Interim Futility Monitoring Assessing Immune Therapies With a Potentially Delayed Treatment Effect

Edward L. Korn and Boris Freidlin

Author affiliations and support information (if applicable) appear at the end of this article.

Published at jco.org on June 27, 2018.

Corresponding author: Edward L. Korn, PhD, Biometric Research Program, National Cancer Institute, MSC9735, Bethesda, MD 20892; e-mail: korne@ctep.nci.nih.gov.

Published by American Society of Clinical Oncology

0732-183X/18/3623w-2444w/\$20.00

ABSTRACT

Purpose

Introduction of new immune therapies that may have a delayed beneficial effect necessitates re-evaluation of traditional clinical trial designs in oncology. A key design feature of randomized trials is interim futility monitoring, which allows stopping early if the accruing data convincingly demonstrate that the experimental treatment is detrimental or is unlikely to be shown superior to the standard treatment. The appropriateness of futility monitoring is frequently questioned when the effect of the experimental treatment may be delayed (eg, in trials of many immune agents). We examine the advisability of using futility monitoring when there is potential for a delayed treatment effect and make recommendations concerning its use.

Methods

We evaluated the loss of statistical power when using some common futility interim monitoring rules and a new proposed conservative rule via simulation under varying amounts of treatment-effect delay and varying accrual periods. We also considered scenarios where the experimental treatment starts out being worse than the standard treatment but ends up being better, as may sometimes be the case when an immune therapy is compared with an active standard therapy.

Results

Some standard methods of futility monitoring can result in an unacceptable loss of power when there is a delayed treatment effect, especially if the accrual period is rapid or the experimental treatment is initially worse. The proposed conservative futility rule has a negligible loss of power in the situations considered.

Conclusion

Although care must be taken with the choice of futility monitoring when there is a delayed treatment effect, inclusion of appropriate rules can reduce the exposure of patients to ineffective therapies without reducing the probability of correctly identifying effective treatments.

J Clin Oncol 36:2444-2449. Published by American Society of Clinical Oncology

INTRODUCTION

If it becomes clear during an ongoing randomized clinical trial that the experimental-treatment arm is no better than the standard-treatment arm (or possibly worse), then formal interim futility monitoring allows one to stop the trial early¹; this limits the number of patients beginning an ineffective therapy (if the trial is still accruing) and allows patients in the trial to discontinue it. Because of its importance, futility monitoring is typically included in the design of randomized clinical trials.² However, a valid concern with using futility monitoring is when there is the potential for a delayed treatment effect. In this situation, traditional implementation of futility

monitoring may unacceptably reduce the ability to detect a truly effective experimental treatment (ie, trial will have reduced power).^{2,3} Intuitively, an early interim look at the trial data is likely to be driven by the events that occurred shortly after randomization when there may be no treatment benefit, making it more likely to stop incorrectly for futility.

The possibility of a delayed treatment effect has become more relevant to cancer trials with the introduction of immune agents like the checkpoint inhibitors (anti-cytotoxic T-lymphocyte antigen 4 [CTLA-4] and inhibitors to programmed death 1 [PD-1] or its ligand [PD-L1]). Biologic rationale and empirical evidence suggest that these immune agents may take a few months to show benefit.^{4,5} (This is opposed to standard

DOI: <https://doi.org/10.1200/JCO.2018.77.7144>

chemotherapeutic agents which directly kill cancer cells.) For example, Figure 1A shows the overall survival (OS) for ipilimumab plus dacarbazine versus placebo plus dacarbazine for previously untreated metastatic melanoma⁶; the OS curves overlap until they begin to separate at approximately 4 months. It is also possible that the experimental treatment will be initially worse than the standard treatment and then better later; for example, Figure 1B displays the results of the CheckMate 057 trial⁷ of nivolumab versus docetaxel for previously treated metastatic nonsquamous non-small-cell lung cancer. The potential for this situation is more likely when the experimental treatment (E) is being compared with an active standard treatment (E ν A), rather than when it is being added to a standard treatment (E + A ν A) or being compared with a placebo (E ν P). Although a delay is not uncommon for these immune-related agents, it is not universal. For example, Figure 1C displays the OS for nivolumab versus docetaxel for previously treated advanced squamous-cell non-small-cell lung cancer with no apparent delay in treatment effect.⁸ Finally, it should be noted that these agents do not work in all settings: Figures 1D and 1E display the OS results of two trials of pembrolizumab in multiple myeloma⁹; both trials were stopped early on the basis of negative interim results.

