncology and Hematology, Princess Margaret Cancer Centre–University Health Network, Department of Medicine, University of Toronto, Toronto, Canada. ³Cancer Therapy Evaluation Program, National Cancer Institute, Bethesda, Maryland. ⁴Patient Advocates In Research, Danville, California. ⁵University of Washington, Seattle, Washington.

Corresponding Author: Patrick Forde, Johns Hopkins University School of Medicine, Room G92, Baltimore, 1650 Orleans Street, MD 21287-1000. Phone: 410-955-3974. E-mail: pforde1@jhmi.edu

doi: 10.1158/1078-0432.CCR-16-3065

©2017 American Association for Cancer Research.

Traditional Endpoints

Overall survival (OS), defined as the time from treatment initiation to death, remains the gold-standard clinical

endpoint for oncology cytotoxic clinical trials. OS measures both the effect of treatment and the survival impact of treatment-related adverse events (irAE). The outcome is clear and measurable, and the benefit of longer survival is irrefutable. However, analyses of OS often require large sample sizes and long patient follow-up to demonstrate benefit, particularly for diseases in which the natural history of the disease unfolds slowly. Furthermore, OS comparisons may be confounded by cross-over within a trial and/or subsequent therapeutic interventions. Another common efficacy endpoint in phase II and some phase III registration clinical trials has been PFS, defined as the duration of time from start of therapy to time of first documented tumor progression or death due to any cause. However, unlike OS, the outcome is not easily measured, and bias can be introduced in PFS assessment, due to dependence on whether an adequate comparator is used (19). The ORR, usually defined as the proportion of patients achieving a complete response (CR) or partial response (PR) based on stringent imaging criteria, is also commonly used in oncology clinical trials. Unlike PFS and OS, which require an active

comparator (e.g., existing standard therapy, placebo), ORR is frequently used in single-arm trials to demonstrate measurable tumor response without requiring direct comparison with a control group. Most novel drug approvals for oncology therapies have been based on demonstration of PFS or OS benefit over existing therapies. More recently, the majority of accelerated approvals have been based on ORR (20); these approvals are conditional and require subsequent confirmation of benefit, such as PFS or OS, in larger and/or randomized studies.

Challenges associated with I-O clinical trial endpoints

As the mechanisms of action and response patterns for I-O agents can differ substantially from conventional therapeutics, I-O disrupts the relatively stable model of traditional cancer drug development (Table 1). Specific challenges exist for developing clinical trial endpoints for I-O agents and reflect the following uncertainties and difficulties in effective clinical trial designs for this class of drug. From the outset, immunocompetent animal models for preclinical development

Table 1. Endpoint assessment in I-O clinical trials

Endpoint	Definition	Advantages	Limitations
Early-phase clinical trials			
MTD	The highest dose of a treatment that does not cause unacceptable side effects	<ul style="list-style-type: none"> Accurately captures tumor response patterns in conventional cytotoxic therapies 	<ul style="list-style-type: none"> Difficult to evaluate nonlinear dose-response relationship of I-O agents, Does not evaluate prolonged treatment effect, Does not capture delayed onset/long lasting side effects
Minimum effective dose	The minimum dose of a treatment that can produce a therapeutic response	<ul style="list-style-type: none"> Allows for evaluation of unique dose-response and dose-toxicity correlations seen with I-O agents 	<ul style="list-style-type: none"> Has not been extensively validated in I-O clinical trials
ORR	Fraction of patients achieving response	<ul style="list-style-type: none"> Can be used in single-arm studies May represent a surrogate for OS given the delayed response pattern seen with I-O agents 	<ul style="list-style-type: none"> Does not capture patients with durable stable disease, Does not differentiate outcome based on depth of response (CR vs. PR)
Late-phase clinical trials			
PFS	Time from treatment to tumor progression or death, whichever occurs first	<ul style="list-style-type: none"> Most practical Is not influenced by subsequent therapy 	<ul style="list-style-type: none"> Dependence on comparator arm, Subject to measurement error, observer bias, assessment schedule, and missing data, Might not accurately predict outcome as several IO trials showed OS benefit but no difference in PFS, Does not capture pseudoprogression, Remains a secondary alternative to OS given the response patterns of I-O agents
OS	Time from treatment initiation to death	<ul style="list-style-type: none"> Remains the golden standard of efficacy for I-O agents Measures both treatment and toxicity effects on survival 	<ul style="list-style-type: none"> Requires large sample size and long follow-up, which can delay approval of a potentially effective treatment, Delay in benefit seen with I-O agents may lead to loss in statistical power Conventional proportional hazards model does not appropriately capture survival kinetics
Milestone survival	Survival probability at a given time point	<ul style="list-style-type: none"> Takes into account the possibility of delayed treatment effect Potential surrogate for OS 	<ul style="list-style-type: none"> Requires extensive validation Does not take into account the totality of the OS data
Surrogate endpoints			
Biomarker/correlative endpoints	Endpoints that do not represent direct clinical benefit but may predict outcome	<ul style="list-style-type: none"> May better capture the mechanism of action and biological response to I-O agents 	<ul style="list-style-type: none"> Require extensive validation, standardization and development of analytical framework
QoL endpoints			
HRQOL endpoints		<ul style="list-style-type: none"> Represent patient-centered endpoints that have been shown to impact survival Allow for evaluation of effect of immune-related toxicities Quantify symptom or functional changes 	<ul style="list-style-type: none"> Absence of extensively validated HRQOL tools Applicability may be limited by poor compliance

Abbreviations: CR, complete response; HRQOL, health-related quality of life; PR, partial response; QoL, quality of life.

Downloaded from <http://aacrjournals.org/clinccancerres/article-pdf/23/17/4959/2038227/4959.pdf> by guest on 12 February 2025

of I-O drugs have either not been available or have been of limited utility, as the model context does not closely resemble the intact human immune system which is present in the majority of cancer patients who enroll in immunotherapy trials, although newer immunocompetent xenograft models are under development. This is a significant limitation and requires the development of I-O models/model systems that can be used to accurately perform preclinical drug screening, toxicity, and efficacy prediction. Moreover, effective correlative pharmacodynamic (PD) biomarkers are still in development for many of these agents in various malignancies. Pharmacodynamics deals with the biochemical and physiological effect of an administered drug on a patient.

