Chapter 6
Response-Adaptive Allocation for Binary Outcomes: Bayesian Methods from the BASS Conference

Roy T. Sabo

6.1 Introduction

Outcome- or response-adaptive allocation methods are used to adjust randomization probabilities in clinical trials based on observations from previously accrued patients. These methods aim to achieve one of several allocation goals, which have included maximizing statistical power, balancing for covariates, and maximizing treatment benefit. In the latter case, adaptive allocation strategies aim to treat patients as ethically as possible, often by minimizing the expected number of treatment failures. These "optimal designs" achieve this minimization through algorithms and functions of success probabilities in each group of subjects.

Though much of the conceptual and theoretical work in adaptive allocation methods has been conducted in the frequentist framework, Bayesian methods are a natural fit for conducting outcome-adaptive allocation in practice. These methods are more easily adaptable to small sample cases and are generally more flexible than are frequentist alternatives. For instance, frequentist allocation approaches generally require an initial *lead-in* period where allocation probabilities are held constant in order to overcome small-sample irregularities in proportion estimates. Some researchers have introduced scaling parameters into allocation algorithms that restrict allocation in early phases of a trial and gradually allow increasing adaptation, but even these approaches cannot account for situations where a treatment group has no observed successes, which would result in no allocation to that group. Bayesian methods can overcome these difficulties in several ways, most notably through informative prior specification or through replacing success proportion estimates with posterior or predictive probabilities of treatment superiority. Bayesian methods are also more readily adapted to account for situations where allocation ratios are desired to adapt

R. T. Sabo (✉)
Department of Biostatistics, Virginia Commonwealth University,
830 East Main Street, Richmond, VA 23298-0032, USA
e-mail: roy.sabo@vcuhealth.org

based on information from multiple outcomes, as joint distributions between multiple outcomes can be estimated through a posterior distribution in a straightforward manner.

In this Chapter we provide two examples of Bayesian approaches to outcome-adaptive allocation. The first overcomes the necessity of a *lead-in* by eliciting an informative yet skeptical prior that exhibits decreasing influence on the posterior as more patients enter a trial. This approach – dubbed the *Decreasingly Informative Prior* approach – was the subject of a 2013 presentation at the Biopharmaceutical and Applied Statistics Symposium (BASS) as well as a subsequent publication Sabo (2014). The second method presents an approach to base allocation upon two outcomes simultaneously, such as in trials where both treatment efficacy and toxicity are important. This approach was the subject of a 2012 BASS presentation and subsequent publication Sabo et al. (2013). In both cases we focus on two- and three-group clinical trials with binary outcomes. A general review of response-adaptive allocation will be provided in the next section, while the Bayesian approach will be covered in Sect. 6.3. The decreasingly informative prior approach will be discussed in Sect. 6.4, while the two-outcome approach will be presented in Sect. 6.5.

6.2 Response-Adaptive Allocation

6.2.1 Optimal Allocation

Rosenberger et al. (2001) derived optimal allocation weights for two-group trials with binary outcomes, with the goal to minimize the expected number of treatment failures. These weights are given below in Eq. 6.1.

$$w_1 = \frac{\sqrt{p_1}}{\sqrt{p_1} + \sqrt{p_2}},$$
$$w_2 = 1 - w_1, \tag{6.1}$$

where p_j is the proportion of successfully treated patients in group j ($j = 1, 2$), and where weight w_j is the probability the next patient will be allocated into the jth treatment group. In practice the unknown estimates p_1 and p_2 are replaced with the current sample proportions $\hat{p}_1$ and $\hat{p}_2$, which could lead to the awkward scenario in early phases of a trial where the weights given in Eq. 6.1 are incalculable due to no events being observed in either of the two groups.

Optimal allocation ratios for three-group trials were established numerically by Tymofyeyev et al. (2007) and in closed-form by Jeon and Hu (2010). These optimal allocation ratios depend upon the relative magnitudes of the success proportions in each group and a constant $B \in (0, 1/3)$, which is a lower allocation bound selected by the investigator (Jeon and Hu recommend selecting $0 < B \leq 1/3$ to prevent situations where a treatment ends up with no patients). We present them here with

minor corrections due to typos in the original manuscript. Let p_1, p_2 and p_3 be the true efficacy rates of treatments 1, 2 and 3, and let $\mathbf{w}^* = (w_1^*, w_2^*, w_3^*)^T$ denote the vector of optimal allocation proportions. Then for $p_1 > p_2 > p_3$, $B \in (0, 1/3)$, and $q_j = 1 - p_j$, $j = 1, 2, 3$, the allocation rates are

$$w_1^* = l_2^{-1}(l_1 + l_3 B) \tag{6.2}$$
$$w_2^* = B$$
$$w_3^* = 1 - B - w_1^*,$$

where,

$$l_1 = \frac{a(p_1 - p_3) + b(p_2 - p_3) + d}{p_3 q_3},$$

$$l_2 = \frac{b(p_1 - p_2) + c(p_1 - p_3) - d}{p_1 q_1} + l_1,$$

$$l_3 = \frac{a(p_1 - p_2) - c(p_2 - p_3) + d}{p_2 q_2} - l_1,$$

$$a = -\frac{Bq_2 - (B - 1)q_3}{p_1 q_1}$$

$$b = -\frac{B(q_3 - q_1)}{p_2 q_2}$$

$$c = \frac{Bq_2 - (B - 1)q_1}{p_3 q_3}$$

$$d = \sqrt{-ab(p_1 - p_2)^2 - ac(p_1 - p_3)^2 - bc(p_2 - p_3)^2}.$$

If $w_1^* > B$ and $w_3^* > B$ then (Eq. 6.2) is the optimal solution. If $w_1^* \leq B$, the solution is $\mathbf{w}^* = (B, B, 1 - 2B)^T$. If $w_3^* \leq B$, the solution is $\mathbf{w}^* = (1 - 2B, B, B)^T$. When $p_1 = p_2 > p_3$ the solution is:

$$w_1^* = w_2^* = \frac{\sqrt{p_1}}{2(\sqrt{p_1} + \sqrt{p_3})}, \; w_3^* = \frac{\sqrt{p_3}}{\sqrt{p_1} + \sqrt{p_3}},$$

provided $w_j^* \geq B \; \forall \, j$. If $B > \frac{\sqrt{p_1}}{2(\sqrt{p_1} + \sqrt{p_3})}$, the solution is $\mathbf{w}^* = (B, B, 1 - 2B)^T$. If $B > \frac{\sqrt{p_3}}{\sqrt{p_1} + \sqrt{p_3}}$, the solution is $\mathbf{w}^* = ((1 - B)/2, (1 - B)/2, B)^T$. When $p_1 > p_2 = p_3$ the solution is:

$$w_1^* = \frac{\sqrt{p_1}}{\sqrt{p_1} + \sqrt{p_3}}, \; w_2^* = w_3^* = \frac{\sqrt{p_3}}{2(\sqrt{p_1} + \sqrt{p_3})},$$

provided $w_j^* \geq B \; \forall \, j$. If $B > \frac{\sqrt{p_3}}{2(\sqrt{p_1} + \sqrt{p_3})}$, the solution is $\mathbf{w}^* = (1 - 2B, B, B)^T$.

