Published in final edited form as:

Stat Med. 2013 September 20; 32(21): 3752–3765. doi:10.1002/sim.5807.

Response-Adaptive Decision-Theoretic Trial Design: Operating Characteristics and Ethics

Ari M. Lipsky, MD PhD^{1,2} and Roger J. Lewis, MD PhD^{1,3}

¹Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, CA

²Gertner Institute for Epidemiology and Health Policy Research, Tel Hashomer, Israel

³Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA

Abstract

Adaptive randomization is used in clinical trials to increase statistical efficiency. In addition, some clinicians and researchers believe that using adaptive randomization leads necessarily to more ethical treatment of subjects in a trial. We develop Bayesian, decision-theoretic, clinical trial designs with response-adaptive randomization and a primary goal of estimating treatment effect, and then contrast these designs with designs that also include in their loss function a cost for poor subject outcome. When the loss function did not incorporate a cost for poor subject outcome, the gains in efficiency from response-adaptive randomization were accompanied by ethically concerning subject allocations. Conversely, including a cost for poor subject outcome demonstrated a more acceptable balance between the competing needs in the trial. A subsequent, parallel set of trials designed to control explicitly type I and II error rates showed that much of the improvement achieved through modification of the loss function was essentially negated. Therefore, gains in efficiency from the use of a decision-theoretic, response-adaptive design using adaptive randomization may only be assumed to apply to those goals which are explicitly included in the loss function. Trial goals, including ethical ones, which do not appear in the loss function are ignored and may even be compromised; it is thus inappropriate to assume that all adaptive trials are necessarily more ethical. Controlling type I and II error rates largely negates the benefit of including competing needs in favor of the goal of parameter estimation.

Keywords

Bayesian statistics; decision theory; adaptive trial design; adaptive randomization; clinical trial ethics

1. Introduction

Adaptive or flexible clinical trial designs [1] have gained increasing attention over the last years among researchers [2], industry [3], and regulatory bodies [4 – 7]. These designs may allow more efficient use of resources [8] and, in trials which are adaptive by design, may allow for the preservation of the overall type I error rate [6, 9].

In trials that use response-adaptive randomization, a type of dynamic allocation, the randomization proportions used to allocate future subjects to treatment arms are adjusted based on the previously acquired data. Prior authors have promoted this approach from an ethical perspective, seeing in them a compromise between collective and individual ethics,

and thus helping to balance the needs of society to learn against the needs of enrolled subjects to receive effective treatments [10, 11]. Trials using response-adaptive randomization have been described as ‘ethically justified in desperate medical situations and may be morally required [11]’. One of the earliest response-adaptive trials, which involved the treatment of premature infants with extracorporeal membrane oxygenation, was designed specifically to address ethical concerns [12]. Unfortunately, the unique design and the very uneven subject allocation led to contentious debate about the trial’s scientific interpretation [13].

Most clinical trials, however, focus on parameter estimation and hypothesis testing rather than improving the outcomes of individual subjects. The impact of using an adaptive design focused on parameter estimation on secondary goals, such as the effective treatment of individuals, is unclear if those goals are not considered in the design of the trial. It seems unlikely that gains in efficiency afforded by an adaptive approach will improve all trial characteristics. Our aim is to address this issue using one form of adaptive design, Bayesian response-adaptive randomization in a decision-theoretic framework, as this framework allows an explicit definition of the trial’s goals in the form of the loss function.

We will first describe our design of Bayesian, decision-theoretic, group sequential clinical trials for diseases with dichotomous outcomes; the design allows the use of response-adaptive randomization, if desired. We then compare a series of four trial designs: The first two trials, one non-response-adaptive and one response-adaptive, have the explicit goal of parameter estimation; other goals, while potentially important, are not specifically included in the loss function. In the second two trials, again one non-response-adaptive and one response-adaptive, we include the goal of improving individual subjects’ welfare by adding a fixed cost for each subject in the trial who suffers the worse of the two possible outcomes. Using these four trial designs, we explore the efficiency gains provided by, and the trade-offs resulting from, the inclusion of response-adaptive randomization. We then highlight specific interim data patterns for which decisions are appreciably different between trial designs. Finally, we examine the impact of constraining the design of these four trials to achieve specific frequentist type I and II error rates.

2. Bayesian decision-theoretic trial design tool

2.1. Previous work

In earlier work, we described Bayesian decision-theoretic clinical trials comparing two treatments for a disease with binary outcomes using a simple terminal decision loss function appropriate for one-tailed hypothesis testing [14]. We extended this in later work to two-tailed trials and incorporated a quadratic loss function for estimation of the treatment effect [15]. These designs, however, were limited in several important ways. First, we considered the probabilities of success in the two arms to be *a priori* independent. Second, we restricted the trials to 1:1 subject randomization and we assumed that this randomization resulted in exactly equal allocation of subjects to the two treatment groups. Third, we only included in the decision loss function the costs pertaining to subject enrollment and incorrect conclusions regarding treatment efficacies in a hypothesis testing framework.

We describe below an extension of the previous work which reparameterizes the probabilities of success in the two arms so that the treatment group efficacy is relative to (and not independent of) that of the control group efficacy; assigns each individual subject to a trial arm with a particular probability and allows response-adaptive randomization; and includes the cost of a subject suffering the worse of the two possible outcomes.

