Published in final edited form as:

Commun Stat Theory Methods. 2017 ; 46(6): 2823–2836. doi:10.1080/03610926.2015.1053929.

Optimal and lead-in adaptive allocation for binary outcomes: a comparison of Bayesian methodologies

Roy T. Sabo and Ghalib Bello

Department of Biostatistics, Virginia Commonwealth University, 830 East Main Street, Richmond, VA 23298-0032, U.S.A

Abstract

We compare posterior and predictive estimators and probabilities in response-adaptive randomization designs for two- and three-group clinical trials with binary outcomes. Adaptation based upon posterior estimates are discussed, as are two predictive probability algorithms: one using the traditional definition, the other using a skeptical distribution. Optimal and natural lead-in designs are covered. Simulation studies show: efficacy comparisons lead to more adaptation than center comparisons, though at some power loss; skeptically predictive efficacy comparisons and natural lead-in approaches lead to less adaptation but offer reduced allocation variability. Though nuanced, these results help clarify the power-adaptation trade-off in adaptive randomization.

Keywords

Adaptive Randomization; Bayesian Methods; Clinical Trials; Predictive Probability

1 Introduction

Optimal outcome-adaptive randomization designs for trials with binary outcomes are available in two- and three-group cases (Rosenberger *et al.*, 2001; Tymofeyev, Rosenberger and Hu, 2007; Jeon and Hu, 2010), where the objective criterion is to minimize the number of treatment failures. In practice, these designs require the estimation of unknown, treatment-specific success rates with sample-based estimators. Though sample proportions are often used for this purpose, Bayes estimators (such as posterior means or modes) could also be used, which may help avoid small-sample deficiencies of the maximum likelihood estimators. Though not technically appropriate for use in optimal designs, estimators based on comparative treatment efficacy have also been proposed (Thall and Wathen, 2007), with estimates based on the posterior probability that one treatment is more successful than another. Thall and Wathen also introduced (in the same study) the concept of a “natural lead-in”, where allocation weights are restricted from changing too much early in a trial, and are gradually granted flexibility to change as more patients accrue.

With respect to the efficacy-based Bayes estimators, predictive probabilities could also be used. While posterior probabilities condition an event on the evidence and variability inherent in the prior information and the currently observed data, predictive probabilities go a step further and also condition on variability due to unobserved data from subjects not yet accrued into the trial. The benefits of using predictive or conditional probabilities in adaptive clinical trial designs have been outlined in several places – notably in Berry (2004 and 2006) – especially with respect to predicting the likelihood of events, event monitoring (Spiegelhalter, Freedman and Blackburn, 1986), sample size estimation (Spiegelhalter and Freedman, 1986; Lecoutre, Derzko and Grouin, 1995) and re-estimation, and early-termination decisions (Herson, 1979; Choi, Smith and Becker, 1985; Lee and Liu, 2008). Since adaptive allocation procedures based on predictive probabilities would account for greater uncertainty due to unaccrued subjects, such methods may be anticipated to adapt more slowly than methods based on posterior estimates and could behave more like natural lead-in methodologies rather than optimal designs. While optimal adaptation methods based on posterior estimates may lead to greater treatment successes, the less-imbalanced designs from predictive-probability-based and lead-in methods may lead to greater statistical power. The trade-off between power and the potential to allocate more patients to a better treatment using outcome-adaptive allocation with Bayesian estimation remains unclear.

The remainder of this manuscript is outlined as follows. Section 2 presents a brief description of outcome-adaptive allocation procedures in the two- and three-group cases, using both optimal and natural lead-in designs. In Section 3 we discuss Bayes methodologies as used in adaptive allocation, including posterior estimates (mean, mode and efficacy), and present two algorithms using predictive probabilities: one using the traditional conceptualization and another using treatment skepticism to drive the prediction. We motivate these methods by focusing on the conjugate *beta-binomial* model for each treatment group, as outlined in Section 3.1.1 of Ashby, Hutton and McGee (1993). Simulation studies are presented in Section 4, where we focus on the within-study behavior of the allocation weights, as well as end-of-study characteristics like power and error rates. We end with a brief summary and discussion in Section 5. For purposes of clarity, the optimization criterion to which “optimal allocation” refers is to minimize the number of treatment failures, as mentioned in Rosenberger *et al.* (2001).

2 Optimal and Natural Lead-In Designs

2.1 Optimal Allocation in Two- and Three-Group Trials

Rosenberger *et al.* (2001) derived optimal allocation weights for two-group studies with binary outcomes, as shown below in Equation 1,

$$\begin{aligned} w_1 &= \frac{\sqrt{p_1}}{\sqrt{p_1} + \sqrt{p_2}}, \\ w_2 &= 1 - w_1, \end{aligned} \quad (1)$$

where weight w_j ($j = 1, 2$) is the probability that the next patient will be allocated into the j^{th} treatment group, and p_1 and p_2 are the (unknown) success proportions in each group. Tymofeyev, Rosenberger and Hu (2007) outlined a numerical method to obtain optimal

allocation weights in the three-group case, while Jeon and Hu (2010) found closed-form expressions under various circumstances of both the marginal probabilities (p_1 , p_2 and p_3) and the term $B \in (0, 1/3)$, which serves as a lower bound for the allocation weights.

The optimal designs in both the two- and three-group cases require estimators of the marginal probabilities. As originally conceived, the population proportion p_j is replaced with the corresponding sample proportion $\hat{p}_j$ of the number of observed successes out of the number of accrued subjects in arm j . Bayes estimators of these rates can also be used, as is discussed below in Sections 3.1 and 3.2.

