OXFORD

COMMENTARY

Design Issues in Randomized Clinical Trials of Maintenance Therapies

Boris Freidlin, Richard F. Little, Edward L. Korn

Affiliations of authors: Biometric Research Branch (BF, ELK) and Clinical Investigations Branch, Cancer Therapy Evaluation Program (RFL), Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD.

Correspondence to: Boris Freidlin, PhD, Biometric Research Branch, National Cancer Institute, 9609 Medical Center Drive, Room 5W102, MSC 9735, Bethesda, MD 20892 (e-mail: freidlinb@ctep.nci.nih.gov).

Abstract

A potential therapeutic strategy for patients who respond (or have stable disease) on a fixed-duration induction therapy is to receive maintenance therapy, typically given for a prolonged period of time. To enable patients and clinicians to make informed treatment decisions, the designs of phase III randomized clinical trials (RCTs) assessing maintenance strategies need to be such that their results will provide clear assessment of the relevant risks and benefits of these strategies. We review the key aspects of maintenance RCT designs. Important design considerations include choice of first-line and second-line therapies, minimizing between-arm differences in follow-up schedules, and choice of the primary endpoint. In order to change clinical practice, RCTs should be designed to accurately isolate and quantify the clinical benefit of maintenance as compared with the standard approach of fixed-duration induction followed by the second-line treatment at progression. To accomplish this, RCTs need to utilize an overall survival (or quality of life) endpoint or, in settings where this is not feasible, endpoints that incorporate the effects of the subsequent line of therapy (eg, time from randomization to second progression or death). Toxicity and symptom information over both the study treatment (maintenance) and the second-line treatment should also be collected and reported.

Maintenance therapies are based on introducing additional treatment (typically lasting until progression) for patients who have a response or stable disease (SD) after (a fixed duration of) first-line therapy. These strategies are broadly categorized into: 1) the switch-maintenance approach, where, after a standard first-line therapy, patients are switched to a different agent until progression, and 2) the continuation-maintenance approach, where a component of the first-line is continued past its standard duration until progression (1). These basic RCT trial designs assessing these maintenance approaches are displayed in Figure 1, A and B, respectively. These designs isolate the benefit of using a new agent (Figure 1A) or continuing administration of a component of the first-line regimen (Figure 1B) in responding/SD patients relative to the standard of care. It is also possible to use an induction/maintenance trial design to evaluate an overall treatment strategy that combines the addition of a new agent to a fixed-duration first-line treatment (induction) as well as continuation of that agent in maintenance (Figure 1C). More complex multistage

and/or multi-arm designs can also be employed for maintenance evaluation as will be described in the next section.

To provide definitive evidence of the clinical benefit of a maintenance strategy, the following four trial-design issues should be considered: 1) choice of first-line therapy, 2) choice of second-line therapy, 3) potential between-arm differences in follow-up schedules, and 4) choice of primary endpoint. We discuss these in turn in this Commentary.

Choice of First-Line Therapy

In the switch-maintenance or continuation-maintenance designs, the first-line therapy is a standard-of-care first-line therapy (Figure 1, A and B). For the induction/maintenance trial design, the new agent is incorporated into the first-line treatment on the experimental arm. Note that unlike the switch-maintenance or continuation-maintenance designs that randomize responding/SD patients after induction, randomization for the
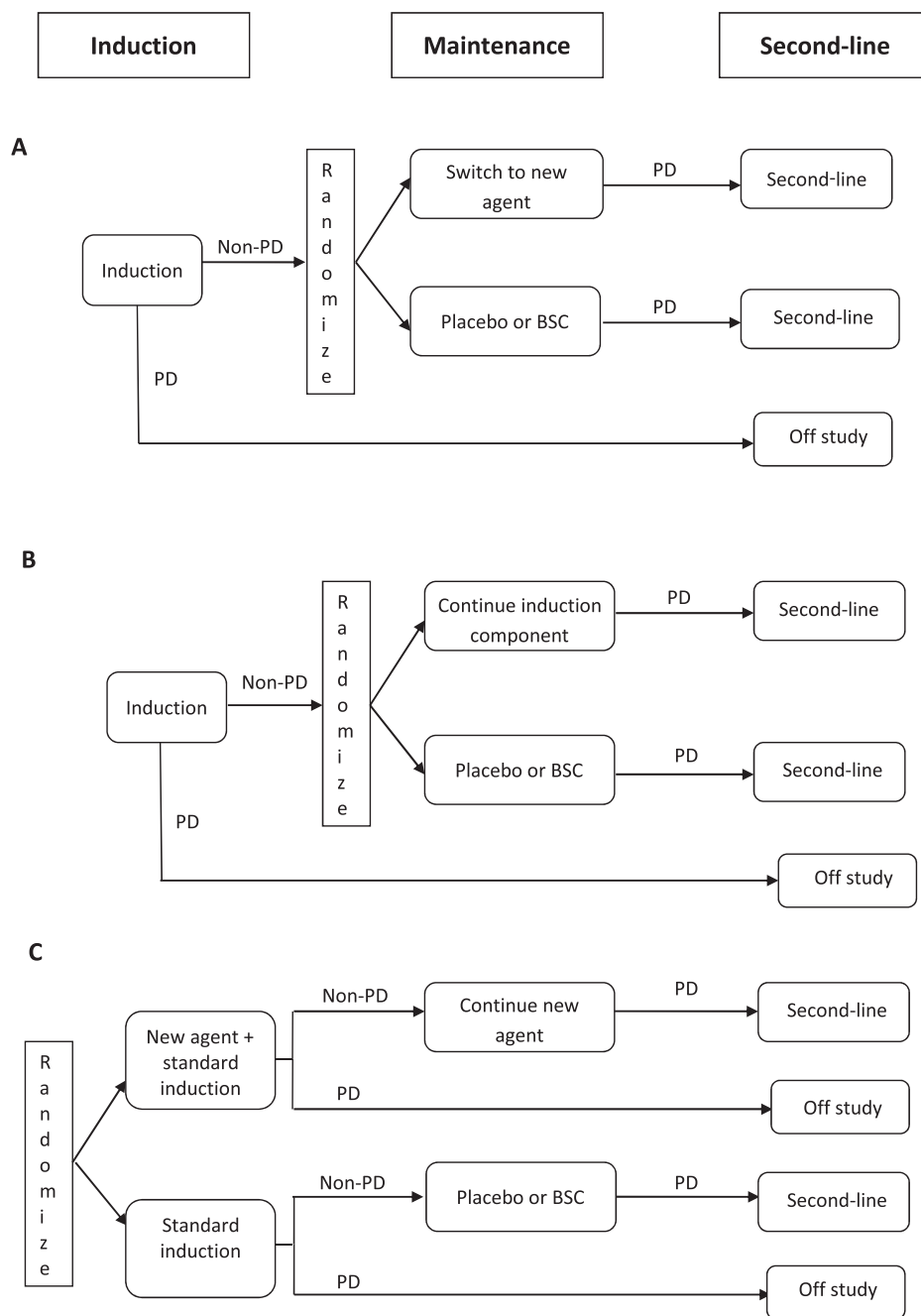
Figure 1. Commonly used randomized clinical trial designs evaluating maintenance. **A)** Switch-maintenance. **B)** Continuation-maintenance. **C)** Induction/maintenance. BSC = best supportive care; PD = progressive disease.

