COMMENTARY

Milestone Survival: A Potential Intermediate Endpoint for Immune Checkpoint Inhibitors

Tai-Tsang Chen

Affiliations of author: Department of Global Biometric Sciences, Bristol-Myers Squibb, Wallingford, CT; Department of Biostatistics, Columbia University, New York, NY.

Correspondence to: Tai-Tsang Chen, PhD, 5 Research Parkway, 2FW-311, Wallingford, CT 06410 (e-mail: tai-tsang.chen@bms.com).

Abstract

Recent advancements in cancer immunotherapies offer diverse strategies for cancer treatment. Among the most promising approaches is the blockade of immune checkpoint molecules to activate antitumor immunity. With targeted immunotherapies of new mechanisms of action come greater challenges in study design and statistical analysis, as well as the need for refining clinical trial endpoints. The long-term survival and delayed clinical effects demonstrated by these therapies could result in substantial prolongation of study duration and loss of statistical power if these key attributes are not accounted for in the study design and statistical analyses. In the Brookings Conference on Clinical Cancer Research held in Washington, DC, in November 2013, several intermediate clinical endpoints, including milestone overall survival, were proposed for the evaluation of cancer immunotherapies to take into account the possibility of delayed treatment effect and to better characterize the clinical activity profile of such agents, particularly immune checkpoint inhibitors. In this manuscript, the use of milestone survival is described as a potential efficacy endpoint for immune checkpoint inhibitors in late-stage drug development that could potentially mitigate the challenge of accelerating the drug development process when the strength of this class of agents is derived from long-term follow-up.

Recent advancements in the molecular understanding of how the immune system can be modulated and stimulated to fight against cancer have led clinical trial researchers to reassess whether the current paradigm of drug development is optimal to harvest the strengths of newly discovered immuno-oncology agents. Among the most promising of these novel approaches is the blockade of immune checkpoint molecules. Immune checkpoints refer to a plethora of inhibitory pathways hardwired into the immune system that are crucial for maintaining self-tolerance and for modulating the duration and amplitude of physiological immune responses in peripheral tissues in order to minimize collateral tissue damage (1). An immune response can subsequently be induced by blocking inhibitory immune checkpoints (2). One such example is ipilimumab, the first-in-class fully human monoclonal antibody (IgG1) that blocks cytotoxic T lymphocyte–associated protein 4 (CTLA-4, also known as CD152).

Ipilimumab has demonstrated a statistically significant improvement in overall survival (OS) in two phase III randomized, controlled clinical trials either alone or in combination with dacarbazine (DTIC) in patients with previously treated and treatment-naïve metastatic melanoma (3,4). The survival curves did not separate until approximately four months with a survival probability leveling off at 20% in the study comparing ipilimumab with and without gp100 vaccine vs gp100 vaccine alone in patients with previously treated advanced melanoma (3). A similar phenomenon was observed in the phase III study involving treatment-naïve patients (4) with updated five-year long-term survival data (5), as well as in phase II ipilimumab clinical trials (6,7). More recently, a pooled analysis of long-term survival data up to 10 years among phase II and phase III clinical trials in 1861 patients with unresectable or metastatic melanoma demonstrated a median OS of 11.4 months (95% confidence interval [CI] = 10.7 to 12.1), with a three-year survival rate estimated to be 22% (95% CI = 20% to 24%) (8).

Another example is nivolumab, a fully human IgG4 antibody that selectively blocks the interaction of the programmed death 1 (PD-1) receptor with its ligands PD-L1 and PD-L2. An

---

approximate three-month delayed treatment effect was also observed in a phase III nivolumab study involving previously untreated metastatic melanoma patients without a BRAF mutation. At the time of the analysis, the one-year survival rate was 73% (95% CI = 66% to 79%) in patients treated with nivolumab, as compared with 42% (95% CI = 33% to 51%) in those treated with dacarbazine (9).

Overall survival remains the gold standard for demonstrating efficacy in oncology clinical trials. It is defined as the time between the date of random assignment and the date of death. For patients without documentation of death, OS is censored on the last date the patients were known to be alive. The most commonly used statistical methods for time-to-event analyses have been the log-rank test and Cox regression analysis (10,11). These standard analyses have maximal statistical power under the proportional hazards assumption. One appealing characteristic of the log-rank statistic is that it does not require any assumption of the shape of the survival curve or the distribution of survival times. While it may serve as an advantage in assessing the efficacy of traditional chemotherapies or targeted therapies, it does not necessarily capture the key attributes of immunotherapies such as long-term survival. Based on the kinetics of the survival effect, the time to final OS analysis may continue to lengthen. The long-term survival and delayed clinical effects could result in substantial prolongation of study duration and loss of statistical power if the trial was designed based on the conventional exponential distribution assumption (12). This poses a challenge to accelerating the drug development process when the strength of this class of agents is derived from long-term follow-up.