6.2.2 Natural Lead-In

Thall and Wathen (2007) model the root for the two-group case as an increasing function of the observed sample size $(n/2N)$, where n is the number of observed patients and N is the planned total sample size. Here, the weighting algorithm becomes

$$w_1 = \frac{p_1^{n/2N}}{p_1^{n/2N} + p_2^{n/2N}}, \tag{6.3}$$
$$w_2 = 1 - w_1.$$

This approach has the effect of acting as a *natural* lead-in, since it forces equal weights at the beginning of a trial and gradually allows more adaptation as the trial continues. In addition, as $n \to N$ the weights in Eq. 6.3 approach the same structural form as those given in Eq. 6.1.

In the three-group case Hu and Zhang (2004) introduced an allocation function based on the doubly adaptive biased coin design (Eisele 1994), which is given as follows

$$w_j = \frac{w_j^* \left((w_j^* \sum_{i=1}^3 n_i)/n_j \right)^\gamma}{\sum_{k=1}^3 w_k^* \left((w_k^* \sum_{i=1}^3 n_i)/n_k \right)^\gamma} \tag{6.4}$$
$$j = 1, 2, 3,$$

where n_j is the current observed sample size in group j, w_j^* is the current optimal allocation weight (Eq. 6.2) in group j, and γ is a tuning parameter for calibrating the degree of randomness of the allocation probability function. By setting $\gamma = (N - (n + 1))/n$ we again achieve a *natural* lead-in that forces equal allocation early in the trial, and approaches the optimum allocation rates found in Eq. 6.2 as $n \to N - 1$ (Bello and Sabo 2016).

6.3 General Bayesian Approach

In the Bayesian framework proportions p_j for treatment groups $j = 1, \ldots, k$, are assigned a common prior distribution $\pi(\theta_0)$, where $\pi(.)$ is some distributional form and θ_0 is some fixed value. The prior distributions are combined with likelihood distributions $p(y_j | p_j, n_j)$ for each treatment group, where $p(.)$ 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 in n_j subjects. The specific choice of prior and likelihood are then synthesized into a posterior distribution for parameter p_j

$$P(p_j|y_j, n_j, \theta_0) \propto p(y_j|p_j, n_j)\pi(\theta_0), \ j = 1, \ldots, k. \tag{6.5}$$

We can slightly generalize this framework by establishing a hierarchical posterior distribution for any parameter θ as follows

$$\theta \sim P(\theta|y) = \frac{p(y|\theta, n)\pi(\theta|\theta_0, n, N)g(\theta_0|\lambda)}{\int p(y|\theta, n)\pi(\theta|\theta_0, n, N)g(\theta_0|\lambda)}, \tag{6.6}$$

where y are the observed data, $p(.)$ is the likelihood function, $\pi(.|\theta_0, n, N)$ is the prior information on θ, and $g(.)$ is a hyperprior on θ_0 with hyperparameter λ. This posterior can be used to estimate the mean or mode success rate in each group, which can then be used in Eqs. 6.1 or 6.2.

6.3.1 Posterior Estimates and Probabilities

As an alternative to posterior means or modes, Huang et al. (2007) and Thall and Wathen (2007) replaced success probabilities with probabilities of greater treatment response. Here we calculate the posterior probability that p_1 is greater than p_2, so that allocation weights increase in favor of treatment 1 as evidence of its superiority accumulates. While similar to using success rates directly, these probabilities tend to provide quicker and greater adaptation. In two-arm trials (Thompson 1933; Thall and Wathen 2007) we need only calculate one probability

$$P_1 = P(p_1 > p_2|y, n, \theta_0) \tag{6.7}$$
$$P_2 = 1 - P_1,$$

where $y = (y_1, y_2)$ and $n = (n_1, n_2)$. In three-arm trials (Bello and Sabo 2016; Sabo and Bello 2017) we calculate three probabilities

$$P_1 = [(p_1 > p_2) \cap (p_1 > p_3)|y, n, \theta_0], \tag{6.8}$$
$$P_2 = [(p_2 > p_1) \cap (p_2 > p_3)|y, n, \theta_0],$$
$$P_3 = [(p_3 > p_1) \cap (p_3 > p_2)|y, n, \theta_0]$$

where $y = (y_1, y_2, y_3)$ and $n = (n_1, n_2, n_3)$. In practice, these posterior probabilities can be used in place of the unknown population success rate for the corresponding group.

Predictive probabilities could also be used in adaptive allocation (Sabo and Bello 2017). Many predictive probability approaches in clinical trials use the current posterior probability distribution (as given in Eq. 6.6) 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, and the resulting predictive distributions are used to calculate the probability of interest. Using the standard formulation of the

predictive distribution produces similar mean or mode estimates to those obtained from simulating from the posterior distribution, since both the posterior and predictive posterior distributions have the same center. An alternative approach, as outlined in Sabo and Bello (2017), 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 $\pi(\theta_0)$ to conservatively account for uncertainty in the non-accrued patients. 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. In essence, the predictive probability distribution is used to simulate responses y_j^* for the remaining n_j^* subjects in treatment j. Direct sampling or markov-chain monte carlo methods (with T iterations) can be used to estimate predictive probabilities for between-treatment comparisons as

$$P_1 = P(p_1 > p_2 | \theta_0, y, y^*, n, n^*) = \sum_{t=1}^{T} I(p_1 > p_2)/T \qquad (6.9)$$

$$P_2 = 1 - P_1,$$

in two-group studies, and as

$$P_1 = P\left[(p_1 > p_2) \cap (p_1 > p_3) | \theta_0, y, y^*, n, n^*\right] = \sum_{t=1}^{T} I\left[\bigcap_{i=2}^{3}(p_1 > p_i)\right]/T, \ (6.10)$$

$$P_2 = P\left[(p_2 > p_1) \cap (p_2 > p_3) | \theta_0, y, y^*, n, n^*\right] = \sum_{t=1}^{T} I\left[\bigcap_{i=1, \neq 2}^{3}(p_2 > p_i)\right]/T,$$

$$P_3 = P\left[(p_3 > p_1) \cap (p_3 > p_2) | \theta_0, y, y^*, n, n^*\right] = \sum_{t=1}^{T} I\left[\bigcap_{i=1}^{2}(p_3 > p_i)\right]/T,$$

in three-group studies. The predictive probabilities given in Eqs. 6.9 and 6.10 can then be incorporated in two- and three-group optimal designs in the same manner as the posterior efficacy comparisons.

6.4   Example 1: The Decreasingly Informative Prior Approach

This method was presented at BASS in 2013 and much of the following passages originally appeared in Sabo (2014). Lead-in and natural lead-in methods are designed to prohibit or constrain adaptation of allocation weights in early stages of a trial, when

estimates may be unreliable due to small sample sizes. Alternatively, one could use a posterior distribution to provide estimators that do not change much in early parts of a trial. Under the Bayesian framework, we could elicit decreasingly informative priors (DIP) which are mass or density functions that are functions of observed (n) and planned (N) sample sizes. These functions would also serve as skeptical priors in that they would be centered around some value θ_0 indicative of treatment equivalence when sample sizes are small. However, information is incrementally transferred to likelihood as n increases, making the prior decreasingly informative.

