STATISTICS IN ONCOLOGY

Outcome-Adaptive Randomization: Is It Useful?

Edward L. Korn and Boris Freidlin

See accompanying editorial on page 606

ABSTRACT

Outcome-adaptive randomization is one of the possible elements of an adaptive trial design in which the ratio of patients randomly assigned to the experimental treatment arm versus the control treatment arm changes from 1:1 over time to randomly assigning a higher proportion of patients to the arm that is doing better. Outcome-adaptive randomization has intuitive appeal in that, on average, a higher proportion of patients will be treated on the better treatment arm (if there is one). In both the randomized phase II and phase III settings with a short-term binary outcome, we compare outcome-adaptive randomization with designs that use 1:1 and 2:1 fixed-ratio randomizations (in the latter, twice as many patients are randomly assigned to the experimental treatment arm). The comparisons are done in terms of required sample sizes, the numbers and proportions of patients having an inferior outcome, and we restrict attention to the situation in which one treatment arm is a control treatment (rather than the less common situation of two experimental treatments without a control treatment). With no differential patient accrual rates because of the trial design, we find no benefits to outcome-adaptive randomization over 1:1 randomization, and we recommend the latter. If it is thought that the patient accrual rates will be substantially higher because of the possibility of a higher proportion of patients being randomly assigned to the experimental treatment (because the trial will be more attractive to patients and clinicians), we recommend using a fixed 2:1 randomization instead of an outcome-adaptive randomization.

J Clin Oncol 29:771-776. Published by the American Society of Clinical Oncology

INTRODUCTION

Randomized clinical trials have long been recognized as a key tool for advancing medical knowledge. Random treatment assignments help to remove any bias due to systematic pretreatment differences in patient populations and allow an inference concerning the causal relationship between the treatment and the outcome. In a typical randomized clinical trial comparing an experimental treatment to a control treatment, each patient has an equal chance of being assigned to either treatment arm. Sometimes, to make the trial more attractive to patients, 2:1 randomization is used, in which the probability that a patient is assigned to the experimental treatment arm is 2/3. In outcome-adaptive randomization, the probability that the next patient receives the experimental treatment is not fixed throughout the trial but changes on the basis of the accruing outcome data. In particular, the probability of being assigned to the experimental treatment arm increases when the accruing outcome data suggest that the experimental treatment is better than the control treatment. The appeal of this approach is that it appears that fewer patients would get the inferior treatment.[1-3]

Outcome-adaptive randomization is sometimes included under the umbrella of "Bayesian clinical trials," but there is nothing inherently Bayesian about it. As noted by Berry,[4] regardless of the rationale used to design a trial, trial designs can be modified to ensure they have the usual ("frequentist") operating characteristics (type 1 error, power). Berry further notes on page 34 "the Bayesian approach served as a tool to build a frequentist design having good properties, such as small average sample size, fewer participants in the trial assigned to ineffective therapy and so on, with a consequent benefit for medical research."[4] Biswas et al[5] report that 20 trials conducted at the M. D. Anderson Cancer Center in 2000 to 2005 used outcome-adaptive randomization.

In this article, we focus on comparing a standard treatment with an experimental treatment, with the archetype being treatment A versus treatment A + B. One-sided questions like this are typical of phase III trials and also of randomized phase II screening trials.[6] For trials of this nature, if one is interested in getting trial results as quickly as possible, then a standard 1:1 randomization is generally the best[7] (assuming the accrual rate is independent

of the trial design). A 1:1 randomization approximately provides the most information about the between-arm treatment effect for a given total sample size. This is intuitively apparent and can also be shown formally by using decision analysis.[8] This leaves open the important question of what trial design is best for the patients enrolled on the trial. In this article, we will compare some outcome-adaptive randomization trial designs with standard randomized trial designs. To focus on the type of randomization, we will insist that all the trial designs have the same operating characteristics and the same interim monitoring rules.