Shea et al. provided a thorough review of multiple endpoints used in regulatory approvals, demonstrating that the US Food and Drug Administration has exercised considerable flexibility in the approval of oncology drugs in the past decade (13). Nevertheless, some of the frequently considered surrogate endpoint candidates, such as best overall response or progression-free survival, may not always be optimal in assessing the efficacy of immune checkpoint inhibitors. Higher response rates, durable tumor regression, and prolonged disease stabilization have been observed following blockade of PD-1 (14–17) or its ligand PD-L1 (18,19), whereas ipilimumab has demonstrated overall and long-term survival benefit despite having lower response rates (3,20). With the newly acquired knowledge and understanding of cancer immunotherapeutic agents, the need for new endpoints in immuno-oncology drug development has been recognized by many (21–24) as we begin to address the unique characteristics of immunotherapeutic agents with novel mechanisms of action.

In the Brookings Conference on Clinical Cancer Research held in Washington, DC, in November 2013, several intermediate clinical endpoints were proposed for the evaluation of cancer immunotherapies to take into account the possibility of delayed treatment effect and to better characterize the clinical activity profile of immuno-oncology agents, particularly immune checkpoint inhibitors such as CTLA-4 or PD-1 inhibitors. These proposed endpoints included clinical benefit rates, gated progression-free survival, tumor growth rate, and milestone survival (25,26).

It is recognized that the most clinically meaningful outcome remains OS. Among the proposed intermediate endpoints for the development of immune checkpoint inhibitors, milestone survival is also an OS endpoint with cross-sectional assessment at a prespecified time point. Historically, the role of milestone survival, or survival probability at a given time point, has been descriptive in nature in most clinical trials because it does not

take into account the totality of the OS data. With novel mechanisms of action and unique efficacy attributes introduced by immune checkpoint inhibitors, a reconsideration of the use of milestone survival as a potential efficacy endpoint is warranted. In this Commentary, study design and analysis consideration are discussed using milestone survival as an intermediate endpoint with OS as the primary endpoint. A phase III study in patients with treatment-naïve metastatic melanoma (4) was retrospectively redesigned and analyzed to illustrate how milestone survival could be implemented in late-stage development of immuno-oncology drugs.

Methods

Nonproportional Hazards Cure Rate Model

The number of events needed in randomized clinical trials with time-to-event endpoints is usually estimated based on an exponential distribution assumption in which we assume that anything that affects the hazards does so by the same ratio at all times, ie, proportional hazards. While this assumption is not unreasonable, it clearly does not reflect the kinetics of the survival effects demonstrated by immune checkpoint inhibitors. Based on the clinical data accumulated thus far, we have observed that these novel agents have induced long-term survival in a subset of patients and their treatment effect may not be recognized during the initial stage of the study. One example was a randomized double-blind phase III study comparing ipilimumab with and without gp100 vaccine vs gp100 vaccine alone in patients with previously treated advanced melanoma. The results showed that the survival curves did not separate until approximately four months with a survival probability leveling off at 20% in the experimental arms.

When the entire study population consists of patients who are potentially susceptible and nonsusceptible to the event of interest within a reasonable monitoring time window or during the course of the study, the nonproportional hazards cure rate model (NPHCRM) is a useful alternative for modeling the time-to-event data. The model allows the presence of delayed clinical effects while providing the flexibility of different risk ratio assumptions for susceptible and nonsusceptible populations after separation of survival curves (12). Figure 1A shows a pictorial illustration of a hypothetical NPHCRM between two treatment arms. A special case of NPHCRM is PHCRM when there is no delayed clinical effect (Figure 1B). These two models will serve as the bases for the remainder of the discussion.