The traditional development of cytotoxic chemotherapy, and more recently molecularly targeted therapies, involves phase I escalating doses of an agent to achieve maximum tolerated dose (MTD), followed by expansion at the MTD to more comprehensively profile toxicity and evaluate preliminary efficacy. In contrast, most I-O agents in clinical development are monoclonal antibodies that exhibit a modest dose-response relationship once receptor saturation has been achieved and can have prolonged biological effects even after discontinuation of treatment.

Toxicities associated with I-O agents are due to autoimmune activation and may be unpredictable in terms of timing, severity, and chronicity. Standard phase I clinical trial toxicity evaluation does not fully capture potentially serious, delayed onset, and/or long-lasting side effects and defining a phase II dose and schedule has been challenging given the lack of dose-limiting toxicity (DLT) in many early-phase trials. Due to rapid development of these agents in the race to FDA approval, the optimal dosing and scheduling of these agents have not been well-established prior to later phase studies and assessment in combination. For example, the recent regulatory approval of the anti-PD-1 antibody, nivolumab, was changed to a 240 mg flat dose despite weight-based dosing being used throughout the clinical development of the agent (21). For many of these agents, dose escalation and repeated dosing beyond a biologically effective point are likely to lead to minimal incremental clinical benefit and increased cost to patients and health systems financially and in terms of time spent in clinic and management of side effects.

Furthermore, I-O agents demonstrate response kinetics that differ significantly from traditional cytotoxic and molecularly targeted agents. Responses may occur within the first few cycles of therapy; however, some may be delayed and result in pseudoprogression, a phenomenon where a radiographic "tumor flare" precedes clinical response.

Lastly, as with other cancer therapies, differential efficacy of I-O agents within a specific tumor subtype may occur, where a subgroup of patients appears to derive most of the clinical benefit. Defining which subgroups derive most benefit is one of the major challenges facing cancer drug development in oncology in general and specifically in I-O. As with other cancer therapies, patient selection remains a challenging area in defining appropriate efficacy endpoints for participants in IO trials as multifaceted clinical (e.g., smoking status in non-small cell lung cancer) and biomarker (e.g., PD-L1 status, mutation burden, gene expression) characteristics appear to affect the likelihood of response. This issue was illustrated in a

recent phase III study in the first-line treatment of NSCLC, which randomly assigned patients with PD-L1-positive NSCLC to either platinum doublet chemotherapy or nivolumab (22). Despite stratifying patients by PD-L1 expression level and histology, retrospective analyses suggest that there may have been an imbalance in tumor mutation burden between the treatment arms which may have contributed to a lack of PFS or OS benefit for nivolumab over chemotherapy in this trial (22).

Efficacy endpoints in early-phase immune-oncology clinical trials

I-O agents pose a challenge to the well-established concept of identifying the maximum tolerated dose (MTD) in phase I clinical trials (23). As responses may be delayed, similarly irAEs may be delayed and, in some cases, may not be observed in the first few cycles of therapy (24, 25). In contrast to conventional cytotoxic or targeted agents, an increase in the administered dose above a biologically optimal dose with maximal receptor saturation, does not necessarily equate linearly to increased efficacy or toxicity. Increasing doses do not follow a standard defined clinical dose-response or dose-toxicity relationship; dosing and schedule have been determined more from feasibility rather than PK and PD effects. Determining the minimum effective dose instead of the conventionally used MTD might be more relevant for phase I I-O trials, given that there is not a direct proportional correlation between activity and toxicity according to dose (3, 4, 26, 27).

The Response Evaluation Criteria in Solid Tumors (RECIST) have been the gold standard for assessment of tumor burden dynamics in oncology clinical trials across multiple tumor types (28, 29). This framework was developed to standardize and interpret therapeutic responses seen with cytotoxic chemotherapeutic agents, where an increase in tumor burden, progression of non-target lesions, and/or development of new lesions signifies disease progression. The unique patterns and timing of response to immunotherapy have raised concerns about the validity of applying conventional response assessment platforms to the efficacy of immunotherapeutic agents. More specifically, pseudoprogression was originally described in advanced melanoma treated with ipilimumab and cannot be assessed by the conventional RECIST criteria (30, 31). Therefore, the irRC were devised to capture the unique response patterns observed with immunotherapy in solid tumors (18).

Immune-related response criteria

A set of irRC have been developed (18) upon which assessment of the antitumor response is based on total measurable tumor burden, including all index lesions at baseline and any index and/or new measurable lesions at each subsequent tumor reassessment (Table 2). IrRC implement bidimensional measurements in contrast to the unidimensional measurements of irRECIST. Responses are thus classified as (1) irCR, where all lesions completely disappear in two observations at least 4 weeks apart (2) irPR, where a decrease of $\geq 50\%$ in tumor burden is noted in two observations at least 4 weeks apart, (3) irSD, where criteria for irCR or irPR are not met in the absence of irPD, and (4) irPD, where an increase of $\geq 25\%$ relative to nadir is noted in two observations at least 4 weeks

Table 2. Comparison of response evaluation criteria by RECIST 1.1, irRS, and irRECIST

	RECIST 1.1	irRS	irRECIST
Baseline tumor assessment	Tumor burden calculated as the sum of diameters of all target lesions—unidimensional measurements	Tumor burden calculated as the sum of the products of the longest perpendicular diameters (SPD) of all index lesions—bidimensional measurements	Calculated as the sum of diameters of all target lesions—unidimensional measurements
Subsequent tumor assessment	Tumor burden calculated as the sum of diameters of all target lesions (unidimensional measurements)	Total tumor burden at follow-up includes SPD of the index lesions and any new measurable lesions	Total measured tumor burden at follow-up is calculated as the sum of diameters of all target lesions and new measurable lesions
Target lesions	All measurable lesions up to a maximum of 5 total (max 2 per organ)	All measurable lesions (up to 5 per organ, up to 10 visceral and 5 cutaneous index lesions)	All measurable lesions up to a maximum of 5 total (max 2 per organ)
Evaluation of response			
CR	Disappearance of all target lesions	irCR: Complete disappearance of all lesions, whether measurable or not and no new lesions	irCR: Complete disappearance of all measurable and nonmeasurable lesions
PR	>30% decrease in the sum of diameters of target lesions compared to baseline sum diameters	irPR: >50% decrease in tumor burden relative to baseline	irPR: >30% decrease in tumor burden relative to baseline and no unequivocal progression of new nonmeasurable lesions
PD	>20% increase in the sum of diameters of target lesions compared to nadir, or appearance of new lesions	irPD: >25% increase in tumor burden relative to nadir	irPD: >20% increase in tumor burden and minimum 5 mm absolute increase in tumor burden compared with nadir or irPD for nontarget or new nonmeasurable lesions
SD	Not meeting criteria for PR or PD	irSD: not meeting criteria for irCR or irPR in the absence of irPD	irSD: not meeting criteria for irCR or irPR in the absence of irPD
New lesions	Represent PD	Do not necessarily represent PD	Do not necessarily represent PD
Confirmation of response	Confirmation of response is not mandatory	Any response other than SD has to be confirmed at least 4 weeks apart after the first assessment	Confirmation of response is not mandatory; however, confirmation of irPD is recommended for patients with minimal total measured tumor burden increase especially during the first 12 weeks of treatment

