Review Article

Do immune checkpoint inhibitors need new studies methodology?

Roberto Ferrara[1], Sara Pilotto[2], Mario Caccese[2], Giulia Grizzi[2], Isabella Sperduti[3], Diana Giannarelli[3], Michele Milella[3], Benjamin Besse[1], Giampaolo Tortora[2], Emilio Bria[2]

[1]Department of Medical Oncology, Gustave Roussy, Villejuif, France; [2]U.O.C. Oncology, University of Verona, Comprehensive Cancer Center, Azienda Ospedaliera Universitaria Integrata, Verona, Italy; [3]Regina Elena National Cancer Institute, Roma, Italy

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

Correspondence to: Prof. Emilio Bria, MD. U.O.C. Medical Oncology, University of Verona, Comprehensive Cancer Center, Azienda Ospedaliera Universitaria Integrata, P.le L.A. Scuro 10, 37124 Verona, Italy. Email: emilio.bria@univr.it.

Abstract: Immune checkpoint inhibitors (ICI) have widely reshaped the treatment paradigm of advanced cancer patients. Although multiple studies are currently evaluating these drugs as monotherapies or in combination, the choice of the most accurate statistical methods, endpoints and clinical trial designs to estimate the benefit of ICI remains an unsolved methodological issue. Considering the unconventional patterns of response or progression [i.e., pseudoprogression, hyperprogression (HPD)] observed with ICI, the application in clinical trials of novel response assessment tools (i.e., iRECIST) able to capture delayed benefit of immunotherapies and/or to quantify tumor dynamics and kinetics over time is an unmet clinical need. In addition, the proportional hazard model and the conventional measures of survival [i.e., median overall or progression free survival (PFS) and hazard ratios (HR)] might usually result inadequate in the estimation of the long-term benefit observed with ICI. For this reason, innovative methodologies such as milestone analysis, restricted mean survival time (RMST), parametric models (i.e., Weibull distribution, weighted log rank test), should be systematically investigated in clinical trials in order to adequately quantify the fraction of patients who are "cured", represented by the tails of the survival curves. Regarding predictive biomarkers, in particular PD-L1 expression, the integration and harmonization of the existing assays are urgently needed to provide clinicians with reliable diagnostic tests and to improve patient selection for immunotherapy. Finally, developing original and high-quality study designs, such as adaptive or basket biomarker enriched clinical trials, included in large collaborative platforms with multiple active sites and cross-sector collaboration, represents the successful strategy to optimally assess the benefit of ICI in the next future.

Keywords: Immune checkpoint inhibitors (ICI); long-term benefit; survival analysis; milestone; clinical trial design

Submitted Jan 14, 2018. Accepted for publication Jan 19, 2018.
doi: 10.21037/jtd.2018.01.131
View this article at: http://dx.doi.org/10.21037/jtd.2018.01.131

Introduction

Over the past years, immunotherapy has brought a paradigm shift in the treatment of advanced cancer patients. Nowadays, 26 immunotherapies have gained the approval from regulatory agencies and proofs of benefit have been reported in at least seventeen cancer types (1). In particular, twenty-five indications for six cytotoxic T-lymphocyte antigen-4 (CTLA-4) or programmed death-1 (PD-1) and its ligand PD-L1 inhibitors, have granted Food and Drug Administration (FDA) approval for metastatic solid tumors from March 2011 to August 2017 (2) (*Table 1*). In addition, several combinatorial treatment strategies are currently being tested. Nearly

Table 1 FDA and EMA approved indications for immune checkpoint inhibitors in advanced solid cancers

Drug	Indications	Line	Primary endpoint	FDA approval	EMA approval
Ipilimumab	Melanoma	≥2nd	OS	Approved	Approved
		1st	ORR	Approved	Approved
		Any in combination with nivolumab	PFS	Approved	Approved
Nivolumab	Melanoma	≥2nd	ORR	Approved	Approved
		1st	OS		
	NSCLC[¤]	≥2nd	OS	Approved	Approved
	RCC	≥2nd	OS	Approved	Approved
	SCCHN	≥2nd	OS	Approved	Approved
	UCC	≥2nd	ORR	Approved	Approved
	MSI-H CRC	≥2nd	ORR	Approved	–
	HCC	≥2nd *	ORR	Approved	–
Pembrolizumab	Melanoma	≥2nd	ORR	Approved	Approved
		1st	OS		
	NSCLC[¤]	≥2nd †	ORR	Approved	Approved
		1st ‡	PFS	Approved	Approved
		1st in combination with chemotherapy §	ORR	Approved	–
	SCCHN	≥2nd	ORR	Approved	–
	UCC	≥2nd	OS	Approved	Approved
	MSI-H any histology	≥2nd	ORR	Approved	–
	Gastric/GEJ	≥2nd	ORR	Approved	–
Atezolizumab	NSCLC[¤]	≥2nd	OS	Approved	Approved
	UCC	≥2nd	ORR	Approved	Approved
Durvalumab	UCC	≥2nd	ORR	Approved	–
Avelumab	UCC	≥2nd	ORR	Approved	–
	Merkel carcinoma	≥2nd	ORR	Approved	Approved

*, after progression to sorafenib; †, in patients with PD-L1 expression ≥1%; ‡, in EGFR and ALK wild type patients with PD-L1 expression ≥50%; §, platinum-based chemotherapy; ¤, both squamous and non-squamous histology; NSCLC, non-small cell lung cancer; RCC, renal cell carcinoma; UCC, urothelial carcinoma; CRC, colorectal cancer; MSI-H, microsatellite instability high; SCCHN, squamous cell carcinoma of the head and neck; GEJ, gastroesophageal junction; ORR, overall response rate; PFS, progression-free survival; OS, overall survival; FDA, Food and Drug Administration; EMA, European Medical Agency.

1,502 clinical trials are evaluating PD-1/PD-L1 inhibitors in cancer patients, of these 1,105 are combination studies of anti PD-1/PD-L1 agents with other immunotherapies, targeted therapies, chemotherapies or radiotherapies (1).

Interestingly, the rapidity of clinical trial enrollment and regulatory agencies accelerated approval has left many unsolved issues to explore in the next wave of immuno-oncology trials. Specifically, relevant unanswered questions concern the optimal study design, endpoints and statistical methods for evaluating immunotherapeutic drugs, the appropriate radiological assessment of antitumor responses, the development of predictive biomarkers and

the harmonization of the assays to test these biomarkers in large patient populations. Most of these issues are related to the intrinsic mechanism of action and kinetic of immune checkpoint inhibitors (ICI). Differently from chemotherapy and targeted agents, ICI induce a continuum of biological events that starts early with immune system activation and that procrastinates until the ideal obtainment of a (sometimes) delayed clinical benefit. This peculiar feature should be carefully considered when designing clinical trials with ICI and innovative study methodologies should be applied to appropriately assess the delayed effect of immunotherapeutic agents in terms of responses and survival benefit.

This review will explore the major methodological issues and challenges regarding endpoints, statistical methods, predictive biomarkers assessment and clinical trials design for ICI, focusing in particular on non-small cell lung cancer (NSCLC) patients.

Methodology and endpoints in clinical trials with ICI

In the immune-oncology era, traditional endpoints of randomized clinical trials, such objective response rate (ORR) according to Response Evaluation Criteria in Solid Tumors (RECIST 1.1) (3), median progression-free survival (PFS) and overall survival (OS) have been extensively questioned. In fact, although they are appropriate for assessing the activity of agents able to induce rapid control of tumor growth, such as targeted or cytotoxic therapies, they may be less suitable for treatments, such as immune checkpoint blockade, where tumor control may develop over time. In particular, unconventional response patterns as pseudoprogression (4) or dissociated responses (5), have been recently characterized in different tumor types treated with ICI and they are generally not adequately identified using traditional RECIST 1.1. Similarly, median PFS based on RECIST 1.1 potentially underestimates the activity of ICI in patients with prolonged stable disease or unconventional responses, and median OS or hazard ratios (HR) are largely suboptimal to capture the key attributes of immunotherapeutic agents, such as delayed clinical effect and long-term survival (6). For these reasons, alternative response evaluation criteria, namely irRC (7), irRECIST (8) and iRECIST (9) and innovative statistical models, such as milestone analysis (10,11), weighted log-rank test (12), restrictive mean survival time (13) and Weibull distribution (14) are currently under development in order to assess delayed effect and prolonged survival of ICI.

Moreover, the conventional oncology specific frameworks, traditionally used to estimate the value of cancer drugs, should also be modified, taking in account the new concept of durable clinical benefit. The American Society of Clinical Oncology (ASCO) has recently published an update of the original framework, incorporating the evaluation of long-term survival. Specifically, bonus points were awarded if the experimental regimen resulted in at least a 50% relative improvement in the percentage of patients alive at a time point corresponding to twice the median OS or PFS point for the control regimen and if at least 20% of patients receiving the control regimen were alive at this time (15). This novel ASCO framework could properly assess the clinical benefit of immunotherapies and recently it was used to review FDA approvals for ICI (2). Interestingly, only 3 out of 23 indications examined gained the long tail bonus points, namely second line ipilimumab and first line nivolumab for metastatic melanoma, second line nivolumab for squamous advanced NSCLC. Considering that 9 out of 23 approvals achieved the 50% improvement in patients alive in the experimental regimen compared with the standard treatment but did not receive bonus points because less than 20% of patients were alive in the control arm, definitive conclusions on where setting the bar to define a significant survival improvement with new immune-oncology agents are difficult to be drawn (16).