The potential loss of power with using futility monitoring with immunotherapy trials has been noted^{3,10-12} and has even apparently led to Bristol-Myers Squibb omitting futility monitoring from their trials involving ipilimumab.¹³ The purpose of this article is to examine critically via simulation the benefits and risks of using various types of futility monitoring when there is a delayed treatment effect. Included is a new futility monitoring rule for use when there is a potential delayed treatment effect in an immunotherapy trial, which is shown to have excellent properties.

METHODS

We performed simulations of trial designs to evaluate the effects of futility interim monitoring on their statistical operating characteristics. Our initial simulations use the scenario of Chen³: 680 patients are uniformly accrued and randomly assigned 1:1 over 34 months (ie, no ramping up of accrual), and the final analysis (using a one-sided .025-level log-rank test) is done when 512 events have been observed. There is no loss to follow-up (censoring). Because any negative effects of futility monitoring with a delayed treatment effect will be more pronounced with a shorter accrual period (because a higher proportion of the events will occur early, during the delay interval), we also consider an accrual period of 12 months. When there is no delay in treatment effect, this design has 90% power to detect a (constant) hazard ratio (HR) of 0.75, with the final analysis occurring at approximately 47 months after the first patient was randomly assigned (at approximately 34 months with a 12-month accrual period). The standard treatment arm is assumed to be exponential, with a median survival of 12 months. The survival curves are displayed in Figure 2A. To assess the benefits of futility monitoring, we also consider the situations when experimental and standard treatments are equally effective (null hypothesis) and when the experimental treatment is worse than the standard treatment (constant HR, 1.30).

Chen³ considers a 3-month delay before the treatment effect begins with the HR of 0.75 constant after the delay. As is well known,^{3,14-16} a delayed treatment effect reduces the power of the trial (eg, from 90% to 72% in the setting described above). To better isolate the potential negative consequences of interim futility monitoring, we decrease the HR (to 0.693) so that the trial will continue to have 90% power when there is no futility

monitoring (Fig 2B). We also consider a delay of 6 months (HR, 0.620 after the delay; Fig 2C), and a crossing-hazards situation where the experimental treatment is initially worse than the control treatment of the first 3 months (HR, 1.30) but then outperforms it (HR, 0.630 after 3 months; Fig 2D).

An additional consideration in designing trials for evaluation of immune therapies is the possibility that a certain proportion of patients will be cured.^{3,5} This is not uncommon with standard chemotherapies (especially in less-advanced disease settings), and it is well known that the required sample size will need to be increased to achieve the required number of events in a reasonable amount of time.^{17,18} However, a cure rate would be expected in addition to increase the negative consequences of interim futility monitoring, because a higher proportion of events would be expected to occur during the delay interval. To investigate this, we performed simulations with the same 12 scenarios above (six survival settings and two accrual periods) but with all patients alive 28 months after random assignment being cured. This corresponds to a cure rate of 20% for the standard treatment arm and 30% for the experimental treatment arm when there is no delayed treatment effect. To maintain 90% power with a trial of reasonable length, the sample size is increased to 800; the final analysis continues to be performed when there are 512 events.

We consider two commonly used futility monitoring rules and a new proposal designed to accommodate a possible delayed treatment effect. Each monitoring rule has two interim analyses. The first is the commonly used Wieand rule,¹⁹ which recommends stopping for futility when one half the expected number of events have been observed (50% information; in this case, 256 events) if the observed HR > 1 (ie, the experimental treatment looks worse than the standard treatment by any amount). We apply a second interim analysis at 75% information with the same futility criteria: stop if the observed HR > 1 . The second futility rule we consider is quite aggressive (ie, easier to stop) and is based on an O'Brien-Fleming β spending function.²⁰ There are different implementations of this approach; we use the one that was used in the trial of tremelimumab versus chemotherapy in advanced melanoma²¹ (see Discussion): stop for futility if the Z-statistic (associated with testing the hazard ratio) < 0.011 at one-third information (first analysis) or if the Z-statistic < 0.864 at two-thirds information (second analysis). These correspond to approximately stopping if the HR > 0.998 at one-third information or HR > 0.913 at two-thirds information in our simulations.

