COMMENTARY

Adaptive Clinical Trials: Advantages and Disadvantages of Various Adaptive Design Elements

Edward L. Korn, Boris Freidlin

Affiliation of authors: Biometric Research Program, National Cancer Institute, Bethesda, MD.

Correspondence to: Edward L. Korn, PhD, Biometric Research Program, MSC 9735, National Cancer Institute, Bethesda, MD 20892 (e-mail: korne@ctep.nci.nih.gov).

Abstract

There is a wide range of adaptive elements of clinical trial design (some old and some new), with differing advantages and disadvantages. Classical interim monitoring, which adapts the design based on early evidence of superiority or futility of a treatment arm, has long been known to be extremely useful. A more recent application of interim monitoring is in the use of phase II/III designs, which can be very effective (especially in the setting of multiple experimental treatments and a reliable intermediate end point) but do have the cost of having to commit earlier to the phase III question than if separate phase II and phase III trials were performed. Outcome-adaptive randomization is an older technique that has recently regained attention; it increases trial complexity and duration without offering substantial benefits to the patients in the trial. The use of adaptive trials with biomarkers is new and has great potential for efficiently identifying patients who will be helped most by specific treatments. Master protocols in which trial arms and treatment questions are added to an ongoing trial can be especially efficient in the biomarker setting, where patients are screened for entry into different subtrials based on evolving knowledge about targeted therapies. A discussion of three recent adaptive clinical trials (BATTLE-2, I-SPY 2, and FOCUS4) highlights the issues.

We frequently hear claims that adaptive clinical trial designs should be used because these novel designs can evaluate treatments faster with fewer patients. In this Commentary, we examine this claim. In particular, what are the benefits and costs of various adaptive design elements in terms of improving the efficiency of drug development?

The US Food and Drug Administration (FDA) guidance on adaptive clinical trials (1) defined them as “a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study.” We focus here on adaptive elements that have been used in cancer treatment trials: well-established interim monitoring that allows stopping trial arms early based on accruing outcome data (which will be briefly reviewed); adaptive trials with biomarkers that allow adjusting the study population; master protocols that allow adding treatment arms or patient subgroups during the trial; and outcome-adaptive randomization in which treatment assignment probabilities are changed

during the trial according to which treatment arm is doing better. We conclude with a more detailed description of three recent adaptive clinical trials with biomarkers: BATTLE-2 (2), I-SPY 2 (3), and FOCUS4 (4).

Interim Monitoring

Interim monitoring of accruing outcome data, reviewed by an independent data monitoring committee (5,6), is motivated by both ethical and resource considerations as it allows one to stop a trial or treatment as soon as the scientific question addressed by the trial or related to that treatment is answered.

Stopping for Superiority or Futility

Group sequential designs (7–13), in which the clinical outcome data is repeatedly assessed over time, are a well-accepted technique to allow stopping a trial (or some of its treatment arms) as soon as the relevant clinical question has been answered.

With electronic data capture, there is the possibility of increasing the number of interim analysis looks. Once interim monitoring commences, there is very little cost in having frequent monitoring from then on (14), so one can analyze the data more often and stop earlier (when appropriate). Potential drawbacks that should be considered when designing an interim monitoring plan are that if a trial stops early, then there may be very little or no information about the longer-term effects of the treatments or the effects of treatments on secondary end points, and the treatment effects will be estimated less precisely than if the trial was not stopped early.

Phase II/III Trial Designs

The traditional path for clinical development of new cancer therapies involves screening for preliminary evidence of activity in phase II studies followed by a definitive evaluation of the promising therapies in confirmatory phase III trials. A phase II/III trial is designed as a phase III trial but with an interim “phase II look” to assess whether the experimental treatment is active enough to continue the trial to its phase III sample size (15). The advantage of a phase II/III trial over a separate phase II and phase III trial is speed, in that the phase II patients can be included in the phase III analysis and one does not have to wait for the phase III protocol development after the phase II results become available. Phase II/III trials are most useful when the phase II end point (eg, response rates or progression-free survival) can be obtained earlier than the definitive phase III end point (eg, overall survival) while still being able to screen out inactive treatments. What the phase II end point should be, how high the bar should be set for that end point, and whether there should be an accrual suspension while waiting for the phase II data to mature are the key design considerations for phase II/III trials (16).