Outcome-adaptive randomization should not be confused with randomization that adapts to help achieve covariate balance between the treatment arms.[9] This latter type of commonly used randomization uses accruing covariate information (but not outcome information) to modify the randomization. For simplicity, we will use the term "adaptive randomization" in this article to refer only to outcome-adaptive randomization.

We will first compare trial designs with fixed unbalanced randomization. Although this type of randomization is not adaptive, it demonstrates some of the important issues that arise because of unbalanced randomization that are also present in adaptive randomization. Phase II adaptive-randomization trial designs and phase III adaptive-randomization trial designs are then considered. For the adaptive-randomization phase III trial designs, we will insist on using a randomization and analysis strategy that controls for possible time trends that could influence outcomes, such as changes in the prognostic mix of the patient population or in ancillary care or salvage treatment over time. We end with a discussion of some remaining issues.

FIXED UNBALANCED RANDOMIZATION

In this section, we consider the situation in which more patients are randomly assigned to the experimental arm than to the control arm, but the probability ($> 1/2$) that a patient is randomly assigned to the experimental arm is fixed throughout the trial. A trial with an unbalanced randomization will generally require a larger sample size than one with a 1:1 randomization. For example, Table 1 considers the situation for a randomized phase II screening trial designed to detect an improvement in response rate from 20% for the control treatment to 40% for the experimental treatment. The operating characteristics of the trial designs (using a normal approximation to the difference in

proportions) are set at typical phase II levels: the type I error (probability of declaring the experimental treatment better than the control when it is not) has been set at 10%, and the power (probability of declaring the experimental treatment better than the control when it has a 40% response rate and the control has a 20% response rate) has been set at 90%. The trial with a balanced (1:1) randomization requires a total of 132 patients, and the trial with a 2:1 randomization ratio requires a total of 153 patients (Table 1). Another way to say this is that the information about the treatment effect obtained from 66 patients in each treatment arm is the same as obtained with 102 patients in the experimental arm and 51 in the control arm. Although higher randomization ratios are uncommon in fixed randomization designs, they are used in adaptive-randomization designs. Therefore, it is worth noting that with ratios of 4:1 and 9:1, the required sample sizes get much larger, for example, 380 patients for 9:1 randomization (Table 1).

When thinking about adaptive randomization, the sample size of the trial is not the only consideration. The two parameters that are often used to evaluate relative merits of such designs are the expected number of nonresponders and the probability a patient will be a responder.[8,10] Because these parameters depend on the true response rates in the treatment arms, they are shown for several alternatives in Table 1. Consider the trial design alternative effect (40% v 20% response rate), which is bolded in Table 1. The 1:1 randomization trial has, on average, 92.4 nonresponders and a 30.0% chance of response for participants. The 2:1 randomization has 102.0 nonresponders and a 33.3% response probability, the 4:1 randomization has 134.4 nonresponders and a 36.0% response probability, and the 9:1 randomization has 235.6 nonresponders and a 38.0% response probability.

In theory, given the true response rates, it is possible to determine the randomization ratio that minimizes the expected number of nonresponders.[10] For 40% versus 20%, this ratio is $\sqrt{2}$:1 (1.41:1). Note that even with this randomization ratio that is optimal in terms of the number of nonresponders, the actual reduction in the number of nonresponders as compared with 1:1 randomization is small: 92.2 versus 92.4 (Table 1). Some[10,11] have suggested adaptive randomization to target this optimal randomization ratio. However, since optimal ratio is so close to 1:1 and the gains are so small even when the ratio is known, we will not consider this type of adaptive randomization. Instead, we consider

Table 1. Average Proportion of Responders and No. of Nonresponders for Various Randomized Phase II Trial Designs Using Fixed Unbalanced Randomization