induction/maintenance design takes place before the first-line treatment (Figure 1C). For example, the ESCAPE trial (2) randomly assigned first-line metastatic non–small cell lung cancer (NSCLC) patients between the experimental arm (induction with chemotherapy+sorafenib followed by sorafenib maintenance) vs the control arm (induction chemotherapy+placebo followed by placebo maintenance). Use of the induction/maintenance design should be justified, as it confounds induction and maintenance roles and thus makes it impossible to isolate the degree to which maintenance contributed to any observed benefit (1). Furthermore, because the design is based on comparing all randomly assigned patients regardless of whether they received maintenance therapy, the ability to detect a maintenance treatment effect is reduced as compared with the other maintenance designs. When one needs to assess the benefits of a new therapy added to induction vs used as maintenance, one possible approach is to use a three-arm trial design that includes experimental arms with and without maintenance (Figure 2A). For example, GOG218 (3) evaluated the role of bevacizumab in first-line treatment of advanced ovarian cancer by randomly assigning patients between the experimental arm 1 (induction with chemotherapy+bevacizumab followed by bevacizumab maintenance), experimental arm 2 (induction with chemotherapy+bevacizumab followed by placebo maintenance), and the control arm (induction with chemotherapy+placebo followed by placebo maintenance).
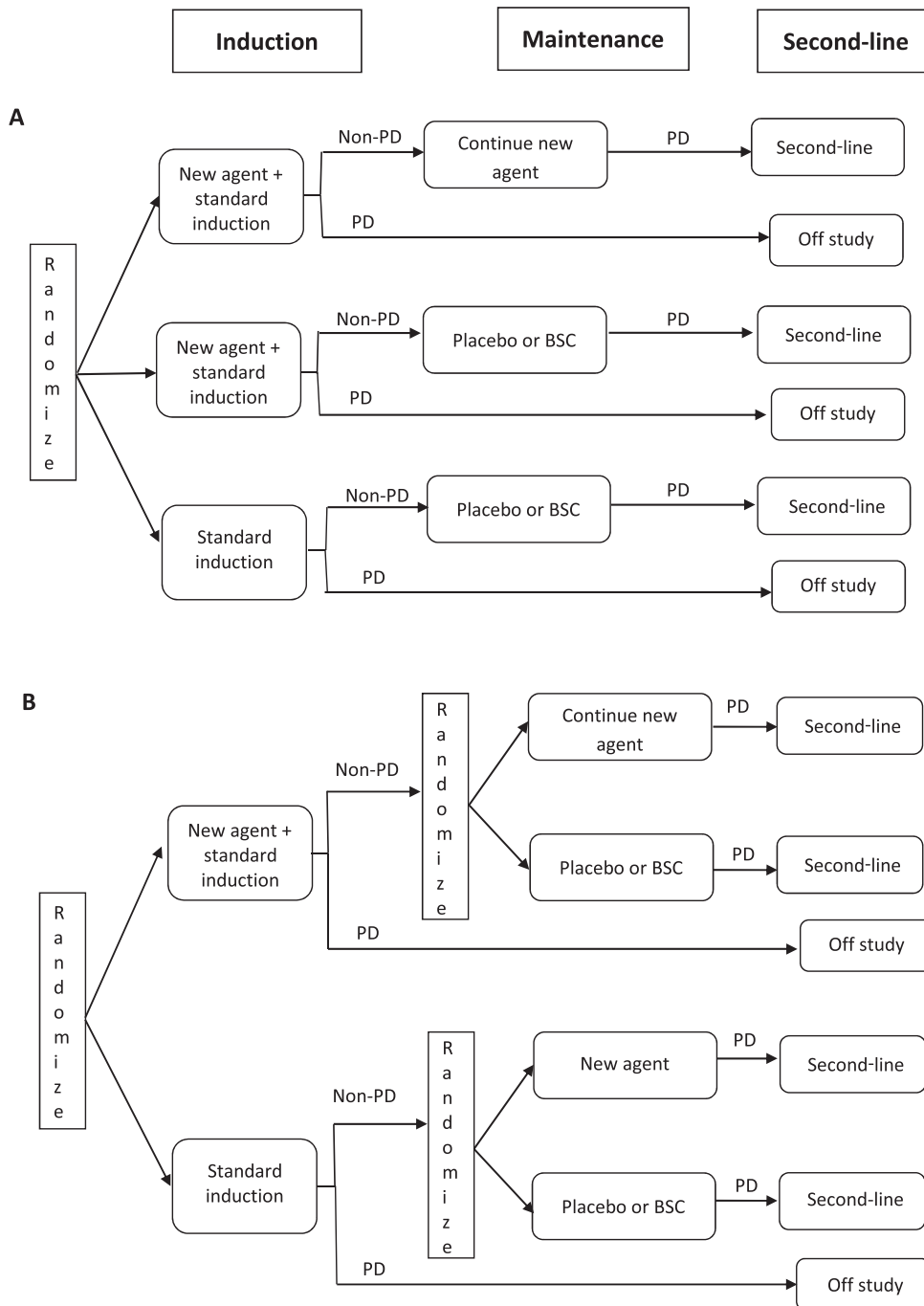
Figure 2. More complex maintenance trial designs. **A)** Three-arm design. **B)** Two-stage randomization design. BSC = best supportive care; PD = progressive disease.