A potential disadvantage of a phase II/III trial is that one is eliminating the flexibility of modifying the phase III trial design based on the results of the phase II trial (1,17). In addition, committing to the question to be asked in phase III at an earlier time locks in a sponsoring organization that may want to keep its options for future trials open longer.

Multi-Arm Trials

A trial with multiple experimental arms and a single control arm can be an efficient way to test multiple treatments in a single disease setting (18). Superiority and futility monitoring comparing each experimental arm to the control arm can be used to drop experimental arms or stop the trial. Some designs also allow the possibility of dropping experimental arms at an interim analysis when there is a more promising experimental arm in the trial (15,19).

Adaptive Trials With Biomarkers

The use of biomarkers related to the patient’s tumor offers the possibility of choosing a treatment that is most likely to work for that patient. A range of clinical trial designs is available for development of such targeted treatments and their associated biomarkers, even with a single biomarker. (Here we focus on the study design; for a review of methods for development and validation of diagnostic assays, see reference 20.) The choice of the design should depend on the preexisting evidence that the treatment benefit is restricted to the biomarker-positive

individuals (21). For example, if the evidence were very strong, then one would conduct the trial with eligibility restricted to biomarker-positive patients (enrichment design). At the other extreme, if little is known about the biomarker (or it has not even been developed yet), one can sometimes perform an analysis of its predictive ability on specimens from a previously conducted clinical trial (22) or plan on developing and validating the biomarker using the data from a proposed trial (23). Some phase II studies (eg, BATTLE-2 and I-SPY 2, discussed below) are explicitly designed to develop predictive biomarkers for subsequent validation.

The commonly occurring middle ground is when there is strong evidence that the treatment is more likely to be effective in the biomarker-positive than the biomarker-negative subgroup (if it works at all), but the evidence is not compelling enough to rule out a meaningful benefit in the biomarker-negative subgroup. In this case, a biomarker-stratified design would be used in which all patients are enrolled with plans to assess the treatment effect in both the biomarker-positive and biomarker-negative patient subsets (24,25).

The use of interim monitoring in the context of biomarker/agent development can be particularly effective as it allows one to adapt the trial to the most promising patient subpopulation. For example, futility monitoring of a biomarker-negative subgroup could suggest that the eligibility for a phase III trial be changed to only biomarker-positive patients when it appears that the new treatment is ineffective in the biomarker-negative subgroup. Alternatively, in a trial that initially limits eligibility to biomarker-positive patients, if interim results are promising, then the study eligibility can be expanded to include patients with a wider range of biomarker values. This type of analysis can also be incorporated into a phase II or II/III trial, where the further testing of the agent may be limited to biomarker-positive patients or expanded to all patients (26).

Master Protocols

As mentioned above, a multi-arm trial design with a fixed number of treatment arms in a specific disease setting can use interim monitoring to make decisions about whether to discontinue arms early for efficacy or futility. The master protocol approach (27) allows adding new treatment arms (in existing or new patient subgroups) to an ongoing trial in which several treatments are already being tested; a trial using a master protocol is sometimes known as a “platform” trial.

One type of master protocol is used to develop multiple treatments in a given disease setting. For example, the STAMPEDE trial (28) evaluates various agents (added to a standard hormone therapy backbone) for advanced prostate cancer by dropping treatment arms when lack of sufficient activity is demonstrated and adding new treatment arms when new promising treatments become available. The advantage of this approach over starting new trials for additional treatments is that it allows the use of the existing multi-arm trial, leading to a quicker start to testing newly available treatments. However, the staggered entry of new experimental arms reduces the statistical efficiency of the multi-arm design as the patients randomly assigned to these new arms can only be compared with the contemporaneously randomized control-arm patients. Moreover, the challenges of conducting a multi-arm trial with a fixed number of treatment arms are present (and may even be exacerbated) when new treatment arms are added in a master protocol: convincing multiple industry partners to participate,

how to share information between the industry partners, funding of the trial, and extra regulatory complexities (27).