Response Rates by Treatment Arm		1:1 (n = 132)		$\sqrt{2}$:1 (n = 135)		2:1 (n = 153)		4:1 (n = 210)		9:1 (n = 380)	
Control Arm	Experimental Arm	P (responders) %	No. of Nonresponders	P (responders) %	No. of Nonresponders	P (responders) %	No. of Nonresponders	P (responders) %	No. of Nonresponders	P (responders) %	No. of Nonresponders
0.2	0.2	20.0	105.6	20.0	108.0	20.0	122.4	20.0	168.0	20.0	304.0
0.2	0.3	25.0	99.0	25.8	100.1	26.6	112.2	28.0	151.2	29.0	269.8
0.2	**0.4**	**30.0**	**92.4**	**31.7**	**92.2**	**33.3**	**102.0**	**36.0**	**134.4**	**38.0**	**235.6**
0.2	0.5	35.0	85.8	37.6	84.3	40.0	91.8	44.0	117.6	47.0	201.4

NOTE. One-sided type 1 error = 10%, power = 90% at 20% v 40% response rates; results based on 500,000 simulations. Characteristics of trial designs corresponding to the trial alternative hypothesis are in bold type. P (responders) % is the average proportions of responders given as a percentage.

Table 2. Average Proportion of Responders, No. of Nonresponders, and Overall Proportion Treated on the Experimental Arm for Various Randomized Phase II Trial Designs, Some of Which Use Adaptive Randomization

						Adaptive Randomization (N = 140)					
		Fixed Sample Size				Capped at 80% Assignment Probability			Capped at 90% Assignment Probability		
Response Rates by Arm		1:1 (n = 132)		2:1 (n = 153)				Overall % Treated on Experimental			Overall % Treated on Experimental
Control Arm	Experimental Arm	P (responders) %	No. of Nonresponders	P (responders) %	No. of Nonresponders	P (responders) %	No. of Nonresponders	Arm	P (responders) %	No. of Nonresponders	Arm
0.2	0.2	20.0	105.6	20.0	122.4	20.0	112.0	50.0	20.0	112.0	50.0
0.2	0.3	25.0	99.0	26.6	112.2	26.0	103.6	59.7	26.0	103.6	60.3
0.2	**0.4**	**30.0**	**92.4**	**33.3**	**102.0**	**33.2**	**93.5**	**66.2**	**33.7**	**92.9**	**68.2**
0.2	0.5	35.0	85.8	40.0	91.8	41.0	82.6	69.9	42.1	81.1	73.6

NOTE. Adaptive randomization uses the method of Thall and Wathen[12] but with no early stopping. One-sided type 1 error = 10%, power = 90% at 20% v 40% response rates; results based on 500,000 simulations. Characteristics of trial designs corresponding to the trial alternative hypothesis are in bold type. P (responders) % is the average proportions of responders given as a percentage.

adaptive randomization methods that have the potential for achieving much higher randomization ratios.

ADAPTIVE RANDOMIZATION IN PHASE II TRIALS

Adaptive randomization increases the probability of assigning patients to the treatment arm that appears to be doing better. There are many ways to do this. We consider the method of Thall and Wathen[12]:

Assignment probability to experimental treatment arm

$$= \frac{[P(E > C)]^a}{[P(E > C)]^a + [1 - P(E > C)]^a} \quad (1)$$

where $a = 1/2$ and $P(E > C)$ is the posterior probability that the experimental treatment is better than the control treatment estimated from the data seen so far (using uniform prior distributions[13]). For example, if $P(E > C)$ equaled 0.05, 0.10, 0.3, 0.5, 0.7, 0.9, or 0.95, then patients would be assigned to the experimental treatment with probability 0.19, 0.25, 0.4, 0.5, 0.6, 0.75, or 0.81, respectively.

