Received: 20 December 2019 | Revised: 23 July 2020 | Accepted: 24 July 2020

DOI: 10.1002/sim.8734

RESEARCH ARTICLE

Statistics in Medicine WILEY

Bayesian enhancement two-stage design with error control for phase II clinical trials

Huaqing Jin | **Guosheng Yin**

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong SAR, China

Correspondence
Guosheng Yin, Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China.
Email: gyin@hku.hk

Phase II clinical trials make a critical decision of go or no-go to a subsequent phase III studies. A considerable proportion of promising drugs identified in phase II trials fail the confirmative efficacy test in phase III. Recognizing the low posterior probabilities of H_1 when accepting the drug under Simon's two-stage design, the Bayesian enhancement two-stage (BET) design is proposed to strengthen the passing criterion. Under the BET design, the lengths of highest posterior density (HPD) intervals, posterior probabilities of H_0 and H_1 are computed to calibrate the design parameters, aiming to improve the stability of the trial characteristics and strengthen the evidence for proceeding the drug development forward. However, from a practical perspective, the HPD interval length lacks transparency and interpretability. To circumvent this problem, we propose the BET design with error control (BETEC) by replacing the HPD interval length with the posterior error rate. The BETEC design can achieve a balance between the posterior false positive rate and false negative rate and, more importantly, it has an intuitive and clear interpretation. We compare our method with the BET design and Simon's design through extensive simulation studies. As an illustration, we further apply BETEC to two recent clinical trials, and investigate its performance in comparison with other competitive designs. Being both efficient and intuitive, the BETEC design can serve as an alternative toolbox for implementing phase II single-arm trials.

KEYWORDS

BET design, binary endpoint, cancer, phase II trial, posterior error rate, Simon's design

1 | INTRODUCTION

The phase II clinical trial is an essential and fundamental step for preliminary assessment of the drug's efficacy.[1] The goals of such trials are to screen out non-promising drugs and carry promising ones into large-scale phase III clinical trials which are typically long-term and costly. Phase II trials can be single-arm,[2] or two-arm involving randomized comparison,[3,4] and some designs even adopt a single-to-double arm transition scheme.[5-7] Although randomized two-arm phase II trials seem to be more preferable in terms of sample efficiency,[8] due to the limited sample size, many phase II trials are still designed as single-arm studies with the adaptive feature for futility stopping.[9] Thus, we focus on single-arm trials with a binary endpoint under the hypotheses framework,

$$H_0 : p \le p_0 \quad \text{vs} \quad H_1 : p \ge p_1, \tag{1}$$

where p is the response rate of the investigational drug, p_0 is the clinically uninteresting response rate, and p_1 represents the target desirable response rate.

In the frequentist paradigm, Gehan[10] proposed a two-stage procedure for cancer phase II trials, where the new treatment would be abandoned if no response is observed at stage 1; otherwise, the trial moves to stage 2 for enrolling more patients. Fleming[11] presented a multistage design which allows early stopping for both futility and efficacy. The most commonly used two-stage design was proposed by Simon,[2] which controls the test size at p_0 and evaluates power at p_1 under a hypothesis testing framework. Ensign et al[12] proposed a three-stage design, which reduces the sample size and increases the probability of early termination when the response rate p is small. However, Ensign's design restricts the rejection region in the first stage to be zero response, and the sample size at least 5. To relax these restrictions, Chen[13] presented another three-stage design by extending Simon's two-stage design. Shuster[14] developed a two-stage design which is minimax in the sense that it has the smallest globally maximized expected sample size. Lin and Shih[15] considered the uncertainty in targeting the alternative hypothesis to study the power at the planning stage and presented an adaptive two-stage design with a new optimal criterion to reduce the expected total sample size. Chen and Shan[16] introduced the optimal and minimax three-stage design with an additional efficacy stopping rule. Mander and Thompson[17] modified Simon's design by minimizing the expected sample size under the alternative hypothesis and allowing the efficacy stopping. Englert and Kieser[18] adopted the branch-and-bound algorithm and developed an adaptive design that allows the sample size of the second stage to depend on the number of responses observed in the first stage. Observing that the sample size of the second stage is a monotonic function of the number of the responses in the first stage, Shan et al[19] proposed an optimal adaptive design which can maintain the type I and II error rates.

