



Published in final edited form as:

Stat Med. 2009 May 30; 28(12): 1680–1689. doi:10.1002/sim.3578.

Using short-term response information to facilitate adaptive randomization for survival clinical trials

Xuelin Huang¹, Jing Ning¹, Yisheng Li¹, Elihu Estey², Jean-Pierre Issa², and Donald A. Berry^{1,*†}

¹Department of Biostatistics, The University of Texas, M. D. Anderson Cancer Center, Unit 1411 1400 Pressler Street, Houston, TX 77030, U.S.A.

²Department of Leukemia, The University of Texas, M. D. Anderson Cancer Center, Houston, TX 77030, U.S.A.

SUMMARY

Increased survival is a common goal of cancer clinical trials. Owing to the long periods of observation and follow-up to assess patient survival outcome, it is difficult to use outcome-adaptive randomization in these trials. In practice, often information about a short-term response is quickly available during or shortly after treatment, and this short-term response is a good predictor for long-term survival. For example, complete remission of leukemia can be achieved and measured after a few cycles of treatment. It is a short-term response that is desirable for prolonging survival. We propose a new design for survival trials when such short-term response information is available. We use the short-term information to ‘speed up’ the adaptation of the randomization procedure. We establish a connection between the short-term response and the long-term survival through a Bayesian model, first by using prior clinical information, and then by dynamically updating the model according to information accumulated in the ongoing trial. Interim monitoring and final decision making are based upon inference on the primary outcome of survival. The new design uses fewer patients, and can more effectively assign patients to the better treatment arms. We demonstrate these properties through simulation studies.

Keywords

adaptive randomization; Bayesian; clinical trial design; efficacy; leukemia

1. INTRODUCTION

Motivated by a cancer trial, we propose a new design for randomized clinical trials with information available on a short-term response to treatment that is a good predictor of long-term survival. Before describing our proposed design, we introduce its context.

Clinical trials involving new drugs are commonly classified into four phases: phase I trials assess the toxicity of a new treatment; phase II trials test whether the treatment has any anti-disease activity; phase III trials compare the new treatment with a standard treatment; and phase IV trials assess outcome and side-effects of the new treatment through long-term follow-up studies. The phase II ‘activity’ trials as mentioned above are also called phase IIa trials. Often

phase IIb trials are conducted to further evaluate the level of efficacy of the new treatment. The sample size of a phase II cancer trial is usually between 30 and 200 patients. If phase II studies show good potential for a new treatment, then phase III trials are conducted. Phase III trials are usually registered with government agencies (the U.S. Food and Drug Administration or its counterpart in another country). Phase III trials involve sample sizes ranging from hundreds to thousands of patients. If the benefit of the new treatment is confirmed in phase III trials, the treatment will usually receive government approval for its market release. Phase IV trials are then post-market studies to evaluate the long-term side-effects of the new treatment.

Traditionally, a phase IIa or IIb trial is a single-arm trial involving an experimental drug only. However, many experimental drugs showing promise in such phase II trials fail in subsequent phase III trials, effectively wasting a substantial amount of biomedical and human resources. From 1991 to 2000, for example, the failure rate of phase III therapeutic trials conducted by the 10 largest pharmaceutical companies in the United States and Europe was as high as 45 per cent. In the field of oncology, the failure rate of phase III trials was even higher, at 59 per cent [1]. This high failure rate, coupled with the great expense and long duration of phase III trials, has led many pharmaceutical companies and research centers to conduct more phase IIb clinical trials in the hope of better evaluating treatment candidates for phase III clinical trials. In a phase IIb trial, it is not sufficient for a new treatment to show 'activity'. It must show some 'superiority' over the standard treatment to be a candidate for phase III trials. This inevitably introduces comparisons into phase IIb trials. Many phase IIb trials have only a single arm and compare the efficacy of a new treatment to historical data from a standard treatment. Such comparisons can be biased, however, due to the differing patient populations. This was one of the reasons for high failure rates of phase III trials. It is also the reason why randomized phase IIb trials are becoming more common. Our proposal uses Bayesian techniques to conveniently incorporate prior experience and historical information, and can incorporate interim monitoring rules. It can be used for both randomized phase IIb and phase III trials, but in reality it might be more suitable for the former. This is because the conduct of a phase III trial is highly regulated by government agencies, so its design must follow the approved patterns. However, pharmaceutical companies and clinical investigators have flexibilities in the conduct of phase II trials. Certainly, good novel designs will eventually be adopted by government agencies. That just takes time.