The estimator (1) of the assignment probability can be unstable at the beginning of the trial because there is little data at that point to estimate $P(E > C)$. One possibility is to have a run-in period with the randomization probability being 1/2 before starting to use (1); this approach will be used when discussing phase III trials. The approach we will use here for phase II trials is the one given by Thall and Wathen[12]: Use formula (1) but with $a = n/(2N)$, where n is the current sample size of the trial and N is the maximum sample size of the trial. This approach yields assignment probabilities closer to 1/2 earlier in the trial. For example, if the current estimate of $P(E > C)$ is 0.9, the probability of assignment to the experimental treatment arm would be 0.57, 0.63, or 0.70 if the trial was one quarter, one half, or three quarters completed, respectively. From a practical perspective, one would in addition want to prevent the assignment probability from becoming too unbalanced, that is, being greater than 0.8 or 0.9 (extreme imbalances can create problems with the study interpretation if there are time trends; see Adaptive Randomization of Phase III Trials). We

considered two versions of the adaptive design with the probability of arm assignment capped at 0.8 and 0.9.

Table 2 displays the results for the adaptive randomization and 1:1 and 2:1 fixed randomization using the same phase II operating characteristics as described in the Fixed Unbalanced Randomization section. (No early stopping is allowed in this set of simulations to simplify the comparison of the designs.) The adaptive approach requires a total of 140 patients compared with 132 patients required for a fixed 1:1 randomization. Under the null hypothesis (response rates are 20% in both arms), the probability of response for a study participant is the same for all designs (20%). However the adaptive randomization designs have higher numbers of nonresponders compared with the 1:1 randomization (112.0 v 105.6; first row of data in Table 2). When the new treatment is beneficial, the adaptive randomization provides a slightly higher probability response: 33.2% or 33.7% versus 30% under the design alternative (third row of data in Table 2). At the same time, the adaptive design continues to result in a higher number of nonresponders than 1:1 randomization except when the treatment effect exceeds the design alternative. With respect to limiting the probability of arm assignment in adaptive randomization results, Table 2 suggests that there is no meaningful difference between capping the probability at 0.8 versus at 0.9. Therefore, we will cap the assignment probability at 0.8 in the following discussion.

Trials with adaptive randomization frequently have interim monitoring based on the assignment probability. For example, Faderl et al[2] suggest stopping their trial and declaring the experimental treatment better than the control treatment if $P(E > C) > p_{stop}$, where $p_{stop} = 0.95$; Giles et al[14] use $p_{stop} = 0.85$ in a similar manner. These investigators also suggest stopping the trial if $P(E > C) < p_{stop}$ and declaring the control treatment better. However, this type of symmetric inefficacy/futility monitoring is inappropriate for the type of one-sided question we are considering here.[15] Instead, for simplicity, we will not consider inefficacy/futility monitoring in the simulations. If the trial reaches a maximum sample size without stopping, we declare that the experimental treatment does not warrant further study.

Table 3 displays the results of the simulations that use early stopping for the adaptive randomization and 1:1 fixed randomization. The maximum sample sizes (190 and 208 for fixed randomization and

Table 3. Average Sample Size, Proportion of Responders, and No. of Nonresponders for Fixed 1:1 and Adaptive Randomized Phase II Trial Design

Response Rates by Treatment Arm		Fixed 1:1 (maximum sample size = 190)			Adaptive Randomization Capped at 80% Assignment Probability (maximum sample size = 208)		
Control Arm	Experimental Arm	Average sample size	P (responders) %	No. of Nonresponders	Average sample size	P (responders) %	No. of Nonresponders
0.2	0.2	177.9	20.3	142.3	194.3	20.2	155.5
0.2	0.3	135.2	25.9	101.4	147.6	26.3	109.7
0.2	**0.4**	**78.4**	**31.4**	**54.8**	**83.7**	**32.1**	**57.3**
0.2	0.5	43.3	36.6	28.2	45.3	37.1	28.8