Objective overall response rate

Although it is a common belief that survival represents the main endpoint for regulatory agency approvals, 15 (60%) out of 25 FDA approvals for ICI were based on ORR as primary endpoint (2). Interestingly, in some patients treated with ICI, initial disease progression assessed by conventional tumor response criteria, such as WHO criteria (17) or RECIST 1.1 (3), may be followed by prolonged clinical stabilization or partial/complete responses. This phenomenon defined as pseudoprogression is caused by T-cell tumor infiltration as a result of immune activation and was described both with anti-PD-1/PD-L1 and anti-CTLA-4 agents in advanced melanoma patients (18,19) and with PD-1/PD-L1 inhibitors in advanced renal cell carcinoma (RCC) (20) and NSCLC patients (5,21,22). The emerging of pseudoprogression and dissociated responses to ICI brought to the development in 2009 of immune related response criteria (irRC) (7). The key differences compared to RECIST criteria were the

introduction of bidimensional measurements (sum of products of the two largest perpendicular diameters), the inclusion of new lesions [usually classified as progressive disease (PD) according to RECIST 1.1] in the total tumor burden and the requirement of confirmation of PD on two consecutive scans at least 4 weeks apart. Subsequently, unidimensional irRC (irRECIST), which used the longest diameter measurements as in RECIST, demonstrated high concordance compared to bidimensional irRC, bypassing the methodological issues linked to the use of bidimensional measurements (8). Finally, the RECIST working group has recently developed a guideline for the use of modified RECIST (named iRECIST) in order to establish a common framework for the management of data from clinical trials with ICI (9). As irRECIST, iRECIST introduced the concept of immune unconfirmed PD (iUPD) which consents to reset the bar if RECIST progression is followed at the next assessment by tumor shrinkage. Basically, the main difference between irRECIST and iRECIST regards the new lesions, which are incorporated into the sum of target lesions in irRECIST while in iRECIST are recorded separately. However, high concordance has been recently reported between irRECIST and iRECIST in a retrospective study including advanced NSCLC patients treated with anti PD-1/PD-L1 agents. Interestingly, for only ~4% of NSCLC patients there was a mismatch between irRECIST and iRECIST, where iRECIST interpretation as iUPD led to unnecessary continuation of immunotherapy (5). To date, few clinical trials have used irRC/irRECIST as secondary response endpoints (23-25) and none has used iRECIST as response criteria to define their endpoints. Therefore, the regulatory agencies continue to base the approvals of new ICI on RECIST 1.1 defined outcomes. In the future, the integration and validation of irRECIST/iRECIST in clinical trials will be of paramount importance in order to provide to immuno-oncologists a practical and reliable tool to face the dilemma about whether and when continue immunotherapy beyond progression.

Another emerging challenge for immunotherapy trials is represented by the evaluation of accelerated tumor growth under ICI, a phenomenon known as hyperprogression (HPD) and recently described in 9% of advanced cancer patients (26), in 29% of head and neck cancers (27) and in 14% of NSCLC patients treated with ICI (28). Although each study used different methodologies to assess HPD, all of them highlighted the importance of measuring tumor growth speed on consecutive computed topography (CT) scans, before the start and during immunotherapy treatment. Retrospective evaluation of HPD in published randomized studies is actually difficult because the CT scans data before immunotherapy start are usually not captured. Therefore, a prospective assessment of HPD in adequately designed clinical trials, which collect CT scans before and during ICI and adopt innovative radiologic tools to quantify tumor kinetics and dynamics over time, will provide a confirmatory evidence regarding this rapid and atypical phenomenon. Finally, the use of ORR as a surrogate endpoint for OS in trials with ICI remains an unsolved question. A meta regression analysis of seventeen randomized trials testing ICI showed a weak but statistically significant correlation between the treatment effect on the ORR and the treatment effect on survival outcomes (i.e., OS and PFS) and suggested that the activity of ICI in terms of ORR explain ~50% of the effects detected in survival (29). Conversely, a systematic review of ten clinical trials evaluating PD-1/PD-L1 inhibitors in advanced NSCLC failed to show a significant correlation between response and survival (30). Considering that ICI activity potentially leads to prolonged disease stabilization and/or unconventional responses, it is likely that disease control rate (DCR), including both responses and tumor stabilization for at least 6 months of treatment (clinical benefit), may be a more clinically relevant surrogate endpoint for survival compared to ORR. The potential future validation of ORR or clinical benefit as surrogate endpoints for survival may consent an earlier analysis of trial data, allowing less expensive and prolonged studies and, most of all, rapidly addressing progressive patients towards other treatments.

PFS

Before the coming of ICI in the cancer treatment scenario, PFS has traditionally been considered a reasonable endpoint for new drug approval in a series of solid tumor, including lung cancer. Unlike OS, PFS is not influenced by post-progression therapies and it can provide an earlier assessment of efficacy and a direct measure of treatment effect, avoiding bias related to crossover (31). In many scenarios (32), and in particular in locally advanced lung cancer, a significant correlation between PFS and OS has been demonstrated (33). As a general rule, being OS the sum of PFS and survival post progression (SPP), the longer is the SPP, the lower is the chance that PFS and OS correlate (34). Regarding oncogene addicted NSCLC patients treated with targeted agents, the ratio between the HR for PFS and the

Table 2 HR for PFS and OS and HR PFS/OS rate for the main randomized phase II and III clinical trials of single agent ICI in advanced NSCLC patients

Trial	Phase	Line	ICI	Treatment arms	Patients (n)	mOS (month)	HR OS	mPFS (month)	HR PFS	HR PFS/OS
Checkmate 017 (37)	III	≥2nd	Nivolumab	Nivolumab 3mg/kg q2 weeks; docetaxel	135; 137	9.2; 6	0.59	3.5; 2.8	0.62	1.05
Checkmate 057 (38)	III	≥2nd	Nivolumab	Nivolumab 3mg/kg q2 weeks; docetaxel	292; 290	12.2; 9.4	0.73	2.3; 4.2	0.92	1.26
Checkmate 026 (39)	III	1st	Nivolumab	Nivolumab 3mg/kg q2 weeks; platinum-based CT	271; 270	14.4; 13.2	1.02	4.2; 5.9	1.15	1.12
KEYNOTE 010 (25)	II/III	≥2nd	Pembrolizumab	Pembrolizumab 10 mg/kg q3 weeks; pembrolizumab 2 mg/kg q3 weeks; docetaxel	346; 345; 343	12.7; 10.4; 8.5	0.71*	4; 3.9; 4	0.88*	1.23
KEYNOTE 024 (40)	III	1st	Pembrolizumab	Pembrolizumab; platinum-based CT	154; 151	30; 14.2	0.63	10.3; 6	0.50	0.79
POPLAR (41)	II	≥2nd	Atezolizumab	Atezolizumab 1,200 mg q3 weeks; docetaxel	144; 143	12.6; 9.7	0.77	2.7; 3	0.98	1.27
OAK (42)	III	≥2nd	Atezolizumab	Atezolizumab 1,200 mg q3 weeks; docetaxel	425; 425	13.8; 9.6	0.73	2.8; 4	0.95	1.31

*, comparison pembrolizumab 2 mg/kg q3 weeks versus docetaxel; ICI, immune checkpoint inhibitors; q, every; mOS, median overall survival; HR, hazard ratio; mPFS, median progression free survival; CT, chemotherapy.

HR for OS has usually been inferior than 1, indicating that a larger benefit in PFS can translate in smaller advantages in OS. In the immune-oncology era this paradigm has been revolutionized (35). In fact, ICI usually induce a delayed clinical benefit that is not always adequately captured by PFS based on conventional RECIST 1.1, whereas it significantly improves OS. Therefore, in randomized trials testing ICI in advanced NSCLC patients, HR for PFS is generally lower than HR for OS and the ratio between them is higher than 1 (range, 1.05–1.31), with the exception of the study comparing pembrolizumab to platinum based chemotherapy in NSCLC patients with PD-L1 expression ≥50% (KEYNOTE-024), in which a large PFS improvement was observed in the immunotherapy arm (36). In *Table 2* are reported the HR for PFS and OS and the HR PFS/OS rate for the main randomized phase II and III trials of single agent ICI in advanced NSCLC patients.