2.2 Allocation Using a Natural Lead-In

Complications can arise in early phases of an adaptively randomized clinical trial due to low accrual. Sample proportions may vary considerably during early phases, or may be zero-valued if no successes are observed (whether due to low efficacy or delayed observation). Thall and Wathen (2007) provided an elegant solution by making the radical from the optimal two-group design (Equation 1) an increasing function of the observed (n) sample size relative to the planned total (N):

$$\begin{aligned} w_1 &= \frac{p_1^{n/2N}}{p_1^{n/2N} + p_2^{n/2N}}, \\ w_2 &= 1 - w_1. \end{aligned} \quad (2)$$

This “natural lead-in” approach begins at equal allocation and slowly phases in any adaptation as the obtained sample size approximates the planned sample size. Though the original authors used Bayes estimators of the probability of treatment efficacy, the use of success probability estimators in each group will lead to allocation weights that eventually approximate the optimal two-group design as the trial nears completion.

For the three-group optimal allocation weights provided by Tymofyeyev, Rosenberger and Hu (2007) and Jeon and Hu (2010), a simple natural lead-in approach is to raise each “optimal” allocation weight (w_j) to the n/N power, such that

$$\begin{aligned} w_1^* &= \frac{w_1^{n/N}}{w_1^{n/N} + w_2^{n/N} + w_3^{n/N}}, \\ w_2^* &= \frac{w_2^{n/N}}{w_1^{n/N} + w_2^{n/N} + w_3^{n/N}}, \\ w_3^* &= 1 - w_1^* - w_2^*. \end{aligned} \quad (3)$$

Here w_j is the optimal three-group allocation rate for the j^{th} group, and w_j^* is the updated allocation rate for the j^{th} group. These weights begin at equal allocation when $n = 0$ ($w_j^* = 1/3, j = 1, 2, 3$) and gradually shift toward the optimal allocation as $n \rightarrow N$.

3 Bayes Estimators In Adaptive Allocation

In the Bayes framework, the parameters $p_j, j = 1, \dots, k$, for each treatment group (which here represent proportions) are assigned a prior distribution $\pi(\phi)$, where $\pi(\cdot)$ is some distributional form and ϕ is some hyper-parameter(s). These prior specifications need not be identical, though we will assume as such for simplicity. The prior distributions are combined with likelihood distributions $p(y_j|p_j, n_j)$ for each treatment group, where $p(\cdot)$ is some distributional form, n_j is the number of observed patients in treatment group j , and y_j is the number of “successful” events observed. The specific choice of prior and likelihood are then synthesized into a posterior distribution

$$P(p_j|y_j, n_j, \phi) \propto p(y_j|p_j, n_j)\pi(\phi), j=1, \dots, k, \quad (4)$$

which is our estimate of the probability distribution of parameter p_j .

3.1 Posterior Estimates and Probabilities

For conjugate models or simple distributions, posterior means or modes can be calculated directly from the posterior distribution listed in Equation 4. These estimators may be directly incorporated into the two- and three-group optimal designs discussed in Section 2.1 or the lead-in designs in Section 2.2. In more complicated scenarios, such as in comparing success rates between treatments, integration or markov chain monte carlo (MCMC) methods can be used to obtain the joint distribution (or sampling distribution) of any two parameters. In two-arm trials we need only calculate one probability

$$\begin{aligned} P_1 &= P(p_1 > p_2 | y, n, \phi) \\ P_2 &= 1 - P_1, \end{aligned} \quad (5)$$

where $y = (y_1, y_2)$ and $n = (n_1, n_2)$. In three-arm trials we calculate three probabilities

$$\begin{aligned} P_1 &= [(p_1 > p_2) \cap (p_1 > p_3) | y, n, \phi], \\ P_2 &= [(p_2 > p_1) \cap (p_2 > p_3) | y, n, \phi], \\ P_3 &= [(p_3 > p_1) \cap (p_3 > p_2) | y, n, \phi], \end{aligned} \quad (6)$$

where $y = (y_1, y_2, y_3)$ and $n = (n_1, n_2, n_3)$. In practice, these posterior probabilities are used in place of the unknown population success rate for the corresponding group. As the optimal and lead-in designs discussed in Section 2 are based on success proportions (p_1 , p_2 and p_3), the use of posterior probabilities of efficacy comparisons between groups in Equations 5 and 6 in these allocation approaches is *ad hoc*.

3.2 Predictive Probabilities

To use predictive probabilities in adaptive allocation, we begin with the posterior distributions given in Equation 4. Many predictive probability approaches in clinical trials use the current posterior probability distribution as the new prior, and combine this

information with some likelihood for the patients who have yet to accrue or whose outcomes are currently unobserved. These predictive distributions (one for each treatment) allow the calculation of the probability of interest. Several studies (Herson, 1979; Lecoutre, Derzko and Grouin, 1995; Lee and Liu, 2008) use the *beta-binomial* distribution to model uncertainty due to unobserved binary outcomes as part of the derivation of predictive probabilities. As explained in Lee and Liu (2008), these methods rely upon some threshold value θ_T in the computation of the predictive probability, the choice of which may be seen as arbitrary. We must also note that the predictive probability models from these studies were not used for adaptive allocation.

The problem with calculating predictive probabilities in adaptive allocation is that simulating from the predictive distribution produces similar estimates to those obtained from simulating from the posterior distribution (results which are verified below in Section 4).

This is due to p_j^* and p_j (see Steps 1–6 below) being simulated from sampling distributions with the same center, which can be seen by adapting a commonly known result of expectations ($E(Y) = E_\theta(E(Y|\theta))$) as follows: $E(p) = E_Y(E(p|Y))$. An alternative approach relies upon the *re-use* of skeptical prior information to calculate predictive probabilities. Rather than assume that future patients will behave similarly to patients already accrued into the trial, we return to our skeptical assumptions expressed in the prior distribution $\pi(\phi)$ to conservatively account for uncertainty in the non-accrued patients (more on the use of skeptical priors will be provided in Section 3.3). The rationale for using this skeptically predictive approach is to avoid the assumption that there are no time-based biases in patient accrual or treatment effectiveness, an issue raised by Korn and Freidlin (2011) in their critique of outcome-adaptive allocation.