Overall Survival as the Primary Endpoint

The statistical power is determined based on the number of events when a randomized comparative clinical trial is designed using a time-to-event endpoint such as OS as the primary endpoint. In general, the target cumulative event rate of approximately 80%, ie, four-fifths of the randomized patients becoming events by the end of the study, usually yields a good balance between the size and length of the trial. When a subset of patients is expected to be event free, a sufficient number of patients needs to be randomly assigned to ensure the total number of events is reachable, because the number of patients at risk is smaller and the events can only be expected from the susceptible population. In addition, the delayed clinical effect observed in most randomized immunotherapy trials leads to the loss of statistical power because the delay at the beginning of the trial
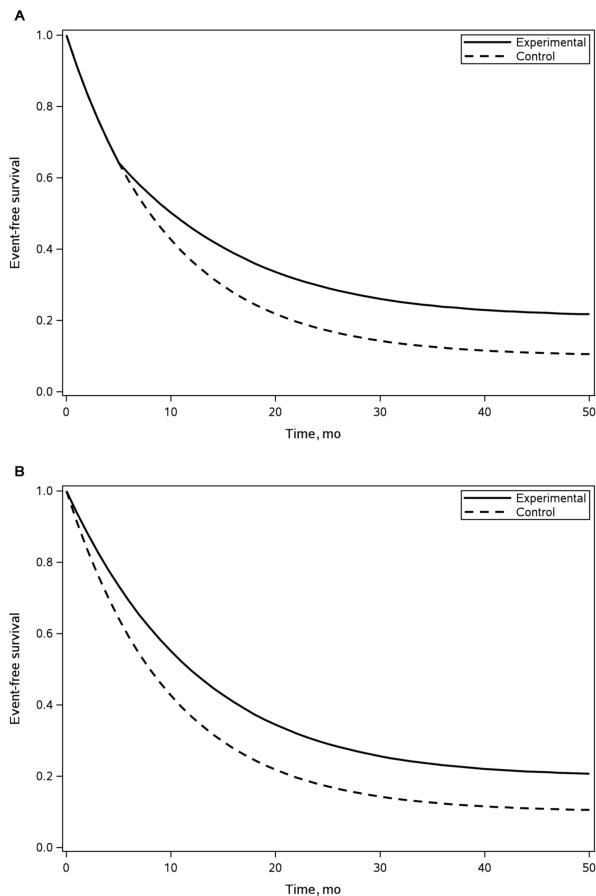
Figure 1. Graphical presentation of Kaplan-Meier survival curves with combinations of long-term survival and delayed clinical effect. The graphs show two hypothetical scenarios of overall survival outcomes where the **top and bottom curves** represent a novel immuno-oncologic agent and a control treatment, respectively. **A)** Nonproportional hazards cure rate model with long-term and delayed effects. **B)** Proportional hazards cure rate model with long-term survival.

Figure 2. The impact of delayed clinical effect on sample sizes. Assuming a median overall survival of 11 months and long term survival probability of 20% in the control arm with a 1:1 randomization ratio, the **graph** shows an increase in the sample size with different magnitudes of delayed clinical effects, ranging from no delay (proportional hazards cure rate model) to four months of delay under nonproportional hazards cure rate model for different treatment effects when the cumulative event rate is kept at 80% in the susceptible population. The corresponding sample sizes under conventional exponential assumption without delayed or long-term survival effects are highlighted with **star symbols**. HR = hazard ratio.

Table 1. Proportion of susceptible population under PHCRM*

	Proportion of long-term survivors in the control arm					
Risk ratio	0.05	0.10	0.15	0.20	0.25	0.30
0.5	0.86	0.79	0.73	0.68	0.63	0.58
0.6	0.89	0.82	0.76	0.71	0.66	0.61
0.7	0.91	0.85	0.79	0.74	0.69	0.63
0.8	0.93	0.87	0.82	0.76	0.71	0.66

* PHCRM = proportional hazards cure rate model.

will offset any treatment benefit that follows. Therefore, the number of events will also need to be increased to compensate for this loss of statistical power.

The following example is used to illustrate the impact of long-term survival and delay of clinical effects on study conduct. Assuming median OS in the control arm is 11 months with long-term survival probability of 20%, Figure 2 shows the sample size increase in studies with 1:1 randomization ratio, with different magnitudes of delayed clinical effects ranging from no delay (PHCRM) to four months of delay under NPHCRM for different treatment effects, ie, risk ratio from 0.6 to 0.8 with an increment of 0.05, in the entire population and an 80% cumulative event rate in the susceptible population. These designs yield reasonable total study durations ranging from two and a half to three years, with an average minimum follow-up time of one to two years when the risk ratio is below 0.7. If one wishes to retain the 80% cumulative event rate, the minimal follow-up duration will decrease as the risk ratio increases with similar accrual rate. The corresponding sample sizes under conventional exponential assumption without delayed or long-term survival effects are marked with star symbols to serve as references. It quickly becomes apparent that the size of phase III randomized clinical trials with OS as the primary endpoint grows increasingly
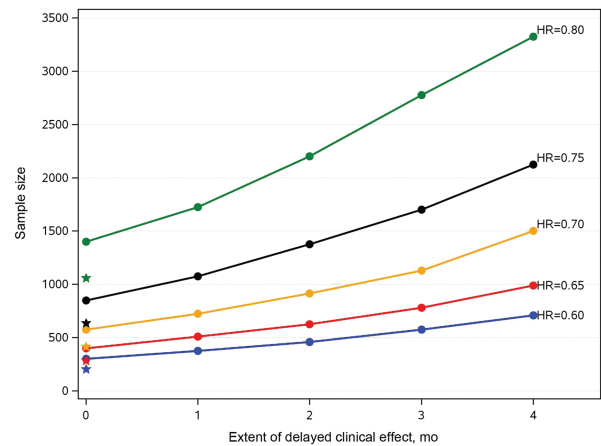
unmanageable if one does not assume a larger treatment effect or smaller delayed clinical effect.