Along the similar lines, various single-arm phase II trial designs have been proposed under the Bayesian paradigm.[20] Thall and Simon[21] presented a Bayesian single-arm design based on the posterior probability which monitors the binary outcomes continuously. Tan and Machin[22] proposed the Bayesian single threshold design (STD) and double threshold design (DTD), which do not need a distribution for the response rate of the investigational drug or the explicit specification of the utility or loss function. Wang et al[23] developed a Bayesian version of Simon's design under which both the frequentist and Bayesian error rates can be controlled. Based on the Bayesian predictive probability and the minimax criterion, Lee and Liu[24] proposed a Bayesian phase II design which allows continuous monitoring of the trial outcomes using predictive distributions. To inherit merits from both the frequentist and Bayesian methods, Dong et al[25] presented a Bayesian-frequentist two-stage design which allows early acceptance and rejection of the null hypothesis as well as controlling both the frequentist and Bayesian error rates.

A salient issue with promising drugs identified in single-arm phase II trials is the potential high failure rate of the subsequent large-scale phase III studies.[26] If the single-arm phase II trial directly leads to a subsequent phase III study, an incorrect decision on rejection of H_0 may result in the failure of the large-scale phase III trial. Gan et al[27] investigated 235 phase III randomized cancer trials published in 10 medical journals and found that only 38% of them achieved significant results. In another investigation, Mandel et al[28] claimed that 91% phase III studies in glioblastoma failed to show an improvement in overall survival although the prior phase II studies had declared success. Such high failure rates of phase III trials cause enormous waste of the time and resources, while the reasons behind are complex.[29] It may be mainly due to the small sample size in phase II trials and a lack of correlation between the tumor response and the survival endpoints used in phase III trials.[30] Another possible reason, as pointed out by Shi and Yin,[31] is the extensive use of Simon's design[29,32,33] and its low posterior probability of H_1 when accepting the drug under some circumstances.

Constructed under a Bayesian framework, the Bayesian enhancement two-stage (BET) design[31] renders a good control over the posterior probability of H_0 when carrying the trial to the second stage and that of H_1 when declaring the drug as promising. However, to control the variance, the BET design utilizes criteria based on the highest posterior density (HPD) interval, that is, the narrowest one among all Bayesian credible intervals. From an intuitive and practical perspective, such criteria lack transparency and interpretability, and thus they do not have a clear range of values to choose from. To circumvent this issue, we adopt the posterior false positive and false negative error rates in Lee and Zelen,[34] which are the counterparts of the type I and type II error rates in the Bayesian framework. The posterior error rates are defined as the posterior probabilities of the true situation being the opposite of the outcome of the trial. Based on these concepts, we replace the constraints on the HPD interval lengths with the posterior error probabilities when rejecting the drug at stage 1 and stage 2. Unlike the BET design which mainly focuses on reducing the posterior error rates under the minimally required response number and using the lengths of HPD intervals to control the variance, we propose the BET design with

error control (namely, BETEC) by explicitly controlling both posterior error rates when rejecting and accepting the drug respectively. While inheriting the merits of the BET design, the BETEC design is much easier to interpret and implement in practice.

The rest of this article is organized as follows. In Section 2, we present the BETEC design and its variant, the δ-BETEC design, and discuss their relationships with the BET design. We present the simulation studies of the BETEC and δ-BETEC designs in Section 3. Section 4 applies the BETEC and δ-BETEC designs to two trial examples to further assess their performances. We provide a brief discussion in Section 5.

2 | METHODOLOGY

2.1 | One-stage single-arm trial

To facilitate understanding the differences between the frequentist and Bayesian designs, we begin our discussion with the one-stage single-arm phase II trial. Focusing on the hypotheses in Equation (1), when the number of responses y out of n subjects is larger than or equal to r, we declare the drug as promising; otherwise the drug is nonpromising. For the one-stage trial, we aim to find the threshold r and the sample size n to control the error rates of the trial.

Under the frequentist framework, we assume the response rate p is a constant. To find the design parameters (r,n), we control the type I and II error rates,

$$\sup_{p\in H_0} \Pr(y \geq r|p,n) = \Pr(y \geq r|p_0,n) \leq \alpha, \quad \sup_{p\in H_1} \Pr(y < r|p,n) = \Pr(y < r|p_1,n) \leq \beta,$$

where α and β are the nominal frequentist type I and II error rates, respectively.

From a Bayesian perspective, the response rate p is assumed to be random, and we typically impose a Beta prior on p. Based on the posterior distribution of p, we can obtain the values of (r, n) by controlling the posterior error rates,

$$\sup_{y\geq r} \Pr(H_0|y,n) = \Pr(H_0|r,n) \leq \alpha^*, \quad \sup_{y<r} \Pr(H_1|y,n) = \Pr(H_1|r-1,n) \leq \beta^*,$$

where α^* and β^* are the Bayesian maximum tolerable false positive and false negative error rates, respectively.