In studies evaluating ICI, the median PFS does not consistently reflect the long-term benefit of treatment. For instance, in phase III trials evaluating ICI in pretreated advanced NSCLC, long-term responses are observed in a proportion of patients (15–20%) that is similar or inferior compared to progressing patients (33–44%) (37,38,42) and for this reason, the median PFS (ranging from 2 to 4 months) (25,37,38,42) will definitively underestimate the effect of ICI in responders. The PFS rate at 1–3 years could be an alternative survival measure and a potential surrogate endpoint for OS benefit (43). In this regard, a retrospective analysis of NSCLC patients treated with PD-1/PD-L1 inhibitors in a single institution, showed that PFS rate at 2 years significantly correlated with longer OS (30). Although choosing PFS as primary endpoint for studies evaluating ICI is questionable because it cannot adequately capture a delayed survival benefit, the need to use PFS in the approval process of new immunotherapeutic agents is inevitably increasing. In fact, the proved efficacy of immunotherapy in different disease settings makes unethical the absence of cross over in trial designs, and for this reason a significant improvement in OS would be a difficult goal to achieve in the next generation randomized studies with ICI.

OS

Traditionally, OS is considered the gold standard among efficacy endpoints in clinical trials and median OS is often quoted as the primary or secondary endpoint of interest. However, median OS may not be the best endpoint for therapies with potential long-term benefit. This observation

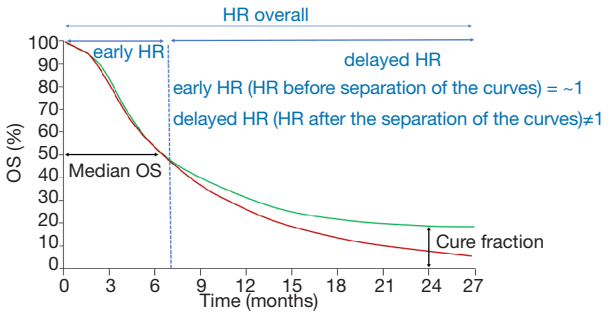

Figure 1 Hypothetical survival curves of an immune checkpoint inhibitor (green line) associated with a long-term benefit compared to a standard non-immunotherapeutic agent (i.e., cytotoxic chemotherapy) (red line).

was reported for the first time in clinical trials evaluating cancer vaccines, such as the phase III study comparing sipuleucel-T, an autologous active cellular immunotherapy, to placebo in advanced prostate cancer patients, where the effect on survival was not evident for the first 8 months of treatment (44). Similarly, also phase III trials of CTLA-4 or PD-1/PD-L1 inhibitors in advanced melanoma (45) and NSCLC (38,42) patients showed delayed separation of survival curves, or even a cross between them with an initial better survival outcome for chemotherapy compared to ICI, as observed in Checkmate 057, a phase III trial comparing nivolumab versus docetaxel in pretreated non-squamous NSCLC patients (38). Recently, an update of the phase I CA209-003 trial testing nivolumab in 129 previously-treated NSCLC patients showed that 5-year OS was 16% for squamous and 15% for non-squamous patients (46), however the 9.9 months of median OS did not adequately estimate the durable benefit demonstrated by the plateaus in the tails of the survival curves. In *Figure 1* is reported the hypothetical survival curve of a treatment (i.e., immunotherapy) that leads to long-term survival in a small proportion of patients (green line) compared to a standard therapy, potentially a cytotoxic agent, (red line) not associated with a prolonged survival benefit. Median OS, calculated as the time point after initiation of the treatment at which 50% of patients are still alive, clearly does not provide any information concerning the minor proportion of patients who occupies the tail of the curves (cure fraction). Therefore, median OS neither differentiates the proportion of patients alive or dead after 50% of patients have died nor reflects the survival time of the patients who are alive after the median OS is reached. In addition, the delayed clinical effect observed with ICI leads to the loss of statistical power if the trial is designed based on conventional proportional hazard model assumption (12). According to the proportional hazard model, HR is equal to 1 in the first part of the curves (early HR) and it becomes unequal to 1 after the separation of the curves (delayed HR). To demonstrate a statistically significant difference in OS, the delta between these two HRs should be high, in fact the HR after the separation of the curves must compensate the lack of separation during the first months of treatment (47) (*Figure 1*). However, the number of events required to have a large delta value should also increase and the study risks to definitively result as underpowered. In this regard, a recent report by the Institute for Clinical and Economic Review (ICER) highlighted the difficulty in using a proportional hazard model in studies evaluating ICI in advanced NSCLC patients (48). In particular, the ICER analysis stated that the existence of two populations in the immune-oncology arms of the trials, a majority who does not respond to ICI and has a high hazard for survival and a minority with sustained responses and low hazard for progression and mortality, makes difficult the use of proportional hazard models for survival analysis. Notably, survival curve statistic that optimally captures the benefit offered by a particular therapy can differ according to the class of drugs or the clinical context (35). As an example, traditional statistical methods (log rank and Cox model) and survival measures (median OS and HRs) can be usually applied for drugs that start to work early (the OS curve separate since the beginning) and continue to be more active compared to the control arm along the treatment is administered with the assumption that anything affecting the hazard does so by the same ratio at all times (*Figure 2A*). Median OS but not HR could be used for non-proportional risk models with absence of long-term survivors, as observed in trials evaluating targeted agents (*Figure 2B*). In fact, the initial large benefit driven by the target agent is entirely capture by the median survival, however with the emerging of resistance this difference disappears and the survival curves cross at a certain time, making the assumption of proportional hazards not applicable in this case. For drugs with delayed benefit, which lead to prolonged survival in a relatively small subset of patients, following a non-proportional risk model (*Figure 2C*), neither median OS, nor HR are appropriate and alternative statistical methods and survival measures should be reported.

Milestone survival analysis is a cross sectional assessment of OS at a pre-specified and clinically meaningful timepoint, using Kaplan Meier survival probabilities. Milestone

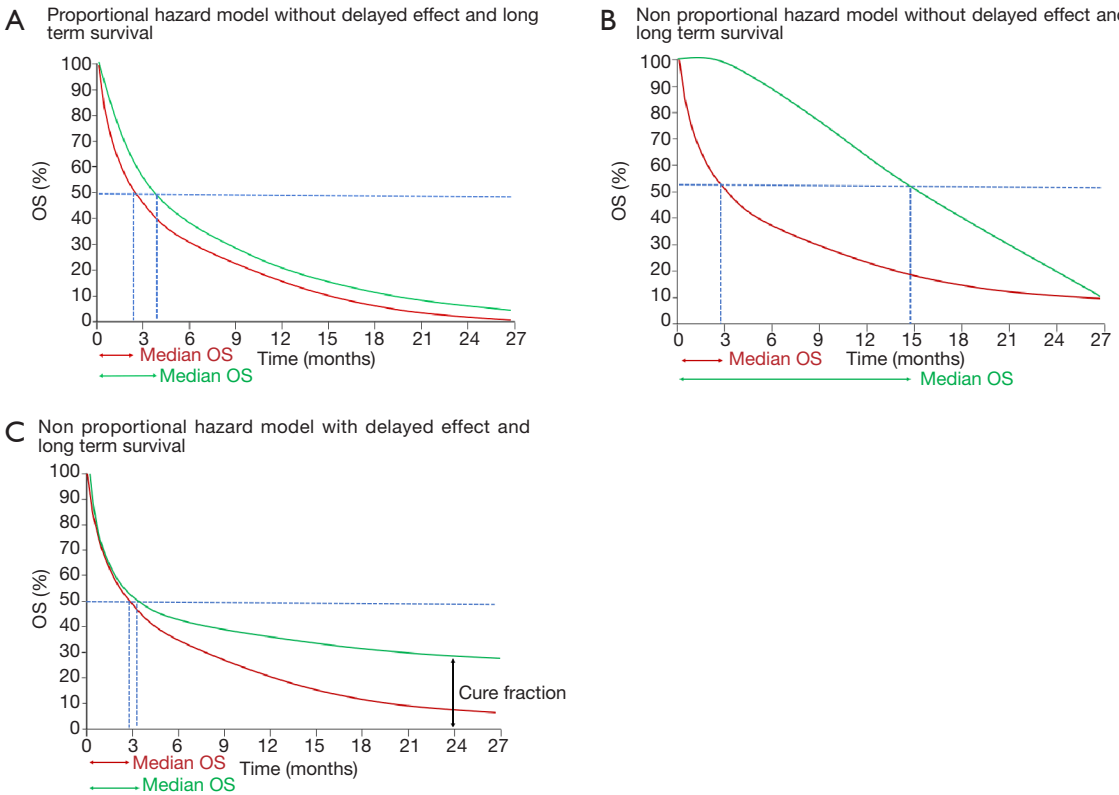

Figure 2 Kaplan-Meier survival curves according to the class of drugs. (A) Drug starts to work early and continues to be more active compared to the control arm following the proportional hazard assumption, without long-term survival benefit; (B) treatment starts to work early, however resistance occurs with absence of prolonged survival benefit (non-proportional hazard model); (C) treatment starts to work late, however a long-term survival benefit is observed (non-proportional hazard model).