The steps and general process of calculating predictive probabilities using both the traditional and skeptically-predictive methods are the same, with the only difference being the distribution that generates the binomial proportions $p_j (j = 1, \dots, k)$ used to simulate the unobserved outcomes. In both cases, predictive probabilities are obtained by generating approximations of the sampling distributions for the p_j using the following six-step process:

Step 1: Simulate p_j^* using posterior distribution from Equation 4 (traditional method) or skeptical prior $\pi(\phi)$ (skeptically-predictive method).

Step 2: Calculate remaining number of subjects in each treatment:

$$n_j^* = (n - \sum_i n_i) / k, j = 1, \dots, k \text{ These values may need to be appropriately rounded.}$$

Step 3: Simulate future number of successes for each treatment:

$$y_j^* \sim \text{binomial}[n_j^*, p_j^*], j = 1, \dots, k.$$

Step 4: Update posterior distribution for p_j using observed ($y = (y_1, \dots, y_k)$) and future ($y^* = (y_1^*, \dots, y_k^*)$) outcomes, replacing the y_j and n_j expressions in Equation 4 with $y_j + y_j^*$ and $n_j + n_j^*$, respectively.

Step 5: Simulate $p_j (j = 1, \dots, k)$ from the updated posterior distributions, calculating

$$I[\cap_{i=1, \neq j}^k (p_j > p_i)] \text{ for each } j = 1, \dots, k, \text{ where } I(\cdot) \text{ is an indicator function.}$$

Step 6: Repeat steps 1–5 T times, where T is an appropriately large number.

For two-group studies, the predictive probabilities for between-treatment comparisons are estimated as

$$P_1 = P(p_1 > p_2 | \phi, y, y^*, n, n^*) = \sum_{t=1}^T I(p_1 > p_2) / T, \\ P_2 = 1 - P_1 \quad (7)$$

For three-group studies, the predictive probability for each group is estimated as

$$P_1 = P[(p_1 > p_2) \cap (p_1 > p_3) | \phi, y, y^*, n, n^*] = \sum_{t=1}^T I\left[\bigcap_{i=2}^3 (p_1 > p_i)\right] / T, \\ P_2 = P[(p_2 > p_1) \cap (p_2 > p_3) | \phi, y, y^*, n, n^*] = \sum_{t=1}^T I\left[\bigcap_{i=1, \neq 2}^3 (p_2 > p_i)\right] / T, \\ P_3 = P[(p_3 > p_1) \cap (p_3 > p_2) | \phi, y, y^*, n, n^*] = \sum_{t=1}^T I\left[\bigcap_{i=1}^2 (p_3 > p_i)\right] / T. \quad (8)$$

The predictive probabilities given in Equations 7 and 8 can then be incorporated in two- and three-group optimal and lead-in designs in the same manner as the posterior efficacy comparisons.

3.3 Beta-Binomial Model for Binary Outcomes

The conjugate *beta-binomial* pair form a natural choice for prior and likelihood, respectively, when the outcome of interest is the proportion of successful outcomes observed in a given trial. Selecting some common skeptical value for the prior information in both treatments ($\phi = p_s$) such that the mode of the prior is centered at that value yields

$\text{beta}[1 + \tilde{n}p_s, 1 + \tilde{n}(1 - p_s)]$ prior distributions for each treatment, where $\tilde{n}$ can be viewed as the amount of researcher-provided information (in a “number of subjects” sense) within the effective sample size $(1 + \tilde{n}p_s + 1 + \tilde{n}(1 - p_s) = 2 + \tilde{n})$ contained within the prior. Choosing *binomial* $[n_j, p_j]$ likelihood functions yields $\text{beta}[1 + \tilde{n}p_s + x_j, 1 + \tilde{n}(1 - p_s) + (n_j - x_j)]$ posterior distributions for each treatment group, according to Equation (4). The updated posterior distributions used in Step 4 for the predictive probabilities (including the unobserved data y^* and n^*) then becomes

$\text{beta}\left[1 + \tilde{n}p_s + x_j + x_j^*, 1 + \tilde{n}(1 - p_s) + (n_j + n_j^* - x_j - x_j^*)\right]$. Note that large values of $\tilde{n}$ will result in slower adaptation, as the prior information will outweigh the observed data as long as $\tilde{n} > n$ (or $\tilde{n} < n^*$) (see Sabo, 2014), though this choice is less important in the natural lead-in methods (Equations 2 and 3), since the adaptation in these cases is restricted by n . For simplicity, we set $\tilde{n} = 1$, which implies that the posterior means and modes for each group are $(x_j + 1 + p_s)/(n_j + 3)$ and $(x_j + p_s)/(n_j + 1)$, respectively.

4 SIMULATION STUDIES

4.1 Simulation Template

Simulations are used to compare the performance of each of the adaptive allocation methodologies previously discussed, in both two- and three-group trials, in optimal and natural lead-in designs, using estimators of center and treatment efficacy, and for the latter using both posterior and predictive probabilities. For both two- and three-group trials, we create two simulation templates that reflect moderate treatment differences: one with lower success rates and one with moderate success rates close to 0.5. In the two-group case we use $p_1 = 0.25$ and $p_2 = 0.1$ as low success rates and $p_1 = 0.55$ and $p_2 = 0.4$ as moderate success rates, while in the three-group case we use $p_1 = 0.25$, $p_2 = 0.15$ and $p_3 = 0.1$ as low success rates and $p_1 = 0.55$, $p_2 = 0.45$ and $p_3 = 0.4$ as moderate success rates. These choices represent instances of low outcome variability (low success rate) and high outcome variability (moderate success rate). Due to the symmetric behavior of binomial rates about 0.5, we will not focus on larger success rates (> 0.5).