An alternative approach that allows one to isolate the induction and maintenance questions is a two-stage randomization design (Figure 2B). At the first stage patients are randomly assigned between standard induction plus the new agent and the standard induction alone. After the induction, patients eligible for maintenance (SD or better) are randomly assigned between maintenance with the new agent vs no maintenance (4–6). Because the second stage randomization is limited to a subset of study patients, special statistical methods (inverse–probability-of-treatment-weighting [7]) are required to provide robust unbiased evidence on the induction question. This design is a special (simple) case of SMART designs (8), which can be used for the initial evaluation of adaptive multistage treatment strategies (9). Note that the two-stage randomization is different from the two-by-two factorial design where both induction and maintenance randomizations are performed concurrently before induction. While these designs are sometimes used (10), they could be inefficient because a substantial proportion of patients may not receive the assigned maintenance therapy, leading to a loss in power of the trial to detect maintenance effects (11). Note that the three-arm design (Figure 2A) is essentially a two-by-two design without the standard-induction-followed-by-maintenance arm and thus has the same inefficiency issue.

The induction/maintenance design (Figure 1C) could be used in a randomized phase II study to screen a new therapy that is expected to provide maximum effect with induction and maintenance administration. A positive signal from the phase II trial would need to be followed by a two-stage randomization (Figure 2B) or a three-arm phase III (Figure 2A) design to delineate the best administration strategy. The (two-arm) induction/maintenance design (Figure 1C) would generally be inappropriate for the definitive phase III evaluation of a new induction/maintenance therapy, unless there exists a clear biologic rationale that the activity of the experimental agent added to standard induction, if it works at all, requires a prolonged continuous administration.

Choice of Second-Line Therapy

The second-line treatment that the control-arm patients receive at progression should be representative of the current standard of care. Depending on the setting, the standard of care could include the maintenance agent (if it has established benefit in the second-line setting), an approved agent similar to the maintenance agent, a completely different therapy, or best supportive care. If the maintenance agent is a standard second-line treatment, then the design should prespecify its use as the second-line therapy on the control arm, otherwise the trial may be difficult to interpret (12). For example, SATURN (13) compared maintenance erlotinib to placebo in advanced NSCLC, but only 21% of the control-arm patients received a subsequent-line epidermal growth factor receptor (EGFR) TKI (72% received some second-line treatment). It is thus impossible to distinguish whether the positive observed treatment effect in that trial is attributable to the maintenance treatment of erlotinib or the fact that less than one-third of the control-arm patients who received second-line therapy received erlotinib or another EGFR TKI (14). Would the trial still have shown a benefit if all 72% of control-arm patients received standard second-line erlotinib?

When the maintenance agent is not a standard of care second-line agent, the trial design should ideally prespecify the same standard second-line therapy (different from the maintenance agent) to be used at progression for both arms. Prespecification of the same second-line therapy (or the same algorithm for its selection in case of a molecularly guided therapy) will ensure that any treatment differences observed are because of the maintenance treatment and not because of an inadvertent imbalance between the treatment arms in the use of various second-line therapies. (Note that any observed between-arm differences in the second-line therapies actually received in a trial that prespecifies the same second-line therapy reflect the differential effect of the two strategies on the delivery of the subsequent treatment and are appropriately captured by the design.) Furthermore, when studies are conducted in countries with limited access to standard second-line therapies, it is possible that the control-arm patients would not receive second-line therapy equivalent to that given in more resource rich settings. Results from such international studies may not be easily generalizable to countries with broad access to effective standard second-line therapies (15). Regardless of what is specified in the trial design, the frequency/types of the second-line (and preferably the subsequent lines) therapies should be recorded for each arm.

For trials in which the maintenance agent has no established benefit in the subsequent treatment lines, there is no scientific rationale or clinical necessity for the control-arm patients to cross over to the maintenance agent at progression (even though it is often argued that the presence of crossover may improve the trial accrual). Such a crossover confounds the evaluation of OS and thus undermines assessment of the true clinical impact of the new treatment strategy (16–18).

Between-Arm Differences in Follow-up Schedules

Differences in follow-up schedules can lead to a biased assessment of between-arm differences in non-OS endpoints. (This is not an issue in placebo-controlled trials.) Although the concern with differences in follow-up schedule is applicable to (non-placebo controlled) trials in general (19–21), it can be especially problematic in maintenance trials: Patients on the control arm may not be followed as closely as the patients receiving maintenance therapy, resulting in progressions identified later in the control vs maintenance patients. This in turn could also lead to a delay in second-line treatments for control vs maintenance arms. For example, in an NSCLC trial of maintenance docetaxel vs docetaxel at progression (22), 37% of the control-arm patients failed to receive the prespecified second-line docetaxel (often because of symptomatic deterioration before formal progression criteria were met [23]). This may be in part because of less frequent tumor evaluations on the control arm (23,24).

A related methodological issue is that more of the control-arm patients than maintenance-arm patients might start second-line therapy without protocol-defined progression if the control-arm is observation. How to treat the data from these patients in the analysis is problematic; if starting second-line therapy is treated as a progression-free survival (PFS) event, it can make the control-arm event rate appear larger than it really is. On the other hand, censoring these observations is also not statistically valid; the underlying assumption behind censoring is that it tells one nothing about the timing of future progression, an assumption that is unlikely to hold in this case.

Choice of Primary Endpoint

For a maintenance therapy to offer patient benefit, it should improve the patient's overall survival (OS) or his or her quality of life (QOL). In quantifying the QOL impact of a maintenance strategy, in addition to disease-related symptoms, one needs to capture treatment-related toxicity during maintenance and the subsequent lines of therapy (25). However, the need for repeated prolonged collection of information over multiple time points including subsequent line of therapy (that are often given outside of the current study) complicates QOL assessment and increases confounding by informative (outcome related) drop-out (26). Furthermore, the multidimensional nature of QOL becomes especially pronounced in the maintenance setting because of the need to balance the burden of treatment and disease-related symptoms over multiple lines of therapy as well as because of the considerable variability in patient's personal tradeoffs between QOL and OS (27). Therefore, effective QOL assessment may require developing new patient-reported outcome instruments and assessment strategies that would allow reliable quantification of the individual patient overall symptom vs toxicity experience (28). Because of the challenges in assessing QOL and the feasibility of using OS as the primary endpoint (because of the increasing sample sizes required as the study population prognosis improves), other time-to-event endpoints are frequently considered for maintenance trials.