The features that are important to a good clinical trial design include interim analyses and outcome-adaptive randomization. Interim analyses allow the researchers to terminate a trial early when evidence for futility or efficacy is sufficient to make a conclusion. There are many commonly used designs for interim analyses [2–8]. By this feature, the trial can be conducted more efficiently to minimize the duration of the trial, the total number of patients (sample size), and the use of other resources. Minimizing the duration of a trial is very important in the drug discovery race where different drugs in different trials are competing to be the winner. Minimizing the sample size is also critical because many types of cancer are rare. Even in the same cancer center, there might be quite a few trials competing for the same group of patients. The design we propose in this article is motivated by these considerations.

An outcome-adaptive randomization uses unbalanced randomization to assign more patients to the treatment arms that appear to be better than others. Such a randomization is a medically ethical design in that more patients participating in the trial are assigned to the superior treatments as the trial proceeds. There are different outcome-adaptive randomization schemes [9–12].

The choice of the primary endpoint is the first important question when designing a clinical trial. To better answer the research question, the choice of a good efficacy criterion is critical in phase II and III clinical trials. Patients with cancer usually receive a few cycles of treatment,

with each cycle lasting a few weeks or somewhat longer. There are two ways to measure the efficacy of the treatment. One is to look at patient response within the treatment period. Such a response criterion could be, for example, tumor shrinkage. For patients with leukemia, the most commonly used response criterion is complete remission (CR) of the disease, which is currently used in many phase II leukemia trials. Although achieving CR is necessary for prolonging survival, it is not sufficient because patients may relapse shortly after achieving CR. Many chemotherapies have improved CR rates. However, because of their short CR durations, the improvements on CR rate do not translate into significant benefit on survival. Since survival is the ultimate goal of treatment, it is desirable to use survival as the primary endpoint of a cancer trial.

However, the long lag time of many months or years to observe a survival endpoint poses some difficulty when designing and conducting a clinical trial, especially when using outcome-adaptive randomization. In order to conduct such a randomization, we need to be able to compare the outcomes of patients currently in the different treatment arms of the trial, and to use the comparison results to determine the assignment probabilities for future patients. Consequently, it is relatively easy to implement adaptive randomization if the endpoint is readily available shortly after the treatment, as is an endpoint of CR. Since it takes a long time to observe the survival endpoint, adaptive randomization for a survival trial will not work as effectively as for a trial using CR as the endpoint. Studies have been done on outcome-adaptive randomization for trials with delayed response [13–15]. However, most of these studies focused on the mathematical and statistical techniques of dealing with the challenges in such a design. They did not consider the case in which the information on short-term response is available, and thus did not take advantage of such information.

We propose a new clinical trial design to address this issue. We use survival as the primary endpoint, but also incorporate the information about short-term patient response in order to implement a more effective adaptive randomization. We use a Bayesian mixture distribution to model the relationship between the short-term response and long-term survival. By connecting the short- and long-term responses, our proposed design allows a comprehensive evaluation and comparison of the treatment arms of the trial. We use the posterior distributions to set up early stopping criteria and implement an outcome-adaptive patient allocation algorithm. We calibrate the criteria and algorithm through extensive simulations to achieve desirable operating characteristics. Such a Bayesian approach of designing clinical trials has been used by many researchers [16–22,24].

Tamura *et al.* [23] reported a case study of an adaptive clinical trial for the treatment of outpatients with depressive disorder. Owing to the time lag to observe the true response, they used a surrogate response for the adaptive randomization. The true response was ignored, even after its information became available. In contrast, we model the relationship between the surrogate (the short-term response) and the true survival response. Before substantial amount of information on the long-term survival (the true response) is available, our model uses primarily historical survival information. The model is updated constantly during the trial and, in particular, updated immediately after each ‘event’ (disease resistance, progression, relapse, or death) is observed.