analysis is usually conducted in a first cohort of randomized patients rather than in the whole population, and it provides long-term survival information in patients with sufficient follow-up, while the entire study continues with OS as primary endpoint. The main characteristic of milestone model is the requirement of a sufficient follow-up, long enough to allow robust estimation of the survival rates. In fact, the milestone analysis should not be conducted until all the patients have met the minimal follow-up time (11). Milestone analysis is based on the assumption deriving from cure rate models that the study population includes a distinct subset of patients who are "cured" and are represented by the tail in the Kaplan Meier curve (49). In addition, milestone outcome better reflects patients hopes and it is much more informative compared to median OS and HR because it answers to patients' primary interest: the potential rate of cure with a specific treatment (10). The application of milestone in the immune-oncology trials could be useful to avoid wrong interpretations of survival data deriving from early interim analyses. As an example, an interim analysis of the phase III trial comparing first line tremelimumab to chemotherapy in advanced melanoma patients (50) showed no OS benefit, however, an extended follow-up revealed a potential separation of the curves, supporting the use of a milestone model to estimate the true survival benefit (51). Challenges with the milestone survival analysis are represented by the choice of the sample size cohort (in fact milestone does not account for the totality of the OS data), the selection of an optimal threshold for the type I error rate at the time of milestone analysis, the difficulty in maintaining study integrity and blinding prior to the final OS analysis and in assessing the post milestone treatment impact on survival (11). Besides that, the most important concern of the milestone analysis is the identification of a meaningful milestone time point. In fact, the survival analysis could be imprecise if the milestone timepoint is too early (too many events censored and no differences in survival) or too late (the set of patients at risk

is small) (10). To avoid the latter condition, a milestone timepoint at which 10% to 20% of the patients in a Kaplan-Meier survival curve remains at risk has been previously proposed (52).

Another novel survival tool currently emerging is the RMST or t-year mean survival time. RMST is a robust statistical procedure able to quantify treatment effect on survival, regardless of the model assumption. Visually, RMST is the area under the survival curve within a specific time window. As milestone analysis, RMST require that the follow-up duration is pre-specified and fixed (53-55). RMST has been recently applied to Checkmate 057 (38). In this study, the median OS in the nivolumab arm (12.2 months) does not adequately capture the long-term survival benefit estimated to be 18% at 3 years (56), similarly the HR of 0.73 cannot be used as a survival measure considering that the 2 survival curves were similar for the first 6 months of treatment and the proportional hazard assumption was not valid. With a follow-up duration of 24 months, the RMST is 13 months for nivolumab versus 11.3 months for docetaxel with a statistically significant difference of 1.7 months (95% CI, 0.4–3.1; P=0.01) in favor of nivolumab. Graphically this difference is represented by the area between the two Kaplan-Meier curves. In Checkmate 057, RMST of 13 months for the nivolumab arm means that NSCLC patients receiving nivolumab and followed for 24 months would survive for an average of 13 months (13).

The weighted log rank test is an additional statistical tool that can be used for non-proportional survival models with delayed clinical benefit and long-term survival. Basically, a weighted log rank test avoids loss of statistical power because it reduces the statistical weight of the early time period, during which survival curves might be similar (12,57). Weighted log ranks have been proposed as novel statistical methods for studies with ICI. However, the a priori definition of the weights is usually difficult because the point at which the survival curve diverge cannot be easily predicted at the start of the trial. Finally, besides weighted log rank, the Weibull distribution represents another parametric survival model, which could provide additional useful information in evaluating ICI effect on survival. In the Weibull model, the survival time depends on the shape parameter of survival curves and hazard is not a constant but a function of time (14,58). The Weibull model can fit well to the immune-oncology clinical trials because it takes into account the different shapes of survival curves and their variation during time. The Weibull model also allows the inclusion of covariates of survival times, useful to describe long tailed distributions.

Safety

Conventional safety analysis using the 3+3 dose escalation design and the first two cycles as a dose limiting toxicity (DLT) assessment period (59) might not accurately describe the safety profile of ICI. Remarkably, grade 3–4 immune related adverse events (irAE) with PD-1/PD-L1 inhibitors are infrequent, for example in pretreated NSCLC they range from 7% to 20% (25,37,38,42), and toxicities are observed usually late (from months to years) (60). Therefore, the phase II dose is often determined on the base of pharmacokinetic (PK)/pharmacodynamic profile or of the maximal administered dose rather than on the maximal tolerated dose (MTD). Future studies should introduce longer DLT periods (≥6 weeks) before escalade to higher doses in order to adequately select the phase II appropriate dose.

Regarding anti CTLA-4 agents, irAE can rapidly become life threatening (61), for this reason, in some protocols it is recommend to hold the anti-CTLA-4 until recovering from grade 2 toxicities. Although these irAE decrease drug exposure and definitively influence treatment dose intensity, they are not formally classified as DLT because they are not grade 3–4 toxicities. Finally, late irAE presenting several months after the last dose of ICI have been described (62), highlighting the importance to incorporate longer follow-up periods (up to 1 year) in clinical trials evaluating ICI in order to capture late post discontinuation toxicities and to characterize their effect on subsequent anticancer therapies. For these reasons, innovative phase I trial designs should better define the timing and the best way to assess DLT and MTD for ICI.

PD-L1 as predictive biomarker in NSCLC patients: challenges and methodological issues

A challenging unmet need for clinical trials evaluating ICI is the absence of reliable biomarkers of response to immunotherapeutic agents, able to identify before the treatment initiation which patients are more likely to experience clinical benefit. Although several biomarkers are currently being tested in different disease settings (63), most of the available data from clinical trials with PD-1/PD-L1 inhibitors evaluated the predictive role of PD-L1 expression. Interestingly, across different tumor types, a relevant number of patients with PD-L1 positivity (40–50%) does not achieve objective response to anti PD-1/PD-L1

therapies, in addition 15% of patients negative for PD-L1 expression, experience objective responses (64). However, a significant correlation between PD-L1 expression and ORR to anti PD-1/PD-L1 agents was reported by a sensitivity analysis of trials investigating PD-1/PD-L1 inhibitors in different cancer types (65) and by a meta-analysis in NSCLC patients (66).

In particular, in advanced lung cancer, the predictive role of PD-L1 did not clearly emerged from clinical trials evaluating nivolumab in pretreated NSCLC patients (37,38,46,67), although a post hoc analysis from Checkmate 057 suggested an improved benefit for patients with PD-L1 positive tumors at the threshold levels of 1%, 5% and 10% (38). Similarly, PD-L1 expression was not predictive of nivolumab benefit in first line setting (39,68). On the contrary, the phase II (KEYNOTE-010) (25) and III (KEYNOTE-024) (36) development of pembrolizumab in NSCLC was restricted to PD-L1 positive patients (threshold 1% for KEYNOTE-010 and 50% for KEYNOTE-024) due to the higher ORR (45% for PD-L1 ≥50%, 16.5% for PD-L1 in the range of 1–49% and 10.7% for PD-L1 <1%) observed in the expansion cohort of a phase I trial (KEYNOTE-001) (69). In the case of the anti-PD-L1 antibody atezolizumab, PD-L1 expression was evaluated on both tumor cells (TC) and immune cells (IC) (70). Although the survival benefit with atezolizumab appeared to correlate with PD-L1 expression on TC and IC in pretreated NSCLC patients (POPLAR trial) (41), this finding was not confirmed in the phase III OAK trial which showed a significant OS benefit in favor of atezolizumab compared to docetaxel regardless of PD-L1 expression (42). In first line setting, the development of atezolizumab followed a different strategy and two phase II trials, BIRCH (71) and FIR (72) tested atezolizumab only in PD-L1 positive (TC and IC) patients. Finally, early trials evaluating durvalumab and avelumab showed a higher response rate in patients with PD-L1 expression on TC ≥25% (ATLANTIC trial) (73) and ≥1% (Javelin solid tumor trial) (74), respectively.

A comprehensive characterization of the putative predictive role of PD-L1 or other biomarkers, such as tumor mutational burden (75,76), IFN-γ mRNA expression (77), tumor infiltrating lymphocytes (78,79), serum circulating factors (80), gut microbiota (81) across several tumor types (63) and in NSCLC patients (82), has been recently reported elsewhere and it goes far beyond the aim of this review. Therefore, we will provide an insight on some methodological and practical issues regarding PD-L1 assessment in advanced NSCLC such as concordance between PD-L1 immunohistochemistry (IHC) assays, variability of PD-L1 assessment on TC and IC and the impact of spatial and temporal tumor heterogeneity on PD-L1 expression. Considering that each PD-1/PD-L1 inhibitor has its own PD-L1 diagnostic test and that different level of PD-L1 expression have been evaluated for correlation with clinical outcome in trials in NSCLC patients, several harmonization studies have recently tried to reduce the high variability of the assays. The Blueprint PD-L1 IHC Assay Comparison Project reported high concordance for PD-L1 level detection between 28-8 (IHC test for nivolumab), SP263 (IHC test for nivolumab, pembrolizumab, durvalumab) and 22C3 (IHC test for pembrolizumab) assays, whereas lower PD-L1 expression was detected by SP142, the companion diagnostic test for atezolizumab (83). These findings suggest a potential risk of false negative results when the antibody SP142 is used to detect PD-L1 on tumor samples. On the bases of Blueprint study, PD-L1 expression was re-evaluated with the 22C3 IHC assay in 400 tumors from OAK trial. Surprisingly, atezolizumab was superior to docetaxel in all subgroups, including tumors with less than 1% PD-L1 expression on TC (84). However, these data are still a matter of debate because PD-L1 <1% was found in 55% of the tumors, whereas the expected rate of negative tumors is around 30%, making it likely that the PD-L1 <1% population included false-negative tumors.