Each simulated trial begins at equal allocation ($w_j = 1/2$ for two-groups; $w_j = 1/3$ for three-groups), and the weights are allowed to adapt after the first subject is allocated and its outcome is observed. For the three-group case, we place some restrictions on the adaptation by setting the minimum allocation ratio as $B = 0.2$ in order to minimize the number of patients allocated away from the second group. We generate a *uniform*[0,1] random number for each simulated patient in each trial, and that patient is allocated accordingly based on the current allocation weights. The outcome for that patient is simulated as a *bernoulli*(p_j) random variable based on the efficacy of the j^{th} treatment. The allocation ratio is then updated using all available outcomes, and we repeat this process for $n = 1, \dots, N$ simulated patients, which represents a complete trial. Total sample sizes of $N = 200$ (two-group / low success rates), $N = 352$ (two-group / moderate success rates), $N = 345$ (three-group / low success rates), and $N = 618$ (three-group / moderate success rates) were chosen in order to approximate 80% power in balanced designs using a chi-square test with one degree of freedom in the two-group case and two degrees of freedom in the three-group case; we also assumed two-sided alternative hypotheses and a 5% type-I error rate. For each of the four cases listed above, we use each of the following adaptive randomization methods: posterior mean (Post Mean), posterior mode (Post Mode), posterior efficacy comparisons (Post Eff), skeptically predictive efficacy comparisons (Skep Pred), and traditional predictive efficacy comparisons (Trad Pred); each approach is used in both the optimal and natural lead-in designs. A balanced design is also used for comparison, leading to a total of 11 trial-types for each case. For each adaptive allocation method we use prior distributions that are both skeptical and mildly informative $\text{beta}[1 + p_S, 2 - p_S]$, where the distributions for both treatment groups are centered on a mode of $p_S = p_2$ or $p_S = p_3$, reflecting the skeptical belief of no efficacy difference between treatments. A total of $T = 5,000$ simulations are run for each set of parameter values and each adaptation method.

The total number of successful outcomes and the total number of patients allocated to each group are summarized across all simulated trials for each method in each case. Empirical power and error rates are also estimated, with power estimated as the proportion of

simulations where the null hypothesis was properly rejected, and with the error rate estimated as the proportion of simulations where the null hypothesis was improperly rejected, where each test is based on the corresponding two-sided chi-square test with a 5% significance level. In addition to those end-of-study measures, within-study mean allocation ratios between the first group and all others are summarized across all simulations at the 25th, 50th, 75th and 100th percentiles of patient accrual. Means and standard errors are provided for each measure except power and error rates, which are summarized with proportions.

4.2 Two-Sample Case

Table 1 presents simulation results for the low success rate case using optimal allocation. Compared to the balanced case, which averaged 35 successes, each of the adaptive methods yielded greater expected successes. Using posterior means, posterior modes and the skeptically predictive efficacy comparisons yielded modest increases, while the efficacy comparisons based on both posterior and predictive probabilities yielded the greatest increase (about 8 successes, on average). Naturally, those two approaches experienced greater variability in their success rates, and suffered from power losses (68% and 70%, respectively). By contrast, the posterior mean and mode and the skeptically predictive approach did not experience much power loss. The within-study variability in allocation ratios for the two efficacy-comparison methods were greater than those for the posterior mean and skeptically predictive approaches. The allocation ratio variability for the posterior mean and mode approaches actually decreased over time, unlike all others. Note that the results for the posterior and traditional predictive efficacy comparison methods are nearly identical, supporting our position that these two approaches are practically equivalent. Another peculiarity to note is that the allocation ratios for the posterior mode method actually decrease over time. This is due in part to the low efficacy rates assumed ($p_1 = 0.25$ and $p_2 = 0.1$), where the difference $\delta = 0.15$ means more than if the rates were closer to 0.50. Another reason is the closed-form expression for the posterior mode in this case ($x_j + 0.1)/(n_j + 1)$, which is near zero in the second group when no successes are observed, thus driving up the allocation ratios and their variability.

Results for the natural lead-in allocation designs with low success rates are found in Table 2. These methods provided smaller increases in the expected number of successes than their optimal design counterparts, but the efficacy-comparison-based methods still had substantial increases over the balanced case (Table 1); these three methods also regained much of the power lost in the optimal design case. Interestingly, the use of a natural lead-in altered the variability trajectory in the allocation ratio from the posterior mean and mode approaches, as here the variability of the allocation ratios increases over a trial, where in the optimal design case the variability of the allocation ratios decreased. Closer inspection shows that allocation ratio variabilities are lower at the start of the trial in the natural lead-in case because the adaptation is constrained by low values of n/N ; thus the increased variability observed in early the unconstrained optimal design cases is not seen. As in the optimal design scenario, the skeptically predictive adaptation method had the lowest variability in allocation ratios for the first three-quarters of the trial, but that variability quickly increased over the last phase as

the actual data began to overwhelm the skeptical beliefs from the predictive prior distribution.

The moderate success rate results using optimal design methods are found in Table 3, where we see that the posterior mean and mode methods provide little improvement in success rates over the balanced case, which provided 167.1 successes on average. In contrast, adaptation using the efficacy-based comparisons yielded a substantial increase in expected successes, again doing so at the cost of power ($\sim 73\%$). The skeptically predictive approach had a modest increase in expected successes (~ 6.2), but maintained the desired level of power and had the lowest measures of variability in all aspects among the three efficacy comparison methods.

After applying the natural lead-in adaptation approaches to the moderate success rate case (Table 4), we see that each method – as expected – had less adaptation than in the optimal design case except for skeptically predictive approach, which here provided similar results with slightly less variability. The two efficacy-based comparison methods regained much of the power lost in the optimal design case, but at the cost of a reduced number of expected successes, which here numbered ~ 10.5 patients more than in the balanced case.

4.3 Three-Sample Case

End-of-study results for optimal design methods in the low success rate case are provided in Table 5. Each approach averaged more successes than the balanced case (~ 57.5), with the efficacy comparison methods providing a larger increase than the measure-of-center-based methods. Unlike in the two-group case, the loss in power for all methods was slight, though there was a slight increase in error rates. The variability in the expected number of successes and group-specific sample sizes was somewhat more homogeneous in this three-group case than it was in the two-group case. Note that in some adaptive cases the average number of patients allocated to the third group was larger than that for the second group, which is an artifact of the optimal three-group optimal design and is not an error in the simulations. Increasing the minimum allocation limit B (discussed in Section 2.1) could help alleviate this phenomenon, but would come at the expense of adaptation.