The new design is described in Section 2 with simulation studies in Section 3 to evaluate its performance. The article is closed with a summary and discussion in Section 4.

2. A NEW DESIGN FOR RANDOMIZED CLINICAL TRIALS

Our proposed design is motivated by a real randomized phase II trial for acute myelogenous leukemia. For confidentiality considerations, we do not name the specific treatments evaluated

in the trial. Both the choices of the design parameters and the clinical scenarios for use in the simulation study are similar to those for the real trial.

Many current leukemia trials simply classify patient response as CR or no CR. We classify short-term response into four categories based on patient status at the end of treatment period: (1) resistance to treatment or death, (2) stable disease, (3) partial remission (PR), and (4) CR. We may assign scores such as 0, 1, 2, and 3, respectively, to these four responses. Such a scoring system should work better than the simple classification of CR/no CR. However, as the values of the four categories may not be equally spaced, we believe that it would be better to use the mean progression-free survival time of each category. We define the progression-free survival time as the elapsed time from treatment to resistance, disease progression, relapse, or death, whichever happens first. For simplicity we call it survival time, but note that it does not measure the overall survival time from treatment to death. Because patients will seek other treatments after their diseases progress, the degree of relevance between current treatment and their overall survival times may not be high. In addition, the information on overall survival may be hard to obtain. Hence it is more appropriate to use progress-free survival than the overall survival as the endpoint for a trial.

We use historical information to give each category an informative prior distribution for its corresponding progress-free survival time. Anyone uncomfortable with the use of informative prior distributions in clinical trial design should note that we do not use the informative prior distributions to *a priori* favor either treatment arm as far as comparison is concerned; rather, we use them as a more reasonable scoring system for the different patient responses. Moreover, we dynamically update the scoring system according to the information being accumulated in the ongoing trial.

A patient is assigned to receive either treatment A or B, using an adaptive procedure that bases assignment probabilities on the results observed among the preceding patients. As efficacy data accrue, patient assignment to the two regimens becomes unbalanced in favor of the better treatment. We describe our model and the adaptive randomization scheme below.

Let $x = a$ or b correspond to treatment A or B, respectively, and n_x represent the number of patients treated in arm x . If the patient i in arm x has a short-term response in the k th category, $k = 1, \dots, 4$, then denote this by $S_{x,k,i} = 1$ and $S_{x,j,i} = 0$ for $1 \leq j \leq 4, j \neq k$. We assume the vectors $(S_{x,1,i}, \dots, S_{x,4,i})$ are independent and identically distributed across $i = 1, \dots, n_x$, and each follows a multinomial distribution $\text{Multi}(1, p_{x,1}, \dots, p_{x,4})$, with $p_{x,k}$ representing the probability of a patient in arm x having a short-term response in the k th category. Further, denote by $T_{x,i}$ the progression-free survival time of patient i in treatment arm $x, i = 1, \dots, n_x$. Conditional on the short-term response being in the k th category, we assume $T_{x,i}$ follows an exponential distribution with rate $\lambda_{x,k}$. Under the above specification, each $T_{x,i}$ has a mixture of exponential distributions. For simplicity, we assume that the short-term response is observed immediately after treatment. We present our model below.