2.2 | The BET design

As a Bayesian design, the BET design[31] imposes a Beta-Binomial model on the response rate p. It is characterized by four parameters (r_1, n_1, r, n) via the posterior probabilities of H_0 and H_1 and the HPD intervals. Let y_1 and y_2 denote the numbers of responses observed in the first and second stages, respectively. In the first stage, the sample size is n_1 and if $y_1 \geq r_1$, the trial would proceed to the second stage, otherwise the trial is terminated early for futility. In the second stage, $n_2 = n - n_1$ new subjects are enrolled. If at the end of the trial the total number of responses $y = y_1 + y_2$ reaches r, the drug is considered as promising; otherwise, the drug is announced as non-promising. The basic idea behind the BET design is that when the drug is declared as promising, the posterior probability of $p > p_0$ at stage 1 and that of H_1 at stage 2 should be adequately large. Therefore, it controls $\Pr(p > p_0|y_1, n_1) > \pi_1$ and $\Pr(p > p_1|y_1 + y_2, n) > \pi_2$, where π_1 and π_2 are prespecified probability cutoffs. To determine the design parameters (r_1, n_1, r, n), Shi and Yin[31] proposed the following constraints,

$$\begin{aligned} \text{Stage 1}: &\quad l_p(\pi_1|r_1,n_1) < \ell_1, \quad \Pr(p > p_0|r_1,n_1) > \pi_1, \\ \text{Stage 2}: &\quad l_p(\pi_2|r,n) < \ell_2, \quad \Pr(p > p_1|r,n) > \pi_2, \end{aligned}$$

where $l_p(\pi|r,n)$ is the length of the HPD interval for p when observing r responses among n subjects with coverage probability π, and ℓ_1 and ℓ_2 are prespecified design parameters.

Compared with the one-stage Bayesian design, the BET design has two stages, thus it enjoys the adaptive feature of futility stopping which helps to reduce the ethical risk. To mitigate the high failure rate of phase III trials, the

BET design sets a more stringent criterion when declaring the drug as promising, that is, it requires a high posterior probability (at least π_2) of $p > p_1$ to guarantee the efficacy of the drug to be high enough, while the one-stage Bayesian design only requires a high posterior probability (at least $1-\alpha^*$) of $p > p_0$. Although it can reduce the false positive rate significantly, such a strict criterion also makes the BET design fail to adopt the posterior false negative error constraint, that is, $\Pr(H_1|r-1,n) \le \beta^*$, as there is typically no feasible solution. As a result, to limit the variance of the trial, the BET design utilizes criteria based on HPD intervals to find a suitable sample size which however sacrifices the transparency and interpretability.

2.3 | The BETEC design

To impose a more stringent efficacy evaluation in single-arm phase II trials while preserving the interpretability, we propose the BETEC design. We adopt a Beta-Binomial model and assume the response rate p follows a Beta prior distribution, that is, $p \sim \text{Beta}(a,b)$, where $a=b=1$ corresponds to a uniform prior distribution.

In the first stage, via the conjugacy property of the Beta-Binomial model, we can derive the posterior distribution of p, $p|(y_1,n_1) \sim \text{Beta}(a+y_1, b+n_1-y_1)$. If the number of responses y_1 is greater than or equal to the cutoff r_1, we carry on the trial to the second stage; otherwise, we terminate the trial early for futility. In stage 1, we aim to screen out the drugs whose response rates are below p_0. When the trial proceeds to the second stage, we expect a high posterior probability of $p > p_0$, that is,

$$\min_{y_1 \ge r_1} \Pr(p > p_0|y_1,n_1) = \Pr(p > p_0|r_1,n_1) > \pi_1, \tag{2}$$

where π_1 is a prespecified cutoff probability and the equality is due to the monotonicity of $\Pr(p > p_0|r_1,n_1)$ with respect to y_1 as shown in the Appendix. The posterior probability $\Pr(p > p_0|r_1,n_1)$ under the Beta-Binomial model can be written as

$$\Pr(p > p_0|r_1,n_1) = \int_{p_0}^{1} \frac{p^{a+r_1-1}(1-p)^{b+n_1-r_1-1}}{B(a+r_1, b+n_1-r_1)} \mathrm{d}p,$$

where $B(a,b)$ is the Beta function with parameters a and b.