Continuing with results for the optimal design approaches in the low success rate case, Table 6 shows the within-study allocation ratios between the first group and both the second and third groups. Here we see the variability in the ratios is somewhat consistent across methods at each point in the trial, though the skeptically predictive approach has lower variance during early phases of the trial. The two efficacy comparison-based methods have larger allocation ratios than other methods throughout the trial, but this difference gradually decreases.

The natural lead-in approaches provide less adaptation and fewer successes than the optimal designs in the low success rate case (Table 7), with the decrease more pronounced for the posterior and traditional predictive methods than for the other three. The variabilities for each method lower than those from the corresponding optimal design in Table 5, though the skeptically predictive design still has the lowest variability. The estimated power and error rate proportions are also comparable and are not meaningfully different from the desired

levels. In Table 8 we see the estimated allocation ratios for each of the natural lead-in methods for the low success rate case. Here we see that the allocation ratio variabilities are everywhere low and are relatively homogeneous across methods.

Results for the moderate success rate case using three-group optimal adaptive allocation designs are found in Table 9, where we see that while the posterior measure-of-center methods offer only a modest increase in the expected number of successes from the balanced case (289.1), the efficacy-based-comparison models provide at least 9 more successes on average, with the posterior-based and traditional predictive-based efficacy comparisons offering an improvement of 16. In this case none of the adaptive methods suffered a power loss, and all error rates were small. The variabilities for all measures were again homogeneous across each method, though the final sample sizes for the adaptive methods were more than three times as variable as for the balanced case. The skeptically predictive approach again had the lowest variability in all measures among the adaptive designs.

Within-study allocation ratios for the three-group, moderate success rate, optimal design case are provided in Table 10. Here we see that the allocation ratios for the posterior and traditional predictive efficacy-based comparison methods are more variable than the ratios for other methods at the beginning of the trial, are nearly equal at the mid-way point of patient accrual, and are thereafter lower throughout the remainder of the trial.

Using the three-group natural lead-in design for the moderate success rate case (Table 11), we see the expected number of successes for most methods are slightly less than what was obtained using the optimal design (Table 9). The variabilities for each end-of-study measure were fairly homogeneous across the approaches, and no method experienced a power loss or a substantial increase in error rate compared to the balanced case. As seen in Table 12, the allocation ratios were lower at the beginning of the trial than in the when using the optimal designs, though the ratios converged to the same values by the end of the trial. The ratios were also much less variable using the natural lead-in than when using the optimal design.

5 DISCUSSION

In summary, the adaptive allocation methods using Bayes estimators of comparative efficacy lead to substantial increases in treatment successes as compared to balanced designs in both the two- and three-group cases. While use of these probabilistic estimates are technically *ad hoc* when used in optimal designs (the natural lead-in approaches are *ad hoc* by definition except at the end of a trial), they provide more adaptation than using posterior mean or mode estimators. Though there is a substantial reduction in power for these approaches, the effect is reduced by coupling the efficacy-based comparisons with a natural lead-in component, which slightly reduces the expected improvement. While the use of skeptically predictive probabilities did not lead to substantial adaptation using optimal designs and for low success rates, they performed similarly to the posterior probability approach when coupled with a natural lead-in or when success rates were modestly large. The results from the previous section also showed that the use of traditional predictive probabilities is practically equivalent to the use of posterior probabilities. In two-group cases, it appears that adaptation through efficacy-based comparisons via posterior probabilities with a natural lead-in is the

best design for low success rate cases, and could also be selected for larger success rates, though those willing to sacrifice a slight decrease in the expected number of treatment successes could select the skeptically predictive approach with a natural lead-in if lower variability is desired. For three-group studies, optimal designs using posterior probabilities are recommended for both low and moderate success rates, as they achieve the greatest number of expected successes while maintaining sufficient power. These results are compatible with those found by Lee, Chen and Yin (2012), who compared Bayesian adaptive allocation (with efficacy comparisons) to fixed allocation in two- and three-sample scenarios.

While we only focused on two parametric simulation templates for the two- and three-group cases, the selections represent instances of low and high variability in the binary outcome response. These choices also reflected a modest efficacy difference between groups, as the rate difference was at most 15%. We purposefully omitted larger effect sizes to avoid providing overly-optimistic evidence of the benefit of adaptive allocation designs, though for such cases the benefits will surely be greater than that provided here. For the three-group lead-in approach, we used a minimum allocation weight of $B = 0.20$. While selecting a lower threshold could have allowed more allocation to the most successful treatment, it would have also allocated more to the worst (and away from the “middle” group), defeating the purpose adaptive allocation altogether. Other types of adaptive allocation approaches exist (See Chang (2008) for a comprehensive list), as do alternative methods to conduct a “lead-in”, though those were omitted to allow focus on more similar types of allocation, predominantly via the focus of optimal designs and lead-in approaches based upon those designs. It is left to future research to compare the performance of the adaptive allocation methods discussed in this manuscript with a more exhaustive complement of designs.