Abbreviations: CR, complete response; irRS, immune-related response criteria; PD, progressive disease; PR, partial response; RECIST, response evaluation in solid tumors; SD, stable disease.

apart. According to these criteria, patients were considered to have irPR or irSD despite emergence of new lesions, as long as the above mentioned thresholds of response were met. Evaluation of the immune-related best overall response (irBOR) is therefore based on changes in total tumor burden from baseline and attempts to account for pseudoprogression, both by allowing the appearance of new lesions, without automatic categorization of such cases as PD, and also confirming apparent initial RECIST progression at a subsequent evaluation. Figure 1 illustrates an unconventional pattern of response to immune checkpoint antibody therapy.

Applying these criteria in a cohort of ipilimumab-treated melanomas, at least 10% of patients with durable clinical benefit were characterized as having progressive disease by WHO response criteria (18). The irRC have been validated in ipilimumab-treated melanoma patients (32, 33); however, the generalizability of these endpoints in other tumor types is, as yet, uncertain. The incidence of pseudoprogression may differ in various tumor types and consequently the use of irRC may differ in the applicability. In addition, as the recent use of RECIST response criteria has largely replaced WHO response criteria, the modified irRECIST have been used in the majority of I-O clinical trials. However, the standardization and definitions have varied across different trials and study sponsors, and have not been validated. Development of a standardized modified irRECIST is a priority for stakeholders in clinical development of I-O therapy to permit meaningful comparison

of I-O therapies and also effective use of irRECIST for regulatory purposes.

Response rates have been a good assessment of activity in early-phase studies and have led to accelerated FDA approval of these agents. For instance, the approval of pembrolizumab in PD-L1-positive (>50% positive tumor cells by immunohistochemistry) second-line previously treated NSCLC was predominantly based on responses observed in an expanded cohort of the Keynote 001 study (7). Pembrolizumab was also approved based on objective responses observed in expanded cohorts of the phase I Keynote 012 trial in head and neck squamous cell cancer trial (34).

However, whether response rates are adequate in I-O trials to reflect clinical benefit of these agents remains a major question. Scenarios to be considered when designing endpoints for I-O clinical trials include the following in addition to the question of pseudoprogression (discussed below). Patients whose tumors initially respond to immune checkpoint antibodies may require discontinuation of therapy due to toxicity and subsequently experience tumor progression. Patients with CR may have better outcomes and chances of remaining in remission than those with PR or SD. Furthermore, many patients with stable disease do not experience a radiographic tumor regression or may demonstrate slow and late tumor regression months to years later. In such cases endpoints such as PFS or OS would better describe the overall benefit to patients. In addition, long-term follow-up of I-O trials and possible

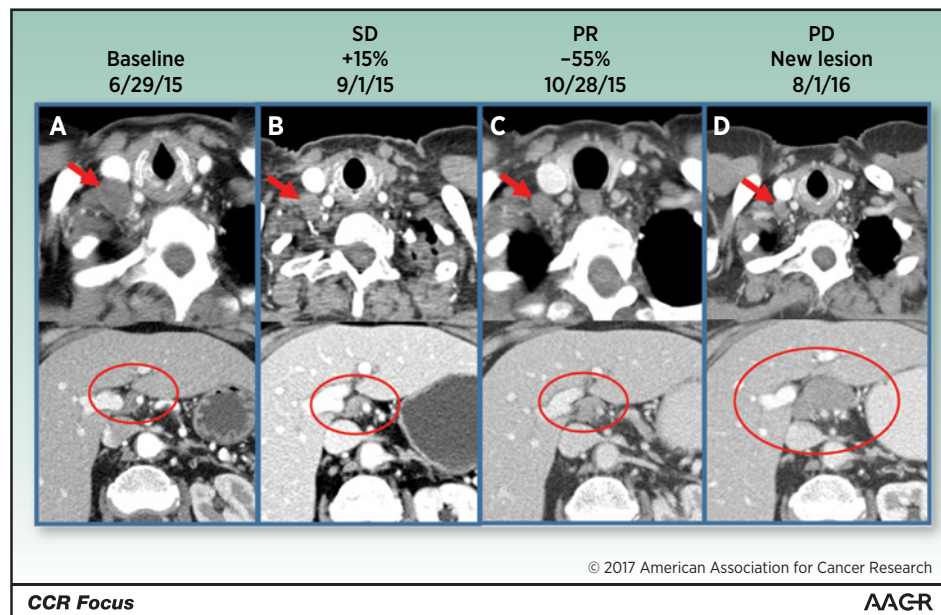


Figure 1.

Radiographic pseudoprogression on anti-PD-1. 74-year-old woman with metastatic carcinoid of the right lung status post middle lobectomy, right hilar radiation therapy with progression on chemotherapy at baseline (A), first follow-up (B), second follow-up (C), and 57-week follow-up (D) after immunotherapy. IV contrast-enhanced CT examination of the chest, abdomen, and pelvis was performed with selected slices of the chest (top images) and abdomen (bottom images). Right supraclavicular adenopathy is shown (A, upper image, red arrow) measuring 16 mm short axis. Node decreases to 7 mm short axis (B–D top images, red arrows) which denotes a normal node by RECIST 1.1 criteria. Patient has additional metastatic disease to include paratracheal adenopathy, numerous right pleural implants, and scattered hepatic metastases. Baseline sum of longest diameters of selected target lesions calculated to 68 mm. First follow-up demonstrated stable disease with 15% increase from baseline. Second follow-up demonstrated partial response with 56% decrease from baseline. Patient continued to sustain partial response until progressive disease with 20 mm short axis porta hepatis adenopathy appearing (D, bottom image, large red circle). A normal 7-mm short axis node is present in this location on prior exams (A–C, bottom images, smaller red circles).

incorporation of long-term survival (e.g., 5-year survival rate) in regulatory approval decisions should be considered. Recent data suggest that from a single-arm study of 129 patients with NSCLC treated with anti-PD-1, up to 16% of patients may be alive, many still responding, at a median of 5-year follow-up (35).