2.2. Trial design

We consider the case of a two-armed, binary-outcome trial, so that each subject experiences either ‘success’ or ‘failure.’ Previously [15], we specified parameters for two independent beta distributions, each describing prior information for the probabilities of success in the control and treatment arms, respectively. Leaving the prior for the control probability of success as a beta distribution, here we express the prior information about the treatment probability of success in terms of the log-odds-ratio. Specifically, we define

$$\theta = \log \left[\frac{p_T(1 - p_C)}{p_C(1 - p_T)} \right],$$

where the true success probabilities in the control and treatment arms are p_C and p_T , respectively, and θ is the (natural) log-odds-ratio of the success probabilities. Our prior information about θ is quantified by specifying the normal distribution. The reparameterization from p_C, p_T to p_C, θ allows us to consider the relative treatment effect explicitly, minimizing dependence on the (incidental) underlying rate of events in the control group. Additionally, since it is the relative effect that is of primary interest in comparison trials, we want to be able to express our uncertainty about it directly. Finally, we further define θ_0 as the minimum value of the log-odds-ratio we want to identify (e.g., the clinically significant difference).

The group-sequential trial is composed of up to B possible blocks of subjects, with each block consisting of M subjects. There are three randomization proportions available at each block, which we label r_i , where i is the index (1:3). r_i is the probability of assigning each of the M subjects in a particular block to the control arm. (The actual probabilities are defined at run-time.) If not all r_i are equal, the trial design allows adaptive randomization.

After $0 \leq j \leq B$ blocks, there are s_K cumulative successes in the control ($k = C$) and treatment ($k = T$) arms, and $f_K = jM - s_K$ cumulative failures. We assume all outcomes are known through the j th block of subjects prior to determining whether to enroll the $(j + 1)$ th block of subjects and, if that block is enrolled, with what randomization proportion. According to the likelihood principle, the current probability distribution depends only on the currently observed data; thus, earlier data patterns or the order in which successes and failures were observed have no effect on the calculation of the posterior probabilities.

At the j th block, there are $M + 1$ ways of partitioning newly enrolled subjects between the control and treatment arms and $jM + 1$ ways the whole population may be partitioned between arms. Within each of those possible partitionings of subjects, there are $(n_C + 1)(n_T + 1)$ possible outcome combinations, where n_C is the cumulative number of subjects in the control arm and n_T is the cumulative number in the treatment arm. We call each unique configuration of treatment assignment and outcomes (n_C, n_T, s_C, s_T) , at any interim analysis or at the end of the trial, a ‘node.’

After each block, we may choose to stop the trial and conclude that the treatment is better than (D_+), equivalent to (D_0), or worse than (D_-) the control. Additionally, if we are not already at the B th block, we may decide to continue the trial by enrolling an additional block of subjects. When response-adaptive randomization is allowed, a separate continuation decision is available for each possible value of r_i .

The terminal part of the loss function depends on the true treatment effect, θ , and the available stop decisions (i.e., D_+ , D_0 , D_-). The full loss function has three components: 1. a per-subject enrollment cost, e ; 2. a failure cost, d , for each subject who experiences the

worse of the two possible outcomes; and 3. a quadratic loss component corresponding to making an error in the final decision regarding the treatment effect. This latter component has two parameters, K_1 and K_0 , which weight the importance of type I and type II errors, respectively; this component is defined as

$$L(\theta, D_0) = K_0 \theta^2 \quad L(\theta, D_+) = \begin{cases} K_1(\theta - \theta_0)^2 & \text{if } \theta < \theta_0 \\ 0 & \text{if } \theta \geq \theta_0 \end{cases} \quad L(\theta, D_-) = \begin{cases} 0 & \text{if } \theta \leq -\theta_0 \\ K_1(\theta + \theta_0)^2 & \text{if } \theta > -\theta_0 \end{cases}.$$

The total loss incurred by making a particular stop decision after enrolling jM subjects is thus $e_j M + d(f_C + f_T) + L(\theta, D_l)$, for $l = +, 0, -$. We integrate over the current θ density, which is based on the prior and any accumulated data, to calculate the expected stopping losses. The welfare of enrolled subjects is captured in the loss function by using a positive failure cost.

At each block $< B$, we must also calculate the expected continuation cost, which is the cost arising from continuing the trial by enrolling an additional block of subjects. For any particular node, there are as many continuation costs as there are unique r_i . Calculating the expected losses incurred by making a continuation decision without approximate methods requires backward induction, as described previously [14, 15]. Each continuation cost is a weighted average of the minimum costs from the nodes in the previously calculated future blocks, where the weights consider two factors. The first factor is the probability of transitioning to a particular pattern of subject treatment assignments. If there are currently x subjects in the control arm, the probability of transitioning to the node which has $(x + \Delta x)$ control subjects is given as $\Delta x \sim \text{bin}(M, r_i)$. Because this distribution depends on r_i , the transition probabilities are different for each of the allowed randomization ratios. (If the r_i are all equal, then there is only one possible continuation decision.) The second factor is the probability of obtaining a particular distribution of additional successes for the newly enrolled subjects in the control and treatment arms. If there are currently s_C and s_T successes, the probability of having a total of $(s_C + \Delta s_C)$ and $(s_T + \Delta s_T)$ successes after enrolling the next block is the expectation of $\text{bin}(\Delta s_C | \Delta x, p_C) \text{bin}(\Delta s_T | M - \Delta x, p_T)$, taken over the uncertainty in p_C and p_T . The probability of transitioning to a particular node in the next block, for each allowed randomization ratio, is thus,

$$\text{bin}(\Delta x | M, r_i) \int_0^1 \int_0^1 \text{bin}(\Delta s_C | \Delta x, p_C) \text{bin}(\Delta s_T | M - \Delta x, p_T) \pi(p_C, p_T | n_C, s_C, n_T, s_T) dp_C dp_T,$$

for each r_i ; ($\pi(\cdot)$ is also conditional on the prior information for p_C, p_T .)