References

- Ashby D, Hutton JL, McGee MA. Simple Bayesian analyses for case-control studies in cancer epidemiology. *Journal of the Royal Statistical Society, Series D*. 1993; 42(4):385–397.
- Bello G, Sabo RT. Outcome-adaptive allocation with natural lead-in for three-group trials with binary outcomes. *Statistics in Biopharmaceutical Research*. (Submitted).
- Berry DA. Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science*. 2004; 19(1):175–187.
- Berry DA. Bayesian clinical trials. *Nature Reviews*. 2006; 5:27–36.
- Chang, M. A daptive design theory and implementation using SAS and R. New York: Chapman & Hall/CRC; 2008.
- Choi SC, Smith PJ, Becker DP. Early decision in clinical trials when the treatment differences are small. Experience of a controlled trial in head trauma. *Controlled Clinical Trials*. 1985; 6:280–288. [PubMed: 4075806]
- Herson J. Predictive probability early termination plans for phase II clinical trials. *Biometrics*. 1979; 35(4):775–783. [PubMed: 526523]
- Hu F, Zhang LX. Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *The Annals of Statistics*. 2004; 32:268–301.
- Jeon Y, Hu F. Optimal adaptive designs for binary response trials with three treatments. *Statistics in Biopharmaceutical Research*. 2010; 2(3):310–318.
- Korn EL, Freidlin B. Outcome-adaptive randomization: is it useful? *Journal of Clinical Oncology*. 2011; 29(6):771–776. [PubMed: 21172882]

- Lecoutre B, Derzko G, Grouin JM. Bayesian predictive approach for inference about proportions. *Statistics in Medicine*. 1995; 14:1057–1063. [PubMed: 7569500]
- Lee JJ, Chen N, Yin G. Worth adapting? Revisiting the usefulness of outcome-adaptive randomization. *Clinical Cancer Research*. 2012; 18(17):4498–4507. [PubMed: 22753588]
- Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. *Clinical Trials*. 2008; 5:93–106. [PubMed: 18375647]
- Rosenberger WF, Stallard N, Ivanova A, Haper CN, Ricks ML. Optimal adaptive designs for binary response trials. *Biometrics*. 2001; 57:909–913. [PubMed: 11550944]
- Sabo RT. Adaptive allocation for binary outcomes using decreasingly informative priors. *Journal of Biopharmaceutical Statistics*. 2014 In Press.
- Spiegelhalter DJ, Freedman LS, Blackburn PR. Monitoring clinical trials: conditional or predictive power. *Controlled Clinical Trials*. 1986; 7:8–17. [PubMed: 3956212]
- Spiegelhalter DJ, Freedman LS. A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. *Statistics in Medicine*. 1986; 5:1–13. [PubMed: 3961311]
- Thall PF, Wathen JK. Practical Bayesian adaptive randomization in clinical trials. *European Journal of Cancer*. 2007; 43(5):859–866. [PubMed: 17306975]
- Tymofyeyev Y, Rosenberger WF, Hu F. Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association*. 2007; 102(477):224–234.

Table 1

Optimal design results for two-group case with $p_1 = 0.25$, $p_2 = 0.1$ and $N = 200$.

	E(Successes)	E(n_1)	E(n_2)	r_{25}	r_{50}	r_{75}	r_{100}	Power	Error
Balanced	35.0	99.9	100.1	–	–	–	–	0.80	0.00
(SE)	5.44	6.92	6.92	–	–	–	–	–	–
Post Mean	37.8	118.5	81.5	1.5	1.6	1.6	1.6	0.81	0.00
(SE)	5.69	10.96	10.96	0.42	0.40	0.35	0.32	–	–
Post Mode	39.2	127.7	72.3	2.6	2.4	2.1	2.0	0.78	0.00
(SE)	6.12	19.97	19.97	1.86	1.87	1.69	1.47	–	–
Post Eff	43.2	154.2	45.8	3.7	6.3	8.7	10.6	0.68	0.00
(SE)	6.63	17.71	17.71	2.99	5.24	6.76	7.45	–	–
Skep Pred	38.2	120.7	79.3	1.1	1.3	1.9	12.2	0.79	0.00
(SE)	5.90	8.38	8.38	0.08	0.17	0.58	8.23	–	–
Trad Pred	43.2	154.4	45.6	3.6	6.4	8.9	10.8	0.70	0.00
(SE)	6.82	18.00	18.00	2.87	5.28	6.82	7.52	–	–

Table 2

Natural lead-in results for two-group case with $\nu_1 = 0.25$, $p_2 = 0.1$ and $N = 200$.

	E(Successes)	E(n_1)	E(n_2)	r_{25}	r_{50}	r_{75}	r_{100}	Power	Error
Post Mean	36.5	110.2	89.8	1.1	1.2	1.4	1.6	0.80	0.00
(SE)	5.57	8.27	8.27	0.08	0.15	0.22	0.29	–	–
Post Mode	36.8	112.5	87.5	1.2	1.3	1.5	1.7	0.81	0.00
(SE)	5.46	9.68	9.68	0.18	0.31	0.48	0.67	–	–
Post Eff	41.0	138.7	61.3	1.4	2.6	5.5	12.0	0.77	0.00
(SE)	6.20	13.16	13.16	0.32	1.13	3.13	8.12	–	–
Skep Pred	37.5	116.6	83.4	1.0	1.1	1.7	12.1	0.80	0.00
(SE)	5.81	7.86	7.86	0.02	0.08	0.39	8.12	–	–
Trad Pred	40.6	138.0	62.0	1.4	2.6	5.4	11.8	0.75	0.00
(SE)	6.15	13.21	13.21	0.31	1.13	3.06	7.89	–	–

Table 3

Optimal design results for two-group case with $p_1 = 0.55$, $p_2 = 0.4$ and $N = 352$.

	E(Successes)	E(n_1)	E(n_2)	r_{25}	r_{50}	r_{75}	r_{100}	Power	Error
Balanced	167.1	176.0	176.0	–	–	–	–	0.81	0.00
(SE)	9.38	9.35	9.35	–	–	–	–	–	–
Post Mean	169.3	189.5	162.5	1.2	1.2	1.2	1.2	0.80	0.00
(SE)	9.47	11.58	11.58	0.14	0.10	0.08	0.07	–	–
Post Mode	169.2	190.3	161.7	1.2	1.2	1.2	1.2	0.81	0.00
(SE)	9.38	12.12	12.12	0.16	0.10	0.08	0.07	–	–
Post Eff	182.6	277.4	74.6	5.4	7.7	9.9	11.5	0.73	0.00
(SE)	10.79	35.01	35.01	5.79	6.85	7.69	7.91	–	–
Skep Pred	173.3	216.4	135.6	1.1	1.3	2.2	12.1	0.80	0.00
(SE)	9.70	15.43	15.43	0.08	0.18	0.84	8.01	–	–
Trad Pred	182.5	277.4	74.6	5.3	7.8	10.0	11.6	0.73	0.00
(SE)	10.83	35.54	35.54	5.71	6.95	7.74	8.04	–	–

Table 4

Natural lead-in results for two-group case with $\nu_1 = 0.55$, $p_2 = 0.4$ and $N = 352$.