$$\begin{aligned}
 & x=a, b(\text{treatment indicator}) \\
 & (S_{x,1,i}, S_{x,2,i}, S_{x,3,i}, S_{x,4,i}) \stackrel{\text{i.i.d.}}{\sim} \text{Multi}(1, p_{x,1}, p_{x,2}, p_{x,3}, p_{x,4}), \quad i=1, \dots, n_x \\
 & T_{x,i} \stackrel{\text{i.i.d.}}{\sim} \sum_{k=1}^4 p_{x,k} \text{Exp}(\lambda_{x,k}), \quad i=1, \dots, n_x \\
 & (p_{x,1}, p_{x,2}, p_{x,3}, p_{x,4}) \sim \text{Dir}(\gamma_{x,1}, \gamma_{x,2}, \gamma_{x,3}, \gamma_{x,4}) \\
 & \mu_{x,k} \equiv \frac{1}{\lambda_{x,k}} \sim \text{IG}(\alpha_{x,k}, \beta_{x,k}), \quad k=1, \dots, 4
 \end{aligned}$$

where $\text{Exp}(\lambda_{x,k})$ is the exponential distribution, $\text{Dir}(\gamma_{x,1}, \gamma_{x,2}, \gamma_{x,3}, \gamma_{x,4})$ is the Dirichlet distribution, and $\text{IG}(\alpha_{x,k}, \beta_{x,k})$ is the inverse gamma distribution. The parameterizations for the exponential and inverse gamma distributions are such that their expectations are equal to $\mu_{x,k}$ and $\beta_{x,k}/(\alpha_{x,k} - 1)$, respectively. We assume, *a priori*, independence between $p_{x,k}$ and $\mu_{x,k}$. Based on historical data about the progression-free survival for patients in the four short-term response categories as mentioned above, we assume the parameters $\alpha_{x,k} = 11$ for $k = 1, 2, 3, 4$, and $\beta_{x,1} = 40, \beta_{x,2} = 300, \beta_{x,3} = 750, \beta_{x,4} = 1100, x = a, b$. The amount of information in these prior distributions is approximately equal to that from 11 patients. The prior mean of $\mu_{x,1}$ is $\beta_{x,1}/(\alpha_{x,1} - 1) = 4$. Similarly, the prior means of $\mu_{x,2}, \mu_{x,3}$, and $\mu_{x,4}$ are 30, 75, and 110, respectively, for $x = a, b$. These choices reflect the information about mean progression-free survival time (in weeks) from historical data. We choose these prior distributions based on the following considerations: the priors should be reasonably informative in order to show the difference between response categories, yet they are not so strong that they can be altered by the data in the ongoing trial. We set the parameters for the Dirichlet distributions as $\gamma_{x,k} = 0.5, k = 1, \dots, 4, x = a, b$, resulting in prior distributions that contain reasonably vague information for the response rates.

By the above assumptions, the posterior distributions of $p_{x,k}$ and $\mu_{x,k}$ have closed forms. Suppose that in treatment arm x , we observe $n_{x,k}$ patients with response k for $k = 1, \dots, 4$. Based on this information, the posterior distribution for $(p_{x,1}, p_{x,2}, p_{x,3}, p_{x,4})$ is $\text{Dir}(\gamma_{x,1} + n_{x,1}, \gamma_{x,2} + n_{x,2}, \gamma_{x,3} + n_{x,3}, \gamma_{x,4} + n_{x,4})$. Suppose the observed or censored survival times for the $n_{x,k}$ patients are $t_{x,1}^{(k)}, \dots, t_{x,n_{x,k}}^{(k)}$ with corresponding indicators $\delta_{x,1}^{(k)}, \dots, \delta_{x,n_{x,k}}^{(k)}$ (0 for censored and 1 for observed time to progression). Then the posterior distribution of $\mu_{x,k}$ is IG

$(\alpha'_{x,k} = \alpha_{x,k} + \sum_{i=1}^{n_{x,k}} \delta_{x,i}^{(k)}, \beta'_{x,k} = \beta_{x,k} + \sum_{i=1}^{n_{x,k}} t_{x,i}^{(k)})$. At any given time point during the trial, we use $\text{Pr}(\cdot | \text{data})$ to denote the probability of an event conditional on the currently observed data.