NOTE. Adaptive randomization uses the method of Thall and Wathen[12]. Trials are stopped early for superiority of the experimental treatment if $P(E > C) > 0.984$. One-sided type 1 error = 10%; power = 90% at 20% v 40% response rates; results based on 500,000 simulations. Characteristics of trial designs corresponding to the trial alternative hypothesis are in bold type. P (responders) % is the average proportions of responders given as a percentage.

the adaptive design, respectively) and value of p_{stop} (0.984) were chosen so that the trial designs had type I error of 10% and power of 90% for the alternative of 20% versus 40% response rates. In terms of probability of response for a participant, the two designs perform similarly: the differences are $< 1\%$ across the range of simulated scenarios. When compared by the number of nonresponders, the adaptive design does nontrivially worse (eg, on average 13 more nonresponders in adaptive design under the null hypothesis) except when the treatment effect exceeds the design target alternative.

As an example based on a real trial, consider the adaptive randomized trial of clofarabine plus low-dose cytarabine versus clofarabine (control arm) for acute myeloid leukemia.[2] With the early stopping rules used, 63% of 54 patients in the experimental arm and 31% of 16 patients in the control arm had responses, yielding 31 nonresponders in total (56% response rate; 70 patients). This is a favorable situation for an adaptive randomization design because the response rates are so different between the arms. But even here, a fixed 2:1 randomization would have arguably been a better option: 38 and 19 patients in each arm would have yielded the same precision for estimating the treatment difference if one saw the same response rate difference between the two arms, yielding 27 nonresponders (53% response rate; 57 patients).

ADAPTIVE RANDOMIZATION IN PHASE III TRIALS

In the phase III setting, the use of adaptive randomization introduces several additional issues. Logistically, phase III studies typically use long-term outcomes like overall survival or progression-free survival; that makes adaptive randomization, which requires sufficient accru-

ing outcome information to adapt, difficult to implement. It is possible to use whatever survival information is available to estimate $P(E > C)$ and adapt the randomization imbalance,[16] but this randomization will not adapt as quickly as when the outcome is an (almost) immediately available binary outcome. However, to keep the exposition simple and give adaptive randomization the best chances to work well, we will continue to assume an immediate binary outcome. A more fundamental concern with adaptive randomization, which was noted when it was first proposed,[17-20] is the potential for bias if there are any time trends in the prognostic mix of the patients accruing to the trial. In fact, time trends associated with the outcome due to any cause can lead to problems with straightforward implementations of adaptive randomization. For example, consider a phase III trial testing whether a new therapy results in a 10% absolute increase in 1-year survival (from 80% to 90%) with 90% power and a one-sided 2.5% type I error. Suppose the experimental therapy provides no benefit over the control treatment but the prognostic mix of the patients entering the trial improves over time with a linearly increasing 1-year survival from 80% to 90% for both treatment arms. Then, the adaptive design described in Table 2 will have type I error inflated to 6.7% (from 2.5%).

One approach to the time-trend problem is to perform block randomization with a block-stratified analysis, as described on pp 331-333 of Jennison and Turnbull[21]: The first B patients on the trial are randomly assigned 1:1 between the treatment arms. At that point, $P(E > C)$ is estimated by using (1) with $a = 1/2$, and the calculated assignment probability is used to randomly assign the next B patients. After the outcomes of these patients have been evaluated, $P(E > C)$ is again estimated by using (1) with $a = 1/2$ to randomly assign the next

Table 4. Average Proportion of Responders and No. of Nonresponders for Various Randomized Phase III Trial Designs

1-Year Survival Rates		Fixed Sample Size (1:1) (n = 522; 261:261)		Fixed Sample Size (2:1) (n = 573; 382:191)		Adaptive Randomization: Block-Stratified Analysis and Randomization Capped at 80% Assignment Probability (n = 748; block size = 50)		
Control Arm	Experimental Arm	P (responders) %	No. of Nonresponders	P (responders) %	No. of Nonresponders	P (responders) %	No. of Nonresponders	Overall % Treated on Experimental Arm
0.8	0.8	80.0	104.4	80.0	114.6	80.0	149.6	50.0
0.8	0.85	82.5	91.4	83.3	95.5	83.3	124.6	66.9
0.8	**0.9**	**85.0**	**78.3**	**86.7**	**76.4**	**87.5**	**93.4**	**75.1**