	E(Successes)	E(n_1)	E(n_2)	r_{25}	r_{50}	r_{75}	r_{100}	Power	Error
Post Mean	168.3	183.0	168.9	1.0	1.1	1.1	1.2	0.80	0.00
(SE)	9.35	9.82	9.82	0.03	0.04	0.06	0.07	–	–
Post Mode	168.3	183.2	168.8	1.0	1.1	1.1	1.2	0.80	0.00
(SE)	9.41	9.97	9.97	0.03	0.05	0.06	0.07	–	–
Post Eff	177.5	244.3	107.7	1.4	2.6	5.6	12.1	0.77	0.00
(SE)	9.94	22.71	22.71	0.34	1.15	3.19	8.14	–	–
Skep Pred	171.9	208.2	143.8	1.0	1.1	1.8	12.1	0.80	0.00
(SE)	9.59	13.20	13.20	0.02	0.08	0.51	7.91	–	–
Trad Pred	177.5	244.7	107.3	1.4	2.7	5.6	12.2	0.78	0.00
(SE)	9.88	22.91	22.91	0.33	1.15	3.16	8.11	–	–

Table 5

Optimal design results for three-group case with $p_1 = 0.25$, $p_2 = 0.15$, $p_3 = 0.1$ and $N = 345$.

	E(Successes)	E(n_1)	E(n_2)	E(n_3)	Power	Error
Balanced	57.6	115.1	114.8	115.1	0.80	0.00
(SE)	7.07	8.64	8.74	8.72	-	-
Post Mean	62.2	155.7	87.6	101.7	0.79	0.01
(SE)	7.76	24.49	17.51	18.05	-	-
Post Mode	62.9	159.2	87.7	98.1	0.79	0.01
(SE)	7.83	26.48	18.50	17.11	-	-
Post Eff	65.8	180.6	85.7	78.7	0.79	0.01
(SE)	7.99	25.11	21.82	10.62	-	-
Skep Pred	61.8	149.6	95.1	100.3	0.78	0.01
(SE)	7.58	16.41	16.91	16.36	-	-
Trad Pred	65.7	180.4	86.0	78.6	0.78	0.01
(SE)	7.94	25.38	22.08	10.26	-	-

Table 6

Optimal design allocation ratios for three-group case with $p_1 = 0.25$, $p_2 = 0.15$, $p_3 = 0.1$ and $N = 345$.

	r_{25}			r_{50}			r_{75}			r_{100}		
	1 vs 2	1 vs 3	1 vs 1	1 vs 2	1 vs 3	1 vs 1	1 vs 2	1 vs 3	1 vs 1	1 vs 2	1 vs 3	1 vs 1
Post Mean	1.8	1.6		2.2	1.8		2.3	1.8		2.4	1.8	
(SE)	0.93	0.86		0.71	0.71		0.57	0.64		0.48	0.60	
Post Mode	2.0	1.8		2.2	1.9		2.3	1.9		2.4	1.9	
(SE)	0.90	0.82		0.74	0.73		0.60	0.66		0.50	0.61	
Post Eff	2.3	2.3		2.7	2.7		2.9	2.9		2.9	2.9	
(SE)	1.04	0.93		0.77	0.65		0.57	0.47		0.42	0.34	
Skep Pred	1.3	1.2		1.6	1.4		2.4	2.3		2.9	2.9	
(SE)	0.68	0.64		0.71	0.64		0.70	0.70		0.47	0.37	
Trad Pred	2.3	2.3		2.7	2.7		2.8	2.9		2.9	2.9	
(SE)	1.03	0.93		0.76	0.65		0.58	0.48		0.46	0.36	

Table 7

Natural lead-in results for three-group case with $d_1 = 0.25$, $p_2 = 0.15$, $p_3 = 0.1$ and $N = 345$.

	$E(\text{Successes})$	$E(n_1)$	$E(n_2)$	$E(n_3)$	Power	Error
Post Mean	60.5	139.5	97.7	107.8	0.78	0.02
(SE)	7.39	15.81	12.10	12.79	–	–
Post Mode	60.5	141.5	97.2	106.3	0.79	0.01
(SE)	7.31	15.92	11.82	12.50	–	–
Post Eff	62.6	154.3	97.0	93.7	0.81	0.01
(SE)	7.37	14.52	13.08	8.60	–	–
Skep Pred	62.6	154.2	97.2	93.7	0.80	0.01
(SE)	7.43	14.40	13.17	8.68	–	–
Trad Pred	62.4	153.8	97.4	9.80	0.80	0.01
(SE)	7.42	15.00	13.68	8.73	–	–

Table 8

Natural lead-in allocation ratios for three-group case with $p_1 = 0.25$, $p_2 = 0.15$, $p_3 = 0.1$ and $N = 345$.

	r_{25}			r_{50}			r_{75}			r_{100}		
	1 vs 2	1 vs 3	1 vs 1	1 vs 2	1 vs 3	1 vs 1	1 vs 2	1 vs 3	1 vs 1	1 vs 2	1 vs 3	1 vs 1
Post Mean	1.1	1.1	1.1	1.4	1.3	1.9	1.9	1.6	2.4	2.4	1.8	1.8
(SE)	0.17	0.15	0.15	0.29	0.28	0.38	0.38	0.42	0.49	0.49	0.60	0.60
Post Mode	1.2	1.1	1.1	1.5	1.4	1.9	1.9	1.6	2.4	2.4	1.9	1.9
(SE)	0.17	0.15	0.15	0.29	0.28	0.38	0.38	0.42	0.47	0.47	0.59	0.59
Post Eff	1.2	1.2	1.2	1.6	1.6	2.2	2.2	2.2	2.9	2.9	2.9	2.9
(SE)	0.20	0.16	0.16	0.32	0.24	0.39	0.39	0.29	0.41	0.41	0.32	0.32
Skep Pred	1.2	1.2	1.2	1.6	1.6	2.2	2.2	2.2	2.9	2.9	2.9	2.9
(SE)	0.20	0.16	0.16	0.32	0.24	0.38	0.38	0.29	0.41	0.41	0.32	0.32
Trad Pred	1.2	1.2	1.2	1.6	1.6	2.2	2.2	2.2	2.9	2.9	2.9	2.9
(SE)	0.20	0.16	0.16	0.32	0.24	0.39	0.39	0.30	0.45	0.45	0.36	0.36

Table 9

Optimal design results for three-group case with $p_1 = 0.55$, $p_2 = 0.45$, $p_3 = 0.4$ and $N = 618$.