If the sample size is reduced to a more reasonable size, the consequence is a substantial prolongation of the study duration. In addition, the study may also run the risk of not being able to reach the prespecified number of events because of insufficient patients in the susceptible population. Using PHCRM as an example, the proportion of susceptible population from which the events will come falls below 80% of all randomized patients when the long-term survival rate is at least 20% in the control arm, with risk ratios ranging from 0.5 to 0.8 (Table 1). If the number of events surpassed the size of the susceptible population, the study would not conclude within a reasonable time frame.

Milestone Survival as an Intermediate Endpoint

Milestone survival is defined as the Kaplan-Meier (27) survival probability at a time point defined a priori, such as two years. The word "milestone" is used in order to distinguish from that of "landmark." The landmark method was first introduced by Anderson et al. (28) in an attempt to address the issue of bias in the analysis of survival when the covariate of interest is an on-study measurement such as tumor response status, ie, responders vs nonresponders. The landmark analysis excludes patients whose survival times did not exceed the fixed time point after

initiation of therapy or random assignment. The survival curve using the landmark analysis will result in a plateau at 100% survival probability from time zero up to the chosen time point.

Milestone survival analysis is a cross-sectional assessment of the OS data at the prespecified time point using Kaplan-Meier survival probabilities. The milestone survival analysis is proposed to be conducted in the first cohort of randomized patients, rather than in the entire study population in order to mitigate the need to accelerate the drug development process when the long-term benefit of the agent is of potential interest. The choice of the milestone requires careful consideration because it often represents a clinically meaningful benchmark. It is important to note that the milestone does not necessarily represent long-term survival. It may represent a time point beyond which the researchers believe the treatment benefit is likely to remain stable. For example, multiple phase II and phase III studies have demonstrated that the OS among patients treated with ipilimumab began to level off at least two years after first day of treatment or random assignment. Based on the information amassed so far, clinical researchers may choose the two-year milestone as the time point of interest in the milestone survival analysis.

Once the time point of interest has been determined, although not an absolute requirement, it is preferable to ensure that sufficient follow-up duration among patients be included in the milestone survival analysis. In other words, the milestone survival analysis should not be conducted until at least the milestone duration has elapsed from the time the last patients entered the study in this cohort. The rationale behind this recommended restriction is to ensure the robustness of the milestone survival analysis. If all patients meet the minimal follow-up requirement, ie, milestone duration, the result will not change, because subsequent follow-up in this cohort will only have an impact on the survival curves beyond the milestone. If the milestone survival analysis is to be implemented, that implies the chosen milestone is an important endpoint that can be considered clinical benefit in clinical practice. Any measure that would ensure the robustness of the analysis is recommended.

When the data are being accessed multiple times during the course of the study, the boundaries for declaring statistically significant findings at the interim analyses are usually adjusted to avoid making excessive false-positive or false-negative errors. The most popular approaches are group sequential methods proposed by O'Brien and Fleming (29), Pocock (30), and error-spending functions by Lan and DeMets (31). While the Pocock and O'Brien-Fleming procedures require more conservative levels at the interim analyses, the error-spending functions are flexible in the numbers and times of interim analyses. The Haybittle-Peto (10,32) method, which allows for one user-defined boundary for all interim analyses with the final boundary closer to the original design, is another way of controlling the error rate. The fundamental concept of these proposals is to maintain the overall experiment-wise error rates when repeated evaluations are imposed on the data. Because the milestone survival analysis is considered an interim analysis, a certain level of multiplicity adjustment is warranted. The adaptation of one method over the others could be arbitrary and subjective. It depends on how much confidence the researchers have in the agents and how much risk one is willing to take.

Example

To illustrate how the proposed milestone survival analysis would have benefited an immuno-oncology trial in late-stage development, a phase III clinical trial in metastatic melanoma reported by Robert et al. (4) was retrospectively redesigned and analyzed using NPHCRM based on the outcome of OS analysis. This multi-center, randomized, double-blind phase III study was conducted in patients with treatment-naïve melanoma who received ipilimumab plus dacarbazine vs dacarbazine with placebo.

The primary efficacy analysis was the comparison of OS between two treatment arms. The protocol assumed a median survival of eight months for dacarbazine with placebo. It was estimated that 500 randomized patients (250 per arm) and a total of 416 deaths were needed to provide approximately 90% power to detect a 38% increase in median OS to 11 months for dacarbazine plus 10 mg/kg ipilimumab. This corresponded to a hazard ratio of 0.727, or a 27% reduction in hazard rate. The study was designed based on the proportional hazards assumption. At the time of study design, it was estimated that it would take 17 months to complete the enrollment and another 17 months of follow-up, with a total study duration of approximately three years.