Intuitively, constraint (2) can be satisfied even for a very small sample size n_1, if r_1 is chosen to be close to n_1 enough, that is, (2) alone cannot control the sample size. However, if r_1 is too large, it is likely to falsely reject a drug in the first stage. This is not desirable as such an imprudent decision may neglect some promising drugs. In the first stage, the information collected is very limited, so that we need to be conservative. Therefore, we add another constraint to control the posterior error rate when rejecting the drug for futility. Only when the evidence of futility is strong enough, we would terminate the trial early. We control the posterior probability of $p > p_1$ when an early termination is confirmed by

$$\Pr(p > p_1|y_1 < r_1, n_1) < a_1,$$

where a_1 is a prespecified design parameter. The computation of $\Pr(p > p_1|y_1 < r_1, n_1)$ can resort to Bayes' theorem and the Monte Carlo method. First, we obtain the posterior distribution of p given $y_1 < r_1$, and using Bayes' theorem, we have

$$f(p|y_1 < r_1, n_1) = \frac{\Pr(y_1 < r_1|p,n_1)\pi(p)}{\Pr(y_1 < r_1|n_1)},$$

where $\pi(p)$ is the prior density of p. The probability $\Pr(y_1 < r_1|n_1) = \int_0^1 \Pr(y_1 < r_1|p,n_1)\pi(p)\mathrm{d}p$ can be calculated using the Monte Carlo method, that is, we first draw a large number of samples $\{p^{(j)}\}_{j=1}^N$ from the prior distribution of p, and then compute $\frac{1}{N}\sum_{j=1}^N \Pr(y_1 < r_1|p^{(j)}, n_1)$ as an approximation of $\Pr(y_1 < r_1|n_1)$. Similarly, we can obtain

$$\Pr(p > p_1|y_1 < r_1, n_1) = \int_{p_1}^{1} f(p|y_1 < r_1, n_1)\mathrm{d}p.$$

If the trial does not show strong evidence of futility, we enroll more patients in the second stage. Our goal is to carry the promising drug forward and reduce the failure rate of a successive phase III trial, and thus we pay more attention to

determining whether $p > p_1$ or not. Similar to the discussion of stage 1, we hope the drug has an adequately high posterior probability of $p > p_1$ in order to claim it to be promising. Therefore, we adopt the following constraint,

$$\min_{y \geq r} \Pr(p > p_1 | y, n) = \Pr(p > p_1 | r, n) = \int_{p_1}^{1} \frac{p^{a+r-1}(1-p)^{b+n-r-1}}{B(a+r, b+n-r)} \mathrm{d}p > \pi_2, \tag{3}$$

where π_2 is a prespecified cutoff probability.

Similar to stage 1, restricting $\Pr(p > p_1 | r, n)$ alone cannot calibrate reasonable values for parameters (r,n). With r adequately close to n, constraint (3) can always be satisfied for almost any n. Hence, we still consider to restrict the posterior error rate when rejecting the drug in the entire trial (including both stage 1 and stage 2). Denote $\mathbb{R}$ as the event of rejecting the drug after the entire trial, we directly restrict the posterior probability of $p > p_1$ under $\mathbb{R}$,

$$\Pr(p > p_1 | \mathbb{R}) < a_2,$$

where a_2 is a prespecified design parameter. The computation of $\Pr(p > p_1 | \mathbb{R})$ is similar to that of $\Pr(p > p_1 | y_1 < r_1, n_1)$. Based on the above constraints, we can search for the minimal sample size (n_1 or n) as well as the corresponding minimum required number of responses (r_1 or r) which satisfy the constraints. The algorithm for optimal parameters (r_1, n_1) is shown in Algorithm 1, where $n_{\min}$ is the minimal sample size our search starts with at stage 1 and we typically set $n_{\min} = 1$. The maximal sample size we search up to at stage 1 is denoted as $n_{\max}$, which is typically set as $n_{\max} = 50$. The algorithm searching for (r,n) can be developed similarly.

As suggested by Shi and Yin,[31] we typically set $\pi_2 > \pi_1$. At stage 2, our priority is to identify the truly promising drugs so as to decrease the failure rate in subsequent phase III trials, while the posterior error probability when falsely claiming the drug nonpromising is less concerned. Consequently, we may choose a relatively large a_2 to decrease the sample size n. Stage 1 should be conservative due to the limited information, so generally we set $a_1 < a_2$ and keep π_1 not very large.