6.4.1 Decreasingly-Informative Prior Model

An alternative to the natural lead-in approach discussed in Thall and Wathen (2007) is the concept of a built-in lead-in component achieved by making the prior distributions functions of *non-accrued* patients. We first assume skeptical prior distributions for each treatment group by centering the efficacy rates around the same value p_0. To simultaneously keep the mode of the prior distribution at p_0 while also accounting for the accruing data, where $\pi()$ is the *common* distributional form of the priors for parameters p_j, $j = 1, \ldots, k$, we make these priors to be functions of the hypothesized value p_0 and the unobserved non-accrued subjects $N - n$ such that $\pi() = \pi(p_0, n, N)$, where N is the total planned sample size, and $n = \sum_{j=1}^{k} n_j$ is the total number of accrued patients.

Say we have binary outcomes in k groups and that we want to model those outcomes using the beta-binomial conjugate pair. Based on the general Bayesian set-up in Eq. 6.6, we could model outcomes in group j as $y_j \sim f(n_j, p_j) = $ binomial(n_j, p_j). The DIP for the group j success rate could be modeled as $p_j \sim \pi(p_0, n, N) = \text{beta}[1 + p_0(N - n), 1 + (1 - p_0)(N - n)]$, where the skeptical value p_θ is chosen as a single value or given its own hyperprior. This hyperprior could take any number of suitable forms, including $p_0 \sim \text{U}[\delta_1, \delta_2]$, where $0 \leq \delta_1 < \delta_2 \leq 1$ are suitably chosen upper and lower bounds for p_0, or even $p_0 \sim \text{beta}[1 + \delta_1, 1 + \delta_2]$ where δ_1 and δ_2 are chosen to elicit diffuse support for p_0. In either case, by parameterizing the priors with $a = 1 + p_0(N - n)$ and $b = 1 + (1 - p_0)(N - n)$, the desired mode is achieved

$$\text{mode} = \frac{a - 1}{a + b - 2} = \frac{p_0(N - n)}{p_0(N - n) + (1 - p_0)(N - n)} = p_0.$$

These prior distributions can be combined with likelihood functions for each treatment group to obtain posterior distributions for each parameter or a joint distribution of all parameters may be obtained. While using a hyperprior for p_0 may lead to a non-closed-form posterior, selecting a particular value for p_θ combined with *beta* priors and *binomial*(n_i, p_i) likelihoods will lead to closed-form posterior distribution for the group j success rate $p_j \sim \text{beta}[1 + y_k + p_0(N - n), 1 + (n_j - y_j) + (1 - p_0)(N - n)]$. Regardless of the choices of prior and likelihood

and also between using posterior means, modes or efficacy comparisons, allocation weights are calculated using the optimal formulations found in Eqs. 6.1 and 6.2, not with Eqs. 6.3 and 6.4 since we are attempting to mimic the effect of a natural lead-in. At the beginning of a trial, the posterior estimates and probabilities depend only upon the skeptical prior information and are centered at the same value p_0, meaning that the allocation weights are equal. As more patients accrue into the trial, the prior information becomes increasingly less important relative to the accrued data. Thus, like the natural lead-in approach, the use of decreasingly-informative prior distributions forces the adaptation to move slowly during early parts of a trial and allows for more sensitive adaptation during latter parts of a trial.

6.4.2   Simulation Study for DIP Model

We performed a simulation study to compare the relative performance of Thall and Wathen's natural lead-in (TW) method with that of the decreasingly-informative prior (DIP) method of adaptive allocation in both two- and three-group trials. For the two-group case we assume that the first treatment has some superior true level of efficacy to the second treatment (i.e. $p_1 > p_2$), while in the three-group case we assume that $p_1 > p_2 > p_3$. In both cases we expect the first group of simulated patients to outperform those from the other groups, and thus expect both procedures to randomize more patients into the first treatment. For each new patient we simulate a random number $u \sim U[0, 1]$ to allocate between groups using Eqs. 6.1 (DIP) or 6.3 (TW) in the two-group case or Eqs. 6.2 (DIP) or 6.4 (TW) in the three-group case. The binary outcome for each patient is then probabilistically simulated based on a treatment-specific Bernouli distribution with success rate p_1, p_2 or p_3. The TW or DIP allocation ratios are then recalculated based on all currently available outcomes, and the process is repeated until the total number of patients is achieved, which is selected to attain at least 80% power in the balanced case.

For the TW procedure we have assumed a non-informative $beta(1, 1)$ prior distribution on the efficacy proportion in each group. For the DIP procedure we examine situations where we select a particular prior value for p_0 and also where we select a non-informative hyperprior on that value. In the former case three values of p_0 are used to represent different realistic scenarios: one where we correctly guess the null hypothesized value, a second where we guess the null hypothesized value incorrectly by understating its value, and a third where we overstate its value. For the hyperprior case we select a diffuse and non-informative $U[0, 1]$ hyperprior in order to mimic the situation where we make no assumptions about the underlying efficacy about either group. We also investigate the use of either posterior means or posterior efficacy comparisons to calculate the allocation probabilities. Each trial was simulated 1000 times for each set of parameter values, from which we measure end-of-trial treatment-specific sample sizes (with standard deviations), empirical power, error rates, and allocation probabilities.

In Table 6.1 we see the results from two-group trials with a true effect-size of $\delta = 0.2$. In this case – which reflects overwhelming evidence of superiority for the first treatment – we see that the TW procedure maintains the highest power, though the DIP procedure is close when the pre-selected skeptical value p_0 is near the actual success rate in the second group. We also see that the methods provide similar allocation (in terms of final sample size), though the DIP method often does so with less variability than the natural lead-in approach. Figure 6.1 shows the average allocation probabilities for both groups throughout the trial. Here we see that the adaptation gradually increases with sample size, which is similar though not identical between the different approaches.

In Tables 6.2 and 6.3 we see comparisons in the two group case with a smaller effect size ($\delta = 0.15$) and where we now formulate the DIP procedure with a diffuse hyperprior. For both the TW and DIP methods we present the use of posterior means to calculate allocation weights in Table 6.2, and the use of efficacy comparisons in

Table 6.1 Simulation summaries for two group case (DIP with point mass). $*$ indicates correct choice of prior

True efficacy		$p_1 = 0.5$	$p_2 = 0.3$	$N = 200$
		DIP		
	TW	$p_0 = 0.2$	$p_0 = 0.3^*$	$p_0 = 0.4$
$\%(n_1 > n_2)$	99.7%	98.4%	99.4%	99.6%
Power	80.9%	71.8%	78.6%	79.7%
$\hat{n}_1$	144.7	153.8	148.3	143.0
$\hat{n}_2$	55.3	46.2	51.7	57.0
(SD)	(16.37)	(20.07)	(16.60)	(14.16)
True efficacy		$p_1 = 0.7$	$p_2 = 0.5$	$N = 200$
		DIP		
	TW	$p_0 = 0.4$	$p_0 = 0.5^*$	$p_0 = 0.6$
$\%(n_1 > n_2)$	98.9%	98.2%	98.5%	98.6%
Power	79.1%	75.8%	79.8%	77.5%
$\hat{n}_1$	143.6	150.0	146.8	142.9
$\hat{n}_2$	56.4	50.0	53.2	57.1
(SD)	(17.80)	(19.08)	(17.14)	(15.32)
True efficacy		$p_1 = 0.9$	$p_2 = 0.7$	$N = 200$
		DIP		
	TW	$p_0 = 0.6$	$p_0 = 0.7^*$	$p_0 = 0.8$
$\%(n_1 > n_2)$	99.9%	99.3%	99.8%	99.9%
Power	95.3%	92.2%	94.7%	93.8%
$\hat{n}_1$	153.0	154.0	153.0	151.5
$\hat{n}_2$	47.0	46.0	47.0	48.5
(SD)	(15.76)	(15.70)	(14.09)	(13.34)