The study was initiated in August of 2006, and the final analysis did not take place until March of 2011. The total number of events at the time of the final analysis was still two events less than the prespecified 416 events. During the last two years of blinded study monitoring, it was clear that the event rate had decreased drastically, which contributed to the prolongation of the study. The long-term survival phenomenon was confirmed upon unblinding of the data set. Figure 3A shows the Kaplan-Meier survival curves at the time of final analysis. Not only did the ipilimumab/dacarbazine arm demonstrate the long-term survival effect, but the dacarbazine-alone arm also showed a similar phenomenon. The delayed clinical effect was also observed, although it was less prominent compared with that of the second-line phase III study (3) in which the OS curves completely overlapped during the first four months.

The study was redesigned retrospectively using NPHCRM, assuming a median OS of nine months and a 10% long-term survival rate in the dacarbazine plus placebo arm, with a delayed clinical effect of four months between the two arms. The same accrual rate led to the entire study duration of approximately four and a half years, which was consistent with what was actually observed. The interim analysis using two-year milestone survival as an intermediate endpoint was performed when the first 300 randomized patients had been followed for at least two years. All randomized patients contributed to the final analysis. A nominal significance level of 0.025 was implemented at the interim analysis based on the difference of the two-year Kaplan-Meier survival probabilities using complementary log-log transformation (33), while the nominal significance level at the final analysis was 0.0328 based on log-rank test statistic. These boundaries ensure an experiment-wise type I error of less than 0.05.

Figure 3B shows the Kaplan-Meier survival curves among first 300 randomly assigned patients with at least 24 months of follow-up in this retrospective real-time analysis, ie, the data from these 300 patients were analyzed by redefining the last patient last visit date as if the analysis were implemented prospectively. Note that the censoring times among these 300 patients who were still on study at the time of the milestone survival analysis were clustered beyond 24 months with minimal loss to follow-up prior to the milestone. This was because of the fact that all patients were given a chance to be followed for at least the milestone duration of two years, and an effort was put into minimizing the number of patients who were lost to follow-up.

The estimated two-year Kaplan-Meier OS probabilities were 14.1% and 24.9% for the dacarbazine plus placebo and
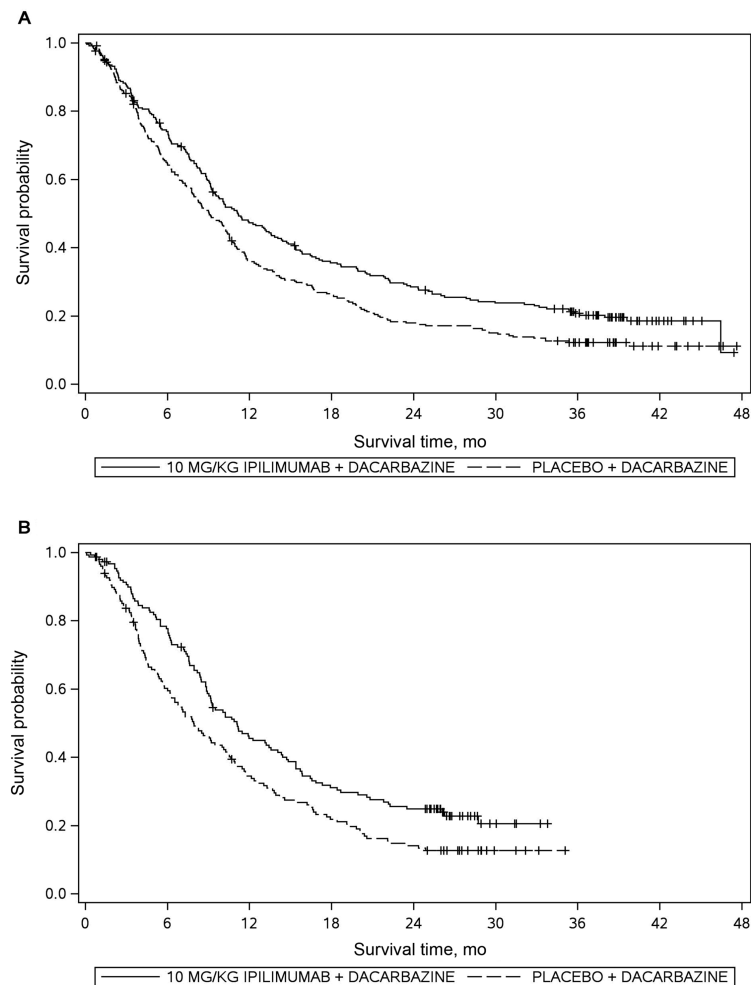
Figure 3. Survival analysis of a phase III study in treatment-naïve metastatic melanoma (4). **A)** Final survival analysis (n = 502). **B)** Iinterim milestone survival analysis (n = 300).

dacarbazine plus Ipilimumab arms, respectively, and the difference of the OS probabilities yielded a nominal P value of .021. The experimental treatment would have been declared efficacious approximately one to one and a half years prior to the time of the actual final analysis, had the milestone survival analysis been implemented in the study design. The final analysis from the same study confirmed the result of the milestone survival analysis with an OS hazard ratio between the two arms of 0.72 (95% CI = 0.59 to 0.87, P = .0009), indicating a 28% risk reduction of death in the ipilimumab plus dacarbazine group compared with the active control group.