Additionally, many I-O trials allow continued therapy despite radiologic progression if patients are deriving clinical benefit, and anecdotal reports of continued stability or response post focal radiation have been reported. Terms such as "unconventional pattern of response" have been used to describe such cases; however, this varies across trials and sponsors. Recent data suggest that conventional response after initial progression on anti-PD-1 immunotherapy is uncommon. In a pooled retrospective analysis of three large multi-center studies of anti-PD-1 immunotherapy in the treatment of metastatic NSCLC, the authors identified 535 patients treated with single-agent anti-PD-1, of whom 121 received treatment past RECIST tumor progression (8). Among all 535 patients treated with anti-PD-1 therapy, the conventional RECIST partial response rate following treatment past progression was 1.9% (10 of 535) or 8.3% (10 of 121) in the treatment past progression subgroup (36).

Any potential benefit of treatment past tumor progression should be weighed against the possibility of negative outcomes that include clinical deterioration on ineffective therapy that may

preclude further treatment options, treatment-related toxicity and, if widely applied in clinical practice, the high financial cost of continued I-O therapies past progression with relatively low likelihood of benefit.

Based on experience with targeted agents it may be that patients with low volume progression, for example, isolated progression in one lesion with stability or response in all other lesions, are those most likely to derive benefit from treatment past RECIST progression; however, this has yet to be confirmed from trial data. Mechanisms of acquired resistance to I-O agents continue to be elucidated and may ultimately assist in defining those patients who may have benefit from treatment past progression (37, 38).

In general, we favor the continued use of conventional RECIST as the primary response evaluation system for I-O trials, with the addition of irRECIST where feasible as a secondary assessment, until there are data across tumor types and I-O agents correlating irRECIST with long-term efficacy outcomes including OS.

Efficacy endpoints in late-phase I-O clinical trials

OS remains the gold-standard clinical efficacy endpoint in cancer clinical trials, including those involving I-O agents; however, OS has some statistical challenges. Reporting hazard ratios or median survival may not adequately represent the treatment benefit due to nonproportional hazards of survival

and long-term survival with I-O agents. These considerations are described in more detail in the statistical design section below. Competitive development of these agents has required efficiency; waiting for measures of OS improvements can take years, thus delaying access to a potentially effective treatment with observed efficacy and in an area of high unmet medical need. As immunological effects can linger for months or even years, the full effect on survival of these agents for the individual may not be fully captured as they receive subsequent therapy or crossover to other agents. Milestone survival has been proposed as an intermediate endpoint and represents the survival probability at a given time point; such endpoints could provide long-term survival information, while the entire study continues with a primary endpoint of OS (39).

Biomarker and correlative endpoints as predictive or prognostic markers for endpoint analysis

Surrogate endpoints are a substitute for clinical efficacy endpoints used in clinical trials when definitive clinical endpoints might not be feasible. Unlike clinical endpoints, surrogate endpoints do not represent direct clinical benefit but may be prognostic of clinical outcome. There is an increased appreciation of the value of correlative immune endpoints. Inclusion of biomarkers in the design and interpretation of I-O trials is recommended. Integration of such endpoints may be important for understanding the underlying biological response to immunotherapy agents, may increase our ability to both interpret clinical trial outcome, and rationally design subsequent clinical trials when the initial approach succeeds or fails. Establishment of such endpoints requires harmonized specimen collection methods, assay standardization, and validation and development of a methodological framework for analysis (40).

Similar to the targeted therapy paradigm, the success of I-O approaches seems to depend on choosing patient populations most likely to benefit. The use of a systems-based approach to biomarker evaluation is critical as the immune system is plastic and use of single biomarkers has proven insufficient for selection of patients most likely to derive clinical benefit (41). PD-L1 assessed by immunohistochemistry has been used as a potential biomarker to select patients for immune checkpoint therapies given the higher response rates observed in this group of patients (5, 7, 8). However, PD-L1 expression has not been sufficient to fully explain therapeutic outcomes and durable clinical benefit observed in patients with PD-L1 nonexpressing tumors (40, 42, 43). There are also differences in PD-L1 antibody assays, heterogeneous expression, lack of concordance regarding assessment on tumor, and/or infiltrating cells and cutoff levels that are not well established.

Somatic mutational density may confer long-term benefit from immune checkpoint blockade (13, 44–46) as somatic genomic alterations are foreign to the immune system and could represent tumor-specific antigens capable of inducing antitumor immune responses (47, 48). The predictive efficacy of T-cell tumor immune infiltrates has been shown in a variety of tumor types, and the immunoprofile of the tumor may represent a robust predictor of clinical outcome (49, 50). To this end, analysis of sequential tumor biopsies may more accurately reflect the evolving tumor microenvironment. These

approaches require further large-scale validation incorporated in the design of I-O trials. For example, one proposed endpoint for early-phase I-O trials has been the use of tumor CD8⁺ T-cell infiltration on sequential tumor biopsies at a defined time point while on therapy (51). Such approaches have particular challenges due to tumor heterogeneity and possible lack of correlation with clinical outcomes that may limit utility and they require large-scale prospective validation. A large number of other predictive biomarkers, including PD-L2, T-cell receptor repertoire, serum lymphocyte, and inflammatory markers of response, and tissue immune gene signatures, are being evaluated.

PD biomarkers should capture the mechanism of action of immunotherapeutic agents and guide the determination of the ideal dose and schedule of administration in early-phase clinical trials; however, these markers are still being explored and need technical and analytical validation. Analysis of T-cell responses is important for understanding the association between antitumor immune responses and clinical benefit. These approaches primarily study the functionality and phenotype of T cells by means of flow cytometry (52), fluctuations in levels of proinflammatory cytokines (53, 54), and cytokine-release associated toxicity (55), and may be considered for inclusion as correlative endpoints. Further study of effective PD biomarkers for response in early-phase I-O trials will be the key for future determination of I-O early efficacy endpoints and faster clinical development of this class of drug. Comprehensive discussion of biomarker approaches is beyond the scope of this paper; however, we recommend the companion article in this series for an in-depth review of the topic (56).