The low performance of SP142 in detecting PD-L1 on TC was additionally confirmed in other 3 harmonization trials (85-87), and the high degree of concordance between 28-8, SP263 and 22C3 were consistent across several different studies (85-90). However, conflicting results emerged from recent analyses showing a higher expression of PD-L1 with SP142 test (91) and a lower expression with 22C3 (92) and 28-8 (91) antibodies.

Differently from PD-L1 expression on TC, PD-L1 IHC on IC is characterized by a greater variability and low interobserver concordance. These discordant results might be due to the co-existence of both cytoplasmic and membranous PD-L1 staining in IC (93) and to the lack of pre-specified criteria for assessment of PD-L1 staining on IC. Besides the variability among assays and between PD-L1 assessment on TC or IC, another critical issue is represented by the spatial and temporal heterogeneity of PD-L1 expression (94). In this regard, KEYNOTE-010 comparing pembrolizumab to docetaxel in PD-L1 ≥1% pretreated NSCLC patients (25), showed that the prevalence of PD-L1 levels ≥50% was similar in archival or rebiopsy

samples (~40–45%) and clinical outcomes in patients with PD-L1 ≥50% did not differ between archival and new samples (95). A superimposable result was reported in an exploratory analysis of the ATLANTIC trial (73), a phase II study of durvalumab in pretreated NSCLC patients, which showed high concordance between fresh biopsies acquired 3 months before treatment compared to older tumor samples (96). Similarly, spatial heterogeneity does not seem to strongly influence PD-L1 expression. In fact, a retrospective study (97) and an exploratory analysis of the ATLANTIC trial (96) showed good concordance for PD-L1 expression between primary tumor and metastatic samples. Finally, regarding PD-L1 intratumor heterogeneity, data are conflicting with some studies showing high concordance of PD-L1 staining between different samples of the same tumor site, tested with the same PD-L1 IHC assay (98), and other studies reporting discordant PD-L1 staining from matched specimens (99).

Characterizing inter and intratumor heterogeneity of PD-L1 expression and overcoming the hurdles of inter-assays variability and of discordant TC-IC stains represent important issues that need to be addressed in future clinical trials with ICI in cancer patients.

Clinical trials design for ICI

The traditional clinical trials designs have been widely reshaped by the advent of ICI, with changes concerning all the different phases of drug development.

Regarding phase I trials, considering the low rate of grade 3–4 toxicities and the relatively absence of DLT for PD-1/PD-L1 inhibitors, alternative designs, such as modified toxicity probability interval design, have been recently developed (100,101). According to this model the proportion of targeted DLT can be less than 17% (in classical 3+3 dose escalation phase I trials the targeted proportion of DLT is 17–33% of patients). PK and pharmacodynamic properties of ICI have been explored only in a limited number of phase I trials (102,103). For example, regarding nivolumab, doses from 0.1 to 10 mg/kg demonstrated 64–70% PD-1 receptor occupancy on CD3+ T cells (103) and the initial FDA approved dose of 3 mg/Kg every 2 (q2) weeks was subsequently changed to a flat dosing of 240 mg q2 weeks based on population PK demonstrating comparability of safety and efficacy for most disease indications (104). A model based PK analysis in different cancer types reported that an alternative flat dosing of 480 mg q4 weeks resulted in similar exposure, efficacy and safety as the 3 mg/kg q2 weeks (105). A better knowledge and interpretation of PK data of phase I trials are of paramount importance considering that advanced melanoma (106), RCC (107) and NSCLC (22) patients may experience prolonged responses after treatment discontinuation and that responses may happen also after rechallenge with the same drug (108). In this regard, a phase III/IV trial (Checkmate 153) comparing continuous nivolumab to observation after 1 year of nivolumab in advanced NSCLC patients recently showed an improvement in PFS (not reached versus 10.3 months, HR =0.42; 95% CI, 0.25–0.71) in patients receiving continuous treatment (109). Despite these hypothesis-generating results, additional data and innovative phase I trials with a deeper insight in the pharmacological properties of ICI are urgently needed.

As previously reported for some targeted agents (110), phase I trial testing ICI had to face the issue of answering to multiple clinical questions in a shorter timeframe, with the final aim of reducing the development time from phase I to registration by regulatory authorities. Remarkably, adaptive and basket designs with biomarker enrichment strategies have led to approval of several ICI, revolutionizing the traditional drug development paradigm based on 3 or more steps (phase I, phase II and phase III). In adaptive design, modifications of the trial are prospectively planned, so that changes may take place while the study is ongoing. The main goal of adaptive design trials is to learn and address several hypotheses at one time in order to speed up the development of the compound (111). One example is KEYNOTE-001, a phase I trial which led to FDA approval of pembrolizumab both in advanced melanoma and NSCLC, in a timeframe <4 years (112). With 1,245 patients enrolled, KEYNOTE-001 is the largest phase I trial to date. Its adaptive design (at least 8 protocol amendments including, among others, modification of the primary endpoint from irRC to RECIST 1.1, addition or abandoning of specific cohorts, increasing sample size for certain cohorts), allowed the trial to simultaneously generate several efficacy data rather than starting different studies for each clinical question (112). Basket designs have been successfully adopted in clinical trial evaluating ICI. Basket trials test the effect of a drug on a single target in a variety of cancer types. In basket studies, the investigators can separately analyze the responses of patients by tumor types, and choose to expand or close patient cohorts according to the benefit of the experimental treatment (111). KEYNOTE-059 is an example of a successful and innovative phase II basket

trial, testing pembrolizumab in patients with high level or microsatellite instability (MSI-H) or deficiency in mismatch DNA repair (dMMR). In ~150 patients with MSI-H or dMMR, pembrolizumab showed an ORR of ~40% and responses were observed regardless of tumor histology, leading to the first histology-agnostic FDA approval of a cancer treatment in USA (1,113). In this regard, the histology-independent benefit of ICI clearly differentiates them from targeted agents. In fact, basket trials are not always reliable in oncogene addicted disease, considering that the simple molecular abnormality (i.e., BRAF mutation) does not imply the efficacy of specific inhibitors (i.e., vemurafenib) and that tumor response strongly depend on the disease context (114).

In a recently published guidance, FDA highlights the key role of the enrichment trial design to identify specific subgroups of patients who would benefit from experimental treatments, encouraging physicians to widely adopt this strategy in clinical trials (115). Both KEYNOTE-001 and KEYNOTE-059 trials imply a biomarker based enrichment design. In the expansion cohort of KEYNOTE-001 in advanced NSCLC (69), the cut off selection for PD-L1 positivity and its validation provided the bases for phase II (25) and III (36) trials testing pembrolizumab in PD-L1 positive NSCLC patients. On these premises, the phase III trial comparing first line pembrolizumab to platinum based chemotherapy was specifically designed in EGFR/ALK wild type NSCLC patients with PD-L1 expression ≥50%, and the most updated results showed significant improvements in the ORR (45.5% *vs.* 29.8%), median PFS (10.3 *vs.* 6.0 months, HR =0.50; P=0.001) and median OS (30 *vs.* 14.2 months; HR =0.63; P=0.002) in favor of pembrolizumab (40). However, the reliability of biomarker enrichment strategies for ICI is still a matter of debate, considering that PD-L1 is neither a totally specific, nor a sensitive predictive biomarker, and that several others (such as tumor mutational burden or TIL) are currently being validated in clinical trials. For example, in the phase III OAK study, atezolizumab significantly improved OS compared to docetaxel (13.8 *vs.* 9.6 months; HR =0.73, P=0.0003) in pretreated NSCLC patients, regardless of PD-L1 expression on TC or IC (42), also when PD-L1 expression was evaluated with the 22C3 more sensitive diagnostic assay (84). Furthermore, the recently published Checkmate 026 failed to show a significant improvement in PFS (HR =1.15; 95% CI, 0.91–1.45; P=0.25) in advanced NSCLC patients with PD-L1 expression ≥5% (39). Of note, in patients with PD-L1 expression ≥50%, the lack of benefit for nivolumab persisted with an HR for progression or death of 1.07 (95% CI, 0.77–1.49). Overall, results from Checkmate 026 both in the whole population and for those tumors with strongly positive PD-L1 expression are inconsistent with first line nivolumab performance in a phase I trial (68). Besides Checkmate 026, another example of unsuccessful biomarkers enrichment strategy design is represented by MYSTIC trial comparing durvalumab *vs.* durvalumab + tremelimumab *vs.* platinum based chemotherapy in 1,092 treatment-naïve-EGFR/ALK wild-type NSCLC patients with PD-L1 expression ≥25%. The co-primary endpoints were OS and PFS. Results are not yet published, however the trial failed to show superiority in PFS of the combination durvalumab plus tremelimumab compared to platinum based chemotherapy in this PD-L1 enriched population (116). Although it is not possible to perform cross trial comparisons, the conflicting results between KEYNOTE-024 and Checkmate 026 or MYSTIC trials are difficult to attribute to differences in the pharmacologic and biologic properties among ICI, while discrepancies in patients selection, biomarker tests, and PD-L1 expression cut points could have contributed to these discordant findings (117). Results from a confirmatory phase III study of first line pembrolizumab *vs.* platinum based chemotherapy (KEYNOTE-042) in advanced NSCLC patients with PD-L1 expression ≥1% (118), will probably shed more light on the utility of enrichment design and on the performance of different biomarker thresholds for PD-L1 positivity.