Discussion

Cancer immunotherapies demonstrate different efficacy kinetics, such as long-term survival and delayed clinical effect, compared with previous cytotoxic or targeted agents. It is unrealistic to expect future phase III confirmatory trials with increasing long-term survivors to finish in a reasonable time frame based on the conventional event-driven approach. The incremental power towards the end of the trial in this situation will be small, which translates to the prolongation of the study. One can always argue that the study could be

terminated early if the diminution of event rate remains for a period of time, or the prespecification of when the final analysis will occur could be amended to take place earlier. While these are potential options, clinical trial researchers do not necessarily feel comfortable implementing mid-course study design change and risking any loss of statistical power. Alternatively, the number of events in the control arm could be used to inform the timing of the final analysis. If this approach is to be considered, the information will certainly need to be provided by the independent Data Monitoring Committee to prevent the immediate study team from being unblinded. Nevertheless, the issue of trial duration prolongation remains if the control arm also induces long-term survival effect. One such example is ipilimumab, which is now being considered as one of the standard-of-care agents in metastatic melanoma. There exists a dilemma to speed up drug development process for the novel agents such as the immune checkpoint inhibitors when the benefits of these agents are derived from long-term follow-up.

Conventional study design under the proportional hazards assumption in studies with time-to-event endpoints may not be appropriate. In order to capture the long-term survivors, it is desired to enroll only a smaller number of patients who will

have sufficient follow-up at the time of final analysis. Because events can only be observed in the nonsusceptible population, a miscalculation of the study size would potentially either lead to a substantial prolongation of the study or the inability to complete the trial.

It is a logical step to consider other alternatives to address these issues. Several different endpoints have been discussed and proposed. Milestone survival analysis is one such endpoint that could mitigate such issues. A cohort of patients with sufficient follow-up could be included in the analysis with an intermediate endpoint of milestone survival that could provide long-term survival information, while the entire study continues with a primary endpoint of OS. This approach ensures that patients in the analysis have reached sufficient follow-up duration to provide the researchers a first glimpse of long-term efficacy and safety, and the probability of capturing the treatment effect could be higher should the delayed clinical effect be present.

The biggest advantage of the milestone survival analysis is the predictable analysis time. Studies with time-to-event primary endpoints such as OS are event driven because the power depends on the accumulation of the events. If the phenomenon of long-term survivors was not anticipated in the design stage of the protocol, the study duration is likely to lengthen. The example of the phase III study showed that the long-term survival rates of 10% to 20% between two treatment arms led to an additional follow-up of two years. Had the milestone survival analysis been implemented, the analysis would have been conducted two years after the random assignment of the 300th patient, ie, one to one and a half years earlier than the time of actual final analysis. The proposed analysis would have equally benefited two other phase III studies in patients with stage III melanoma in the adjuvant setting and in previously treated advanced melanoma, had the analysis been implemented in the study design stage. Both long-term survival and delayed clinical effect were observed at the time of final analyses of the primary endpoints. Both studies had overlapping survival curves during the first three to four months. The adjuvant study yielded a three-year recurrence-free survival rate of 46% (95% CI = 41% to 51%) (34), and the two-year survival rate was 24% (95% CI = 16% to 32%) (3) for the study in previously treated advanced melanoma.

The milestone analysis also allows direct characterization of survival probability or long-term survival effect, which cannot be captured using the conventional log-rank test statistic or hazard ratio in the Cox regression analysis. In the context of treating advanced melanoma, the survival probability defined as the milestone could represent a time point beyond which the treatment effect is expected to remain stable. It also allows a potential earlier benefit/risk assessment without sacrificing the need for long-term follow-up because a cohort of patients with long-term follow-up can be evaluated while the entire study continues. When the delayed clinical effect is present, the milestone survival analysis could have greater statistical power. For a trial designed with 12 months of median control and eight months of delayed clinical effect with postseparation hazard ratios in the range of 0.5 to 0.6, the interim power of the milestone survival analysis could be 15% to 20% greater than that of the conventional O'Brien-Fleming method based on the same boundary. Finally, both intermediate and final endpoints are survival endpoints, although the former was evaluated cross-sectionally and the latter longitudinally. The confirmation of the survival benefit can be derived from the same study.