Quality of life, patient-reported outcomes

Health-related quality-of-life (HRQOL) endpoints and patient-reported outcomes (PRO) can enrich the assessment of antitumor treatments by augmenting traditional survival and efficacy measures, and should be included in every clinical trial. These patient-centered endpoints have been correlated with prognosis and have become incorporated into clinical trials more frequently over the last decade (57). Several reasons underpin this enthusiasm and change. Multiple treatments for a particular tumor type may yield similar efficacy results and thus HRQOL can be used to differentiate these therapies (58). Patient advocacy groups and quality-of-life research organizations such as the International Society for Quality of Life Research (ISOQOL) have made efforts in clinical trials to focus on the patient experience, their symptoms, and QOL (59). The FDA has recognized the importance of HRQOL, as evidenced by guidance documents for HRQOL endpoints and approved agents with labeled PRO endpoints (60). Finally, the National Cancer Institute (NCI) formed the Symptom Management and Quality of Life Steering Committee, which reorganized their clinical trials to integrate patient-centered endpoints by utilizing the PRO tool for standard reporting with Common Terminology Criteria for Adverse Events (PRO-CTCAE; ref. 61).

Beyond this rationale for anticancer therapy, other specific justifications exist to incorporate HRQOL and PRO endpoints into I-O trials. While single-agent immune treatments (e.g., PD-1 checkpoint inhibitors) may be tolerable, combination

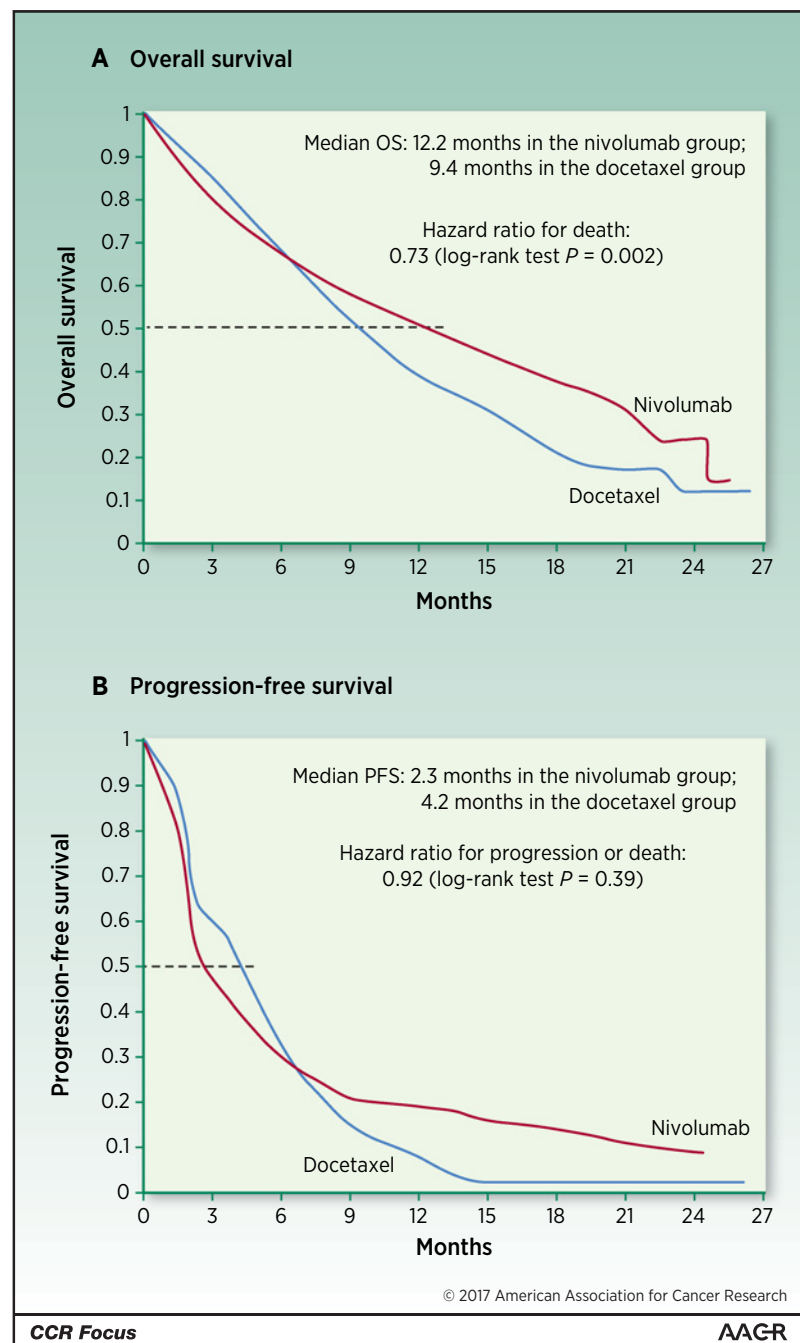
regimens may have more severe AEs (62). How these toxicities impact HRQOL is not known; designing trials with HRQOL endpoints will permit this evaluation. Traditional endpoints may not completely capture the entire benefit or harm from immunotherapy; HRQOL and PRO endpoints can help quantify symptom or functional improvement or decline. The protracted length of immune treatment and off-target immune effects for some patients, compared with relatively predictable and short-lived effects of chemotherapy or molecularly targeted agents, can lead to unexpected chronic toxicities with an impact

on HRQOL. These long-term toxicities may be difficult to capture within the defined follow-up period of a clinical trial. While long-term follow-up for these AEs is a challenge, electronic PRO tools could facilitate this process, allowing patients to complete away from the treatment facility and send to the study team.

Integrating these endpoints into clinical trials face several barriers. Currently no tool exists specifically for I-O and it is imperative to explore whether an IO-specific tool would be useful. Typically, instruments such as the Functional

Figure 2.

Replicates of Kaplan–Meier survival curves of nivolumab vs. docetaxel in advanced nonsquamous non–small cell lung cancer (5). **A**, OS. **B**, PFS. The horizontal dashed line represents 50% survival, and its corresponding time is median OS and PFS, respectively. Hazard ratio was estimated based on a stratified Cox proportional-hazards model. Restricted mean survival time of OS up to 24 months represents life expectancy with treatment of nivolumab for the disease over the next 24 months. It can be estimated as the area under the survival curve up to 24 months. Hasegawa et al. calculated that for OS, the increase of the restricted mean survival time by nivolumab would be 1.7 months compared with docetaxel [95% confidence interval (CI), 0.4–3.1]. For PFS, the estimated increase of restricted mean survival time is 1.3 months (95% CI, 0.2–2.3; ref. 78). Milestone survival is the proportion of subjects who are alive at a prespecified time point based on Kaplan–Meier survival estimate. Milestone analysis is performed at interim analysis in the first set of patients when each of their length of follow-up reaches the milestone time. Milestone survival rate is not shown on the figure that summarizes the final analysis results.