(a) TW

(b) DIP $p_0 = 0.2$
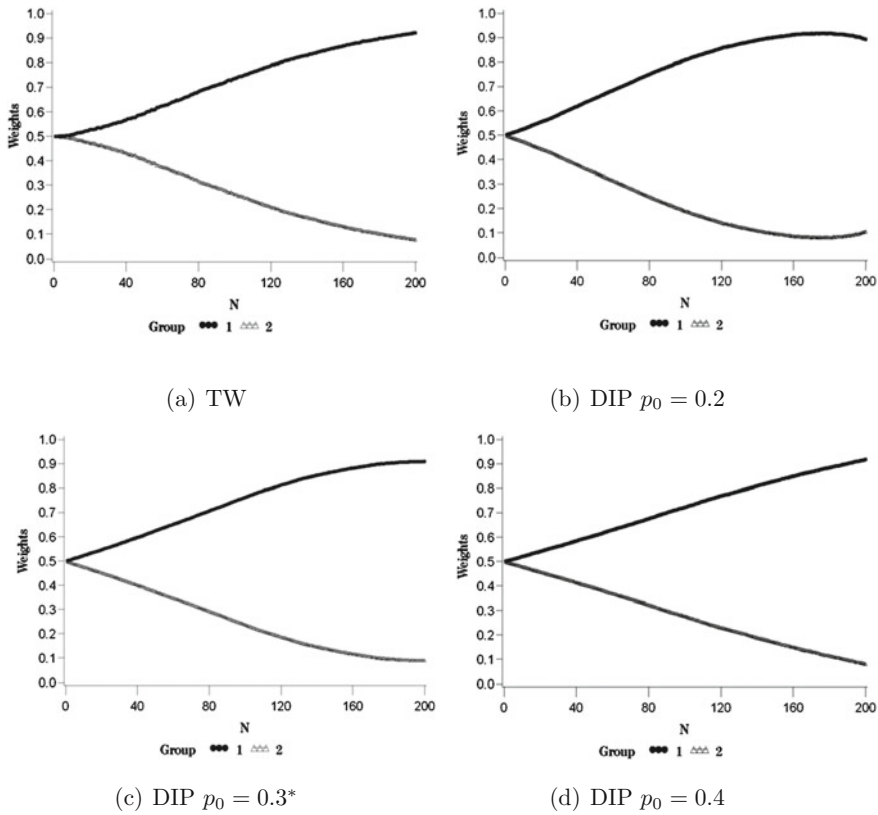
(c) DIP $p_0 = 0.3^*$

(d) DIP $p_0 = 0.4$

Fig. 6.1 Allocation probabilities for two group case (DIP with point mass). $*$ indicates correct choice of prior

Table 6.3. In the posterior mean case (Table 6.2) we see that though both methods provide some adaptation, neither meaningfully increases the expected number of successes from that achieved using balanced allocation. However, when posterior efficacy comparisons are used (Table 6.3), we see that in addition to providing more adaptation, both methods increase the expected number of treatment successes relative that achieved using balanced allocation.

Tables 6.4 and 6.5 present results from three-group trials using either posterior means and efficacy. In this case both posterior formulations provide increased treatment successes relative to balanced allocation. While the natural lead-in approach provides greater adaptation and more treatment successes, the DIP procedure has less variability in these measures.

Table 6.2 Simulation summaries for two group case (DIP with hyperprior; posterior mean)

True efficacy	$p_1 = 0.25$	$p_2 = 0.1$	$N = 200$
	Bal.	TW	DIP
Exp. Succ.	35.1 (3.85)	36.5 (4.28)	36.1 (4.00)
$\hat{n}_1$	100.2 (7.13)	110.6 (7.81)	105.3 (6.89)
$\hat{n}_2$	99.8 (7.13)	89.4 (7.81)	94.7 (6.89)
Power	80.0%	81.2%	80.3%
Error	0.0%	0.0%	0.0%
R_{50}	–	1.24 (0.15)	1.06 (0.04)
R_{75}	–	1.40 (0.22)	1.14 (0.07)
R_{100}	–	1.58 (0.30)	1.53 (0.25)
True efficacy	$p_1 = 0.55$	$p_2 = 0.4$	$N = 352$
	Bal.	TW	DIP
Exp. Succ.	167.2 (7.86)	167.9 (8.92)	167.8 (8.57)
$\hat{n}_1$	175.7 (9.47)	183.1 (10.23)	181.4 (9.65)
$\hat{n}_2$	176.3 (9.47)	168.9 (10.23)	170.6 (9.65)
Power	80.0%	81.3%	80.1%
Error	0.0%	0.0%	0.0%
R_{50}	–	1.08 (0.04)	1.05 (0.03)
R_{75}	–	1.13 (0.06)	1.10 (0.04)
R_{100}	–	1.17 (0.07)	1.17 (0.07)

6.5   Example 2: Accounting for Multiple Outcomes

There may be occasions when both the efficacy and toxicity of a novel treatment
are under investigation, or where there are two important measures of efficacy. In
such situations the meaning of a successful treatment could be defined as being
one that is effective while not inducing toxicity, or is effective in more than one
way. Investigators of such treatments may then want to utilize both outcomes in an
outcome-adpative allocation process. One such method was presented at BASS in
2012, and much of the following passages appeared in Sabo et al. (2013).

6.5.1   Models for Dual Outcomes

We assume that the dual primary outcomes in the trial are dichotomous in nature
(e.g. success or failure). The outcomes are not required to be immediately observable
(though that definitely helps), provided that such delays are not too great with respect
to the pace of patient enrollment and the planned duration of the trial Zelen (1969).
At best, such delays merely prolong the period during which the original allocation

Table 6.3 Simulation summaries for two group case (DIP with hyperprior; posterior efficacy)

True efficacy	$p_1 = 0.25$	$p_2 = 0.1$	$N = 200$
	Bal.	TW	DIP
Exp. Succ.	35.1 (3.85)	40.6 (5.12)	38.5 (4.56)
$\hat{n}_1$	100.2 (7.13)	138.6 (12.85)	122.7 (7.60)
$\hat{n}_2$	99.8 (7.13)	61.4 (12.85)	77.3 (7.60)
Power	80.0%	77.8%	83.1%
Error	0.0%	0.0%	0.0%
R_{50}	–	2.56 (1.11)	1.44 (0.29)
R_{75}	–	5.51 (3.12)	2.24 (0.73)
R_{100}	–	12.2 (8.22)	12.3 (7.78)
True efficacy	$p_1 = 0.55$	$p_2 = 0.4$	$N = 352$
	Bal.	TW	DIP
Exp. Succ.	167.2 (7.86)	178.0 (15.11)	175.1 (12.79)
$\hat{n}_1$	175.7 (9.47)	244.1 (22.60)	226.9 (17.08)
$\hat{n}_2$	176.3 (9.47)	107.9 (22.60)	125.1 (17.08)
Power	80.0%	76.5%	81.5%
Error	0.0%	0.0%	0.0%
R_{50}	–	2.63 (1.12)	1.69 (0.46)
R_{75}	–	5.55 (3.14)	3.24 (1.57)
R_{100}	–	12.3 (8.44)	12.4 (8.13)