In performing the required calculations, we separately record the transition probabilities and minimum node costs, which reduces the overall computational effort (see next section). Specifically, if we define $T_i(\Delta x, \Delta s_C, \Delta s_T)$ as the probability of transitioning with the i^{th} randomization ratio to a future node that has Δx of M additional subjects assigned to the control arm, Δs_C additional control successes, and Δs_T additional treatment successes, and further define $C(\Delta x, \Delta s_C, \Delta s_T)$ as the minimum expected cost in those future nodes, then the expected continuation cost from the current node, for the i^{th} randomization ratio, is:

$$\sum_{\Delta x=0}^M \sum_{\Delta s_C=0}^{\Delta x} \sum_{\Delta s_T=0}^{M-\Delta x} T_i(\Delta x, \Delta s_C, \Delta s_T) C(\Delta x, \Delta s_C, \Delta s_T). \quad (1)$$

The expected continuation cost from a particular node is thus (1) minimized with respect to i . The optimal decision at each node is that which gives rise to the minimum of the expected stopping and (up to 3) continuation costs.

2.3. Computation

We use a Markov chain Monte Carlo (MCMC) algorithm, programmed in C++ (Visual C++ 2005, Microsoft Corp., Redmond, WA), to generate samples from the current probability distributions, and to numerically integrate the loss function over those distributions to determine the decisions that give rise to the minimum expected stopping cost at each node. [The code is available upon request from the authors.] As the program steps backward through the trial, similar integration is performed to calculate the transition probabilities and, by reading in the future minimum node costs, the expected continuation costs are calculated.

Priors for p_C and θ are specified. To facilitate comparisons with standard frequentist designs, we further define $R_K \equiv K_1/K_0$, where R_K represents the relative importance of a type I to a type II error.

At every node, the program records the expected stopping and continuation costs, as well as the decision associated with the minimum expected cost. These are the overall expected costs for the entire trial should that node be reached and a particular decision made. Additionally, the expected number of additional subjects to be enrolled is recorded; this is 0 for stop decisions. The overall expected cost and expected subject enrollment for the entire trial are thus those values given at the null (initial) node.

The output from the trial design software is a list of every possible node with the decision that incurs the minimum expected cost at each node. Though the initial run is computationally intensive, it is possible to vary the cost weights (i.e., K_0 , R_K , d , and e) to create new designs in much less time. This is accomplished by examining (1): The initial run performs all the MCMC calculations, recording both T_i and C assuming that K_0 and R_K are 1, and d and e are 0. Subsequent runs avoid the intensive MCMC calculations by reading in the saved T_i , and scaling and shifting the saved C to the desired values. As with the initial run, subsequent runs must also start at the last block and work backward.

Characteristics of the designed trials (e.g., number of subjects enrolled, cost, and type I and type II error rates) can be determined by simulation under different assumed distributions for p_C and θ [15]. By drawing the ‘true’ values for p_C and θ from their respective prior distributions, we determine through simulation the expected overall cost and number of subjects. This provides an internal consistency check by comparing these simulated values with the expected values from the null node as determined by backward induction. We used the BOA package with R for convergence diagnostics [16, 17], and external consistency checks of expected costs at selected nodes were performed with WinBUGS [18].

3. Application to response-adaptive designs

3.1. Overview

We initially designed four trials, incrementally increasing the complexity of the decision space and cost structure (Table 1, first 4 trial designs). The ‘base’ design included costs of estimation error and subject enrollment (but not a cost of treatment failure, i.e., $d = 0$) in the loss function, and used a single, fixed randomization ratio. The second trial (‘response-adaptive’) added response-adaptive randomization to the base design. The third trial (‘failure cost’) added the cost of treatment failure to the loss function of the base design ($d > 0$). And the fourth trial (‘response-adaptive/failure cost’) included both response-adaptive randomization and a cost for treatment failure in the loss function.

All four trials enrolled up to 4 blocks of 8 subjects. Prior information was minimal, with $\text{beta}(1, 1)$ for the control arm (p_C), and $N(0, 25)$ (i.e., standard deviation = 5, or equivalently, $\text{variance}^{-1} = 0.04$) for the log-odds-ratio (θ). We specified the minimum log-odds-ratio we wished to be able to identify, $\theta_0 = 2.197$, which corresponds to $p_T = 0.75$ if $p_C = 0.25$.

The loss function parameters, K_0 and R_K , were derived empirically for the base design to yield approximate customary frequentist error rates (type I (α) = 0.05; type II (β) = 0.20) upon simulation. For the next three trial designs, these same parameters were held constant, and there was no further attempt to constrain the error rates. The cost of enrollment, e , was set at 1 to ensure that information bore a cost so that the trial would not necessarily proceed to the maximum available subjects, and to provide a cost-unit for appreciating differences between trial designs; here the unit is the cost of enrolling one additional subject. The cost of subject failure, d , was arbitrarily set at 50 in the failure cost and response-adaptive/failure cost trials. Each randomization ratio listed in Table 1 was available at each block.