Assessment of Cancer Therapy-General (FACT-G) or European Organization for Research and Treatment of Cancer–Quality of Life Questionnaire-30 (EORTC QLQC30) are often utilized to measure HRQOL, but may not be sufficiently sensitive for I-O and have never been validated in patients treated with these agents (63, 64). There may also be possible biased PRO responses in open-label studies, poor compliance, and difficulty differentiating tumor-related from irAEs. Addressing these challenges will help ensure HRQOL endpoint success when implemented into I-O trials.

Statistical design and analysis considerations for I-O clinical trials

While cancer immunotherapy trials use PFS and OS as endpoints to evaluate antitumor response and efficacy, the unique characteristics of the antitumor response induced by immunotherapy and pattern of survival curves require tailored and novel statistical design considerations and analysis of late-stage trials (Fig. 2). Delayed separation or crossing of survival curves is observed in many randomized clinical trials of I-O agents (5, 6, 8, 30, 65–70), and stable plateaus at the tail represent long-term survival in trials or cohorts with long-term follow-up (30, 71–76). This indicates deviations from the most commonly used standard model, with proportional hazards assumption that the hazard ratio of the investigational arm versus the control arm is constant over time. The delay in benefit may lead to substantial loss of statistical power, thus reducing the ability to detect the effect if the trial uses conventional calculation of required events based on proportional hazards assumption (77–79).

Alternative statistical models to account for the delayed separation of survival curve or long-term survival can improve the planning and analysis of immunotherapy trials. For instance, piecewise exponential models describe different hazard ratios before and after a specified timing of delayed onset of survival benefit (80). Cure rate models are useful to explicitly model a cured fraction of patients when scientific rationale for the presence of long-term survivors is strong (78, 81). It is recommended that the required number of events calculated under these models reflects the unique survival kinetics for the design to yield adequate statistical power. The presence and timing of the delayed response also have an impact on the timing of the interim analysis and trial duration (82), highlighting the importance of tailoring study design to cancer immunotherapy characteristics.

Milestone survival is the proportion of subjects who are alive at a prespecified time point based on Kaplan–Meier survival estimate. It may be considered as an endpoint for interim analysis to allow accelerated drug approval, while OS remains the primary endpoint representing long-term benefit and will be analyzed using log-rank test or other statistical methods to account for nonproportional hazard. The primary endpoint of OS is characterized by the survival cure over the entire follow-up time, but not just the median of the curve.

Various statistical methodologies for analyzing survival data featuring nonproportional hazards can be applied to I-O trials. Weighted log-rank test potentially yields greater

power than the traditional log-rank test by putting more weight on the difference after separation of the survival curves (83). One could also consider modeling time-varying hazard ratios (84). In order to overcome the difficulty in interpreting the hazard ratio when proportional hazards assumption is violated, alternative measures such as restricted mean survival time and milestone survival have been proposed to quantify the treatment benefit (85, 86). The choice of the primary measure should fit the objective of the trial.

Conclusions

Early-phase clinical trials

Phase I trials of novel I-O agents should aim to select a dose and schedule that makes pharmacodynamic sense, avoiding additive dosing, at the minimum dose and schedule that is biologically active. To meet this goal, correlative genomic and immunologic endpoints should be explored in early-phase studies and correlated with both toxicity and efficacy outcomes.

The optimal go/no-go decision point to choose for I-O phase IB/II trials is yet to be fully defined; however, it is likely this will depend on the patient population. For example, even a single objective response to a novel I-O agent in a population that has a very low expected response rate merits intensive correlative investigation of tumor and other biospecimens obtained from the responding patient. Such correlative studies may lead to further clinical studies in defined populations of patients. Tumors that are already resistant or refractory to anti-PD-(L)1 therapy represent such a population.

Late-phase clinical trials

As with other anticancer therapies, an improvement in OS in a well-controlled randomized study represents the gold-standard demonstration of efficacy for a novel I-O compound. Given the extended duration of responses seen with I-O, it is possible that ORR may represent a surrogate for OS in I-O trials; however, this is yet to be shown conclusively. Other surrogates such as PFS have been accepted for regulatory purposes and continue to be evaluated in I-O studies. The role of immune-related response continues to be elucidated, and the ongoing development of irRECIST is very welcome. Similarly, assessment of HRQOL and PRO in I-O trials is recommended in all phases of development.

Clinical development of I-O agents presents both challenges and opportunities for drug development. The absence of a clear dose–efficacy relationship may allow minimization of dose escalation cohorts in phase I and early assessments of efficacy in expanded phase Ib trials across tumor types. Defining the optimal dose and schedule may be more challenging than with traditional cytotoxics, and use of endpoints such as target receptor occupancy or changes in immune characteristics of serial on-treatment tumor biopsies should be considered. Later phase trials should focus on efficacy assessments with real-life applicability. While novel endpoints, such as immune-related response, and biomarker-based correlative endpoints have been proposed, further refinement is needed and may come about through consensus guidelines.

Disclosure of Potential Conflicts of Interest

M. Yarchoan reports receiving commercial research grants from Bristol-Myers Squibb and Exelixis. L.Q.M. Chow is a consultant/advisory board member for Amgen, AstraZeneca/Medimmune, Bristol-Myers Squibb, Genentech, Merck, Novartis, Pfizer, and Seattle Genetics. P.M. Forde is a consultant/advisory board member for AstraZeneca, BMS, Boehringer,

Celgene, EMD Serono, Merck, and Novartis. No potential conflicts of interest were disclosed by the other authors.

Received April 10, 2017; revised June 26, 2017; accepted July 10, 2017; published online September 1, 2017.