	$E(\text{Successes})$	$E(n_1)$	$E(n_2)$	$E(n_3)$	Power	Error
Balanced	288.7	206.3	205.8	205.9	0.80	0.00
(SE)	12.38	11.74	11.70	12.05	-	-
Post Mean	294.5	261.2	159.0	197.8	0.80	0.01
(SE)	13.36	40.13	31.64	37.48	-	-
Post Mode	294.6	261.5	159.0	197.6	0.80	0.02
(SE)	13.29	41.27	32.05	37.21	-	-
Post Eff	304.8	333.9	149.1	135.0	0.80	0.01
(SE)	13.40	46.44	40.91	17.77	-	-
Skep Pred	297.6	282.4	159.7	175.9	0.81	0.01
(SE)	13.09	33.99	28.32	26.13	-	-
Trad Pred	304.8	335.6	147.2	135.3	0.81	0.01
(SE)	13.27	43.32	36.90	18.72	-	-

Table 10

Optimal design allocation ratios for three-group case with $p_1 = 0.55$, $p_2 = 0.45$, $p_3 = 0.4$ and $N = 618$.

	r_{25}			r_{50}			r_{75}			r_{100}		
	1 vs 2	1 vs 3	1 vs 1	1 vs 2	1 vs 3	1 vs 1	1 vs 2	1 vs 3	1 vs 1	1 vs 2	1 vs 3	1 vs 1
Post Mean	1.7	1.5	1.5	1.9	1.5	1.5	2.0	1.5	1.5	2.1	1.5	1.5
(SE)	0.71	0.67	0.60	0.60	0.63	0.51	0.51	0.57	0.43	0.53	0.53	0.53
Post Mode	1.7	1.5	1.5	1.9	1.5	1.5	2.0	1.5	1.5	2.1	1.5	1.5
(SE)	0.72	0.68	0.60	0.60	0.62	0.52	0.52	0.57	0.46	0.53	0.53	0.53
Post Eff	2.5	2.5	2.8	2.8	2.8	2.9	2.9	2.9	2.9	2.9	2.9	2.9
(SE)	0.98	0.85	0.72	0.61	0.53	0.44	0.38	0.30	0.30	0.30	0.30	0.30
Skep Pred	1.5	1.3	2.0	1.6	1.6	2.7	2.5	2.9	2.9	2.9	2.9	2.9
(SE)	0.70	0.65	0.66	0.64	0.62	0.66	0.40	0.32	0.32	0.32	0.32	0.32
Trad Pred	2.5	2.5	2.8	2.8	2.8	2.9	2.9	2.9	2.9	2.9	2.9	2.9
(SE)	0.95	0.86	0.68	0.59	0.47	0.39	0.34	0.28	0.28	0.28	0.28	0.28

Table 11

Natural lead-in results for three-group case with $d_1 = 0.55$, $p_2 = 0.45$, $p_3 = 0.4$ and $N = 618$.

	$E(\text{Successes})$	$E(n_1)$	$E(n_2)$	$E(n_3)$	Power	Error
Post Mean	291.7	238.4	175.8	203.8	0.81	0.01
(SE)	12.55	22.82	19.59	23.87	–	–
Post Mode	292.0	238.8	176.4	202.8	0.80	0.01
(SE)	12.60	22.67	19.69	23.77	–	–
Post Eff	297.4	277.8	172.8	167.4	0.82	0.01
(SE)	12.33	22.39	20.17	12.06	–	–
Skep Pred	297.5	278.3	172.3	167.3	0.82	0.01
(SE)	12.58	22.97	21.10	12.16	–	–
Trad Pred	297.6	278.0	172.5	167.5	0.82	0.01
(SE)	12.60	23.01	20.87	12.27	–	–

Table 12

Natural lead-in allocation ratios for three-group case with $p_1 = 0.55$, $p_2 = 0.45$, $p_3 = 0.4$ and $N = 618$.

	r_{25}			r_{50}			r_{75}			r_{100}		
	1 vs 2	1 vs 3	1 vs 2	1 vs 3	1 vs 2	1 vs 3	1 vs 2	1 vs 3	1 vs 2	1 vs 3	1 vs 2	1 vs 3
Post Mean	1.1	1.1	1.4	1.2	1.7	1.3	2.1	1.5				
(SE)	0.15	0.14	0.25	0.26	0.33	0.38	0.43	0.54				
Post Mode	1.1	1.1	1.4	1.2	1.7	1.3	2.1	1.5				
(SE)	0.15	0.14	0.25	0.26	0.34	0.38	0.43	0.53				
Post Eff	1.2	.2	1.6	1.6	2.2	2.2	2.9	2.9				
(SE)	0.19	0.15	0.30	0.23	0.36	0.27	0.38	0.30				
Skep Pred	1.2	1.2	1.6	1.7	2.2	2.2	2.9	2.9				
(SE)	0.20	0.16	0.30	0.23	0.37	0.28	0.39	0.30				
Trad Pred	1.2	1.2	1.6	1.7	2.2	2.2	2.9	2.9				
(SE)	0.20	0.15	0.30	0.23	0.36	0.27	0.40	0.31				