Denote the mean of $T_{x,i}$ by $\mu_{x,i}$, $x = a, b$. Given $(p_{x,1}, \dots, p_{x,4}, \mu_{x,1}, \dots, \mu_{x,4})$, we have

$\mu_x = \sum_{k=1}^4 p_{x,k} \mu_{x,k}$. We continuously update and compare the distributions of μ_a and μ_b , incorporating data from all patients. We use simulation to evaluate the posterior probability $p = \text{Pr}(\mu_a > \mu_b | \text{data})$, and use this probability in the adaptive randomization. We assign patients to treatment A with probability p and to treatment B with probability $q = 1 - p$. If at any point during the trial (including at the end of the trial) $p > p_U$ (or $p < p_L = 1 - p_U$), then treatment A (or B) is selected as the superior treatment. If the maximum number of patients is enrolled, and at the end of the trial, still none of the arms is selected as superior, then the trial is inconclusive. The cut-off value p_U and the maximum sample size are calibrated in the simulation studies to achieve desirable operating characteristics. That is to say, we keep adjusting these numbers in our simulation studies until we get satisfactory operating characteristics for the design.

3. A SIMULATION STUDY

We use simulations to evaluate the performance of the above adaptive randomization procedure under different clinical scenarios (5000 simulations per scenario). For the simulations, we set the accrual rate to one patient per week. The maximum number of patients is 120. After the initial 120 weeks of enrollment time, there is an additional follow-up period of 40 weeks. The distributions of progression-free survival time (in weeks) are shown in Table I. In scenario 1, the outcomes of the two arms A and B have the same distributions, namely the same probabilities for CR, PR, stable disease, and resistance or death, and the same progression-free survival time distributions for patients falling in each of the four short-term response categories. Simply put, this is a scenario of the null hypothesis. In this scenario, by choosing $p_L = 0.025$ for the proposed design, the probability of selecting arm A (or B) is 4.6 per cent (4.8 per cent) by simulations. These correspond to one-sided type I errors in a frequentist design. We use the

same p_L in all of the scenarios, so that the selection probabilities for a treatment arm can be interpreted as the power in a frequentist design at a two-sided significance level of roughly 0.10. In scenario 2, treatment B has higher CR and PR rates than treatment A, but the CR/PR durations are the same for the two arms. As a result, treatment B has a higher mean survival time than treatment A. Simulations of this scenario indicate that the chance of selecting arm A (or B) as the better treatment is about 0.02 per cent (59.0 per cent). Using adaptive randomization to assign more patients to the better treatment arm, on average, arm A has about 16 patients, and arm B has 71 patients. In scenario 3, arm B has higher CR and PR rates, and also longer CR/PR durations. Simulations of this scenario indicate that the probability of selecting arm A as the superior treatment is 0.02 per cent, and that for selecting arm B is 97.6 per cent. On an average, arm A has about 11 patients and arm B has 51 patients.

For the above design, if one is concerned that the inferior arm has too few patients and thus may not have sufficient amount of information, then a simple remedy would be to use equal randomization for the first, say 30, patients, and start adaptive randomization at the 31st patient. The simulation results under this modification are also presented in Table I. It can be seen that the results are of the same patterns as above, and now every treatment arm has a sufficient number of patients. By using a less aggressive adaptive randomization, in some cases, the numbers of total patients are actually reduced, and the power is greater. This gain is due to the relatively more balanced patient distributions. The price paid is that slightly more patients are assigned to the inferior treatment arm.

We compare our proposed design with an alternative design for survival trials that also uses outcome-adaptive randomization, but not the information about short-term patient responses. For convenience, we call it a common design, although actually it is also a relatively new design that has not been commonly used in practice yet. In this design, we assume the survival times for patients in the two treatment arms have exponential distributions with mean parameters μ_a and μ_b , respectively. The prior distributions of μ_a and μ_b are assumed to be $IG(\alpha, \beta)$ with $\alpha = 2$ and $\beta = 60$. This is a very vaguely informative prior distribution with mean $\beta/(\alpha - 1) = 60$, which is roughly equal to the mean survival time in scenario 1 mentioned above. We compute the posterior probability $p = \Pr(\mu_a > \mu_b \mid \text{data})$, and use it as we described previously to determine patient assignment probabilities, early termination rules, and final decision rules.