Finally, due that multiple studies evaluating ICI have a low enrollment target (76 patients per trial on average for investigator initiated studies) (1), it is unrealistic that small single center trials will recruit enough patients to produce high quality results. Furthermore, the main pitfall accompanying the entering in the clinic of many different ICI will probably be the absence of direct comparisons between different compounds, tested in different clinical settings and in distinct patients' populations.

Recently, FDA summarized examples of collaborative and novel trial designs that could allow more questions to be efficiently addressed in a single multicenter trial (119). A promising example is the LUNG MAP program using a common biomarker screening platform to classify molecular subgroups of patients and assign them to specific matched targeted therapies (120). Similarly for ICI, collaborative platforms, coordinated both by pharma companies and non-profit organizations, including studies with multiple arms and hundreds of active sites, will help to avoid excessive

data fragmentation and duplication and will provide the background for the development of high-quality designed clinical trials, where sharing of findings and resources can ultimately lead to accelerated scientific innovation.

Conclusions

Besides the paradigm shift in cancer treatment, the advent of ICI has also raised several questions regarding the most appropriate endpoints, statistical models, biomarker assessment methodologies and clinical trial designs. Specifically, a more extensive use of iRECIST to assess antitumor responses and the replacement of traditional statistical methods (log rank and Cox proportional hazard model) and survival measures (median OS and HR) with new models (such as milestone analysis or RMST) able to capture delayed survival benefit and long-term tails are key issues to address in the next wave of trials with immunotherapies. Regarding predictive biomarkers, in particular PD-L1 expression, the integration and harmonization of the existing assays are critical to reduce variability and provide a reliable test to identify responders or patients who should be early switched to different treatments. Furthermore, a single biomarker may not mirror the real systemic immunological landscape of the patient. Therefore, translating to ICI the idea of targeted therapies, for which one biomarker is usually enough to predictive the drug benefit, appears an unrealistic objective. Moreover, innovative study designs such as adaptive or basket and biomarker enriched clinical trials, which may address different hypotheses at one time, potentially identifying molecular subgroups of patients with increased benefit from ICI, represent a promising strategy to pursue. In conclusion, building large collaborative platforms of clinical trials and selecting the appropriate bars to assess the clear health benefit and value of ICI represent the major challenges for the future research in the immune-oncology field.

Acknowledgements

Funding: Dr. Roberto Ferrara was the recipient of the grant DUERTECC/EURONCO (Diplôme Universitaire Européen de Recherche Translationnelle Et Clinique en Cancérologie) for 2016–2017, and of the IASCL (International Association for the Study of Lung Cancer) Young Investigator Award for 2017–2018. Dr. Sara Pilotto was supported by a Fellowship Award of the International Association for the Study of Lung Cancer. Prof. Emilio Bria received honoraria or speakers' fee from MSD, Astra-Zeneca, Celgene, Pfizer, Helsinn, Eli-Lilly, BMS, Novartis, Roche, and he received research support from A.I.R.C. (Associazione Italiana Ricerca sul Cancro, grants no. 14282 and no. 20583), I.A.S.L.C. (International Association for the Study of Lung Cancer), L.I.L.T. (Lega Italiana per la Lotta contro i Tumori), Fondazione Cariverona, Astra-Zeneca, Roche and Open Innovation.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Tang J, Shalabi A, Hubbard-Lucey VM. Comprehensive analysis of the clinical immuno-oncology landscape. Ann Oncol 2018;29:84-91.
2. Ben-Aharon O, Magnezi R, Leshno M, et al. Association of Immunotherapy With Durable Survival as Defined by Value Frameworks for Cancer Care. JAMA Oncol 2018;4:326-32.
3. 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.
4. Chiou VL, Burotto M. Pseudoprogression and Immune-Related Response in Solid Tumors. J Clin Oncol 2015;33:3541-3.
5. Tazdait M, Mezquita L, Lahmar J, et al. Patterns of responses in metastatic NSCLC during PD-1 or PDL-1 inhibitor therapy: Comparison of RECIST 1.1, irRECIST and iRECIST criteria. Eur J Cancer 2018;88:38-47.
6. Chen TT. Statistical issues and challenges in immuno-oncology. J Immunother Cancer 2013;1:18.
7. Wolchok JD, Hoos A, O'Day S, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. Clin Cancer Res 2009;15:7412-20.
8. Nishino M, Giobbie-Hurder A, Gargano M, et al. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. Clin Cancer Res 2013;19:3936-43.
9. Seymour L, Bogaerts J, Perrone A, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol 2017;18:e143-52.

10. Hellmann MD, Kris MG, Rudin CM. Medians and Milestones in Describing the Path to Cancer Cures: Telling "Tails." JAMA Oncol 2016;2:167-8.
11. Chen TT. Milestone Survival: A Potential Intermediate Endpoint for Immune Checkpoint Inhibitors. J Natl Cancer Inst 2015;107:djv156.
12. Fine GD. Consequences of Delayed Treatment Effects on Analysis of Time-to-Event Endpoints. Drug Information Journal 2007;41:535-9.
13. Pak K, Uno H, Kim DH, et al. Interpretability of Cancer Clinical Trial Results Using Restricted Mean Survival Time as an Alternative to the Hazard Ratio. JAMA Oncol 2017;3:1692-6.
14. Carroll KJ. On the use and utility of the Weibull model in the analysis of survival data. Control Clin Trials 2003;24:682-701.
15. Schnipper LE, Davidson NE, Wollins DS, et al. Updating the American Society of Clinical Oncology Value Framework: Revisions and Reflections in Response to Comments Received. J Clin Oncol 2016;34:2925-34.
16. Schnipper LE, Schilsky RL. Are Value Frameworks Missing the Mark When Considering Long-term Benefits From Immuno-oncology Drugs? JAMA Oncol 2018;4:333-4.
17. Miller AB, Hoogstraten B, Staquet M, et al. Reporting results of cancer treatment. Cancer 1981;47:207-14.
18. Hodi FS, Hwu WJ, Kefford R, et al. Evaluation of Immune-Related Response Criteria and RECIST v1.1 in Patients With Advanced Melanoma Treated With Pembrolizumab. J Clin Oncol 2016;34:1510-7.
19. Long GV, Weber JS, Larkin J, et al. Nivolumab for Patients With Advanced Melanoma Treated Beyond Progression: Analysis of 2 Phase 3 Clinical Trials. JAMA Oncol 2017;3:1511-9.
20. Escudier B, Motzer RJ, Sharma P, et al. Treatment Beyond Progression in Patients with Advanced Renal Cell Carcinoma Treated with Nivolumab in CheckMate 025. Eur Urol 2017;72:368-76.
21. Kazandjian D, Keegan P, Suzman DL, et al. 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.
22. Gettinger SN, Horn L, Gandhi L, et al. Overall Survival and Long-Term Safety of Nivolumab (Anti-Programmed Death 1 Antibody, BMS-936558, ONO-4538) in Patients With Previously Treated Advanced Non-Small-Cell Lung Cancer. J Clin Oncol 2015;33:2004-12.
23. Lynch TJ, Bondarenko I, Luft A, et al. Ipilimumab in combination with paclitaxel and carboplatin as first-line treatment in stage IIIB/IV non-small-cell lung cancer: results from a randomized, double-blind, multicenter phase II study. J Clin Oncol 2012;30:2046-54.
24. Wolchok JD, Kluger H, Callahan MK, et al. Nivolumab plus ipilimumab in advanced melanoma. N Engl J Med 2013;369:122-33.
25. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 2016;387:1540-50.
26. Champiat S, Dercle L, Ammari S, et al. Hyperprogressive Disease Is a New Pattern of Progression in Cancer Patients Treated by Anti-PD-1/PD-L1. Clin Cancer Res 2017;23:1920-8.
27. Saâda-Bouzid E, Defaucheux C, Karabajakian A, et al. Hyperprogression during anti-PD-1/PD-L1 therapy in patients with recurrent and/or metastatic head and neck squamous cell carcinoma. Ann Oncol 2017;28:1605-11.
28. Ferrara R, Caramella C, Texier M et al. MA10.11 Hyperprogressive Disease (HPD) Is Frequent in Non-Small Cell Lung Cancer (NSCLC) Patients (Pts) Treated with Anti PD1/PD-L1 Agents (IO). IASLC 18th World Conference on Lung Cancer. 2017.
29. Roviello G, Andre F, Venturini S, et al. Response rate as a potential surrogate for survival and efficacy in patients treated with novel immune checkpoint inhibitors: A meta-regression of randomised prospective studies. Eur J Cancer 2017;86:257-65.
30. Shukuya T, Mori K, Amann JM, et al. Relationship between Overall Survival and Response or Progression-Free Survival in Advanced Non-Small Cell Lung Cancer Patients Treated with Anti-PD-1/PD-L1 Antibodies. J Thorac Oncol 2016;11:1927-39.
31. Soria JC, Massard C, Le Chevalier T. Should progression-free survival be the primary measure of efficacy for advanced NSCLC therapy? Ann Oncol 2010;21:2324-32.
32. Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. Int J Technol Assess Health Care 2014;30:312-24.
33. Mauguen A, Pignon JP, Burdett S, et al. Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: a re-analysis of meta-analyses of individual