One of the challenges with the milestone survival analysis, similar to all interim analyses in studies with group sequential design, is the difficulty in maintaining study integrity post milestone analysis, ie, unblinding prior to final OS analysis. Crossover after unblinding of the data will prevent researchers from definitively quantifying or even demonstrating the survival benefit of the experimental treatment. Regardless of whether the study is unblinded at the time of milestone analysis, the analyses should be conducted by a Data Monitoring Committee to maintain study integrity. In addition, the determination of the size of analysis cohort could be challenging as it is related to the accrual duration. The inclusion of all patients in the milestone survival analysis may be more efficient with speedy enrollment, as all patients will have fairly similar follow-up duration.

Another difficult task for the researchers is in milestone selection. The crossover of Kaplan-Meier curves observed in a phase III trial of ipilimumab in metastatic castration-resistant prostate cancer (CRPC) highlighted the challenge of selecting the optimal milestone (35). If the milestone had been placed at a time before the treatment took effect, the potential immunotherapy treatment benefit in CRPC would never have been observed. This approach lends itself well when prior data are available to enable an understanding of appropriate milestone time point selection. A two-year milestone was chosen in the retrospective real-time analysis of the phase III study based on earlier clinical data in the ipilimumab development program in which the OS appeared to begin to stabilize beyond two years. Researchers would also have to decide how confident they are in the milestone survival endpoint in order to determine the interim type I error rate they are willing to spend. The interim boundary chosen in the retrospective study design shown in the example was 0.025. Had a more stringent boundary been chosen, the study would have continued to the final survival analysis. Although it would have taken longer, the outcome of a statistically significant and clinically meaningful improvement in OS would remain unchanged.

The milestone survival analysis is also a cross-sectional analysis and only considers a given time point as the primary endpoint; hence the analysis does not account for the totality of OS data. Any potential diminishing treatment effect post milestone in the same analysis cohort, or worse performance of the experimental arm in the complementary cohort, could lead to an increased false-positive rate. These are legitimate concerns when milestone survival is being implemented. One could consider a sensitivity analysis of analyzing the same cohort of patients with predefined minimal length of follow-ups using the conventional log-rank test or Cox regression model. Other potential sensitivity analyses, including those conducted on all randomized patients at the time of the interim analysis, can also be considered to add another layer of assurance.

The proposed milestone survival endpoint was discussed in the context of a randomized comparative setting for an earlier or accelerated approval. If preliminary data are promising, milestone survival could be used as an endpoint in a single-arm study. Given the long-term survival effect potentially induced by newly approved immunotherapeutic agents, more focus on future trial conduct should be placed on improving the long-term survival rate. It is not an inconceivable notion to have the milestone survival serve as the primary endpoint in confirmatory phase III studies when more data become available in this

class of agents. Depending on the anticipated accumulating data at the time of the analyses, different milestones can be defined for the interim and final analyses. For instance, it is more sensible to examine the one-year milestone survival if patients are projected to have only a minimum of one-year follow-up at the time of the interim look, while the two-year milestone survival can be assessed at the time of final analysis when all patients have had sufficient follow-up duration.

Future research will be conducted to assess the operating characteristics of the milestone survival analysis. These include the optimal size of the milestone survival analysis cohort and boundary at the time of interim monitoring, the efficiency introduced by the milestone analysis compared with the traditional group sequential methods, the impact of the magnitude of treatment effect post milestone in the analysis cohort and that of the performance in the complementary cohort on the final OS analysis. Milestone survival analysis can also be applied retrospectively to completed phase III studies, regardless of whether the primary endpoint was met (eg, ipilimumab phase III CRPC study), in order to better understand the performance of the proposed methodology relative to that of conventional approaches.

Conclusion

As our knowledge of cancer immunology advances, it is important to evaluate the efficiency of the conventional drug development process and consider novel endpoints, statistical designs, and analyses so that the benefit of these new therapeutic agents can be fully captured. Immune checkpoint inhibitors are an example of the importance of understanding the mechanism of action and the disease characteristics before embarking upon the research. An intermediate endpoint of milestone survival was proposed in an attempt to accelerate the drug development process, while mitigating the long-term survival and delayed clinical effect phenomenon introduced by efficacious immune-checkpoint inhibitors. For the clinical studies in late-stage drug development that involve single agents or in various combinations with immune checkpoint inhibitors, the proposed milestone survival is an endpoint that is worth considering.