With tumor biomarkers, there are additional benefits of a master protocol (29). In an umbrella trial, patients with one tumor type are enrolled in different treatment arms depending on molecular characteristics of their tumor. An example of an umbrella trial is Lung-MAP (30) for patients with advanced squamous cell lung cancer. As new targeted agents are developed for this disease setting, treatment arms (and possibly control arms) can be added to these trials for patients with tumors with the appropriate molecular targets. In a basket trial, patients are enrolled in various treatment arms based on the molecular characterization of their tumor regardless of its histology. As new actionable targets with their associated drugs are discovered, they are added to the trial. An example of a basket trial is the National Cancer Institute Molecular Analysis for Therapy Choice (NCI-MATCH) trial (31). A key part of umbrella and basket trials is the screening component, which is used to direct patients to the treatments from which they are most likely to benefit. Note that both umbrella and basket designs allow adding single experimental treatment arms, or experimental arms with corresponding control arms for randomized comparisons.

Master protocols with biomarker-defined treatments are a new paradigm and are highly efficient as many of the biomarker-defined subgroups may represent small proportions of the patient population, which can be captured by the screening component of the trial. This is a great advantage to patients who, by having their tumor screened once, can potentially find a trial and treatment relevant to their tumor's molecular characteristics. The adaptive elements of stopping treatment arms based on unfavorable or very favorable interim monitoring results and adding treatment arms when they become relevant further increase the efficiency of the master protocol trial design. It should be noted that the evidentiary requirements for adding a new treatment to a master protocol, which may be different for early phase and definitive studies, should be clearly defined. This will help to avoid the potential pressure to add new treatment arms with weak credentials to keep the master protocol active.

Outcome-Adaptive Randomization

With outcome-adaptive randomization, the accruing outcome data of an ongoing trial is used to adjust the randomization ratio so that a higher proportion of patients are randomly assigned to the treatment arm(s) that appear to be doing better. Although outcome-adaptive randomization continues to be widely promoted as a novel design approach, it is not a new concept; play-the-winner treatment assignments were first proposed in 1969 (32,33). (It has been reported to have been used in 44 trials conducted by the MD Anderson Cancer Center as of August 2011 (34).) Outcome-adaptive randomization will generally assign a higher proportion of patients to treatment arms that are more effective (if there are any). However, it has a number of drawbacks that raise questions about whether it is useful or appropriate (35). The first is that any time trends in the prognostic characteristics of the patient population enrolling in the trial will bias the results of the trial. For example, if in the earlier part of the trial, patients are randomly assigned equally to the experimental and control treatment arms, but are randomly assigned 9:1 in favor of the experimental arm later in the trial, then an improving prognostic pool of patients being randomly assigned in the trial will translate into a bias in favor of the

experimental arm. (Special statistical methods (35,36) can overcome this bias, but they result in substantial power loss.) Because of this potential bias, outcome-adaptive randomization is inappropriate for long-term definitive phase III trials. What about earlier phase randomized trials, which may be finished more quickly than a phase III trial (leading to less potential bias) and where error rates are sometimes relaxed? Even here, outcome-adaptive randomization can still lead to problems, as demonstrated by the investigators in two phase II trials (37,38) acknowledging the interpretation limitations of their results because of the outcome-adaptive randomization. It should be noted that this bias problem with outcome-adaptive randomization has long been known (39).

A second problem with outcome-adaptive randomization is its statistical inefficiency due to having an unequal number of patients on the treatment arms. For example, a trial with 50 patients on each arm will provide a more precise estimate of the treatment effect than a trial with 90 patients on the experimental arm and 10 patients on the control arm. This means that to get the same amount of information about the treatment effect, trials using outcome-adaptive randomization will have to be larger (and take longer) than trials with equal randomization. Outcome-adaptive randomization will therefore delay getting new effective treatments to the clinical community and will also expose more patients to ineffective treatments in clinical trials that use it. For example, we estimated that in the (first) BATTLE trial (40) the outcome-adaptive randomization led to a trial that was 74% larger, with potentially 65% more patients for whom treatment failed (progressive disease) than if a fixed-randomization trial with interim monitoring had been used (41).

A third problem with outcome-adaptive randomization is that even though it will generally put more patients on the better treatment arm, it will occasionally put a moderately larger proportion of patients on the worse treatment arm (42). This cannot happen if equal randomization is used with, as is typical, any sort of block randomization. Finally, we note that there have been ethical issues raised concerning using outcome-adaptive randomization, although this is a controversial subject (41,43–49).