The base and response-adaptive trials were designed first. The cost of failure was then added to each trial so that the same T_i and scaled C (as explained above) were used in creating the failure cost and response-adaptive/failure cost trials, respectively. This helped minimize random differences (e.g., due to random number generator sequences and rounding errors) between trial designs. Nonetheless, to further check consistency, all four trials were also designed independently: there were no differences in action decisions at any nodes among the trial design pairs.

Three additional trials were designed (one of each type; the base design already met the design criteria) with K_0 and R_K selected to yield error rates of approximately $\alpha = 0.05$, and $\beta = 0.20$. (Table 1, last three trial designs.)

Expected enrollment characteristics (including total enrollment and outcomes) and costs (including total cost and cost components) were simulated with p_C and θ randomly sampled from their respective prior distributions. Type I error rates were simulated by setting p_C to 0.5 and θ to 0, and Type II error rates were simulated by setting p_C to 0.25 and θ to 2.197 (which implies $p_T = 0.75$). The expected enrollment characteristics and costs from the above fixed p_C , θ simulations were also recorded; however, we will always be referring to the expected values obtained under random sampling of p_C , θ —and not those from fixed simulation—unless otherwise stated. For all simulations, the observed expectation of θ and its variance were recorded at each node where a trial ended. The observed θ s from the simulations were used with SAS 9.2 (SAS Institute, Cary, NC) to compute $\text{var}(\theta)$ among the θ values at the stopping nodes. Separately, the $\text{var}(\theta)$ data at each stopping node from the random simulation were used to compute the simulated predictive distribution of $\text{var}(\theta)$, which is the expected certainty in θ given our priors. All simulations were performed using 100,000 virtual trials.

To help ensure fair comparisons of costs between designs that did and did not include a cost of failure, a corrected cost was calculated *post facto* for the latter designs: 50 cost units were added to the original simulated cost for each expected subject failure as determined by simulation.

3.2. Output characteristics: fixed estimation error parameters

The upper half of Table 2 displays the simulated type I and II error rates, and the simulated and predicted sample sizes and costs, for each of the first four trial designs. We note that the predicted and simulated values demonstrate good agreement.

Compared to the base design, the expected cost in the response-adaptive design is lower (by 10%), and the expected enrollment falls slightly (1%). Further, there is a sizable reduction in the type I error (36%), though there is some increase in the type II error (12%); this is in part due to our choice of R_K . The predictive distribution of $\text{var}(\theta)$ decreases (i.e., the expected certainty increases) with the response-adaptive design (from 5.5 to 5.1; 8%) (Figure 1A).

In the upper half of Table 2, we see that adding the cost of subject failure to the base and response-adaptive trials, respectively, results in a substantial decrease in expected enrollment (45%), with concomitant increases in both type I and II error rates (three-and-a-half-fold) and the predictive distribution of $\text{var}(\theta)$ (25%). Adding response-adaptive randomization to the failure cost design decreases the expected cost (9%) and the type I error (37%), and increases type II error (12%); enrollment falls slightly (1%) and the predictive distribution of $\text{var}(\theta)$ decreases (6%).

The corrected costs are shown in the right-hand column of Table 2. The failure cost and response-adaptive/failure cost designs have lower corrected costs than the base and response-adaptive designs, respectively.

3.2.1. Subject allocation and successful outcomes—The base and failure cost designs divide the subjects evenly between the arms of the trial (50% in treatment arm), whereas adding response-adaptive randomization to the two trials preferentially assigns subjects to the treatment arm (66% and 70%, respectively) (Figure 2A, ignoring the shading).

In the base and response-adaptive trials, 50% of the subjects are expected to have successful outcomes (Figure 2A, shading), whereas adding a failure cost results in higher overall success rates (58% and 62%, respectively), despite our neutral prior information. In absolute terms, the expected number of subjects who experience the worse outcome in the four trials, respectively, is 12.4, 12.2, 5.7, and 5.2. The addition of failure cost results in an over 50% reduction in the number of poor outcomes. The trade-offs in terms of subject enrollment, assignment, and success rates on one hand, and frequentist error rates on the other, are apparent in Figure 2A.

The inclusion of failure cost affects the block in which simulated trials tend to reach stop decisions (Figure 3A). The two trials that do not consider failure cost generally stop closer to the last possible block, whereas the two trials that do consider failure cost have a greater percentage of stop-decisions earlier in the course of the trials.

Figure 4A displays the percentage of the total expected cost attributable to each of the three components of the loss function. Note that because $e = 1$, the total enrollment cost is necessarily equal to the simulated number of subjects in the trial.

3.2.2. Select nodes—In addition to the global differences described above among the four trial designs, there are also important differences at the nodal level; here, we highlight the actions taken at two example nodes chosen by the authors. The expected costs for each available decision at each of the two nodes are displayed in the upper halves of Tables 3A and 3B, respectively. Note that multiple continuation decisions are only available in the response-adaptive and response-adaptive/failure cost trials.

After the third block is enrolled, one particular node has 24 subjects distributed such that 14 of 16 (88%) assigned to the control arm have experienced a successful outcome, while only 3 of 8 (38%) in the treatment arm have experienced a successful outcome. In the base design, the decision that minimizes the expected cost is to enroll the final block of 8 subjects

with 1:1 randomization. In the response-adaptive design, however, it is more cost efficient to enroll the 8 subjects preferentially (1:3) into the underperforming treatment arm where less information is available. Similar behavior is noted in the node that has 1 of 11 (9%) control subjects and 7 of 13 (54%) treatment subjects experiencing a successful outcome, with the response-adaptive design continuing enrollment with 3:1 allocation.