NOTE. Adaptive randomization uses the method of Thall and Wathen[12] but with no early stopping. One-sided type 1 error = 2.5%, power = 90% at 80% v 90% 1-year survival rates; results based on 500,000 simulations. Characteristics of trial designs corresponding to the trial alternative hypothesis are in bold type. P (responders) % is the average proportions of responders given as a percentage.

B patients, and so on, until the trial ends. The block-stratified analysis of this trial estimates the treatment effect in each block of patients and then averages these effects over the blocks by using the Mantel-Haenszel statistic.[22] This block randomization method and block-stratified analysis eliminate the possibility of time trends confounding the results. However, as will be shown next, they reduce the efficiency of the adaptive randomization.

Table 4 compares phase III designs that use the blocked-adaptive randomization with a blocked-stratified analysis (block size 50) with 1:1 and 2:1 fixed randomization (all without interim monitoring). All designs have a one-sided type I error of 2.5% and 90% power for detecting 80% versus 90% 1-year survival rates (1-year survival for each patient assumed to be known immediately). The blocked-adaptive randomization that uses the block-stratified analysis requires 748 patients. This can be contrasted with 522 patients required in the 1:1 sample-size design. (A large part of the sample size increase is due to the necessity of using a stratified analysis: if an unstratified analysis was used for the adaptive design, then 606 patients would be required.) The adaptive randomization design results in a considerable increase in the number of nonresponders relative to the 1:1 randomization (eg, 149.6 *v* 104.4 under the null hypothesis) while providing a marginal improvement in probability of response (87.5% *v* 86.7% under the design alternative hypothesis). Moreover, use of the adaptive design requires a 43% increase in overall sample size. Table 4 shows that a 2:1 fixed randomization provides improvement in probability of response similar to that of the adaptive randomization design without the substantial increases in the number of nonresponders or much larger sample size.

DISCUSSION

Because of the possibility of time trends, we believe that any phase III trial with adaptive randomization should use block randomization and a block-stratified analysis. Short-term, placebo-controlled, randomized phase II trials that use adaptive randomization would not require stratification. However, for a randomized phase II trial that is not blinded, there is the possibility that the recruitment patterns of the trial could change substantially over the course of the trial because of the knowledge that the randomization is favoring the experimental treatment arm. Therefore, we believe the block-stratified approach is necessary for randomized phase II trials that are not placebo controlled. For both phase II and phase III trials, there is the possibility that knowledge that the randomization ratio is favoring the control treatment arm will drastically diminish accrual, suggesting the advisability of placebo-controlled designs when possible.

Adaptive randomization is inferior to 1:1 randomization in terms of acquiring information for the general clinical community and offers modest-to-no benefits to the patients on the trial, even assuming the best-case scenario of an immediate binary outcome. Our negative conclusions concerning the utility of adaptive randomization should not be applied to adaptive trial design modifications in general. In particular, formal interim monitoring that allows for early stopping for striking superiority or futility/inefficacy has long been an adaptive part of cancer trial designs. Multiarm trials with more than one experimental treatment arm, with interim monitoring applied to the distinct experimental arm-control arm comparisons, are an efficient adaptive way to compare multiple treatments to a control treatment.[23] Finally, the use of biomarker-defined subsets in randomized trials, with interim monitoring applied separately to these subsets, will become more important as cancer treatments become more individualized.[24]

AUTHOR CONTRIBUTIONS

Conception and design: Edward L. Korn, Boris Freidlin
Administrative support: Edward L. Korn, Boris Freidlin
Data analysis and interpretation: Edward L. Korn, Boris Freidlin
Manuscript writing: Edward L. Korn, Boris Freidlin
Final approval of manuscript: Edward L. Korn, Boris Freidlin