BATTLE-2

The BATTLE-2 phase II trial (2) randomized the treatment for patients with advanced non-small cell lung cancer among four treatment arms: erlotinib (arm 1, control), erlotinib+MK-2206 (arm 2), MK-2206+AZD6244 (arm 3), or sorafenib (arm 4). The primary outcome was eight-week disease control rate (DCR). Patients were ineligible if they had epidermal growth factor receptor (EGFR)-sensitizing mutations or ALK gene fusions, and patients who had received prior erlotinib (estimated in the protocol to be 40% of the patients) were randomly assigned only to treatment arms 2–4. The primary analysis specified was a comparison of each of the experimental arm DCRs with the control DCR. The trial was planned with two stages, each with 200 patients. In the first stage, in addition to the between-arm DCR efficacy comparisons, an aim was to identify predictive biomarkers that could be used to guide patient assignments in the second stage. No clear predictive biomarkers were found after the analysis of the stage 1 results, and the second stage of the trial was not started; we discuss only the stage 1 design here.

Equal randomization was used for the first 70 patients, and then outcome-adaptive randomization (based on accruing outcome information adjusted for KRAS-mutation and EGFR

resistance status) was used for the remaining 130 patients. The DCRs for the 186 evaluable patients for the four arms were 32% (6/19), 50% (18/36), 53% (37/70), and 46% (28/61), respectively. Based on these results and a lack of the ability to find predictive biomarkers, the investigators concluded that better biomarker-driven treatments are needed for this patient population (2).

In evaluating the statistical properties of the stage 1 design, note the imbalance in the numbers of patients treated in the four treatment arms (19, 36, 70, and 61). As mentioned earlier, this leads to the outcome-adaptive randomization being inefficient. To assess the magnitude of the inefficiency for this trial, we calculate that the observed numbers of patients treated would have 80% power (with one-sided 10% type I error) to detect treatment differences of 64% vs 30% (arm 2 vs arm 1), 63% vs 30% (arm 3 vs arm 1), and 60% vs 30% (arm 4 vs arm 1). Instead, one could randomly assign 120 patients (approximately 30 in each arm) and have the same 80% power that the 200-patient outcome-adaptive randomization design had for detecting a 60% vs 30% difference in DCRs. (To achieve overall equal arm allocation, one could use a 1:1:1 fixed randomization for the patients with prior erlotinib and a 2:1:1:1 randomization for the patients without prior erlotinib.)

Although the outcome-adaptive randomization leads to a larger and longer trial than fixed randomization, what about its potential benefit for patients in the trial? Indeed, the observed overall DCR was 48% and is slightly higher than the 45% we estimate would be observed if there were equal numbers of patients in each treatment arm. However, the absolute numbers tell a somewhat different story concerning patients with bad outcomes (progressive disease): With the outcome-adaptive randomization, there were 97 patients for whom the treatment failed (out of 186 evaluable patients treated), while with fixed randomization we estimate there would be 66 patients for whom the treatment would fail (out of an estimated 120 that would be treated).

Finally, we note that the futility monitoring in BATTLE-2 was specified (in the protocol) to be quite conservative, apparently stopping the trial only if all three experimental treatment arms look futile as compared with the control arm in all subgroups defined by the biomarkers. One might argue that, because of the outcome-adaptive randomization, if only one experimental treatment arm were doing very poorly, then the probability of being randomized to that arm would become low. However, this would be little comfort to the patients who were randomized to that very poor treatment. Presumably, the independent data and safety monitoring board would step in at some point (if this had been relevant for the accruing data in this trial), but it is preferable to have reasonable futility guidelines as part of a trial design.

I-SPY 2

The I-SPY trial (3) is a randomized phase II using a master protocol in which experimental agents are tested against control treatments in a neoadjuvant setting. The primary outcome is pathological complete response (pCR). Outcome-adaptive randomization was utilized within each of eight randomization subgroups (defined by hormone receptor status, human epidermal growth factor receptor 2 [HER2] status, and high-risk category 1 vs 2 on the 70-gene MammaPrint assay). The primary analyses were to assess the efficacy of the experimental agent within each of 10 (overlapping) biomarker-defined groups ("signatures"), which were composed of differing combinations of

the eight randomization subgroups. We examine here the assessment of standard neoadjuvant chemotherapy plus neratinib (a tyrosine kinase inhibitor of HER2 and EGFR) vs standard neoadjuvant chemotherapy alone (control arm); Among the 347 patients who were randomized, 127 were assigned to a neratinib-containing arm and 84 to a control arm (50). The investigators concluded from the trial that neratinib was highly likely to be beneficial in the signature subgroup of patients who were HER2 positive/hormone receptor negative. Their results also suggest that the neratinib is beneficial for all patients with HER2-positive tumors and not for patients with HER2-negative tumors.