Adding the cost of failure to either the base or response-adaptive design (i.e., failure cost or response-adaptive/failure cost) changes the decision to stop, concluding that the control (first node example) or treatment (second node example) is better.

Though a continuation decision was not chosen in either node under the response-adaptive/failure cost designs, it is interesting to compare the decision costs: In the response-adaptive designs, the decisions that seem least ethical at the nodal level (i.e., that preferentially randomize subjects to apparently poorly performing arms) have the least cost, and vice versa. (In other words, in Table 3A, the cost of continuing is ordered as 3:1 > 1:1 > 1:3, and in Table 3B, the continuation costs are ordered 1:3 > 1:1 > 3:1.) In the response-adaptive/failure cost designs, the lowest cost among the continuation decisions is that which seems to make the most ethical sense (when considering only the subjects in the trial), and the ordering of the continuation decisions is exactly the reverse of that seen without failure cost.

3.3. Output characteristics: targeted type I & II error rates

In the targeted frequentist error rate trials, the addition of response-adaptive randomization to either the base or the failure cost designs results in considerably lower cost (67% and 63% reduction, respectively) and decreased enrollment (15% and 10% reduction, respectively), while preserving the customary frequentist error rates (Table 1, bottom half). In contrast to the fixed estimation error parameter trials, the addition of failure cost to the base or response-adaptive designs increases subject enrollment (2% and 7%, respectively). The predictive distribution of $\text{var}(\theta)$ is essentially stable across the four trials in this group (Figure 1B).

With regard to subject allocation, the response-adaptive designs again preferentially assign subjects to the treatment arm (66% vs 50% for both pairs of designs; Figure 2B). The percentage of successes in the four trials was 50%, 50%, 51%, and 54%, respectively. In absolute terms, we expect 12.4, 10.7, 12.5, and 10.6 failure outcomes in the base, response-adaptive, failure cost, and response-adaptive/failure cost trials, respectively.

With constrained type I and II error rates, all four of the trial designs now exhibit similar curves describing their cumulative distribution of stopping blocks (Figure 3B). As described above, adding failure cost to either the base or the response-adaptive design shifts the curve slightly to the right, requiring additional subjects to be enrolled.

In Figure 4B, it is evident that the contribution of the enrollment cost becomes more influential in the response-adaptive design as compared to the base design, as the former can achieve the desired error rates more efficiently. We also observe that in the trials with failure cost (as compared to the respective trials without failure cost), the proportion of the cost attributable to estimation error is fairly stable, while the cost of subjects experiencing the worse outcome has largely displaced that arising from enrollment.

In Table 2, we observe much higher costs in the designs that include failure cost when the simulated error rates are targeted as compared to when these error rates are not targeted. Correcting the costs no longer gives trials designed with failure cost an edge in terms of lower overall expected cost. We also note the substantial increases in the K_0 s required to achieve these stable error rates (Table 1).

Finally, we note in the bottom halves of Tables 3A and 3B that the decisions at the two example nodes no longer appear more ethical (considering only these individual nodes) with the inclusion of failure cost. Additionally, the trends in the response-adaptive/failure cost continuation costs now mirror those of the response-adaptive trial. More generally, the addition of failure cost to the fixed error rate trials results in trials which do not have appreciably different decision patterns.

3.4. Additional characteristics of the target error rate simulations

To give the reader a better appreciation of the impact of response-adaptive randomization and failure cost, we provide the expected enrollment characteristics and costs under the same fixed parameter assumptions used to simulate the frequentist error rates (Tables 4a and 4b). The trends are not dissimilar from those described above.

4. Discussion

Many trials, including those incorporating response-adaptive randomization, focus on the goal of and error rates associated with parameter estimation through specification of desired type I and II errors. Though the researchers may have additional important goals (such as concern for the subjects) in mind, if the trials are not explicitly designed to take these goals into consideration, the trial designs will remain oblivious to those goals.

We can potentially address this concern by using a decision-theoretic model to design trials which include in their loss functions costs reflecting the major goals of the trial. The tension among these often competing goals is thus addressed more explicitly, and an appropriate balance among the needs is established, given the enumerated costs and their weights. Adding the flexibility which comes with a response-adaptive randomization design can then be used to increase the efficiency by which this balance of goals is achieved. Without the additional consideration of subject welfare, however, a design labeled as response-adaptive—which seems to imply a preference to allocate subjects to the better performing arm—may provide a false sense of beneficence. A decision-theoretic trial that adapts its course based on subject responses will do so to improve efficiency irrespective of subject welfare, unless the concern is specifically included.

As illustrated with the first group of example trial designs, adding response-adaptive randomization to the base trial improves the efficiency of parameter estimation, while slightly decreasing subject enrollment (a secondary goal as evidenced by the low relative cost it was assigned). While the addition of this adaptive feature improves these global characteristics, the improved efficiency does not necessarily translate into more ethical trials. There was only a small reduction in the absolute number of expected failed outcomes due to the expected enrollment decreasing. Further, as we saw in the examples of two specific trial decision points, the more efficient trials led to individual decisions which compromised our concern for the subjects in the study until the cost of subject failure was explicitly included. It is important to recognize, however, that adding the cost of subject failure does not guarantee that an obviously sound decision is made at every node, but that the decision will have arisen from quantitative balancing of all enumerated competing goals.