We use 5000 simulations to evaluate the performance of the common design under the same clinical scenarios we used to evaluate the proposed design. The operating characteristics of the common design are presented in the far right panel of Table I. By choosing a cut-off value $p_L = 0.007$ in the stopping rules, the common design has a type I error rate that is similar to that of the proposed design (see scenario 1). In scenario 2, the proposed design has greater power (59.0 per cent vs 42.9 per cent) than the common design to detect the difference between the two arms, using a smaller number of total patients (87 vs 103). The proposed design is also more efficient than the common design in that it assigns less patients to the inferior treatment arm A (16 vs 26, or 18 per cent vs 25 per cent of the respective total number of patients). In scenario 3, treatment B has higher CR/PR rates, and also longer CR/PR durations. Under this scenario, the proposed design has greater power (97.6 per cent vs 64.8 per cent), requires fewer patients to reach a conclusion (62 vs 93), and assigns a smaller portion of patients to the inferior arm A (11 vs 21, or 18 per cent vs 29 per cent of the respective total number of patients) than the common design. The reduction in the total number of patients required under the proposed design can result in substantial save of resources and time. In addition, the reduction in the number of patients assigned to the inferior treatment arm is ethically very appealing.

Overall, we can see that the proposed design addresses the ultimate treatment goal of prolonging patient survival, while also using early response information to increase the

efficiency of adaptive randomization. When the information on early response such as CR/no CR is available, it would be a waste if not using it.

4. DISCUSSION AND SUMMARY

In order to make the best use of resources and to carefully select candidate treatments for phase III trials, there is an increasing need for phase II trial designs that evaluate the benefit of a new treatment on survival. Many authors have addressed the problem of impressive phase II results for an experimental treatment that later on fails in phase III trials [25–28]. These authors have advocated the use of progression-free survival as the endpoint for phase II cancer trials, and have emphasized the importance of making comparisons and using randomization in phase II trials.

When using survival as the primary endpoint in either randomized phase II or III trials, the information on short-term patient response is valuable and should not be ignored. We have proposed a new statistical design that connects short-term response with long-term survival. It has advantages over traditional designs that evaluate either short-term response or survival, but not both. Traditional designs evaluating short-term response can implement adaptive randomization in an almost real-time fashion, but may fail to address the ultimate concern of patients, which is survival. Other traditional designs using survival as the endpoint can address the ultimate goal of prolonging survival of patients. However, effective adaptive randomization in such designs can be difficult, and may result in excessive patients assigned to the inferior treatment arm. Our proposed design combines the advantages of these two types of traditional designs, and avoids their disadvantages. To the best of our knowledge, the proposed design is the first model-based design that uses short-term response information to facilitate adaptive randomization in survival clinical trials.

The simplest design using both short- and long-term responses (such as CR and survival in cancer studies) would be one that incorporates the short-term response (CR) in outcome-adaptive randomization, and uses the survival outcome (say, a log-rank test) in interim and final analyses for early stopping decisions and the final conclusion. Such a design lacks a model to connect the short- and long-term responses. In some trials where the positive correlation between CR and survival has been well established, such a naïve design may not have serious problems. However, in general, such a design is not well justified because the outcome-adaptive randomization is not based on the evaluation of the primary endpoint. Our proposed design takes care of this problem by modeling the relationship between the short- and long-term responses.

We have not optimized our design for this study. Many authors have considered the optimization of outcome-adaptive randomization methods. Optimization is a complicated problem, and its solution depends on the criteria and constraints for optimality. Rosenberger *et al.* [29] proposed to minimize the expected number of treatment failures in the trial under fixed variance of the test statistic. Cheng and Berry [20] proposed to maximize the number of total successes for patients in the trial and future patients combined. They set a constraint that each arm must have a probability of at least r ($0 < r < 1$) of being chosen for each patient (r is usually a small number such as 0.1). Future research may include further investigation of the optimization of the proposed design.

Acknowledgments

This research was supported by the U.S.A. National Institute of Health grants 1 P50 CA100632 and 1 PO1 CA108631-01.