Table 6.4 Simulation summaries for three group case (DIP with hyperprior, true efficacy: $p_1 = 0.25$, $p_2 = 0.15$, $p_3 = 0.1$, $N = 345$, and $B = 0.2$)

Posterior mean

	Bal.	Natural lead-in		DIP	
E(S)	57.2 (4.2)	62.6 (6.1)		59.0 (4.6)	
Power	79.5%	81.1%		78.5%	
Error	0.0%	1.1%		1.3%	
R_{50}	–	2.93 (2.17)	2.36 (1.79)	1.39 (0.75)	1.29 (0.69)
R_{75}	–	2.30 (1.05)	1.96 (0.98)	1.53 (0.77)	1.32 (0.70)
R_{100}	–	2.18 (0.75)	1.87 (0.73)	2.18 (0.74)	1.83 (0.72)

Posterior efficacy

	Bal.	Natural lead-in		DIP	
E(S)	57.2 (4.2)	67.2 (6.5)		62.5 (5.6)	
Power	79.5%	77.0%		76.9%	
Error	0.0%	0.7%		0.9%	
R_{50}	–	3.68 (2.46)	3.32 (1.44)	1.88 (0.89)	1.81 (0.79)
R_{75}	–	3.07 (0.82)	2.99 (0.61)	2.43 (0.79)	2.37 (0.73)
R_{100}	–	2.93 (0.40)	2.94 (0.32)	2.91 (0.44)	2.93 (0.35)

Table 6.5 Simulation summaries for three group case (DIP with hyperprior, true efficacy: $p_1 = 0.55$, $p_2 = 0.45$, $p_3 = 0.4$, $N = 618$, and $B = 0.2$)

Posterior mean

	Bal.	BS		DIP	
E(S)	288.4 (8.9)	294.4 (14.9)		290.2 (11.6)	
Power	78.8%	81.3%		80.1%	
Error	0.0%	0.7%		1.3%	
R_{50}	–	2.34 (1.50)	2.00 (1.58)	1.44 (0.70)	1.31 (0.67)
R_{75}	–	2.00 (0.84)	1.61 (0.87)	1.66 (0.67)	1.36 (0.65)
R_{100}	–	1.89 (0.61)	1.50 (0.63)	1.90 (0.58)	1.46 (0.61)

Posterior efficacy

	Bal.	BS		DIP	
E(S)	288.4 (8.9)	305.6 (21.0)		302.4 (19.3)	
Power	78.8%	80.2%		80.3%	
Error	0.0%	0.5%		0.8%	
R_{50}	–	3.54 (2.26)	3.20 (1.39)	2.57 (0.76)	2.52 (0.72)
R_{75}	–	3.03 (0.68)	2.99 (0.51)	2.85 (0.52)	2.84 (0.47)
R_{100}	–	2.93 (0.38)	2.95 (0.29)	2.93 (0.40)	2.94 (0.32)

proportions are held constant, and at worst prohibit adaptation until latter stages of the trial, possibly even excluding changes all together (Berry and Eick 1995). The two outcomes do not need to be observed simultaneously in each patient; however, it must be noted that the algorithm would be biased in favor of the observed outcome in such cases. Further, we assume that the total sample size is fixed at some n, and that patients are randomized into one of k treatment groups or arms. This data will then be used to estimate θ_j and λ_j, $j = 1, \ldots, k$, where these parameters represent the mean of the first and second outcomes in each of the k treatments, respectively. Since we are assuming that our observations are dichotomous, these parameters would most likely represent proportions, but could be arranged to represent odds ratios or relative risks.

Bayesian methods can be used to turn the observed data and any beliefs concerning the two outcomes for each treatment into posterior probabilities on the k pairs of parameters in which we are interested. Regardless of how we calculate the posterior probabilities, or of what combinations we use for the two outcomes, we want the allocation weight for treatment j to be proportional to posterior probabilities of "positive" outcomes (e.g. efficacy), and proportional to the complements of "negative" outcomes (e.g. toxicity, futility). In the following subsections, we illustrate three different approaches for estimating allocation proportions. These approaches differ in how the posterior probabilities are calculated, based on whether we compare the outcome parameters directly between treatments or to hypothesized values.

6.5.1.1 Comparisons Between Treatment Arms

We first outline the case where we compare the "success" rates for both the first and second outcomes (θ_j and λ_j, respectively) for treatment j to the corresponding rates in all other treatments. The result of these comparisons are the posterior probabilities $P_{j\ell}^{\theta} = P(\theta_j > \theta_\ell)$ for the first outcome and $P_{j\ell}^{\lambda} = P(\lambda_j > \lambda_\ell)$ for the second outcome, where these comparisons are made for $\ell = 1, \ldots, k$, where $P_{jj}^{\theta} = P_{jj}^{\lambda} = 1$. If the θ_j and λ_j represent "positive" events (implying that larger values of $P_{j\ell}^{\theta}$ and $P_{j\ell}^{\lambda}$ indicate greater likelihoods of positive responses), then the allocation weight for the jth of k treatment arms is defined as

$$
w_j = \frac{\left(\Pi_{\ell=1}^k P_{j\ell}^{\theta} P_{j\ell}^{\lambda} \right)^{c(n)}}{\sum_{i=1}^k \left(\Pi_{\ell=1}^k P_{i\ell}^{\theta} P_{i\ell}^{\lambda} \right)^{c(n)}},
$$

where $c(n)$ is a suitably chosen tuning parameter that can adjust the pace of adaptation (Thall and Wathen 2007; Bello and Sabo 2016). Note that the allocation weight w_j for treatment j is proportional to the product of the posterior probabilities that the success rates for outcomes θ and λ in treatment j are greater than the success rates in every other treatment. Thus, the weight w_j can increase (or decrease) in a number of ways. For example, the allocation weight can increase if the success rate for just one of the outcomes in treatment j is larger than the corresponding rate in just one other treatment (assuming the probabilities for all other comparisons stay constant), or it could increase if treatment j has a higher outcome-one (or outcome-two) success rate than all other treatments; in this latter case the weight may increase more than in the former case. Conversely, w_j can decrease if treatment j is outperformed by another or several other treatments, with respect to outcome one, outcome two, or both.

Note that if one of the outcomes (say the second) were to represent a "negative" outcome (implying that higher rates for the λ_j represented undesirable outcomes, and that larger values of $P_{j\ell}^{\lambda}$ indicate a greater likelihood of that undesirable outcome happening), then we could simply focus on the "positive" complement $1 - P_{j\ell}^{\lambda}$ for each outcome in the allocation weight for the jth of k treatment arms.