The new futility rule considered is a proposed ad hoc modification of the Wieand approach that stops for futility if the HR > 1 at the first interim analysis, occurring when at least 50% of the expected events have occurred and at least two thirds of the observed events have occurred later than 3 months from randomization. The second interim analysis occurs when at least 75% of the expected events have occurred and at least two thirds of the observed events have occurred later than 3 months from randomization and again stops for futility if the HR > 1 . Intuitively, this approach is designed to ensure that the interim analysis is dominated by events occurring outside of the delay interval, so that whether the experimental treatment is futile can be reliably assessed. The particular analysis-time criteria for the proposed futility rule have been chosen so that the rule will have good properties when there is no delayed treatment effect and also when there is a delayed treatment effect of 3 to 6 months (see the simulation in the Results section). With a short accrual time, it can happen that there are no interim analyses because the criteria are not satisfied before the final analysis occurs; this is appropriate if there are insufficient post-delay events observed by the time of the final analysis.

Although we believe that a maximum potential 6-month treatment-effect delay covers most immunotherapy trials, if there is evidence supporting a longer potential delay then the proposed futility rule could be modified by requiring two thirds of the observed events to occur after one half the maximum contemplated delay. For example, if a maximum 10-month delay is considered possible, then the first interim analysis would occur when at least 50% of the expected events have occurred and at least two thirds of the observed events have occurred later than 5 months from randomization. Note that we are not considering delayed treatment effects that are part of the design of the trial. For example, a trial of 2 versus 4 years

Fig 1. Overall survival for arms of some trials testing checkpoint inhibitors. (A) Ipilimumab plus dacarbazine (gold curve) versus placebo plus dacarbazine (blue curve) for previously untreated metastatic melanoma (N = 502); adapted from Robert et al.⁶ (B) Nivolumab (red curve) versus docetaxel (gray curve) for previously treated metastatic nonsquamous non-small-cell lung cancer (N = 582); adapted from Kazandjian et al.⁷ (C) Nivolumab (blue curve) versus docetaxel (gold curve) for previously treated advanced squamous-cell non-small-cell lung cancer (N = 272); adapted from Brahmer et al.⁸ (D) Pomalidomide/low-dose dexamethasone plus pembrolizumab (pembro; red curve) versus pomalidomide/low-dose dexamethasone (gray curve, standard of care [SOC]) for third-line relapsed/refractory multiple myeloma (N = 249; planned accrual = 300); adapted from US Food and Drug Administration.⁹ (E) Lenalidomide/low-dose dexamethasone plus pembrolizumab (red curve) versus lenalidomide/low-dose dexamethasone (gray curve, standard of care [SOC]) for newly diagnosed multiple myeloma (N = 301; planned accrual = 640); adapted from US Food and Drug Administration.⁹

Fig 2. Hypothetical survival curves (gold, experimental; blue, control) for simulations. (A) Hazard ratio (HR), 0.75; no delayed treatment effect. (B) Three-month delayed treatment effect, then HR, 0.693. (C) Six-month delayed treatment effect, then HR, 0.620. (D) Crossing hazards with HR, 1.30 in the first 3 months then HR, 0.628.

of therapy has a 2-year delay of differential treatment effect built into the trial design (because patients in both treatment arms receive the same therapy for the first 2 years). In these situations, one should use different analysis approaches to accommodate the delay (eg, randomizing the treatment assignment after 2 years of treatment²² or, for blinded trials, using a landmark analysis starting at 2 years). (A landmark analysis excludes patients who have events in the first 2 years²³; this should not be confused with a milestone analysis that refers to a survival rate at a pre-determined time point among all patients.¹⁸)

RESULTS

Simulation results are presented in Table 1 for when there is no cure rate. First, consider the scenarios assuming 34-month accrual (top half of table). The first row shows the no-delay scenario (Fig 2A) that the trial was designed to detect with 90% power. In this case, the Wieand and proposed approaches to futility monitoring negligibly reduce the power of the trial; the more aggressive O'Brien-Fleming approach loses approximately 2% power, which we would consider borderline acceptable. The next two rows show the benefits of futility monitoring when the experimental treatment does not work: fewer patients accrued to the trial and the trial ending sooner. Rows 4 to 6 directly address the issue of a delayed treatment effect. The loss of power with the O'Brien-Fleming approach is unacceptable and is approximately 1%, 2%, and 3% for the Wieand approach, with a 3-month delay, 6-month delay,

and crossing hazards, respectively. The proposed futility rule acceptably maintains the power of the trial.