Contract/grant sponsor: U.S.A. National Institute of Health; contract/grant numbers: P50 CA100632; 1 PO1 CA108631-01

REFERENCES

1. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery* 2004;3:711–715.
2. Pocock SJ. Group sequential methods in design and analysis of clinical trials. *Biometrika* 1997;64:191–200.
3. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979;35:549–556. [PubMed: 497341]
4. Kim K, Demets DL. Design and analysis of group sequential tests based on the type I error spending rate function. *Biometrika* 1987;74:149–154.
5. Simon R. Optimal 2-stage designs for phase-II clinical trials. *Controlled Clinical Trials* 1989;10:1–10. [PubMed: 2702835]
6. Berry, DA. Statistical innovations in cancer research. Chapter 33. In: Kufe, DW.; Holland, JF.; Frei, E.; Bast, RC.; Pollock, RE.; American Cancer Society. , editors. *Cancer Medicine*. e.6. London: BC Decker; 2003. p. 465-478.
7. Berry DA. Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science* 2004;19:175–187.
8. Berry DA. Bayesian clinical trials. *Nature Reviews Drug Discovery* 2006;5:27–36.
9. Berry DA, Eick SG. Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Statistics in Medicine* 1995;14:231–246. [PubMed: 7724909]
10. Hu F, Rosenberger WF. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association* 2003;98:671–678.
11. Cheung YK, Inoue LYT, Wathen JK, Thall PF. Continuous Bayesian adaptive randomization based on event times with covariates. *Statistics in Medicine* 2006;25:55–70. [PubMed: 16025549]
12. Cheng Y, Shen Y. Bayesian adaptive designs for clinical trials. *Biometrika* 2005;92:633–646.
13. Bai ZD, Hu F, Rosenberger WF. Asymptotic properties of adaptive designs for clinical trials with delayed response. *The Annals of Statistics* 2002;30:122–139.
14. Biswas A. Generalized delayed response in randomized play-the-winner rule. *Communications in Statistics: Simulation and Computation* 2003;32:259–274.
15. Zhang L, Rosenberger WF. Response-adaptive randomization for survival trials: the parametric approach. *Journal of the Royal Statistical Society, Series C: Applied Statistics* 2007;56:153–165.
16. Thall PF, Simon R. Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics* 1994;50:337–349. [PubMed: 7980801]
17. Rosener GL, Berry DA. A Bayesian group sequential design for a multiple arm randomized clinical trial. *Statistics in Medicine* 1995;14:381–394. [PubMed: 7746978]
18. Tan S-B, Machin D. Bayesian two-stage designs for phase II clinical trials. *Statistics in Medicine* 2002;21:1991–2012. [PubMed: 12111883]
19. Christen JA, Müller P, Wathen K, Wolf J. Bayesian randomized clinical trials: a decision-theoretic sequential design. *The Canadian Journal of Statistics* 2004;32:387–402.
20. Cheng Y, Berry DA. Optimal adaptive randomized designs for clinical trials. *Biometrika* 2007;94:673–689.
21. Atkinson AC, Biswas A. Bayesian adaptive biased-coin designs for clinical trials with normal responses. *Biometrics* 2005;61:118–125. [PubMed: 15737084]
22. Emerson SS, Kittelson JM, Gillen DL. Bayesian evaluation of group sequential clinical trial designs. *Statistics in Medicine* 2007;26:1431–1449. [PubMed: 17066402]
23. Huang X, Biswas S, Oki Y, Issa J-P, Berry DA. A parallel phase I/II clinical trial design for combination therapies. *Biometrics* 2007;63:429–436. [PubMed: 17688495]