Progression-Free Survival

Progression-free survival has become increasingly used in RCTs as the primary endpoint. From the regulatory perspective, a general argument for PFS as a trial endpoint is that it allows a new agent that has better PFS than an old agent to receive timely regulatory approval and become available to patients as another treatment option (29). One could argue that regardless of increased clinical benefit, an agent with better PFS has at least equivalent biologic activity and, most likely, at least equivalent clinical activity. However, in the context of maintenance trials, one is typically not comparing a new maintenance agent to an old one. Instead, one is either: 1) comparing a maintenance agent to observation/placebo or 2) adding a new maintenance agent to the standard maintenance regimen (with the new maintenance agent typically having proven activity in the subsequent lines of therapy). In neither case is one substituting the use of a new agent for a standard agent, but one is instead adding an early/prolonged administration of an agent to a standard treatment option. Therefore, the argument that the maintenance agent is offering an alternative treatment does not generally apply (18).

In selecting a primary endpoint for measuring therapeutic effect in Phase III clinical trials, the use of PFS is often motivated by the fact that PFS is not confounded by subsequent lines of therapy (17,30). However, unless PFS reflects some tangible aspect of treatment-associated clinical benefit, this argument is not valid: Just because one can assess and find between-treatment-arm differences in an endpoint does not mean it is an appropriate endpoint (17). We consider two additional potential justifications for using PFS as a measure of therapeutic effect in the context of trials of maintenance therapies.

First, PFS is used as a surrogate for OS. That is, a PFS benefit of maintenance therapy seen in a RCT will reflect an OS benefit. However, increasing availability of effective subsequent-line therapies (attenuating any PFS/OS relationship) makes this unlikely in many settings (16). This is particularly true for maintenance treatments, where the subsequent-line therapy in the control arm is frequently the maintenance agent (making it a comparison of immediate vs a delayed administration strategy) (18). The fact that a PFS benefit of a particular maintenance therapy might be reflective of an OS benefit for this maintenance therapy if subsequent therapies did not exist is irrelevant in the real world, where such therapies do exist (16). This suggests that in the absence of evidence that PFS predicts OS benefit in the presence of available second-line therapies, the only convincing way to assess OS benefit is to measure it directly. Another possible reason PFS is sometimes considered a surrogate for OS is that there is confusion of the concepts of PFS as an individual-level surrogate (ie, an individual's PFS reflects his/her eventual OS) with PFS as a trial-level surrogate (ie, the trial PFS treatment effect reflects the trial OS treatment effect). An intermediate variable can be a good individual-level surrogate but a poor trial-level surrogate (31); this may be the case with PFS and OS in some settings (32).

A second possible justification for using PFS as a trial endpoint is that an improvement in PFS represents an improvement in patient QOL and therefore represents direct patient benefit. However, progression in solid tumors is defined as an increase in radiographic measurement of tumor size above a predetermined threshold that was determined to represent biologic activity rather than clinical benefit (33,34), and progression in hematologic malignancies is often determined by a change in laboratory values that would not be generally expected to cause symptoms (18,35). Therefore, PFS differences could be driven by radiographic or laboratory changes that may be unnoticeable to the patient and thus do not translate into direct clinical benefit. In view of that, the validity of PFS as a direct measure of QOL depends very much on the disease setting, magnitude of PFS improvement, and toxicity of the maintenance agent. For example, in a setting where progression usually leads to increased disease symptoms, delaying progression may have direct QOL benefits. However, even in this setting, the QOL benefit due to a delayed progression (and the magnitude of the delay) would need to be considered in light of any QOL decrement due to the toxicity of the maintenance agent. For example, Ozols (36) questioned whether a seven-month improvement in median PFS (from 21 to 28 months [37]) because of maintenance therapy consisting of 12 cycles (vs 3 cycles) of intravenous paclitaxel given every 28 days was offering clinical benefit to ovarian cancer patients in complete remission. Furthermore, in indolent disease settings with effective salvage treatments (eg, first-line myeloma or follicular lymphoma), patients developing symptoms would be expected to quickly achieve a prolonged symptom-free remission when treated with salvage therapy.

Another example of the potential role of PFS as the primary outcome in maintenance trials is given by the debate on the best strategy for use of rituximab in follicular lymphoma: After addition of rituximab to induction chemotherapy was shown to improve OS, rituximab maintenance was evaluated and approved for high-tumor-burden follicular lymphoma, in large part based on the PRIMA study (38–40). This study randomly assigned patients in complete or partial response after first-line immunochemotherapy to receive either two years of rituximab maintenance or observation. After a six-year follow-up, the study demonstrated for the maintenance vs observation arm six-year PFS rates of 59.2% vs 42.7% (HR = 0.58, 95% CI = 0.48 to 0.69, $P < .0001$) and no difference in six-year OS rates (87.4% vs 88.7%, HR = 1.027, $P = .885$). Considering that the study also demonstrated no difference in QOL between the arms and a modest increase in toxic effects on the maintenance arm, it would appear that the PRIMA results represent evidence against the clinical benefit of rituximab maintenance (41,42) and, more broadly, evidence that PFS may not accurately measure QOL in indolent-disease maintenance settings.

Regardless of symptom prevention, it has been suggested that maintenance therapy that improves PFS without improvement in OS may provide patient QOL benefit by delaying subsequent more toxic lines of therapy (43). To properly demonstrate a QOL benefit using this justification, however, maintenance trials should collect and compare the relevant QOL data (toxicity, symptoms, treatment duration), not just over the maintenance treatment but also over the second-line treatment (and, if possible, over the subsequent treatment lines). This is needed to ensure that the delay in second-line therapy is not offset by QOL differences in the subsequent lines of therapy.

Other Time-to-Event Endpoints

There is often a concern that prolonged administration of an agent will lead to resistant relapse that may reduce ability of the patients to benefit from the same or similar agents in the future (44,45). Because of this, several alternative endpoints to PFS have been introduced that are designed to assess the overall therapeutic effect of a maintenance (early continuous treatment) strategy vs a delayed (treat at progression) strategy.