With a 12-month accrual period (bottom half of Table 1), the benefits of futility monitoring are more limited (rows 8 and 9), and both the Wieand and O'Brien-Fleming approaches result in an unacceptable loss of power (rows 10 to 12). The loss of power is especially concerning for the O'Brien-Fleming approach (eg, power is reduced from 90% to 54% under the crossing-hazard alternative). The proposed approach maintains acceptable power.

Similar results are observed when a proportion of patients are being cured (Table 2). Futility monitoring provides substantial benefit when the experimental treatment does not work (eg, the average number of patients enrolled is reduced from 800 to 682 when the HR is 1 [row 2]). Both the Wieand and O'Brien-Fleming rules result in unacceptable loss of power with shorter accrual times, whereas the proposed futility rule preserves the trial power (rows 10 to 12).

DISCUSSION

Selecting an appropriate interim futility rule requires achieving a delicate balance between patient safety and public health considerations. On one hand, if the experimental therapy is ineffective, one would want to minimize the number of patients exposed to the therapy. On the other hand, if the new therapy is beneficial, one would want to maximize the probability of detecting the benefit.

Table 1. Simulated Trial Power and Average Duration and Sample Size with Three Methods of Futility Interim Monitoring (see text) Under Six Treatment-Effect Settings and Two Different Accrual Periods

Setting	No Interim Analyses		Wieand Approach		O'Brien-Fleming Approach		Proposed Approach	
	Power	Duration (months), SS	Power	Duration (months), SS	Power	Duration (months), SS	Power	Duration (months), SS
Accrual 34 months								
HR, 0.75 (no delay)	0.900	47.4, 680	0.898	47.1, 678.6	0.879	46.2, 671.7	0.898	47.1, 678.6
HR, 1.00	0.0247	43.9, 680	0.0245	34.1, 602.6	0.0226	28.2, 529.1	0.0245	34.3, 605.3
HR, 1.30	0	41.3, 680	0	25.2, 504.4	0	19.9, 399.3	0	26.1, 526.1
3-month delay; HR, 0.693*	0.901	47.8, 680	0.894	47.3, 676.7	0.846	45.4, 661.2	0.895	47.4, 677.4
6-month delay; HR, 0.620†	0.903	48.3, 680	0.881	47.2, 673.1	0.786	43.9, 644.3	0.885	47.4, 674.4
HR, 1.30; HR, 0.628‡	0.899	48.3, 680	0.872	47.0, 671.7	0.762	43.2, 638.0	0.891	47.8, 677.8
Accrual 12 months								
HR, 0.75 (no delay)	0.899	34.3, 680	0.897	34.1, 680	0.879	33.3, 679.7	0.898	34.2, 680
HR, 1.00	0.0246	30.5, 680	0.0243	21.5, 680	0.0223	16.9, 660.5	0.0245	23.3, 680
HR, 1.30	0	27.7, 680	0	13.7, 680	0	10.9, 607.5	0	19.7, 680
3-month delay; HR, 0.693	0.901	34.8, 680	0.882	33.9, 680	0.794	31.1, 677.3	0.898	34.5, 680
6-month delay; HR, 0.620	0.903	35.4, 680	0.804	32.1, 680	0.626	27.1, 670.4	0.879	34.3, 680
HR, 1.30; HR, 0.628	0.899	35.4, 680	0.781	31.6, 680	0.537	24.7, 665.1	0.895	35.1, 680

NOTE. Maximum sample size = 680; final analysis at 512 events; type 1 error = 0.025; 100,000 simulated data sets per row. HRs are experimental arm hazard over control-arm hazard so that numbers less than 1.00 represent the experimental treatment being better.

Abbreviations: HR, hazard ratio; SS, sample size.

*The HR is 1.00 for 3 months, then is 0.693 for the remaining time.

†The HR is 1.00 for 6 months, then is 0.620 for the remaining time.

‡The HR is 1.30 for 3 months, then is 0.628 for the remaining time.

Commonly used futility rules have been optimized for settings with no treatment delay. If the treatment effect is delayed, the application of many commonly used futility rules may result in an unacceptably high loss of power, because interim results are dominated by the early events occurring before the new treatment started working. (Predictably, this loss of power may become particularly dramatic in cases of rapid patient accrual.) We propose a simple approach to restoring the balance between safety and loss of power by delaying interim analysis times until the data become sufficiently representative of the expected treatment effect.