24. Tamura RN, Faries DE, Andersen JS, Heiligenstein JH. A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *Journal of the American Statistical Association* 1994;89:768–775.
25. Fazzari M, Heller G. The phase II/III transition: toward the proof of efficacy in cancer clinical trials. *Controlled Clinical Trials* 2000;21:360–368. [PubMed: 10913810]
26. Francart J, Legrand C, Sylvester R, Glabbeke MV, van Meerbeeck JP, Robert A. Progression-free survival rate as primary endpoint for phase II cancer clinical trials: application to mesothelioma—the EORTC lung cancer group. *Journal of Clinical Oncology* 2006;24:3007–3012. [PubMed: 16809726]
27. Levin VA, Ictech S, Hess KR. Impact of phase II trials with progression-free survival as end-points on survival-based phase III studies in patients with anaplastic gliomas. *BMC Cancer* 2007;7:106. [PubMed: 17587447]
28. Markman M. Use of progression-free survival as valid endpoint in phase II cancer clinical trials. *Current Oncology Reports* 2007;9:159–160. [PubMed: 17430685]
29. Rosenberger WF, Stallard N, Ivanova A, Harper CN, Ricks ML. Optimal adaptive designs for binary response trials. *Biometrics* 2001;57:909–913. [PubMed: 11550944]

Table 1

Operating characteristics of two survival clinical trial designs with outcome-adaptive randomization.

Arm	Four category response rates				Four category mean survival				Proposed design using short-term response information		Common design ignoring short-term response information	
	p_1	p_2	p_3	p_4	μ_1	μ_2	μ_3	μ_4	Mean # of patients	Chance to be selected	Mean # of patients	Chance to be selected
<i>Scenario 1: A and B have the same CR/PR rates and durations</i>												
A	0.2	0.4	0.1	0.3	4	30	75	110	58 (50 per cent)	0.046	59 (50 per cent)	0.046
B	0.2	0.4	0.1	0.3	4	30	75	110	58 (50 per cent)	0.048	59 (50 per cent)	0.045
Total:									116 (100 per cent)		118 (100 per cent)	
<i>Scenario 2: B has higher CR/PR rates, same CR/PR durations</i>												
A	0.2	0.4	0.1	0.3	4	30	75	110	16 (18 per cent)	0.0002	26 (25 per cent)	0.002
B	0.1	0.1	0.2	0.6	4	30	75	110	71 (82 per cent)	0.590	77 (75 per cent)	0.429
Total:									87 (100 per cent)		103 (100 per cent)	
<i>Scenario 3: B has higher CR/PR rates, longer CR/PR durations</i>												
A	0.2	0.4	0.1	0.3	4	30	75	110	11 (18 per cent)	0.0002	21 (29 per cent)	0.001
B	0.1	0.1	0.2	0.6	6	45	112	165	51 (82 per cent)	0.976	72 (71 per cent)	0.648
Total:									62 (100 per cent)		93 (100 per cent)	
<i>Using equal randomization for the first 30 patients</i>												
<i>Scenario 1: A and B have the same CR/PR rates and durations</i>												
A	0.2	0.4	0.1	0.3	4	30	75	110	58 (51 per cent)	0.048	58 (50 per cent)	0.047
B	0.2	0.4	0.1	0.3	4	30	75	110	56 (49 per cent)	0.054	58 (50 per cent)	0.047
Total:									114 (100 per cent)		116 (100 per cent)	
<i>Scenario 2: B has higher CR/PR rates, same CR/PR durations</i>												
A	0.2	0.4	0.1	0.3	4	30	75	110	23 (31 per cent)	0.0002	29 (29 per cent)	0.004
B	0.1	0.1	0.2	0.6	4	30	75	110	52 (69 per cent)	0.704	70 (71 per cent)	0.477
Total:									75 (100 per cent)		99 (100 per cent)	
<i>Scenario 3: B has higher CR/PR rates, longer CR/PR durations</i>												
A	0.2	0.4	0.1	0.3	4	30	75	110	20 (33 per cent)	0.0002	22 (28 per cent)	0.0004
B	0.1	0.1	0.2	0.6	6	45	112	165	40 (67 per cent)	0.918	58 (82 per cent)	0.832

Arm	Four category response rates				Four category mean survival				Proposed design using short-term response information		Common design ignoring short-term response information	
	p_1	p_2	p_3	p_4	μ_1	μ_2	μ_3	μ_4	Mean # of patients	Chance to be selected	Mean # of patients	Chance to be selected
Total:									60 (100 per cent)		80 (100 per cent)	