Time from randomization to progression on the second-line therapy or death from any cause (PFS2) (Figure 3) has been used in myeloma, colorectal, and ovarian cancers and was suggested by the EMA as a possible endpoint (46,47). (Note that
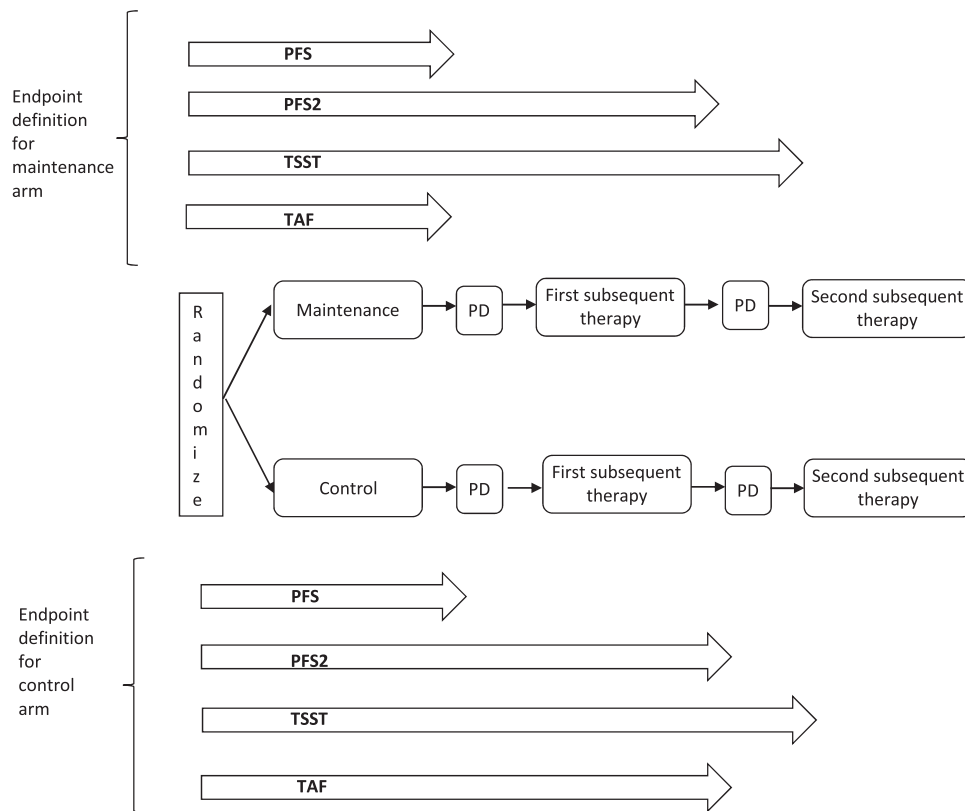
Figure 3. Endpoint definitions: 1) PFS = progression-free survival: time from randomization to disease progression or death whichever occurs first; 2) PFS2 = second progression–free survival: time from randomization to second progression or death whichever occurs first (for patients who do not receive any subsequent treatment either time of death or time of first progression can be used, both approaches should be reported as sensitivity analyses); 3) TSST = time to second subsequent therapy: time from randomization to start of second subsequent therapy or death whichever occurs first (for patients who do not receive any subsequent treatment either time of death or time of first progression can be used, both approaches should be reported as sensitivity analyses); 4) TAF = time to approach failure: for the maintenance arm it is time from randomization to disease progression or death and for the control arm it is time from randomization to second progression or death whichever occurs first (for control-arm patients who do not receive any subsequent treatment either time of death or time of first progression can be used, both approaches should be reported as sensitivity analyses). PD = progressive disease.

even though the baseline time is randomization, in establishing the RECIST documentation for the second progression the first scan that documented the first progression is used as the baseline scan.) By incorporating the treatment effect on both first and second lines of therapy, PFS2 provides a better reflection (than PFS) of the total effect of the maintenance treatment on the patient. Consider metastatic colorectal cancer, (mCRC) where, in order to balance disease control and treatment toxicity (eg, oxaliplatin neurotoxicity), complex therapeutic strategies that include "stop-and-go" and maintenance components are evaluated using PFS2 and other especially developed endpoints (48–51). For example, the CAIRO3 study (52) evaluated the role of maintenance therapy in mCRC patients with response/SD after six cycles of induction with bevacizumab, capecitabine, and oxaliplatin (CAPOX-B) by randomly assigning them between two treatment strategies: 1) continuation maintenance with capecitabine/bevacizumab followed at progression by second-line CAPOX-B or 2) observation followed at progression by second-line CAPOX-B (Figure 4). (Note, however, that only 47% of the maintenance arm patients and 60% of the observation arm patents actually received the protocol-specified second-line therapy.) The primary endpoint was PFS2, with secondary endpoints being PFS and OS. The study results for the maintenance vs observation arms were as follows: 8.5 vs 4.1 months median PFS (hazard ratio [HR] = 0.43, 95% confidence interval [CI] = 0.36 to 0.52, $P < .0001$), 11.7 vs 8.5 months median PFS2 (HR = 0.67, 95%

CI = 0.56 to 0.81, $P < .0001$), and 21.6 vs 18.1 months median OS (HR = 0.89, 95% CI = 0.73 to 1.07, $P = .22$); slightly worse toxicity and QOL profiles were observed on the maintenance arm. Given the OS, QOL, and toxicity outcomes, the implication of these results for clinical practice depends on individual patient preference for treatment breaks vs a small improvement in survival. In that context, the improvement in PFS2 adds confidence in the observed statistically insignificant 3.5-month OS improvement: PFS2 results demonstrate that the initial PFS improvement is sustained in longer-term disease control.

When it is not feasible to ensure consistent follow-up with regular tumor assessments until the time of second progression, time to second subsequent therapy or death (TSST) (Figure 3) is sometimes used to approximate PFS2 (47). For example, an RCT comparing maintenance olaparib vs placebo in relapsed ovarian cancer patients with response/SD after induction chemotherapy reported PFS, TSST, and OS in BRCA-mutated subgroup (53): 11.2 vs 4.3 months median PFS (HR = 0.18, 95% CI = 0.10 to 0.31, $P < .0001$), 23.8 vs 15.2 months median TSST (HR = 0.44, 95% CI = 0.29 to 0.67, $P = .00013$), and 34.9 vs 31.9 months median OS (HR = 0.73, 95% CI = 0.45 to 1.17, $P = .19$). As with PFS2, positive results of a TSST analysis support that the observed statistically nonsignificant difference in OS might be real.