The need to minimize errors in parameter estimation in our response-adaptive/failure cost trial design was thus balanced against the individual subject outcomes, with response-adaptive randomization invoked to improve the efficiency by which the goals were achieved. Trade-offs such as these (e.g., between the strict scientific aims and the individual subjects' welfare), however, are often difficult to acknowledge, let alone formulate and quantify. While generating an appropriate form for the loss function is not a perfect science, there is guidance in the literature [19]. Costs included in the loss function may be incurred

from enrollment, from subjects experiencing the worse outcome, or from considering the application of the imperfect information gained in the trial to the population at large (patient horizon) [20]. Additional costs, reflecting the diverse and often competing needs surrounding trial design, may also be included [19].

It appears that constraining the trials to achieve typical type I and II error rates negates most of the benefit conferred from consideration of competing costs. By virtue of there being lower enrollment with response-adaptive randomization, however, there is also less opportunity to cause harm within the trial. We observed a slightly better chance of a subject experiencing a successful outcome when the failure costs were included, though this is certainly not convincing and is primarily due to dilution of the failed outcomes with additional successful outcomes. At the individual decision points, the inclusion of failure cost does not seem to affect substantively the decisions in our error rate-constrained trials.

An important advantage of the decision-theoretic Bayesian approach, when the relevant costs are incorporated, is transparency. Designing and simulating trials, as we did above, allows us to evaluate the implications of our decisions on both a global and a decision-by-decision basis, making the implications of our cost function more apparent. At least for less complex trials, the decisions made at the individual decision points can be reviewed to ensure that they are consistent with the goals of the trial, considering the competing scientific, regulatory, economic, and ethical needs.

A key limitation of the present work is that the notion of the patient horizon has not been considered explicitly. The loss function should include subjects enrolled in the trial, as well as future patients whose treatment will in part depend on the trial results. The optimal trial design could then be defined as that which maximizes successful outcomes across the entire population, though ethical considerations are often more complicated (e.g., determining how to balance the needs of the enrolled subjects versus those of the hypothetical, as-yet untreated patients in the horizon). A further limitation is that it is not known whether our conclusions hold for trials with different loss function structures, or different prior information and designs more generally. Additionally, though not a limitation, we emphasize that we have focused on one of multiple methodological issues that must be considered when conducting adaptive trials (see, e.g., references 21 and 22).

5. Conclusion

The use of response-adaptive randomization is an important advancement in clinical trials. Using such designs in a decision-theoretic framework, however, only serves to increase the efficiency with which trials fulfill the goals included in the loss function. Any goals not explicitly included will be ignored, and may even be compromised. Constraining trials to achieve pre-specified frequentist error rates curtails the impact of including competing costs.

Acknowledgments

Support: This publication was made possible by Grant Number 1F32RR022167 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH.

References

1. Dragalin V. Adaptive designs: terminology and classification. *Drug Information Journal*. 2006; 40:425–435.
2. Scott CT, Baker M. Overhauling clinical trials. *Nature Biotechnology*. 2007; 25:287–292.

3. Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, Pinheiro J. PhRMA Working Group. Adaptive designs in clinical drug development—an executive summary of the PhRMA Working Group. *Journal of Biopharmaceutical Statistics*. 2006; 16:275–283. [PubMed: 16724485]
4. European Medicines Agency (EMA). Reflection paper on methodological issues in confirmatory clinical trials with flexible design and analysis plan. London: EMA; 2007 Oct. Available from: <http://www.ema.europa.eu/pdfs/human/ewp/245902enadopted.pdf> [accessed 20/03/2011]
5. Gottlieb, S. [accessed 20/03/2011] Deputy Commissioner for Medical and Scientific Affairs, Food and Drug Administration—2006 Conference on Adaptive Trial Design—July 10, 2006. Available from: <http://www.fda.gov/NewsEvents/Speeches/ucm051901.htm>
6. U.S. Food and Drug Administration. Rockville, MD: FDA; 2010 Feb. Guidance for industry: Adaptive design clinical trials for drugs and biologics (draft). Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf> [accessed 20/03/2011]
7. U.S. Food and Drug Administration. Rockville, MD: FDA; 2010 Feb. Guidance for the use of Bayesian statistics in medical device clinical trials. Available from: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf> [accessed 27/02/2013]
8. Berry DA. Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science*. 2004; 19:175–187.
9. Chow, S-C.; Chang, M. *Adaptive Design Methods in Clinical Trials*. Boca Raton, FL; Chapman & Hall / CRC; 2006.
10. Palmer CR, Rosenberger WF. Ethics and practice: alternative designs for phase III randomized clinical trials. *Controlled Clinical Trials*. 1999; 20:172–186. [PubMed: 10227416]
11. Pullman D, Wang X. Adaptive designs, informed consent, and the ethics of research. *Controlled Clinical Trials*. 2001; 22:203–210. [PubMed: 11384785]
12. Bartlett RH, Roloff DW, Cornell RG, Andrews AF, Dillon PW, Zwischenberger JB. Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. *Pediatrics*. 1985; 76:479–487. [PubMed: 3900904]
13. Rosenberger WF, Lachin JM. The use of response-adaptive designs in clinical trials. *Controlled Clinical Trials*. 1993; 14:471–484. [PubMed: 8119063]
14. Lewis RJ, Berry DA. Group sequential clinical trials: a classical evaluation of Bayesian decision-theoretic designs. *Journal of the American Statistical Association*. 1994; 89:1528–1534.
15. Lewis RJ, Lipsky AM, Berry DA. Bayesian decision-theoretic group sequential clinical trial design based on a quadratic loss function: a frequentist evaluation. *Clinical Trials*. 2007; 4:5–14. [PubMed: 17327241]
16. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2009. Available from: <http://www.R-project.org> [accessed 14/08/09]
17. Smith BJ. boa: An R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software*. 2007; 21:1–37.
18. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*. 2000; 10:325–337.
19. Ashby D, Tan SB. Where's the utility in Bayesian data-monitoring of clinical trials? *Clinical Trials*. 2005; 23:197–205. [PubMed: 16279143]
20. Cheng Y, Su F, Berry DA. Choosing sample size for a clinical trial using decision analysis. *Biometrika*. 2003; 90:923–936.
21. Lipsky AM, Greenland S. Confounding due to changing background risk in adaptively randomized trials. *Clinical Trials*. 2011; 8:390–397. [PubMed: 21610005]
22. van der Graaf R, Roes KC, van Delden JJ. Adaptive trials in clinical research: scientific and ethical issues to consider. *JAMA*. 2012; 307:2379–2380. [PubMed: 22692169]