It would appear that any inefficiency due to the outcome-adaptive randomization in this trial would be minor as the ratio of patients treated in the experimental and control arms is close to 1:1. However, this may not be correct as the inefficiency would depend on the imbalance in patients treated in each of the 10 signature subgroups being analyzed; the imbalances could be large and in different directions for some of the subgroups, averaging out to be not that different from 1:1. Additionally, because the outcome-adaptive randomization for this trial was based on patient characteristics as well as the accruing pCR outcome data, there can be prognostic imbalances between the treatment arms. For example, 57% of the patients in the neratinib arm were HER2 positive, but only 28% of the patients in the control arm were ($P < .0002$). This means that any analysis will have to be stratified by the randomization subgroups. Presumably, the statistical modeling used by the investigators incorporates this and any other relevant stratifications, but the conclusions drawn from the trial are only reliable as far as the modeling is correct. Regrettably, the investigators intentionally did not to present the pCR rates for the patients by the treatment arms, instead presenting only their modeling results. Without presentation of the trial data used for the analysis, one cannot evaluate the robustness of the statistical modeling or the investigators' conclusions for the trial.

A strength of this trial is the evaluation of the treatments within biomarker-defined subgroups. This can be difficult in general because of the limited sample sizes within subgroups. It is not clear from the trial report how many patients were in the subgroup with the reported positive finding. In addition, it is not clear what, if any, multiple comparisons adjustment was made for the evaluation of a treatment effect in a nontrivial number of subgroups. Although some statistical philosophies do not believe in control for multiple comparisons, when it comes to the development of new treatments, it is useful to avoid developing therapies that do not work (51).

Finally, as with BATTLE-2, a very conservative futility rule for stopping the trial was specified in the protocol (low probability of success in all 10 signature groups). As it happens, randomization was stopped to two of the eight randomization subgroups by trial's end (the HER2-negative/MammaPrint-category-1 subgroups), suggesting some additional (implicit) futility monitoring was in place. Because no data are given in the trial report, it is impossible to evaluate how well this futility monitoring protected patients from receiving inferior therapies.

FOCUS4

The FOCUS4 trial (4) uses an umbrella design for patients with advanced/metastatic colorectal cancer with stable or responding disease after firstline chemotherapy. Patients are potentially assigned to one of five substudies for consent and

randomization to a targeted agent (vs a control treatment) based on a biomarker categorization of their tumors: BRAF mutant tumors, PIK3CA mutant tumors, KRAS/NRAS mutant tumors, EGFR-dependent tumors, and a nonstratified category. Each substudy uses a (fixed ratio) randomized phase II/III design with progression-free survival as the end point for the phase II evaluations and progression-free survival and overall survival for successive phase III evaluations. At present (www.focus4trial.org), the BRAF and KRAS/NRAS substudies are in development, the PIK3CA substudy (aspirin vs placebo) and the nonstratified category (capecitabine vs active monitoring) are accruing, and the EGFR-dependent substudy (AZD8931 vs placebo) closed after randomly assigning 32 patients because of lack of benefit seen at its preplanned interim analysis (52).

FOCUS4 allows for new substudies to be added when 1) there is information about a new biomarker target with a potentially active associated drug, 2) an agent has shown sufficient activity in its biomarker-defined substudy that it warrants being tested more broadly, and 3) new information (from the trial data or externally) about an existing biomarker-defined subgroup suggests the biomarker categorization should be modified (4). FOCUS4 uses efficient adaptive design elements in a transparent manner for definitively evaluating agents that may be effective only within biomarker-defined subgroups.

Conclusions

Interim monitoring of outcome data to make decisions about closing treatment arms, in its many forms, is an extremely useful adaptive element of clinical trial design. It accelerates public dissemination of important study results and protects patients on trials from ineffective treatments. The increased use of real-time electronic data entry, processing, and analysis should allow for more frequent interim analyses, leading to quicker decisions. Adding treatment arms to an ongoing master protocol is not a minor undertaking (nor is conducting a multi-arm trial simple to begin with), but it is a highly efficient way to proceed when patients are screened into different substudies of the master protocol based on their tumor characteristics. The suboptimal properties of outcome-adaptive randomization have long been known, so it is unfortunate that it is still being used; the outcome-adaptive randomization subjects the trial results to lack of interpretability because of possible time trends in the data or questions about the robustness of the modeling (if used). In addition, adaptive methods that require complex statistical modeling that is neither transparent nor reproducible should be avoided. As the advantages and disadvantages of adaptive design elements may vary considerably over different clinical settings, it is important that their use in a particular application be clearly justified. Finally, reports of trials should provide adequate and transparent presentation of the study design and results to optimize their utility to the clinical community.