References

- Drake CG, Jaffee E, Pardoll DM. Mechanisms of immune evasion by tumors. *Adv Immunol* 2006;90:51–81.
- Mellman I, Coukos G, Dranoff G. Cancer immunotherapy comes of age. *Nature* 2011;480:480–9.
- Brahmer JR, Tykodi SS, Chow LQ, Hwu WJ, Topalian SL, Hwu P, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med* 2012;366:2455–65.
- Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* 2012;366:2443–54.
- 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 Engl J Med* 2015;373:1627–39.
- Brahmer J, Reckamp KL, Baas P, Crino L, Eberhardt WE, Poddubska E, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med* 2015;373:123–35.
- Garon EB, Rizvi NA, Hui R, Leigh N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med* 2015;372:2018–28.
- Fehrenbacher L, Spira A, Ballinger M, Kowanz 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.
- 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.
- Motzer RJ, Rini BI, McDermott DF, Redman BG, Kuzel TM, Harrison MR, et al. Nivolumab for metastatic renal cell carcinoma: results of a randomized phase II trial. *J Clin Oncol* 2015;33:1430–7.
- Sharma P, Callahan MK, Bono P, Kim J, Spiliopoulou P, Calvo E, et al. Nivolumab monotherapy in recurrent metastatic urothelial carcinoma (CheckMate 032): a multicentre, open-label, two-stage, multi-arm, phase 1/2 trial. *Lancet Oncol* 2016;17:1590–8.
- Ansell SM, Lesokhin AM, Borrello I, Halwani A, Scott EC, Gutierrez M, et al. PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. *N Engl J Med* 2015;372:311–9.
- Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* 2015;372:2509–20.
- Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* 2012;12:252–64.
- Topalian SL, Drake CG, Pardoll DM. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* 2015;27:450–61.
- Hodi FS, Hwu WJ, Kefford R, Weber JS, Daud A, Hamid O, et al. Evaluation of immune-related response criteria and RECIST v1.1 in patients with advanced melanoma treated with pembrolizumab. *J Clin Oncol* 2016;34:1510–7.
- Bohnsack O, Hoos A, Ludajic K. Adaptation of the immune related response criteria: irRECIST. *Ann Oncol* 2014;Supplement 4:iv361–iv72.
- Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbe 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.
- Seymour L, Ivy SP, Sargent D, Spriggs D, Baker L, Rubinstein L, et al. The design of phase II clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the national cancer institute investigational drug steering committee. *Clin Cancer Res* 2010;16:1764–9.
- Administration USFD. CDER New Drug Review: 2016. Update2016.
- Administration USFD. Modification of the Dosage Regimen for Nivolumab 2016. Available from: <http://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm520871.htm>.
- Carbone DP, Reck M, Paz-Ares L, Creelan B, Horn L, Steins M, et al. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N Engl J Med* 2017;376:2415–26.
- Postel-Vinay S, Aspeslagh S, Lanoy E, Robert C, Soria JC, Marabelle A. Challenges of phase 1 clinical trials evaluating immune checkpoint-targeted antibodies. *Ann Oncol* 2016;27:214–24.
- Weber JS, Kahler KC, Hauschild A. Management of immune-related adverse events and kinetics of response with ipilimumab. *J Clin Oncol* 2012;30:2691–7.
- Paoletti X, Le Tourneau C, Verweij J, Siu LL, Seymour L, Postel-Vinay S, et al. Defining dose-limiting toxicity for phase 1 trials of molecularly targeted agents: results of a DLT-TARGETT international survey. *Eur J Cancer* 2014;50:2050–6.
- Kohrt HE, Tumeh PC, Benson D, Bhardwaj N, Brody J, Formenti S, et al. Immunodynamics: a cancer immunotherapy trials network review of immune monitoring in immuno-oncology clinical trials. *J Immunother Cancer* 2016;4:15.
- Patnaik A, Kang SP, Rasco D, Papadopoulos KP, Ellassa-Schaap J, Beeram M, et al. Phase I study of pembrolizumab (MK-3475; anti-PD-1 monoclonal antibody) in patients with advanced solid tumors. *Clin Cancer Res* 2015;21:4286–93.
- 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. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000;92:205–16.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47.
- Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* 2010;363:711–23.
- Chiou VL, Burotto M. Pseudoprogression and immune-related response in solid tumors. *J Clin Oncol* 2015;33:3541–3.
- 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.
- Nishino M, Gargano M, Suda M, Ramaiya NH, Hodi FS. Optimizing immune-related tumor response assessment: does reducing the number of lesions impact response assessment in melanoma patients treated with ipilimumab? *J Immunother Cancer* 2014;2:17.
- Chow LQ, Haddad R, Gupta S, Mahipal A, Mehra R, Tahara M, et al. Antitumor activity of pembrolizumab in biomarker-unselected patients with recurrent and/or metastatic head and neck squamous cell carcinoma: results from the phase Ib KEYNOTE-012 expansion cohort. *J Clin Oncol* 2016;34:3838–45.
- Brahmer J, Horn L, Jackman DM, Spigel DR, Antonia S, Hellman M, et al., editors. Five-year follow-up from the CA209-003 study of nivolumab in previously treated advanced non-small cell lung cancer (NSCLC): Clinical characteristics of long-term survivors. AACR; 2017.
- Kazandjian D, Keegan P, Suzman DL, Pazdur R, Blumenthal GM. Characterization of outcomes in patients with metastatic non-small cell lung cancer treated with programmed cell death protein 1 inhibitors past RECIST version 1.1-defined disease progression in clinical trials. *Semin Oncol* 2017;44:3–7.

37. Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, Hu-Lieskovan S, et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N Engl J Med* 2016;375:819–29.
38. Anagnostou V, Smith KN, Forde PM, Niknafs N, Bhattacharya R, White J, et al. Evolution of neoantigen landscape during immune checkpoint blockade in non-small cell lung cancer. *Cancer Discov* 2017;7:264–76.
39. Chen TT. Milestone survival: a potential intermediate endpoint for immune checkpoint inhibitors. *J Natl Cancer Inst* 2015;107.
40. Topalian SL, Taube JM, Anders RA, Pardoll DM. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat Rev* 2016;16:275–87.
41. Sharma P, Allison JP. The future of immune checkpoint therapy. *Science* 2015;348:56–61.
42. Herbst RS, Soria JC, Kowanzet M, Fine GD, Hamid O, Gordon MS, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* 2014;515:563–7.
43. Postow MA, Callahan MK, Wolchok JD. Immune checkpoint blockade in cancer therapy. *J Clin Oncol* 2015;33:1974–82.
44. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA4 blockade in metastatic melanoma. *Science* 2015;350:207–11.
45. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* 2014;371:2189–99.
46. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 2015;348:124–8.
47. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science* 2015;348:69–74.
48. Castle JC, Kreiter S, Diekmann J, Lower M, van de Roemer N, de Graaf J, et al. Exploiting the mutanome for tumor vaccination. *Cancer Res* 2012;72:1081–91.
49. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 2006;313:1960–4.
50. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer* 2012;12:298–306.
51. Huang RR, Jalil J, Economou JS, Chmielowski B, Koya RC, Mok S, et al. CTLA4 blockade induces frequent tumor infiltration by activated lymphocytes regardless of clinical responses in humans. *Clin Cancer Res* 2011;17:4101–9.
52. Beatty GL, Haas AR, Maus MV, Torigan DA, Soulen MC, Plesa G, et al. Mesothelin-specific chimeric antigen receptor mRNA-engineered T cells induce anti-tumor activity in solid malignancies. *Cancer Immunol Res* 2014;2:112–20.
53. Porter DL, Levine BL, Kalos M, Bagg A, June CH. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. *N Engl J Med* 2011;365:725–33.
54. Grupp SA, Kalos M, Barrett D, Aplenc R, Porter DL, Rheingold SR, et al. Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. *N Engl J Med* 2013;368:1509–18.
55. Davila ML, Riviere I, Wang X, Bartido S, Park J, Curran K, et al. Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia. *Sci Transl Med* 2014;6:224ra25.
56. Mehnert JM, Monjabez AM, Beerthuijzen JMT, Collyar D, Rubinstein L, Harris LN. The challenge for development of valuable immuno-oncology biomarkers. *Clin Cancer Res* 2017;23:4970–9.
57. Quinten C, Coens C, Mauer M, Comte S, Sprangers MA, Cleeland C, et al. Baseline quality of life as a prognostic indicator of survival: a meta-analysis of individual patient data from EORTC clinical trials. *Lancet Oncol* 2009;10:865–71.
58. Escudier B, Porta C, Bono P, Powles T, Eisen T, Sternberg CN, et al. Randomized, controlled, double-blind, cross-over trial assessing treatment preference for pazopanib versus sunitinib in patients with metastatic renal cell carcinoma: PISCES Study. *J Clin Oncol* 2014;32:1412–8.
59. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;22:1889–905.
60. Food and Drug Administration UDOHaHS. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. USA: FDA. 2009.</bib>
61. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. Validity and reliability of the US National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *JAMA Oncol* 2015;1:1051–9.
62. Chen TW, Razak AR, Bedard PL, Siu LL, Hansen AR. A systematic review of immune-related adverse event reporting in clinical trials of immune checkpoint inhibitors. *Ann Oncol* 2015;26:1824–9.
63. Cella DF, Tulsy DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. *J Clin Oncol* 1993;11:570–9.
64. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;85:365–76.
65. 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.
66. 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.
67. Wolchok JD, Neyns B, Linette G, Negrier S, Lutzky J, Thomas L, et al. Ipilimumab monotherapy in patients with pretreated advanced melanoma: a randomised, double-blind, multicentre, phase 2, dose-ranging study. *Lancet Oncol* 2010;11:155–64.
68. Ribas A, Kefford R, Marshall MA, Punt CJ, Haanen JB, Marmol M, et al. Phase III randomized clinical trial comparing tremelimumab with standard-of-care chemotherapy in patients with advanced melanoma. *J Clin Oncol* 2013;31:616–22.
69. 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.
70. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet* 2017;389:255–65.
71. Robert C, Thomas L, Bondarenko I, O'Day S, Weber J, Garbe C, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med* 2011;364:2517–26.
72. Schwartztruber DJ, Lawson DH, Richards JM, Conry RM, Miller DM, Treisman J, et al. gp100 peptide vaccine and interleukin-2 in patients with advanced melanoma. *N Engl J Med* 2011;364:2119–27.
73. Kirkwood JM, Ibrahim JC, Sondak VK, Richards J, Flaherty LE, Ernstoff MS, et al. High- and low-dose interferon alfa-2b in high-risk melanoma: first analysis of intergroup trial E1690/S9111/C9190. *J Clin Oncol* 2000;18:2444–58.
74. Maio M, Grob JJ, Aamdal S, Bondarenko I, Robert C, Thomas L, et al. Five-year survival rates for treatment-naïve patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. *J Clin Oncol* 2015;33:1191–6.
75. Eggermont AM, Chiarion-Sileni V, Grob JJ, Dummer R, Wolchok JD, Schmidt H, et al. Adjuvant ipilimumab versus placebo after complete resection of high-risk stage III melanoma (EORTC 18071): a randomised, double-blind, phase 3 trial. *Lancet Oncol* 2015;16:522–30.
76. Schadendorf D, Hodi FS, Robert C, Weber JS, Margolin K, Hamid O, et al. Pooled analysis of long-term survival data from phase II and phase III trials of ipilimumab in unresectable or metastatic melanoma. *J Clin Oncol* 2015;33:1889–94.
77. Hoos A. Evolution of end points for cancer immunotherapy trials. *Ann Oncol* 2012;23Suppl 8:viii47–52.
78. Chen TT. Statistical issues and challenges in immuno-oncology. *J Immunother Cancer* 2013;1:18.
79. Mick R, Chen TT. Statistical challenges in the design of late-stage cancer immunotherapy studies. *Cancer Immunol Res* 2015;3:1292–8.
80. Hoos A, Eggermont AM, Janetzki S, Hodi FS, Ibrahim R, Anderson A, et al. Improved endpoints for cancer immunotherapy trials. *J Natl Cancer Inst* 2010;102:1388–97.

81. Othus M, Barlogie B, Leblanc ML, Crowley JJ. Cure models as a useful statistical tool for analyzing survival. *Clin Cancer Res* 2012;18:3731–6.
82. Hoos A, Britten CM, Huber C, O'Donnell-Tormey J. A methodological framework to enhance the clinical success of cancer immunotherapy. *Nat Biotechnol* 2011;29:867–70.
83. Fleming TR, Harrington DP. *Counting process and survival analysis*. Chichester: John Wiley & Sons. 1991.
84. Saegusa T, Di C, Chen YQ. Hypothesis testing for an extended cox model with time-varying coefficients. *Biometrics* 2014;70:619–28.
85. Hasegawa T, Uno H, Wei LJ. Nivolumab in nonsquamous non-small-cell lung cancer. *N Engl J Med* 2016;374:492–3.
86. Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol* 2014;32:2380–5.