6.5.1.2 Comparisons to Hypothesized Values

As mentioned in Huang et al. (2007), we could compare the "success" rates for each outcome in each treatment to hypothesized values (say p_o^{θ} and p_o^{λ}), should such values exist. For instance, we could compare the efficacy rates for a set of new treatments to a rate of 30% established by a "gold-standard" treatment, or physicians may wish to keep the toxicity rates below a 10% threshold. If such values are available, then the posterior probabilities $P_j^{\theta} = P(\theta_j > p_o^{\theta})$ and $P_j^{\lambda} = P(\lambda_j > p_o^{\lambda})$ can be calculated from the posterior distributions for each outcome in each treatment group. If we

assume that the two outcomes are "positively" valued, then the allocation weight for the jth of k treatment arms is defined as

$$w_j = \frac{\left(P_j^\theta P_j^\lambda\right)^{c(n)}}{\sum_{i=1}^k \left(P_i^\theta P_i^\lambda\right)^{c(n)}} \qquad (6.11)$$

The weights described in Eq. 6.11 are proportional to the likelihood of positive outcomes *in single treatments*. While the treatments in this case are not directly compared with one another, the two outcomes in each group are compared to the same values. Treatments are thus indirectly compared, and superiority of one treatment over the hypothesized value will lead to an increased allocation weight for that treatment when either: such superiority is not as strong or lacking for other treatments, or those treatments are showing inferiority to the hypothesized values. The behavior of allocation weights for ambiguous scenarios would by their nature be difficult to predict.

6.5.1.3 Hybrid Approach

A likely scenario is the case where we want to compare one outcome between treatments and the other outcome within each treatment to a hypothetical standard. This could be the case if we wanted to determine the treatment with the greatest efficacy, provided that it kept toxicity below an allowable threshold. We assume that the first outcome is compared between treatments and the second is compared to a hypothesized value, so for each treatment j we will have $k-1$ posterior probabilities $P_{j\ell}^\theta = P(\theta_j > \theta_\ell)$, $\ell = 1, \ldots, k$ for the first outcome (recall $P_{jj}^\theta = 1$), and one posterior probability $P_j^\lambda = P(\lambda_j > p_o^\lambda)$ for the second outcome. If we assume that both outcomes represent "positive" outcomes, then the allocation weight for the jth of k treatment arms is defined as

$$w_j = \frac{\left(P_j^\lambda \Pi_{\ell=1}^k P_{j\ell}^\theta\right)^{c(n)}}{\sum_{i=1}^k \left[P_i^\lambda \left(\Pi_{\ell=1}^k P_{i\ell}^\theta\right)\right]^{c(n)}}. \qquad (6.12)$$

6.5.2 Simulation Study for Dual Objective Model

We calculate weights w_j for the $j = 1, \ldots, k$ treatment arms based assuming that posterior probabilities are raised to the power $(n/2N)$ as described in Eq. 6.3. By simulating $u \sim U[0, 1]$, we allocate the simulated patient to the jth treatment arm if $\sum_{i=0}^{j-1} w_i < u < \sum_{i=1}^j w_i$, where $w_0 = 0$. At this point we simulate the efficacy and toxicity outcome for the new patient by generating a random outcome from a Bernoulli trial with efficacy probability $(p_e + \delta_j)$, where δ_j is the amount by which

the probability of a successful outcome in the jth treatment arm differs from p_e, and by also generating a random outcome from a second Bernoulli trial with toxicity probability $(p_t + \tau_j)$, where τ_j is the amount by which the probability of a toxic outcome in the jth treatment arm differs from p_t. These new values are combined with the existing data to calculate posterior probabilities of both efficacious outcomes and toxic outcomes, which are in turn used to update the allocation weights, the method of which depends upon whether the performance of the treatment arms are being compared to hypothesized values, each other or both. One simulated clinical trial ends when the maximum sample size of $n = 200$ patients has been fully allocated. This process is repeated $m = 1000$ times for each set of assumed efficacy and toxicity rates.

Here we focus solely upon three-arm studies where efficacy is compared between arms and toxicity is compared to a hypothesized value. We assume informative and skeptical beta prior distributions for the P_j^e and P_j^t ($beta(1.3, 1.7)$ and $beta(1.1, 1.9)$, respectively). While the probability that a given treatment is less toxic than a hypoth-esized value ($P_t = 0.1$) can again be calculated using the posterior distribution of P_j^t, we use direct sampling to calculate $P_{jk}^e = P(P_j^e > P_k^e)$. Assuming treatment groups are independent, we simulate $m = 1000$ values each from the posterior distribu-tions of the P_j^e, $j = 1, \ldots, k$, to obtain $(P_{1,j}^e, \ldots, P_{1000,j}^e)$ and estimate the posterior probability that treatment arm j is more successful than treatment arm k as

$$P_{jk} = P(P_j^e > P_k^e) = \frac{\sum_{i=1}^m I(P_{i,j}^e > P_{i,k}^e)}{1000},$$

where $I()$ is an indicator function.

The average behaviors of allocation weights under various scenarios are found in Fig. 6.2. The first three panels show relatively straightforward scenarios, where (i) there is no efficacy or toxicity differences between the three treatments, (ii) the first treatment is more efficacious than the other two treatments, and (iii) the first treatment is more toxic than the other two treatments. The allocation weights do not change in the first case, skew in favor of the first treatment in the second case, and skew away from the first treatment in the third case. The average sample sizes presented in Table 6.6 for these three cases corroborate the visual results. In the ambiguous case where the first treatment is simultaneously more efficacious ($p_1^e = 0.5$) and toxic ($p_1^t = 0.2$) than the second and third treatments, Fig. 6.2 shows that the allocation weights change little during the trial, and the average numbers of subjects (Table 6.2) allocated between the three treatments (78.1, 60.1 and 61.7, respectively) are not as different as in first three cases. Here efficacy is slightly more meaningful than toxicity because in each treatment there are two inter-arm efficacy comparisons for every toxicity comparison. In the case where the first treatment is more efficacious ($p_1^e = 0.4$) than the second treatment ($p_2^e = 0.3$), which in turn is more efficacious than the third treatment ($p_3^e = 0.2$), more patients (101.0) are allocated to the first treatment than to the second (61.0) and third (38.0), with heavy favoring of treatment one resulting predominantly from the large efficacy difference between the first and
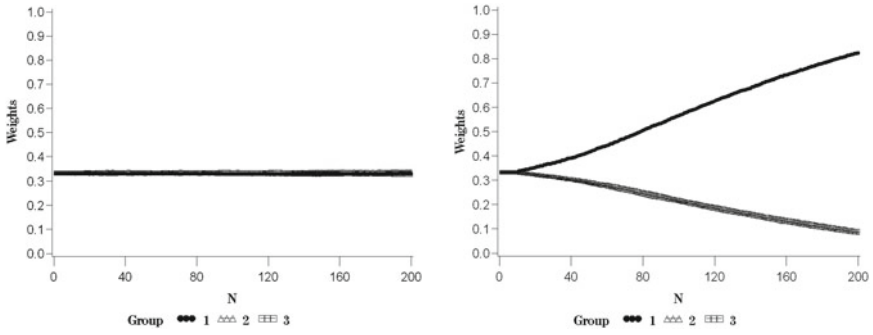
(a) $p_1^e = p_2^e = p_3^e = 0.3, p_1^t = p_2^t = p_3^t = 0.1$