Notes

References

1. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. 2012;12(4):252–264.
2. Melero I, Hervas-Stubbs S, Glennie M, Pardoll DM, Chen L. Immunostimulatory monoclonal antibodies for cancer therapy. *Nat Rev Cancer*. 2007;7(2):95–106.
3. Hodi FS, O'Day SJ, McDermott DF, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med*. 2010;363(8):711–723.
4. Robert C, Thomas L, Bondarenko I, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med*. 2011;364(26):2517–2526.
5. Maio M, Grob JJ, Aamdal S, 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(10):1191–1196.
6. Prieto PA, Yang JC, Sherry RM, et al. CTLA-4 blockade with ipilimumab: long-term follow-up of 177 patients with metastatic melanoma. *Clin Cancer Res*. 2012;18(7):2039–2047.
7. Lebbé C, Weber JS, Maio M, et al. Survival follow-up and ipilimumab retreatment for patients with advanced melanoma who received ipilimumab in prior phase II studies. *Ann Oncol*. 2014;25(11):2277–2284.
8. Schadendorf D, Hodi FS, Robert C, 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; In press.
9. Robert C, Long GV, Brady B, et al. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med*. 2015;372(4):320–330.
10. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer*. 1976;34(6):585–612.
11. Cox DR. Regression Models and life-tables (with discussion). *J Roy Statist Soc B*. 1972;34(2):187–220.
12. Chen TT. Statistical issues and challenges in immuno-oncology. *J Immunother Cancer*. 2013;1:18.
13. Shea MB, Roberts SA, Walrath JC, Allen JD, Sigal EV. Use of multiple endpoints and approval paths depicts a decade of FDA oncology drug approvals. *Clin Cancer Res*. 2013;19(14):1–10.
14. Wolchok JD, Kluger H, Callahan MK, et al. Nivolumab plus ipilimumab in advanced melanoma. *N Engl J Med*. 2013;369(2):122–133.
15. 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(26):2443–2454.
16. Weber JS, Minor DR, D'Angelo SP, et al. A phase 3 randomized, open-label study of nivolumab (anti-PD-1; BMS-936558; ONO-4538) versus investigator's choice chemotherapy (ICC) in patients with advanced melanoma after prior anti-CTLA-4 therapy. *Ann Oncol*. 2014;25(Suppl 4):Abstr LBA3.
17. Hamid O, Robert C, Daud A, et al. Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. *N Engl J Med*. 2013;369(2):134–144.
18. Brahmer JR, Tykodi SS, Chow LQ, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med*. 2012;366(26):2455–2465.
19. Herbst RS, Gordon MS, Fine GD, et al. A study of MPD3280A, an engineered PD-L1 antibody in patients with locally advanced or metastatic tumors [abstract 3000]. *J Clin Oncol*. 2013;31(Suppl):15s.
20. Wolchok JD, Neyns B, Linette G, et al. Ipilimumab monotherapy in patients with pretreated advanced melanoma: a randomised, double-blind, multicentre, phase 2, dose-ranging study. *Lancet Oncol*. 2010;11(2):155–164.
21. 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(23):7412–7420.
22. Hoos A, Eggermont AMM, Janetzki S, et al. Improved endpoints for cancer immunotherapy trials. *J Natl Cancer Inst*. 2010;102(18):1388–1397.
23. Berry DA. The hazards of endpoints. *J Natl Cancer Inst*. 2010;102(18):1376–1377.
24. Ribas A, Hersey P, Middleton MR, et al. New challenges in endpoints for drug development in advanced melanoma. *Clin Cancer Res*. 2011;18(2):336–341.
25. Conference on Clinical Cancer Research–Issue Brief. November 2013. Accessible at: http://www.pharmamedtechbi.com/~/media/Supporting%20Documents/The%20Pink%20Sheet/75/46/FOCR_immunotherapy_issue_brief.pdf. (Accessed April 17, 2015)
26. Dolgin E. Cancer immunotherapy advances spawn calls for new endpoints. *Nat Med*. 2013;19(11):1357.
27. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Amer Statist Assn*. 1958;53(282):457–481.
28. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol*. 1983;1(11):710–719.
29. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3):549–556.
30. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika*. 1977;64(2):191–199.
31. Lan K, DeMets D. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983;70(3):659–663.
32. Haybittle JL. Repeated assessments of results in clinical trials of cancer treatment. *Brit J Radiol*. 1971; 44(526):793–797.
33. Klein JP, Logan B, Harhoff M, Andersen PK. Analyzing survival curves at a fixed point in time. *Stat Med*. 2007;26(24):4505–4519.
34. Eggermont AMM, Chiarion-Sileni V, Grob JJ, et al. Ipilimumab versus placebo after complete resection of stage III melanoma: Initial efficacy and safety results from the EORTC 18071 phase III trial [abstract LBA9008]. *J Clin Oncol*. 2014;32(Suppl):5s.
35. Kwon ED, Drake CG, Scher HI, et al. Ipilimumab versus placebo after radiotherapy in patients with metastatic castration-resistant prostate cancer that had progressed after docetaxel chemotherapy (CA184-043): a multicentre, randomised double-blind phase 3 trial. *Lancet Oncol*. 2014;15(7):700–712.