When evaluating maintenance with an agent that is routinely used as a salvage therapy, a potential concern is that this maintenance strategy could reduce the number of therapeutic
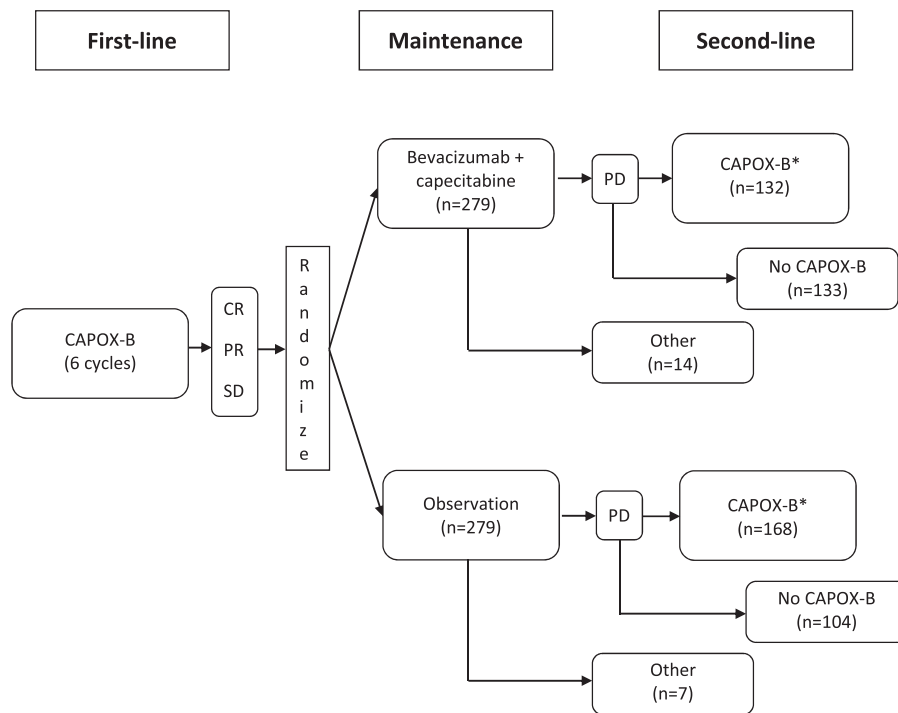
Figure 4. CAIRO3 trial design (52). CAPOX-B = capecitabine+oxaliplatin+bevacizumab. For this study the primary endpoint second progression–free survival (PFS2) was defined as the second progression on CAPOX-B but the first progression for patients who received second-line treatment other than CAPOX-B or no subsequent therapy at all. However, the study also demonstrated that results for the time to second progression on any second-line therapy endpoint were similar to the protocol-defined PFS2 results. **Asterisk** indicates second-line CAPOX-B was administered until progression. CAPOX-B = bevacizumab, capecitabine, and oxaliplatin; CR = complete response; PD = progressive disease; PR = partial response; SD = stable disease; PFS2 = second-progression free survival.

salvage options available for the patient. For example, in a setting with three lines of effective therapy (A, B, and C), using B in maintenance after A may leave the maintenance arm patients with only C as an effective salvage option at progression (while the control arm patients would have B and C available). In this case, a clinically relevant comparison of the ability of the two strategies to control the disease could be based on comparing the time to first progression on the maintenance arm and the time to second progression on the control arm. This endpoint, time to approach failure (TAF) (Figure 3), was described by Rajkumar et al. (18) in the context of comparing an immediate vs a delayed administration strategy in myeloma. In clinical settings where the disease can be controlled with observation and strategic retreatment (at each progression), a comparison of the maintenance and retreatment strategies could be based on comparing the time to first progression on the maintenance arm and the time to the failure to respond to retreatment on the retreatment arm (similarly to the TAF endpoint). For example, in low-burden follicular lymphoma patients responding to induction therapy, the RESORT trial (54) compared a maintenance strategy (a single dose of rituximab every 13 weeks) with a retreatment strategy (observation with no treatment until progression, and then retreatment with four doses of rituximab at each progression). The primary endpoint in this study was time to treatment failure (TTF), with treatment failure defined as no response to rituximab for the retreatment arm and progression for the maintenance arm. The study demonstrated no difference in TTF (three-year failure-free rates 64% and 61%, P = .33 for the maintenance and retreatment arms, respectively) and no difference in QOL. However, there was a considerable difference in PFS (3-year remission rates of 78% and 50% for maintenance

and retreatment arms). In this setting, the TTF endpoint allowed one to quantify the lack of clinical benefit from the maintenance strategy; the study concluded that the retreatment strategy is preferable to maintenance (35).

For evaluating individual patient benefit from a maintenance strategy, the PFS endpoint, by itself, does not accurately capture the relevant therapeutic effect. Endpoints that incorporate outcomes of subsequent lines of therapy (eg, PFS2) allow one to address the concern about the effect of maintenance on the effectiveness of subsequent lines of therapy. Moreover, because these endpoints provide a better (than PFS) quantification of long-term disease control they are likely to provide a better reflection of maintenance effect on OS. Use of these endpoints may require additional logistical considerations and resources associated with longer follow-up. Furthermore, defining these endpoints for patients who never receive subsequent therapy postprogression (after randomization) requires careful consideration. Two possible values in this case are: 1) time of first progression or 2) death—both may introduce bias. Therefore, if the number of these patients is nontrivial, then both approaches should be reported.

Maintenance strategies are focused on a prolonged postinduction administration of agents(s) with established activity in more advanced settings. Ideally, RCTs should be designed to provide a direct assessment of clinical benefit as measured by OS and QOL. When the therapeutic goal of a maintenance strategy is to prolong patient life, the RCT should be sized to detect a clinically meaningful improvement in OS, if such a study can be completed in a timely manner (eg, in the mCRC or metastatic ovarian cancer settings). In indolent disease settings, where timely completion of an OS-targeted study is not feasible,

endpoints incorporating outcomes of the subsequent lines of therapy could be considered to approximate the maintenance effect on OS. When the main therapeutic goal of maintenance is to improve QOL, then the RCT should be designed to provide direct evidence of the net gain in overall QOL from the maintenance strategy.

Finally, as the cost of cancer care has become one of the fastest growing components of US health care spending (55), the implications of maintenance strategy for the sustainability of the public health system cannot be ignored. Therefore, it is important for maintenance RCTs to provide society with an objective estimate of the clinical benefit of maintenance so its cost-effectiveness can be accurately quantified. In particular, the use of endpoints other than OS and QOL should be justified.