REFERENCES

1. Maki RG, Wathen JK, Patel SR, et al: Randomized phase II study of gemcitabine and docetaxel compared with gemcitabine alone in patients with metastatic soft tissue sarcomas: Results of sarcoma alliance for research through collaboration study 002 [corrected]. J Clin Oncol 25:2755-2763, 2007

2. 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 112:1638-1645, 2008

3. Adamina M, Tomlinson G, Guller U: Bayesian statistics in oncology: A guide for the clinical investigator. Cancer 115:5371-5381, 2009

4. Berry DA: Bayesian clinical trials. Nat Rev Drug Discov 5:27-36, 2006

5. Biswas S, Liu DD, Lee JJ, et al: Bayesian clinical trials at the University of Texas M. D. Anderson Cancer Center. Clin Trials 6:205-216, 2009

6. Rubinstein LV, Korn EL, Freidlin B, et al: Design issues of randomized phase II trials and a proposal for phase II screening trials. J Clin Oncol 23:7199-7206, 2005

7. Brittain E, Schlesselman JJ: Optimal allocation for the comparison of proportions. Biometrics 38:1003-1009, 1982

8. Berry DA, Eick SG: Adaptive assignment versus balanced randomization in clinical trials: A decision analysis. Stat Med 14:231-246, 1995

9. Pocock SJ, Simon R: Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 31:103-115, 1975

10. Rosenberger WF, Stallard N, Ivanova A, et al: Optimal adaptive designs for binary response trials. Biometrics 57:909-913, 2001

11. Bandyopadhyay U, Bhattacharya R: Adaptive allocation and failure saving in randomised clinical trials. J Biopharm Stat 16:817-829, 2006

12. Thall PF, Wathen JK: Practical Bayesian adaptive randomisation in clinical trials. Eur J Cancer 43:859-866, 2007

13. Thompson WR: On the likelihood that one unknown probability exceeds another in view of the evidence of the two samples. Biometrika 25:285-294, 1933

14. Giles FJ, Kantarjian HM, Cortes JE, et al: Adaptive randomized study of idarubicin and cytarabine versus troxacitabine and cytarabine versus troxacitabine and idarubicin in untreated patients 50 years or older with adverse karyotype acute myeloid leukemia. J Clin Oncol 21:1722-1727, 2003

15. Freidlin B, Korn EL, Gray R: A general inefficacy interim monitoring rule for randomized clinical trials. Clin Trials 7:197-208, 2010

16. Cheung YK, Inoue LY, Wathen JK, et al: Continuous Bayesian adaptive randomization based on event times with covariates. Stat Med 25:55-70, 2006

17. Cornfield J, Halperin M, Greenhouse SW: An adaptive procedure for sequential clinical trials. J Am Stat Assoc 64:759-770, 1969

18. Weinstein MC: Allocation of subjects in medical experiments. N Engl J Med 291:1278-1285, 1974

19. Byar DP, Simon RM, Friedewald WT, et al: Randomized clinical trials: Perspectives on some recent ideas. N Engl J Med 295:74-80, 1976

20. Wei LJ, Durham S: The randomized play-the-winner rule in medical trials. J Am Stat Assoc 73:840-843, 1978

21. Jennison C, Turnbull BW: Group Sequential Methods With Applications to Clinical Trials. New York, NY, Chapman & Hall, 2000

22. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22:719-748, 1959

23. Freidlin B, Korn EL, Gray R, et al: Multi-arm clinical trials of new agents: Some design considerations. Clin Cancer Res 14:4368-4371, 2008

24. Freidlin B, McShane LM, Korn EL: Randomized clinical trials with biomarkers: Design issues. J Natl Cancer Inst 102:152-160, 2010

■ ■ ■