The melanoma trial of tremelimumab versus chemotherapy mentioned in the Methods section is sometimes taken as a prime

example of the problematic nature of futility monitoring when there is a delayed treatment effect^{3,11}: The trial was stopped for futility at the second interim analysis (when 340 deaths had occurred) and the OS HR was 0.96.²¹ However, little harm was done by this early stopping (performed 8 months after the last patient was enrolled), because the investigators were able to perform the regularly scheduled final analysis when there were 534 deaths, observing an OS HR of 0.88 ($P = .127$).²¹ One could still argue that stopping for futility (and publicly reporting the interim results as futile²⁴) when results would end up non-negligibly positive (even if not statistically significant) is not an optimal strategy. Because it seems unlikely that the HR was > 1 at any time past 50%

Table 2. Simulated Trial Power and Average Duration and Sample Size With Three Methods of Futility Interim Monitoring (see text) Under Six Treatment-Effect Settings (with all patients alive at 28 months cured) and Two Different Accrual Periods

Setting	No Interim Analyses		Wieand Approach		O'Brien-Fleming Approach		Proposed Approach	
	Power	Duration (months), SS	Power	Duration (months), SS	Power	Duration (months), SS	Power	Duration (months), SS
Accrual 34 months								
HR, 0.75 (no delay)	0.900	41.3, 800	0.898	41.1, 797.5	0.879	40.3, 787.0	0.898	41.1, 797.7
HR, 1.00	0.0246	38.1, 800	0.0243	30.2, 673.2	0.023	25.2, 576.8	0.0244	30.7, 682.3
HR, 1.30	0	35.9, 800	0	22.8, 535.7	0	18.1, 426.0	0	25.3, 592.0
3-month delay; HR, 0.68*	0.902	41.7, 800	0.895	41.3, 795.0	0.850	39.8, 773.4	0.899	41.4, 796.9
6-month delay; HR, 0.583†	0.902	42.2, 800	0.883	41.4, 790.1	0.798	38.7, 751.2	0.891	41.6, 793.7
HR, 1.30; HR, 0.6‡	0.901	42.2, 800	0.878	41.3, 788.3	0.778	38.3, 743.4	0.898	42.1, 798.8
Accrual 12 months								
HR, 0.75 (no delay)	0.900	26.7, 800	0.898	26.6, 800	0.879	26.0, 797.9	0.900	26.6, 800
HR, 1.00	0.0249	24.0, 800	0.0244	17.9, 800	0.0225	14.5, 744.8	0.0249	21.5, 800
HR, 1.30	0	21.9, 800	0	12.3, 796.1	0	10.0, 655.9	0	21.3, 800
3-month delay; HR, 0.68	0.901	27.0, 800	0.875	26.2, 800	0.786	24.3, 787.2	0.901	27.0, 800
6-month delay; HR, 0.583	0.901	27.4, 800	0.778	24.7, 800	0.587	21.0, 765.7	0.894	27.1, 800
HR, 1.30; HR, 0.6	0.899	27.4, 800	0.731	24.0, 800	0.506	19.4, 751.5	0.899	27.4, 800

NOTE. Maximum sample size = 800; final analysis at 512 events; type 1 error = 0.025; 100,000 simulated data sets per row. HRs are experimental arm hazard over control-arm hazard, so that numbers < 1.00 represent the experimental treatment being better.

Abbreviations: HR, hazard ratio; SS, sample size.

*The hazard ratio is 1.00 for 3 months, then is 0.68 until 28 months.

†The hazard ratio is 1.00 for 6 months, then is 0.583 until 28 months.

‡The hazard ratio is 1.30 for 3 months, then is 0.6 until 28 months.

information, there would have been no recommendation to stop the trial for futility using either the Wieand rule or our proposed futility monitoring rule.

In summary, although one needs to be cautious with futility interim monitoring when there is a possible delayed treatment effect, there is no need to abandon this important protection for clinical trial participants. In particular, the proposed futility monitoring rule results in a small loss of power regardless of whether the treatment effect is delayed (even with rapid accrual) but offers considerable savings in time and patients treated when the experimental treatment is no better than, or is worse than, the standard treatment.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at jco.org.