Alternative Use of a PFS Endpoint

Phase II/III designs could provide an efficient approach for using an intermediate endpoint to decide adaptively whether to continue accrual to a phase III evaluation (56). This approach can be used in the maintenance setting, where the phase II endpoint can be PFS (looking for activity) and the phase III endpoint can be OS (looking for definitive clinical benefit). For example, in advanced NSCLC, B-24 (57) evaluated cediranib using an induction/maintenance design. The study employed a phase II/III design that required a PFS hazard ratio of less than 0.77 in phase II to continue on to the phase III OS evaluation.

Conclusions

Unlike the regulatory setting, where the question is whether a new drug has enough activity to justify making it available to patients with a given disease, a maintenance question is typically focused on comparing an immediate treatment strategy with a delayed (treat at progression) strategy that often incorporates an active salvage agent. In this context, the relevant therapeutic question, whether a patient presenting with a given disease would derive maximum benefit by using the agent early (in maintenance) vs later at the time of progression, is not fully captured by a PFS endpoint. Therefore, in order to provide definitive evidence to inform clinical practice, RCTs evaluating maintenance should, if possible, use either OS as the endpoint or endpoints directly measuring QOL. In settings where this is not feasible, PFS2 (or similar endpoints that accurately reflect the long-term therapeutic effect for the clinical situation at hand) should be used. Furthermore, to accurately isolate the clinical impact of maintenance, RCT designs should incorporate appropriate first-line and second-line therapies and minimize between-arm differences in follow-up schedules.

References

1. Gerber DE, Schiller JH. Maintenance chemotherapy for advanced non-small-cell lung cancer: new life for an old idea. *J Clin Oncol.* 2013;31(8):1009–1020.
2. Scagliotti G, Novello S, von Pawel J, et al. Phase III study of carboplatin and paclitaxel alone or with sorafenib in advanced non-small-cell lung cancer. *J Clin Oncol.* 2010;28(11):1835–1842.
3. Burger RA, Brady MF, Bookman MA, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. *N Engl J Med.* 2011;365(26):2473–2483.
4. Habermann TM, Weller EA, Morrison VA, et al. Rituximab-CHOP versus CHOP alone or with maintenance rituximab in older patients with diffuse large B-cell lymphoma. *J Clin Oncol.* 2006;24(19):3121–3127.
5. Mateos MV, Oriol A, Martínez-López J, et al. Bortezomib, melphalan, and prednisone versus bortezomib, thalidomide, and prednisone as induction therapy followed by maintenance treatment with bortezomib and thalidomide versus bortezomib and prednisone in elderly patients withuntreated multiple myeloma: a randomised trial. *Lancet Oncol.* 2010;11(10):934–941.
6. Tummarello D, Mari D, Graziano F, et al. A randomized, controlled phase III study of cyclophosphamide, doxorubicin, and vincristine with etoposide (CAV-E) or teniposide (CAV-T), followed by recombinant interferon-alpha maintenance therapy or observation, in small cell lung carcinoma patients with complete responses. *Cancer.* 1997;80(12):2222–2229.
7. Lunceford JK, Davidian M, Tsiatis AA. Estimation of survival distributions of treatment policies in two-stage random assignment designs in clinical trials. *Biometrics.* 2002;58(1):48–57.
8. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Stat Med.* 2005;24(10):1455–1481.
9. Thall PF, Logothetis C, Pagliaro LC, et al.Adaptive therapy for androgen-independent prostate cancer: a randomized selection trial of four regimens. *J Natl Cancer Inst.* 2007;99(21):1613–1622.
10. Pettengell R, Schmitz N, Gisselbrecht C, et al. Rituximab purging and/or maintenance in patients undergoing autologous transplantation for relapsed follicular lymphoma: a prospective randomized trial from the lymphoma working party of the European group for blood and marrow transplantation. *J Clin Oncol.* 2013;31(13):1624–1630.
11. Durrleman S, Simon R. When to randomize? *J Clin Oncol.* 1991;9(1):116–122.
12. Cohen MH, Johnson JR, Chattopadhyay S, et al. Approval summary: erlotinib maintenance therapy of advanced/metastatic non-small cell lung cancer (NSCLC). *Oncologist.* 2010;15(12):1344–1351.
13. Cappuzzo F, Ciuleanu T, Stelmakh L, et al. Erlotinib as maintenance treatment in advanced non-small-cell lung cancer: a multicentre, randomised, placebo-controlled phase 3 study. *Lancet Oncol.* 2010;11(6):521–529.
14. FDA Oncology Drug Advisory Committee Meeting Transcript Dec 2009. http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/UCM197771.pdf. Accessed August 7, 2015.
15. Escudier B, Heng DY, Smyth-Medina A, et al. Considerations for the design of future clinical trials in metastatic renal cell carcinoma. *Clin Genitourin Cancer.* 2014;12(1):1–12.
16. Korn EL, Freidlin B, Abrams JS. Overall survival as the outcome for randomized clinical trials with effective subsequent therapies. *J Clin Oncol.* 2011;29(17):2439–2442.
17. Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? *J Clin Oncol.* 2012;30(10):1030–1033.
18. Rajkumar SV, Gahrton G, Bergsagel PL. Approach to the treatment of multiple myeloma: a clash of philosophies. *Blood.* 2011;118(12):3205–3211.
19. Panageas KS, Ben-Porat L, Dickler MN, et al. When you look matters: the effect of assessment schedule on progression-free survival. *J Natl Cancer Inst.* 2007;99(6):428–432.
20. Freidlin B, Korn EL, Hunsberger S, et al. Proposal for the use of progression-free survival in unblinded randomized trials. *J Clin Oncol.* 2007;25(15):2122–2126.
21. Sridhara R, Mandrekar SJ, Dodd LE. Missing data and measurement variability in assessing progression-free survival endpoint in randomized clinical trials. *Clin Cancer Res.* 2013;19(10):2613–2620.
22. Fidias PM, Dakhil SR, Lyss AP, et al. Phase III study of immediate compared with delayed docetaxel after front-line therapy with gemcitabine plus carboplatin in advanced non-small-cell lung cancer. *J Clin Oncol.* 2009;27(4):591–598.
23. Fidias P, Novello S. Strategies for prolonged therapy in patients with advanced non-small-cell lung cancer. *J Clin Oncol.* 2010;28(34):5116–5123.
24. Gridelli C, de Marinis F, Di Maio M, et al. Maintenance treatment of advanced non-small-cell lung cancer: results of an international expert panel meeting of the Italian association of thoracic oncology. *Lung Cancer.* 2012;76(3):269–279.
25. Fallowfield LJ, Fleissig A. The value of progression-free survival to patients with advanced-stage cancer. *Nat Rev Clin Oncol.* 2011;9(1):41–47.
26. Cella D. Bevacizumab and quality of life in advanced cervical cancer. *Lancet Oncol.* 2015;16(3):241–243.
27. Silvestri G, Pritchard R, Welch HG. Preferences for chemotherapy in patients with advanced non-small cell lung cancer: descriptive study based on scripted interviews. *BMJ.* 1998;317(7161):771–775.
28. Secord AA, Coleman RL, Havrilesky LJ, et al. Patient-reported outcomes as end points and outcome indicators in solid tumours. *Nat Rev Clin Oncol.* 2015;12(6):358–370.
29. Sargent DJ, Hayes DF. Assessing the measure of a new drug: is survival the only thing that matters? *J Clin Oncol.* 2008;26(12):1922–1923.
30. Begg CB. Justifying the choice of endpoints for clinical trials. *J Natl Cancer Inst.* 2013;105(21):1594–1595.
31. Korn EL, Albert PS, McShane LM. Assessing surrogates as trial endpoints using mixed models. *Stat Med.* 2005;24(2):163–182.
32. Blumenthal GM, Karuri SW, Zhang H, et al. Overall Response Rate, Progression-Free Survival, and Overall Survival With Targeted and Standard Therapies in Advanced Non-Small-Cell Lung Cancer: US Food and Drug Administration Trial-Level and Patient-Level Analyses. *J Clin Oncol.* 2015;33(9):1008–1014.
33. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst.* 2000;92(3):205–216.
34. Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol.* 2009;27(17):2874–2880.