(b) $p_1^e = 0.5, p_2^e = p_3^e = 0.3; p_1^t = p_2^t = p_3^t = 0.1$
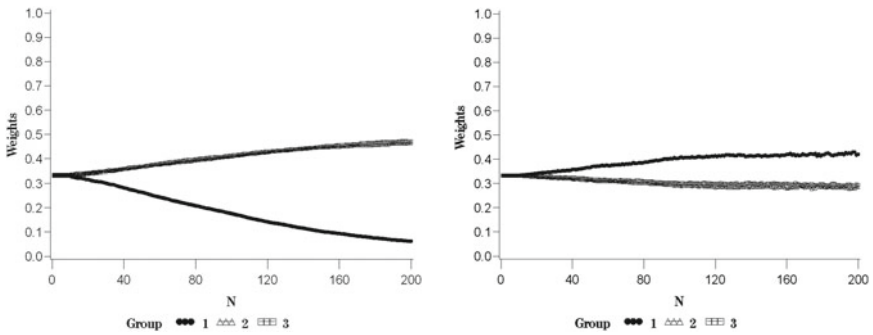
(c) $p_1^e = p_2^e = p_3^e = 0.3; p_1^t = 0.25, p_2^t = p_3^t = 0.1$

(d) $p_1^e = 0.5, p_2^e = p_3^e = 0.3; p_1^t = 0.2, p_3^t = 0.1$
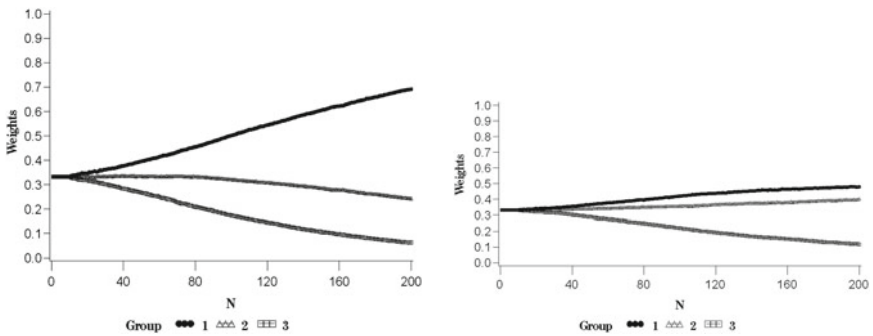
(e) $p_1^e = 0.4, p_2^e = 0.3, p_3^e = 0.2; p_1^t = p_3^t = 0.1$

(f) $p_1^e = 0.4, p_2^e = 0.3, p_3^e = 0.2; p_1^t = 0.15, p_2^t = 0.1, p_3^t = 0.05$

Fig. 6.2 Average allocation weights based on number of accrued patients in 3 treatment arms for given efficacy and toxicity probabilities

Table 6.6 Average sample size (with standard deviation) for 3–arm trials: results from simulation study with $m = 1000$ repetitions with treatment comparisons made between treatments for efficacy and to hypothesized values for toxicity ($p_o^t = 0.1$)

Parameters	Sample size	Standard deviation	Parameters	Sample size	Standard deviation
$p_1^e = 0.3$			$p_1^e = 0.5$		
$p_2^e = 0.3$	$\hat{n}_1 = 66.0$	$SD_1 = 23.1$	$p_2^e = 0.3$	$\hat{n}_1 = 78.1$	$SD_1 = 28.4$
$p_3^e = 0.3$	$\hat{n}_2 = 66.2$	$SD_2 = 23.0$	$p_3^e = 0.3$	$\hat{n}_2 = 60.1$	$SD_2 = 24.4$
$p_1^t = 0.1$	$\hat{n}_3 = 67.8$	$SD_3 = 22.5$	$p_1^t = 0.2$	$\hat{n}_3 = 61.7$	$SD_3 = 25.5$
$p_2^t = 0.1$			$p_2^t = 0.1$		
$p_3^t = 0.1$			$p_3^t = 0.1$		
$p_1^e = 0.5$			$p_1^e = 0.4$		
$p_2^e = 0.3$	$\hat{n}_1 = 113.0$	$SD_1 = 21.9$	$p_2^e = 0.3$	$\hat{n}_1 = 101.0$	$SD_1 = 23.4$
$p_3^e = 0.3$	$\hat{n}_2 = 42.9$	$SD_2 = 17.1$	$p_3^e = 0.2$	$\hat{n}_2 = 61.0$	$SD_2 = 21.8$
$p_1^t = 0.1$	$\hat{n}_3 = 44.1$	$SD_3 = 17.3$	$p_1^t = 0.1$	$\hat{n}_3 = 38.0$	$SD_3 = 15.0$
$p_2^t = 0.1$			$p_2^t = 0.1$		
$p_3^t = 0.1$			$p_3^t = 0.1$		
$p_1^e = 0.3$			$p_1^e = 0.4$		
$p_2^e = 0.3$	$\hat{n}_1 = 37.2$	$SD_1 = 14.7$	$p_2^e = 0.3$	$\hat{n}_1 = 82.7$	$SD_1 = 26.0$
$p_3^e = 0.3$	$\hat{n}_2 = 81.2$	$SD_2 = 25.7$	$p_3^e = 0.2$	$\hat{n}_2 = 72.3$	$SD_2 = 24.0$
$p_1^t = 0.25$	$\hat{n}_3 = 81.6$	$SD_3 = 26.4$	$p_1^t = 0.15$	$\hat{n}_3 = 45.1$	$SD_3 = 18.7$
$p_2^t = 0.1$			$p_2^t = 0.1$		
$p_3^t = 0.1$			$p_3^t = 0.05$		

third treatments. For the other ambiguous case, where treatments one and two are sequentially more efficacious and toxic than treatment three, Fig. 6.2 shows that the weights turn against the third treatment in favor of the first and second (even though it is less toxic, it is also less efficacious than the other two). The weights for the first treatment are slightly higher than those for the second, and both are larger than the weights for the third treatment. The average number of total patients allocated to the first and second treatments (82.7 and 72.3, respectively) are also higher than the average number allocated to the third treatment (45.1). This is again due to the fact that while treatment two is less toxic than treatment one, treatment one is much more efficacious than treatment three. This might be a scenario where we consider different radical exponents for the two outcomes.