Figure 1.

Expected number of subjects and predictive distribution of $\text{var}(\theta)$ in each trial design, from simulation. The left vertical axis corresponds to the expected number of subjects in each of the trial designs; the right vertical axis corresponds to the predictive distribution of $\text{var}(\theta)$. Panel A corresponds to the first set of trials (fixed estimation error parameters), and Panel B corresponds to the second set of trials (fixed simulated frequentist errors). The base trial is the same in both panels.

Figure 2.

Expected number of subjects enrolled in each of the trial designs, by arm and outcome, and frequentist error rates, from simulation. Panel A corresponds to the first set of trials (fixed estimation error parameters), and Panel B corresponds to the second set of trials (fixed simulated frequentist errors). The base trial is the same in both panels.

Figure 3.

Cumulative percentage of stop decisions at each block for each of the trial designs, from simulation. A decision to stop at block 0 is equivalent to not starting the trial; a decision to stop at block 4 is equivalent to enrolling all planned subjects. Panel A corresponds to the first set of trials (fixed estimation error parameters), and Panel B corresponds to the second set of trials (fixed simulated frequentist errors). The base trial is the same in both panels.

Figure 4.

Breakdown of total cost of each trial design by source of cost, from simulation. The sources include cost of subject enrollment (e), cost of an individual subject experiencing the worse (“failure”) outcome (d), and cost of parameter (θ) misestimation. For all trials, $e = 1$. For the two trial designs on the left of each panel, $d = 0$; the two designs on the right of each panel have $d = 50$. Panel A corresponds to the first set of trials (fixed estimation error parameters), and Panel B corresponds to the second set of trials (fixed simulated frequentist errors). See text for specification of cost from parameter misestimation. The base trial is the same in both panels.

Trial design characteristics. The base trial was designed to yield specific simulated type I and II error rates. The estimation error parameters used to design the base trial were then used to design the rest of the fixed estimation error parameter trials; specific simulated type I and II error rates were not targeted. The estimation error parameters for the last three trials were chosen explicitly to give specific simulated type I and II frequentist error rates.

Table 1

	Trial Design	Adaptive Randomization	Failure Cost	Available Randomization Ratios (r_i)	Enrollment Cost (e)	Failure Cost (d)	Estimation Error	
							K_0	R_K
Fixed Estimation Error Parameters	Base	No	No	1:1	1	0	190	3.23
	response-adaptive	Yes	No	1:3 1:1 3:1	1	0	190	3.23
	failure cost	No	Yes	1:1	1	50	190	3.23
	response-adaptive/failure cost	Yes	Yes	1:3 1:1 3:1	1	50	190	3.23
Targeted Type I & II Errors	Base	No	No	1:1	1	0	190	3.23
	response-adaptive	Yes	No	1:3 1:1 3:1	1	0	60	2.95
	failure cost	No	Yes	1:1	1	50	5608	3.18
	response-adaptive/failure cost	Yes	Yes	1:3 1:1 3:1	1	50	2100	2.96

Error rates, enrollment, and costs for all trials. Simulations used fixed values of p_C , θ to determine error rates, and values randomly sampled from the prior to determine expected number of subjects and cost. Backward induction values were reported by the software for node 0. Corrected Cost adds 50 units to the simulated cost for each subject failure if this cost has not already been included in the trial design.

Table 2

	Trial Design	Simulation				Backward Induction		Corrected Cost
		Fixed p_C, θ		$p_C, \theta \sim$ priors		No. Subs	Cost	
		Type I Error	Type II Error	No. Subs	Cost			
Fixed Estimation Error Parameters	Base	0.050	0.199	24.8	250.8	24.8	248.2	870.5
	response-adaptive	0.032	0.225	24.5	220.1	24.5	222.9	831.6
	failure cost	0.182	0.299	13.8	647.3	13.8	645.1	647.3
	response-adaptive/failure cost	0.115	0.336	13.6	587.2	13.6	585.8	587.2
Targeted Type I & II Errors	Base	0.050	0.199	24.8	250.8	24.8	248.2	870.5
	response-adaptive	0.050	0.198	21.2	82.9	21.2	82.5	616.3
	failure cost	0.050	0.200	25.3	7037.5	25.3	7154.2	7037.5
	response-adaptive/failure cost	0.050	0.202	22.7	2622.9	22.7	2680.3	2622.9

No. Subs – Expected number of subjects enrolled.