patients' data. Lancet Oncol 2013;14:619-26.
34. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. J Natl Cancer Inst 2009;101:1642-9.
35. Pilotto S, Carbognin L, Karachaliou N, et al. Moving towards a customized approach for drug development: lessons from clinical trials with immune checkpoint inhibitors in lung cancer. Transl Lung Cancer Res 2015;4:704-12.
36. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med 2016;375:1823-33.
37. 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.
38. 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.
39. 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.
40. Brahmer J, Abreu Rodriguez D, Robinson AG et al. OA 17.06 Updated Analysis of KEYNOTE-024: Pembrolizumab versus Platinum-Based Chemotherapy for Advanced NSCLC With PD-L1 TPS ≥50%. IASLC 18th World Conference on Lung Cancer. 2017.
41. Fehrenbacher L, Spira A, Ballinger M, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. Lancet 2016;387:1837-46.
42. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet 2017;389:255-65.
43. Ascierto PA, Long GV. Progression-free survival landmark analysis: a critical endpoint in melanoma clinical trials. Lancet Oncol 2016;17:1037-9.
44. Kantoff PW, Higano CS, Shore ND, et al. Sipuleucel-T immunotherapy for castration-resistant prostate cancer. N Engl J Med 2010;363:411-22.
45. 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.
46. Brahmer J, Horn L, Jackman D, et al. 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. Cancer Res 2017;77:CT077.
47. Hoos A. Evolution of end points for cancer immunotherapy trials. Ann Oncol 2012;23 Suppl 8:viii47-52.
48. Institute for Clinical and Economic Review. ICER Value Assessment Framework. Available online: https://icer-review.org/methodology/icers-methods/icer-value-assessment-framework
49. Othus M, Barlogie B, Leblanc ML, et al. Cure models as a useful statistical tool for analyzing survival. Clin Cancer Res 2012;18:3731-6.
50. Ribas A, Kefford R, Marshall MA, 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.
51. Hoos A, Britten CM, Huber C, et al. A methodological framework to enhance the clinical success of cancer immunotherapy. Nat Biotechnol 2011;29:867-70.
52. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet 2002;359:1686-9.
53. Trinquart L, Jacot J, Conner SC, et al. Comparison of Treatment Effects Measured by the Hazard Ratio and by the Ratio of Restricted Mean Survival Times in Oncology Randomized Controlled Trials. J Clin Oncol 2016;34:1813-9.
54. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol 2013;13:152.
55. A'Hern RP. Restricted Mean Survival Time: An Obligatory End Point for Time-to-Event Analysis in Cancer Trials? J Clin Oncol 2016;34:3474-6.
56. Felip Font E, Gettinger SN, Burgio MA, et al. Three year follow-up from Checkmate 017/057: Nivolumab versus Docetaxel in patients with previously treated advanced non-small cell lung cancer. Ann Oncol 2017;28:v460-96.
57. Harrington DP, Fleming TR. A Class of Rank Test Procedures for Censored Survival Data. Biometrika 1982;69:553-66.
58. Ying GS, Heitjan DF. Weibull prediction of event times in clinical trials. Pharm Stat 2008;7:107-20.
59. Le Tourneau C, Lee JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. J Natl Cancer Inst 2009;101:708-20.
60. Haanen JB, Carbonnel F, Robert C, et al. Management of

toxicities from immunotherapy: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2017;28:iv119-42.
61. Fecher LA, Agarwala SS, Hodi FS, et al. Ipilimumab and its toxicities: a multidisciplinary approach. Oncologist 2013;18:733-43.
62. Nanda R, Chow LQM, Dees EC, et al. Pembrolizumab in Patients With Advanced Triple-Negative Breast Cancer: Phase Ib KEYNOTE-012 Study. J Clin Oncol 2016;34:2460-7.
63. Nishino M, Ramaiya NH, Hatabu H, et al. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. Nat Rev Clin Oncol 2017;14:655-68.
64. Mandal R, Chan TA. Personalized Oncology Meets Immunology: The Path toward Precision Immunotherapy. Cancer Discov 2016;6:703-13.
65. Carbognin L, Pilotto S, Milella M, et al. Differential Activity of Nivolumab, Pembrolizumab and MPDL3280A according to the Tumor Expression of Programmed Death-Ligand-1 (PD-L1): Sensitivity Analysis of Trials in Melanoma, Lung and Genitourinary Cancers. PLoS ONE 2015;10:e0130142.
66. Aguiar PN, Santoro IL, Tadokoro H, et al. The role of PD-L1 expression as a predictive biomarker in advanced non-small-cell lung cancer: a network meta-analysis. Immunotherapy 2016;8:479-88.
67. 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.
68. Gettinger S, Rizvi NA, Chow LQ, et al. Nivolumab Monotherapy for First-Line Treatment of Advanced Non-Small-Cell Lung Cancer. J Clin Oncol 2016;34:2980-7.
69. Garon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. N Engl J Med 2015;372:2018-28.
70. Gordon MS, Herbst RS, Horn L, et al. PS01.62: Long-Term Safety and Clinical Activity of Atezolizumab Monotherapy in Metastatic NSCLC: Final Results from a Phase Ia Study: Topic: Medical Oncology. J Thorac Oncol 2016;11:S309-10.
71. Peters S, Gettinger S, Johnson ML, et al. Phase II Trial of Atezolizumab As First-Line or Subsequent Therapy for Patients With Programmed Death-Ligand 1-Selected Advanced Non-Small-Cell Lung Cancer (BIRCH). J Clin Oncol 2017;35:2781-9.
72. Spigel DR, Chaft JE, Gettinger SN, et al. Clinical activity and safety from a phase II study (FIR) of MPDL3280A (anti-PDL1) in PD-L1–selected patients with non-small cell lung cancer (NSCLC). J Clin Oncol 2015;33:8028.
73. Garassino M, Vansteenkiste J, Kim JH, et al. PL04a.03: Durvalumab in ≥3rd-Line Locally Advanced or Metastatic, EGFR/ALK Wild-Type NSCLC: Results from the Phase 2 ATLANTIC Study. J Thorac Oncol 2017;12:S10-1.
74. Jerusalem G, Chen F, Spigel D, et al. OA03.03 JAVELIN Solid Tumor: Safety and Clinical Activity of Avelumab (Anti-PD-L1) as First-Line Treatment in Patients with Advanced NSCLC. J Thorac Oncol 2017;12:S252.
75. Snyder A, Wolchok JD, Chan TA. Genetic basis for clinical response to CTLA-4 blockade. N Engl J Med 2015;372:783.
76. Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015;348:124-8.
77. Higgs BW, Morehouse C, Streicher K, et al Relationship of baseline tumoral IFNγ mRNA and PD-L1 protein expression to overall survival in durvalumab treated NSCLC patients. J Clin Oncol 2016;34:3036.
78. Herbst RS, Soria JC, Kowanetz M, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. Nature 2014;515:563-7.
79. Tumeh PC, Harview CL, Yearley JH, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature 2014;515:568-71.
80. Mezquita L, Auclin E, Charrier M, et al. The Lung Immune Prognostic Index (LIPI), a predictive score for immune checkpoint inhibitors in advanced non-small cell lung cancer (NSCLC) patients. Ann Oncol 2017;28:mdx 380.029.
81. Routy B, Le Chatelier E, Derosa L, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science 2018;359:91-7.
82. Grizzi G, Caccese M, Gkountakos A, et al. Putative predictors of efficacy for immune checkpoint inhibitors in non-small-cell lung cancer: facing the complexity of the immune system. Expert Rev Mol Diagn 2017;17:1055-69.
83. Hirsch FR, McElhinny A, Stanforth D, et al. PD-L1 Immunohistochemistry Assays for Lung Cancer: Results from Phase 1 of the Blueprint PD-L1 IHC Assay Comparison Project. J Thorac Oncol 2017;12:208-22.
84. Gadgeel S, Kowanetz M, Zou V, et al. Clinical efficacy of Atezolizumab in PD-L1 subgroups defined by SP142 and 22C3 IHC assay in 2L+ NSCLC: results from the