35. Friedberg JW. End of rituximab maintenance for low-tumor burden follicular lymphoma. *J Clin Oncol*. 2014;32(28):3093–3095.
36. Ozols RF. Maintenance therapy in advanced ovarian cancer: progression-free survival and clinical benefit. *J Clin Oncol*. 2003;21(13):2451–2453.
37. Markman M, Liu PY, Wilczynski S, et al. Phase III randomized trial of 12 versus 3 months of maintenance paclitaxel in patients with advanced ovarian cancer after complete response to platinum and paclitaxel-based chemotherapy: a Southwest Oncology Group and Gynecologic Oncology Group trial. *J Clin Oncol*. 2003;21(13):2460–2465.
38. Salles G, Seymour JF, Offner F, et al. Rituximab maintenance for 2 years in patients with high tumor burden follicular lymphoma responding to rituximab plus chemotherapy (PRIMA): a phase 3, randomized controlled trial. *Lancet*. 2011;377(9759):42–51.
39. U.S. Food and Drug Administration, Rituximab 2011. http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm241928.htm. Accessed August 7, 2015.
40. Salles GA, Seymour JF, Feugier P, et al. Updated 6 year follow-up of the PRIMA study confirms the benefit of 2-year rituximab maintenance in follicular lymphoma patients responding to frontline immunochemotherapy. 2013 ASH Annual Meeting. Abstract 509.
41. Friedberg JW. Rituximab maintenance in follicular lymphoma: PRIMA. *Lancet*. 2011;377(9759):4–6.
42. Haines I. Rituximab maintenance therapy for follicular lymphoma. *Lancet*. 2011;377(9772):1151.
43. Goldberg P. ODAC clarifies standards for maintenance in ovarian cancer. *Cancer Letter*. 2014;40(26):1–5.
44. Ludwig H, Sonneveld P, Davies F, et al. European perspective on multiple myeloma treatment strategies in 2014. *Oncologist*. 2014;19(8):829–844.
45. Gainor JF, Shaw AT. Emerging paradigms in the development of resistance to tyrosine kinase inhibitors in lung cancer. *J Clin Oncol*. 2013;31(31):3987–3996.
46. Mbanya Z, Shkun C. Time to second objective disease progression (PFS2): an emerging clinical trial endpoint with regulatory and reimbursement implications. *Blood*. 2014;124(21):6005.
47. European Medicines Agency. Guideline on the evaluation of anticancer medicinal products in man. 2013. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/01/WC500137128.pdf. Accessed August 7, 2015.
48. Chibaudel B, Tournigand C, Bonnetain F, et al. Therapeutic strategy in unresectable metastatic colorectal cancer: an updated review. *Ther Adv Med Oncol*. 2015;7(3):153–169.
49. de Gramont A Re-challenge and the concept of lines of therapy in metastatic colorectal cancer. *Eur J Cancer*. 2011;47(Suppl 3):S76–S84.
50. Allegra C, Blanke C, Buyse M, et al. End points in advanced colon cancer clinical trials: a review and proposal. *J Clin Oncol*. 2007;25(24):3572–3575.
51. Chibaudel B, Bonnetain F, Shi Q, et al. Alternative end points to evaluate a therapeutic strategy in advanced colorectal cancer: evaluation of progression-free survival, duration of disease control, and time to failure of strategy--an Aide et Recherche en Cancerologie Digestive Group Study. *J Clin Oncol*. 2011;29(31):4199–4204.
52. Simkens LH, van Tinteren H, May A, et al. Maintenance treatment with capecitabine and bevacizumab in metastatic colorectal cancer (CAIRO3): a phase 3 randomised controlled trial of the Dutch Colorectal Cancer Group. *Lancet*. 2015;385(9980):1843–1852.
53. Ledermann J, Harter P, Gourley C, et al. Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a pre-planned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. *Lancet Oncol*. 2014;15(8):852–861.
54. Kahl BS, Hong F, Williams ME, et al. Rituximab extended schedule or re-treatment trial for low-tumor burden follicular lymphoma: eastern cooperative oncology group protocol e4402. *J Clin Oncol*. 2014;32(28):3096–3102.
55. Schnipper LE, Davidson NE, Wollins DS, et al. American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options. *J Clin Oncol*. 2015; 33(23):2563–2577.
56. Korn EL, Freidlin B, Abrams JS, et al. Design issues in randomized phase II/III trials. *J Clin Oncol*. 2012;30(6):667–671.
57. Goss GD, Arnold A, Shepherd FA, et al. Randomized, double-blind trial of carboplatin and paclitaxel with either daily oral cediranib or placebo in advanced non-small-cell lung cancer: NCIC clinical trials group BR24 study. *J Clin Oncol*. 2010;28(1):49–55.

COMMENTARY