AUTHOR CONTRIBUTIONS

Manuscript writing: All authors

Final approval of manuscript: All authors

REFERENCES

- Freidlin B, Korn EL: Monitoring for lack of benefit: A critical component of a randomized clinical trial. *J Clin Oncol* 27:629-633, 2009
- Korn EL, Freidlin B: Inefficacy interim monitoring procedures in randomized clinical trials: The need to report. *Am J Bioeth* 11:2-10, 2011
- Chen TT: Statistical issues and challenges in immune-oncology. *J Immunother Cancer* 1:18, 2013
- Finke LH, Wentworth K, Blumenstein B, et al: Lessons from randomized phase III studies with active cancer immunotherapies—outcomes from the 2006 meeting of the Cancer Vaccine Consortium (CVC). *Vaccine* 25:B97-B109, 2007 (suppl 2)
- Anagnostou V, Yarchoan M, Hansen AR, et al: Immuno-oncology trial endpoints: Capturing clinically meaningful activity. *Clin Cancer Res* 23:4959-4969, 2017
- Robert C, Thomas L, Bondarenko I, et al: Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med* 364:2517-2526, 2011
- Kazandjian D, Suzman DL, Blumenthal G, et al: FDA approval summary: Nivolumab for the treatment of metastatic non-small-cell lung cancer with progression on or after platinum-based chemotherapy. *Oncologist* 21:634-642, 2016
- Brahmer J, Reckamp KL, Baas P, et al: Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med* 373:123-135, 2015
- U.S. Food & Drug Administration: FDA alerts healthcare professionals and oncology clinical investigators about two clinical trials on hold evaluating KEYTRUDA (pembrolizumab) in patients with multiple myeloma. <https://www.fda.gov/Drugs/DrugSafety/ucm574305.htm>
- Hoos A, Eggermont AMM, Janetzki S, et al: Improved endpoints for cancer immunotherapy trials. *J Natl Cancer Inst* 102:1388-1397, 2010
- Menis J, Litière S, Tryfonidis K, et al: The European Organization for Research and Treatment of Cancer perspective on designing clinical trials with immune therapeutics. *Ann Transl Med* 4:267, 2016
- Xu Z, Zhen B, Park Y, et al: Designing therapeutic cancer vaccine trials with delayed treatment effect. *Stat Med* 36:592-605, 2017
- Hoos A, Britten CM, Huber C, et al: A methodological framework to enhance the clinical success of cancer immunotherapy. *Nat Biotechnol* 29:867-870, 2011
- Halperin M, Rogot E, Gurian J, et al: Sample sizes for medical trials with special reference to long-term therapy. *J Chronic Dis* 21:13-24, 1968
- Zucker DM, Lakatos E: Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika* 77:853-864, 1990
- Fine GD: Consequences of delayed treatment effects on analysis of time-to-event endpoints. *Drug Inf J* 41:535-539, 2007
- Sposto R, Sather HN: Determining the duration of comparative clinical trials while allowing for cure. *J Chronic Dis* 38:683-690, 1985
- Chen TT: Milestone survival: A potential intermediate endpoint for immune checkpoint inhibitors. *J Natl Cancer Inst* 107:djv156, 2015
- Wieand S, Schroeder G, O'Fallon JR: Stopping when the experimental regimen does not appear to help. *Stat Med* 13:1453-1458, 1994
- Jennison C, Turnbull BW: *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton, FL, Chapman & Hall/CRC, 2000
- 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* 31:616-622, 2013
- Durrleman S, Simon R: When to randomize? *J Clin Oncol* 9:116-122, 1991
- Anderson JR, Cain KC, Gelber RD: Analysis of survival by tumor response. *J Clin Oncol* 1:710-719, 1983
- Ribas A, Hauschild A, Kefford R et al: Phase III, open-label, randomized, comparative study of tremelimumab (CP-675,206) and chemotherapy (temozolomide [TMZ] or dacarbazine [DTIC]) in patients with advanced melanoma. *J Clin Oncol* 26, 2008 (suppl; abstr LBA9011)

Affiliations

All authors: National Cancer Institute, Bethesda, MD.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Interim Futility Monitoring Assessing Immune Therapies With a Potentially Delayed Treatment Effect

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ifc.

Edward L. Korn

No relationship to disclose

Boris Freidlin

No relationship to disclose