randomized OAK study. Ann Oncol 2017;28:mdx380.001
85. Adam J, Rouquette I, Damotte D, et al. PL04a.04: Multicentric French Harmonization Study for PD-L1 IHC Testing in NSCLC. J Thorac Oncol 2017;12:S11-2.
86. Rimm DL, Han G, Taube JM, et al. A Prospective, Multi-institutional, Pathologist-Based Assessment of 4 Immunohistochemistry Assays for PD-L1 Expression in Non-Small Cell Lung Cancer. JAMA Oncol 2017;3:1051-8.
87. Scheel AH, Dietel M, Heukamp LC, et al. Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. Mod Pathol 2016;29:1165-72.
88. Skov BG, Skov T. Paired Comparison of PD-L1 Expression on Cytologic and Histologic Specimens From Malignancies in the Lung Assessed With PD-L1 IHC 28-8pharmDx and PD-L1 IHC 22C3pharmDx. Appl Immunohistochem Mol Morphol 2017;25:453-9.
89. Ratcliffe MJ, Sharpe A, Midha A, et al. Agreement between Programmed Cell Death Ligand-1 Diagnostic Assays across Multiple Protein Expression Cutoffs in Non-Small Cell Lung Cancer. Clin Cancer Res 2017;23:3585-91.
90. Batenchuk C, Albitar M, Sudarsanam S, et al. Abstract 4015: A comparative study of PD-L1 IHC 22C3 and 28-8 FDA-approved diagnostic assays in cancer. Cancer Res 2017;77:4015.
91. Soo R, Lim J, Asuncion BR, et al. P2.01-027 A Comparison of Five Different Immunohistochemistry Assays for Programmed Death Ligand-1 Expression in Non-Small Cell Lung Cancer Samples: Topic: Proteins in Lung Cancer and Proteomics. J Thorac Oncol 2017;12:S801.
92. Yeh YC, Lin SF, Chiu CH, et al. 398PD PD-L1 status in Taiwanese lung adenocarcinoma patients: Comparison of PD-L1 immunohistochemical assays using antibody clones 22C3, SP142 and SP263 with clinicopathological correlation. Ann Oncol 2016;27:ix123-5.
93. Mahoney KM, Sun H, Liao X, et al. PD-L1 Antibodies to Its Cytoplasmic Domain Most Clearly Delineate Cell Membranes in Immunohistochemical Staining of Tumor Cells. Cancer Immunol Res 2015;3:1308-15.
94. Büttner R, Gosney JR, Skov BG, et al. Programmed Death-Ligand 1 Immunohistochemistry Testing: A Review of Analytical Assays and Clinical Implementation in Non-Small-Cell Lung Cancer. J Clin Oncol 2017;35:3867-76.
95. Herbst RS, Baas P, Perez-Gracia JL, et al. PD1.06 (also presented as P2.41): Pembrolizumab versus Docetaxel for Previously Treated NSCLC (KEYNOTE-010): Archival versus New Tumor Samples for PD-L1 Assessment. J Thorac Oncol 2016;11:S174-5.
96. Midha A, Sharpe A, Scott M, et al. PD-L1 expression in advanced NSCLC: Primary lesions versus metastatic sites and impact of sample age. J Clin Oncol 2016;34:3025.
97. Sakakibara R, Inamura K, Tambo Y, et al. EBUS-TBNA as a Promising Method for the Evaluation of Tumor PD-L1 Expression in Lung Cancer. Clin Lung Cancer 2017;18:527-34.e1.
98. Rehman JA, Han G, Carvajal-Hausdorf DE, et al. Quantitative and pathologist-read comparison of the heterogeneity of programmed death-ligand 1 (PD-L1) expression in non-small cell lung cancer. Mod Pathol 2017;30:340-9.
99. Ilie M, Long-Mira E, Bence C, et al. Comparative study of the PD-L1 status between surgically resected specimens and matched biopsies of NSCLC patients reveal major discordances: a potential issue for anti-PD-L1 therapeutic strategies. Ann Oncol 2016;27:147-53.
100. Postel-Vinay S, Aspeslagh S, Lanoy E, et al. Challenges of phase 1 clinical trials evaluating immune checkpoint-targeted antibodies. Ann Oncol 2016;27:214-24.
101. Ji Y, Wang SJ. Modified toxicity probability interval design: a safer and more reliable method than the 3 + 3 design for practical phase I trials. J Clin Oncol 2013;31:1785-91.
102. Brahmer JR, Tykodi SS, Chow LQM, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. N Engl J Med 2012;366:2455-65.
103. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N Engl J Med 2012;366:2443-54.
104. FDA Opdivo prescribing information 2016. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/label/2016/125554s017s018lbl.pdf
105. Zhao X, Ivaturi V, Gopalakrishnan M, et al. Abstract CT101: A model-based exposure-response (E-R) assessment of a nivolumab (NIVO) 4-weekly (Q4W) dosing schedule across multiple tumor types. Cancer Res 2017;77:CT101.
106. Topalian SL, Sznol M, McDermott DF, et al. Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab. J Clin Oncol 2014;32:1020-30.
107. McKay RR, Martini D, Moreira RB, et al. Outcomes of PD-1/PD-L1 responders who discontinue therapy for immune-related adverse events (irAEs): Results of a cohort of patients (pts) with metastatic renal cell carcinoma (mRCC). J Clin Oncol 2017;35:467.

108. Pollack MH, Betof A, Dearden H, et al. Safety of resuming anti-PD-1 in patients with immune-related adverse events (irAEs) during combined anti-CTLA-4 and anti-PD1 in metastatic melanoma. Ann Oncol 2018;29:250-5.
109. Spiegel DR, McLeod M, Hussain MA, et al. Randomized results of fixed-duration (1-yr) versus continuous nivolumab in patients (pts) with advanced non-small cell lung cancer (NSCLC). Ann Oncol 2017;28:v460-96.
110. Chabner BA. Early Accelerated Approval for Highly Targeted Cancer Drugs. N Engl J Med 2011;364:1087-9.
111. Menis J, Hasan B, Besse B. New clinical research strategies in thoracic oncology: clinical trial design, adaptive, basket and umbrella trials, new end-points and new evaluations of response. Eur Respir Rev 2014;23:367-78.
112. Kang SP, Gergich K, Lubiniecki GM, et al. Pembrolizumab KEYNOTE-001: an adaptive study leading to accelerated approval for two indications and a companion diagnostic. Ann Oncol 2017;28:1388-98.
113. Mullard A. Genetic Biomarker trumps tissue type in landmark oncology approval. Nat Rev Drug Discov 2017;16:447.
114. Hyman DM, Puzanov I, Subbiah V, et al. Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. N Engl J Med 2015;373:726-36.
115. FDA Guidance for industry: Enrichment strategies for clinical trials to support approval of human drugs and biological products, 2012. Available online: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm332181.pdf
116. Peters S, Antonia S, Goldberg SB, et al. 191TiP: MYSTIC: a global, phase 3 study of durvalumab (MEDI4736) plus tremelimumab combination therapy or durvalumab monotherapy versus platinum-based chemotherapy (CT) in the first-line treatment of patients (pts) with advanced stage IV NSCLC. J Thorac Oncol 2016;11:S139-40.
117. Remon J, Besse B, Soria JC. Successes and failures: what did we learn from recent first-line treatment immunotherapy trials in non-small cell lung cancer? BMC Med 2017;15:55.
118. Study of MK-3475 (Pembrolizumab) Versus Platinum-based Chemotherapy for Participants With PD-L1-positive Advanced or Metastatic Non-small Cell Lung Cancer. Available online: https://clinicaltrials.gov/ct2/show/NCT02220894
119. Woodcock J, LaVange LM. Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. N Engl J Med 2017;377:62-70.
120. Herbst RS, Gandara DR, Hirsch FR, et al. Lung Master Protocol (Lung-MAP)-A Biomarker-Driven Protocol for Accelerating Development of Therapies for Squamous Cell Lung Cancer: SWOG S1400. Clin Cancer Res 2015;21:1514-24.

Cite this article as: Ferrara R, Pilotto S, Caccese M, Grizzi G, Sperduti I, Giannarelli D, Milella M, Besse B, Tortora G, Bria E. Do immune checkpoint inhibitors need new studies methodology? J Thorac Dis 2018;10(Suppl 13):S1564-S1580. doi: 10.21037/jtd.2018.01.131