References

1. US Food and Drug Administration. *Draft Guidance for Industry—Adaptive Design Clinical Trials for Drugs and Biologics*. Rockville, MD: U.S. Department of Health and Human Services; 2010.
2. Papadimitrakopoulou V, Lee JJ, Wistuba II, et al. The BATTLE-2 study: A biomarker-integrated targeted therapy study in previously treated patients with advanced non-small-cell lung cancer. *J Clin Oncol*. 2016;34(30):3638–3647.
3. Barker AD, Sigman CC, Kelloff, et al. I-SPY 2: An adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharm Ther*. 2009;86:97–100.
4. Kaplan R, Maughan T, Crook A, et al. Evaluating many treatments and biomarkers in oncology: A new design. *J Clin Oncol*. 2013;31:4562–4568.
5. Ellenberg SS, Fleming TR, DeMets DL. *Data Monitoring in Clinical Trials*. Chichester, UK: Wiley; 2002.
6. Herson J. Coordinating data monitoring committees and adaptive clinical trial designs. *Drug Inf J*. 2008;42:297–301.
7. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika*. 1977;64:191–199.
8. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35:549–556.
9. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983;70:659–663.
10. Ellenberg SS, Eisenberger MA. An efficient design for phase III studies of combination chemotherapies. *Cancer Treat Rep*. 1985;69:1147–1152.
11. Wieand S, Schroeder G, O'Fallon JR. Stopping when the experimental regimen does not appear to help. *Stat Med*. 1994;13:1453–1458.
12. Freidlin B, Korn EL, Gray R. A general inefficacy interim monitoring rule for randomized clinical trials. *Clin Trials*. 2010;7:197–208.
13. Zhang Q, Freidlin B, Korn EL, Halabi S, Mandrekas S, Dignam J. Comparison of futility monitoring guidelines using completed phase III oncology trials. *Clin Trials*. 2016; in press.
14. Freidlin B, Korn EL, George SL. Data monitoring committees and interim monitoring guidelines. *Control Clin Trials*. 1999;20:395–407.
15. Bretz F, Schmidli H, König F, Racine A, Maurer W. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: General concepts. *Biometrical J*. 2006;48:623–634.
16. Korn EL, Freidlin B, Abrams JS, Halabi S. Design issues in randomized phase II/III trials. *J Clin Oncol*. 2012;30:667–671.
17. Cuffe RL, Lawrence D, Stone A, Vandemeulebroecke M. When is a seamless study desirable? Case studies from different pharmaceutical sponsors. *Pharm Stat*. 2014;13:229–237.
18. Freidlin B, Korn EL, Gray R, Martin A. Multi-arm clinical trials of new agents: Some design considerations. *Clin Cancer Res*. 2008;14:4368–4371.
19. Thall PF, Simon R, Ellenberg SS. Two-stage selection and testing designs for comparative clinical trials. *Biometrika*. 1988;75:303–310.
20. Clark GM, McShane LM. Biostatistical considerations in development of biomarker-based tests to guide treatment decisions. *Stat Biopharm Res*. 2011;3:549–560.
21. Freidlin B, Korn EL. Biomarker enrichment strategies: Matching trial design to biomarker credentials. *Nat Rev Clin Oncol*. 2014;11:81–90.
22. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst*. 2009;101:1446–1452.
23. Freidlin B, Simon R. Adaptive signature design: An adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin Cancer Res*. 2005;11:7872–7878.
24. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: Design issues. *J Natl Cancer Inst*. 2010;102:152–160.
25. Freidlin B, Korn EL, Gray R. Marker sequential test (MaST) design. *Clin Trials*. 2014;11:19–27.
26. Freidlin B, McShane LM, Polley MY, Korn EL. Randomized phase II trial designs with biomarkers. *J Clin Oncol*. 2012;30:3304–3309.
27. Redman MW, Allegra CJ. The master protocol concept. *Semin Oncol*. 2015;42:724–730.
28. James ND, Sydes MR, Clarke NW, et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): Survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *Lancet*. 2016;387:1163–1177.
29. Simon R. Genomic alteration-driven clinical trial designs in oncology. *Ann Int Med*. 2016;165:270–278.
30. Herbst RS, Gandara DR, Hirsch FR et al. Lung Master Protocol (Lung-MAP)—a biomarker-driven protocol for accelerating development of therapies for squamous cell lung cancer: SWOG S1400. *Clin Cancer Res*. 2015;21:1514–1524.
31. Conley BA, Doroshow JH. Molecular Analysis for Therapy Choice: NCI MATCH. *Semin Oncol*. 2014;41:297–299.
32. Zelen M. Play the winner rule and the controlled clinical trial. *J Am Statist Assoc*. 1969;64:131–146.
33. Wei LJ, Durham S. The randomized play-the-winner rule in medical trials. *J Am Statist Assoc*. 1978;73:840–843.
34. Lee JJ, Chu CT. Bayesian clinical trials in action. *Stat Med*. 2012;31:2955–2971.
35. Korn EL, Freidlin B. Outcome-adaptive randomization: Is it useful? *J Clin Oncol*. 2011;29:771–776.
36. Simon R, Simon NR. Using randomization tests to preserve type I error with response-adaptive and covariate-adaptive randomization. *Stat Probab Lett*. 2011;81:767–772.
37. Faderl S, Ravandi F, Huang X, et al. A randomized study of clofarabine versus clofarabine plus low-dose cytarabine as front-line therapy for patients aged 60 years and older with acute myeloid leukemia and high-risk myelodysplastic syndrome. *Blood*. 2008;112:1638–1645.
38. Garcia-Manero G, Jabbour E, Borthakur G, et al. Randomized open-label phase II study of decitabine in patients with low- or intermediate-risk myelodysplastic syndromes. *J Clin Oncol*. 2013;20:2548–2553.
39. Byar DP, Simon RM, Friedewald WT, et al. Randomized clinical trials—perspectives on some recent ideas. *N Engl J Med*. 1976;295:74–80.