Using the same simulations from which the previous results were obtained, we have also calculated the percentage of simulations for which each of the three treatment arms had the highest number of allocated patients. These results are found in Table 6.7 and show that the most efficacious and least toxic treatments routinely receive the most patients. Also reported is the proportion of simulated trials (for both the adaptive and balanced allocation procedures) for which the various efficacy and toxicity rates were deemed significantly different between the three possible

Table 6.7 Percentage of larger samples and decisions in favor in 3–arm trials: results from simulation study with $m = 1000$ repetitions with treatment comparisons made between treatments for efficacy and to hypothesized values for toxicity ($p_o^t = 0.1$). Case 1: $p_1^e = p_2^e = p_3^e = 0.3$, $p_1^t = p_2^t = p_3^t = 0.1$. Case 2: $p_1^e = 0.5$, $p_2^e = p_3^e = 0.3$, $p_1^t = p_2^t = p_3^t = 0.1$. Case 3: $p_1^e = p_2^e = p_3^e = 0.3$, $p_1^t = 0.25$, $p_2^t = p_3^t = 0.1$. Case 4: $p_1^e = 0.5$, $p_2^e = p_3^e = 0.3$, $p_1^t = 0.2$, $p_2^t = p_3^t = 0.1$. Case 5: $p_1^e = 0.4$, $p_2^e = 0.3$, $p_3^e = 0.2$, $p_1^t = p_2^t = p_3^t = 0.1$. Case 6: $p_1^e = 0.4$, $p_2^e = 0.3$, $p_3^e = 0.2$, $p_1^t = 0.15$, $p_2^t = 0.1$, $p_3^t = 0.05$

Case comparison	Reject in Favor of		% of	Reject in Favor of		% of
	(Adapt) (%)	(Equal) (%)	Samples	(Adapt) (%)	(Equal) (%)	Samples
	Case 1			Case4		
Eff: 1v2	6.8	4.7	$n_1 > n_2, n_3$	77.2	76.6	$n_1 > n_2, n_3$
Eff: 1v3	6.1	5.1	31.2%	78.9	79.9	49.2%
Eff: 2v3	6.6	5.4	$n_2 > n_1, n_3$	8.0	4.7	$n_2 > n_1, n_3$
Tox: 1v2	7.1	4.4	33.9%	50.8	54.2	22.7%
Tox: 1v3	7.2	5.4	$n_3 > n_1, n_2$	52.7	49.9	$n_3 > n_1, n_2$
Tox: 2v3	7.3	6.4	33.8%	9.2	5.3	27.1%
	Case 2			Case5		
Eff 1v2	77.3	77.8	$n_1 > n_2, n_3$	32.7	32.4	$n_1 > n_2, n_3$
Eff 1v3	75.4	78.2	93.2%	74.7	82.0	79.7%
Eff 2v3	11.6	6.3	$n_2 > n_1, n_3$	36.7	40.2	$n_2 > n_1, n_3$
Tox: 1v2	4.3	5.8	3.2%	6.0	6.2	17.7%
Tox: 1v3	4.4	6.2	$n_3 > n_1, n_2$	7.5	6.7	$n_3 > n_1, n_2$
Tox: 2v3	11.6	5.6	3.3%	12.0	5.1	1.9%
	Case 3			Case6		
Eff 1v2	8.9	5.1	$n_1 > n_2, n_3$	35.3	34.7	$n_1 > n_2, n_3$
Eff 1v3	9.3	4.5	1.5%	81.8	82.9	55.0%
Eff 2v3	5.3	6.3	$n_2 > n_1, n_3$	38.2	43.0	$n_2 > n_1, n_3$
Tox: 1v2	75.3	78.0	48.7%	23.3	20.8	37.3%
Tox: 1v3	75.2	74.3	$n_3 > n_1, n_2$	58.6	61.8	$n_3 > n_1, n_2$
Tox: 2v3	4.9	4.8	48.8%	29.4	31.0	6.9%

treatment pairings (1 vs. 2, 1 vs. 3, and 2 vs. 3) using chi-square tests. The estimated proportions for the adaptive and fixed allocation methods are similar for Cases 1, 2, 3, 4 and 6, and the adaptive allocation method features a slight loss of power compared to the fixed allocation method in Case 5. These Cases show that the benefit of allocating subjects away from less efficacious or more toxic treatments may come at the cost of slightly lower power as compared to the fixed allocation method.

6.6   Discussion

Presented here are examples of adaptive allocation algorithms conducted under the Bayesian analytic framework. These methods – an adaptive allocation algorithm for dual outcomes, and the decreasingly informative prior approach – were originally presented at the BASS conference in 2012 and 2013, respectively. While these are emblematic of Bayesian techniques, they are by no means the only examples in the adaptive allocation literature. One particularly active research area is in covariate-adjusted response-adaptive allocation designs (Bandyopadhyay et al. 2007; Thall and Wathen 2007), where allocation algorithms can be balanced for patient characteristics, or where particular sub-groups can be given separate allocation weights. Another example is adaptive allocation designs for clinical trials with continuous Biswas and Bhattacharya (2016) our survival outcomes Zhang and Rosenberger (2007), which in general require entirely different algorithms and concepts of what constitutes "optimal" treatment outcomes.

References

Bandyopadhyay, U., Biswas, A., & Bhattacharya, R. (2007). A covariate adjusted two-stage allocation design for binary responses in randomized clinical trials. *Statists in Medicine*, *26*, 4386–4399.

Bello, G., & Sabo, R. T. (2016). Outcome-adaptive allocation with natural lead-in for three-group trials with binary outcomes. *Journal of Statistical Computation and Simulation*, *86*, 2441–2449.

Berry, D. A., & Eick, S. G. (1995). Adaptive assignment versus balanced randomization in clinical trials: A decision analysis. *Statistics in Medicine*, *14*, 231–246.

Biswas, A., & Bhattacharya, R. (2016). Response-adaptive designs for continuous treatment responses in phase iii clinical trials: A review. *Statistical Methods in Medical Research*, *25*, 81–100.

Eisele, J. R. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference*, *38*, 249–261.

Hu, F., & Zhang, L. X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *The Annals of Statistics*, *32*, 268–301.

Huang, X., Biswas, S., Oki, Y., Issa, J. P., & Berry, D. A. (2007). A parallel phase i/ii clinical trial design for combination therapies. *Biometrics*, *63*, 429–436.

Jeon, Y., & Hu, F. (2010). Optimal adaptive designs for binary response trials with three treatments. *Statistics in Biopharmaceutical Research*, *2*, 310–318.

Korn, E. L., & Freidlin, B. (2011). Outcome-adaptive randomization: Is it useful? *Journal of Clinical Oncology*, *29*, 771–776.

Rosenberger, W. F., Stallard, N., Ivanova, A., Haper, C. N., & Ricks, M. L. (2001). Optimal adaptive designs for binary response trials. *Biometrics*, *57*, 909–913.

Sabo, R. T., co-authors. (2013). An outcome-adaptive allocation method for clinical trials with dual objectives. *Statistics in Biopharmaceutical Research*, *5*, 67–78.

Sabo, R. T. (2014). Adaptive allocation for binary outcomes using decreasingly informative priors. *Journal of Biopharmaceutical Statistics*, *24*, 569–578.

Sabo, R. T., & Bello, G. (2017). Optimal and lead-in adaptive allocation for binary outcomes: A comparison of methodologies. *Communications in Statistics: Theory and Methods*, *46*, 2823–2836.

Thall, P. F., & Wathen, J. K. (2007). Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments. *Statistics in Medicine*, *24*, 1947–1964.

Thall, P. F., & Wathen, J. K. (2007). Practical bayesian adaptive randomization in clinical trials. *European Journal of Cancer*, *43*, 859–866.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, *25*, 285–294.

Tymofyeyev, Y., Rosenberger, W. F., & Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association*, *102*, 224–234.

Zelen, M. (1969). Play the winner rule and the controlled clinical trial. *Journal of the American Statistician*, *64*, 131–146.

Zhang, L., & Rosenberger, W. F. (2007). Response-adaptive randomization for survival trials: The parametric approach. *Journal of the Royal Statistical Society, Series C*, *34*, 562–569.