Table 3

A. Stopping and continuation decision costs at the node where 14 of 16 subjects (88%) have experienced the successful outcome in the control arm, and 3 of 8 subjects (38%) have experienced the successful outcome in the treatment arm. The stop decisions include control better than treatment ($C > T$), equivalent to treatment ($C = T$), and worse than treatment ($C < T$). The continuation decisions determine the odds with which the next enrolled block is randomized to control:treatment, where non-1:1 randomization is allowed. The minimum cost in each row is boldfaced.									
	Trial Design	Stop			Continue				
		$C > T$	$C = T$	$C < T$	3:1	1:1	1:3		
Fixed Estimation Error Parameters	Base	301.6	1165.6	12699.3		270.0			
	response-adaptive	300.1	1165.6	12699.4	281.0	270.2		264.9	
	failure cost	651.6	1515.6	13049.3		777.1			
	response-adaptive/failure cost	650.1	1515.6	13049.4	743.9	777.3		816.2	
Targeted Type I & II Errors	Base	301.6	1165.6	12699.3		270.0			
	response-adaptive	104.4	383.6	3675.0	104.9	101.9		100.3	
	failure cost	8309.7	34113.0	369129.3		7476.2			
	response-adaptive/failure cost	3197.8	12961.0	128591.9	3055.0	2994.1		2979.5	

B. Stopping and continuation decision costs at the node where 1 of 11 subjects (9%) have experienced the successful outcome in the control arm, and 7 of 13 subjects (54%) have experienced the successful outcome in the treatment arm. The stop decisions include control better than treatment ($C > T$), equivalent to treatment ($C = T$), and worse than treatment ($C < T$). The continuation decisions determine the odds with which the next enrolled block is randomized to control:treatment, where non-1:1 randomization is allowed. The minimum cost in each row is boldfaced.									
	Trial Design	Stop			Continue				
		$C > T$	$C = T$	$C < T$	3:1	1:1	1:3		
Fixed Estimation Error Parameters	Base	11482.5	978.3	406.4		340.8			
	response-adaptive	11429.4	969.8	413.7	333.4	340.7		356.7	
	failure cost	12282.5	1778.3	1206.4		1402.4			
	response-adaptive/failure cost	12229.4	1769.8	1213.7	1431.8	1401.9		1380.6	
Targeted Type I & II Errors	Base	11482.5	978.3	406.4		340.8			
	response-adaptive	3318.0	323.7	136.4	121.3	123.2		127.9	
	failure cost	332567.8	28807.7	12143.5		10117.6			
	response-adaptive/failure cost	116503.9	11313.3	4770.7	4263.9	4294.8		4421.2	

Table 4

A. Simulated trial characteristics with fixed values of p_C and θ that correspond to those used to derive the type I error rates. Corrected Cost adds 50 units for each subject failure if this cost has not already been included in the trial design.						
	Trial Design	$p_C = 0.5, \theta = 0 \ (p_T = 0)$				
		Total Subjects Enrolled	Subjects in Control Arm (%)	Total Failures (%)	Cost	Corrected Cost
Fixed Estimation Error Parameters	base	25.2	12.6 (50.0)	12.6 (49.9)	174.5	804.6
	response-adaptive	25.0	11.6 (46.4)	12.5 (49.9)	120.6	745.1
	failure cost	12.6	6.3 (49.9)	6.3 (49.9)	866.7	866.7
	response-adaptive/failure cost	13.3	5.1 (38.6)	6.6 (49.9)	684.7	684.7
Targeted Type I & II Errors	base	25.2	12.6 (50.0)	12.6 (49.9)	174.5	804.6
	response-adaptive	21.3	9.6 (44.9)	10.6 (49.9)	64.0	595.8
	failure cost	25.6	12.8 (50.0)	12.8 (49.9)	4957.4	4957.4
	response-adaptive/failure cost	22.9	10.3 (44.9)	11.4 (49.9)	2082.4	2082.4

B. Simulated trial characteristics with fixed values of p_C and θ that correspond to those used to derive the type II error rates. Corrected Cost adds 50 units for each subject failure if this cost has not already been included in the trial design.						
	Trial Design	$p_C = 0.25, \theta = 2.197 \ (p_T = 0.75)$				
		Total Subjects Enrolled	Subjects in Control Arm (%)	Total Failures (%)	Cost	Corrected Cost
Fixed Estimation Error Parameters	base	25.6	12.8 (50.0)	12.8 (50.0)	208.4	848.5
	response-adaptive	26.6	11.9 (44.8)	12.6 (47.4)	233.2	863.3
	failure cost	12.5	6.3 (49.9)	6.3 (50.0)	654.3	654.3
	response-adaptive/failure cost	14.5	4.7 (32.2)	5.9 (41.1)	630.0	630.0
Targeted Type I & II Errors	base	25.6	12.8 (50.0)	12.8 (50.0)	208.4	848.5
	response-adaptive	23.3	10.3 (44.3)	11.0 (47.2)	80.6	630.0
	failure cost	26.4	13.2 (50.0)	13.2 (50.0)	6107.5	6107.5
	response-adaptive/failure cost	25.3	9.9 (39.0)	11.2 (44.5)	2625.6	2625.6