40. Kim ES, Herbst RS, Wistuba II, et al. The BATTLE trial: Personalizing therapy for lung cancer. *Cancer Discov*. 2011;1:44–53.
41. Korn EL, Freidlin B. Commentary on Hey and Kimmelman. *Clin Trials*. 2015;12:122–124.
42. Thall P, Fox P, Wathen J. Statistical controversies in clinical research: Scientific and ethical problems with adaptive randomization in comparative clinical trials. *Ann Oncol*. 2015;26(8):1621–1628.
43. Hey SP, Kimmelman J. Are outcome-adaptive allocation trials ethical? *Clin Trials*. 2015;12:102–106.
44. Berry DA. Commentary on Hey and Kimmelman. *Clin Trials*. 2015;12:107–109.
45. Lee JJ. Commentary on Hey and Kimmelman. *Clin Trials*. 2015;12:110–112.
46. Saxman SB. Commentary on Hey and Kimmelman. *Clin Trials*. 2015;12:113–115.
47. Joffe S, Ellenberg SS. Commentary on Hey and Kimmelman. *Clin Trials*. 2015;12:116–118.
48. Buyse M. Commentary on Hey and Kimmelman. *Clin Trials*. 2015;12:119–121.
49. Hey SP, Kimmelman J. Rejoinder. *Clin Trials*. 2015;12:125–127.
50. Park JW, Liu MC, Yee D, et al. Adaptive randomization of neratinib in early breast cancer. *N Engl J Med*. 2016;375:11–22.
51. Korn EL, Freidlin B. The likelihood as statistical evidence in multiple comparisons in clinical trials: No free lunch. *Biom J*. 2006;3:346–355.
52. Adams RA, Brown E, Brown L, et al. FOCUS4-D: Results from a randomised, placebo controlled trial (RCT) of AZD8931 (an inhibitor of signaling by HER 1, 2, and 3) in patients (pts) with advanced or metastatic colorectal cancer (aCRC) in tumours what are wildtype (wt) for BRAF, PIK3CA, KRAS & NRAS. *Ann Oncol*. 2016;27(suppl 6):509.