Algorithm 1. Parameter Search for (r_1, n_1) in the BETEC Design

Input:

Design parameters π_1, a_1, $n_{\min}$ and $n_{\max}$. Typically, $n_{\min} = 1$ and $n_{\max} = 50$.
for $m = n_{\min}, \ldots, n_{\max}$ **do**
  for $y_1 = 1, \ldots, m$ **do**
    Compute $\Pr(p < p_0 | y_1, m)$ and $\sum_{h=0}^{y_1-1} \Pr(p < p_1 | h, m)$
    if $\Pr(p < p_0 | y_1, m) < \pi_1$ and $\sum_{h=0}^{y_1-1} \Pr(p < p_1 | h, m) > a_1$ **then**
      Let $r_1 = y_1$, $n_1 = m$.
      return (r_1, n_1)
    end if
  end for
end for

Output:

Design parameters (r_1, n_1).

2.4 | The δ-BETEC design

In Algorithm 1, to search for the optimal parameters (r_1, n_1, r, n), the Monte Carlo sampling procedure needs to be implemented repeatedly, which is time-consuming. Thus, we provide an alternative to the BETEC design which is called the δ-BETEC design by specifying a δ margin in stage 2. In the δ-BETEC design, we adopt the following constraints to determine parameters (r_1, n_1, r, n),

$$\begin{aligned} \text{Stage 1}: &\quad \Pr(p > p_1 | r_1 - 1, n_1) < b_1, \quad \Pr(p > p_0 | r_1, n_1) > \pi_1, \\ \text{Stage 2}: &\quad \Pr(p > p_1 + \delta | r - 1, n) < b_2, \quad \Pr(p > p_1 | r, n) > \pi_2, \end{aligned}$$

where (π_1, π_2, b_1, b_2) are the prespecified design parameters and δ is a prespecified margin to mimic the two equations in stage 1. In the δ-BETEC design, we choose the default value $\delta = 0.1$, and if $p_1 + 0.1 > 1$, we adopt $\delta = 0.5(1 - p_1)$. In our simulation and real data applications, δ is always selected as 0.1.

Algorithm 2. Parameter Search for (r_1, n_1) in the δ-BETEC Design

Input:

Design Parameters π_1, b_1, $n_{\min}$ and $n_{\max}$. Typically, $n_{\min} = 1$ and $n_{\max} = 50$.
for $m = n_{\min}, \ldots, n_{\max}$ **do**
  for $y_1 = 1, \ldots, m$ **do**
    Compute $\Pr(p < p_0 | y_1, m)$ and $\Pr(p < p_1 | y_1 - 1, m)$
    if $\Pr(p < p_0 | y_1, m) < \pi_1$ and $\Pr(p < p_1 | y_1 - 1, m) > b_1$ **then**
      Let $r_1 = y_1$, $n_1 = m$.
      return (r_1, n_1)
    end if
  end for
end for

Output:

Design parameters (r_1, n_1).

The δ-BETEC design applies the same strategy to control the posterior error rate under the trial outcome of claiming the drug promising. However, for limiting the posterior error rate on the other side, the δ-BETEC design follows a more aggressive strategy by requiring the posterior distribution of p to be largely confined between p_0 and p_1 at stage 1 and between p_1 and $p_1 + \delta$ at stage 2 when the response number reaches the maximal rejecting value (ie, $r_1 - 1$ at stage 1 and $r - 1$ at stage 2). With the conjugate property of the Beta-Binomial model, this strategy greatly simplifies the computation. However, at stage 2, since for moderate sample size n, $\Pr(p > p_1 | r, n) \approx \Pr(p > p_1 | r - 1, n)$, it is impossible to restrict $\Pr(p > p_1 | r - 1, n)$ below a small threshold value b_2 because we require $\Pr(p > p_1 | r, n)$ to be large. Hence, we introduce δ to confine the right tail of the posterior distribution of the response rate p in the second stage. The algorithm to search for the optimal parameters (r_1,n_1) in the δ-BETEC design is given in Algorithm 2. We can search for parameters (r,n) in a similar way.

3 | SIMULATION STUDIES

We conduct extensive simulation studies on the BETEC and δ-BETEC designs to assess their performances under different values of (p_0, p_1). The corresponding R codes are available at https://github.com/JINhuaqing/BETEC. We vary p_0 from 0.05 to 0.3 and fix δ as 0.1 and take a non-informative prior $p \sim \text{Beta}(1,1)$. As discussed in Section 2, we let $\pi_2 > \pi_1$ so we set $(\pi_1, \pi_2) = (0.5, 0.55)$. Table 1 presents the optimal solutions of (r_1, n_1, r, n) under different design parameters. Intuitively, a larger posterior false negative error rate would require a smaller sample size. However, it is worth noting that for the BETEC design, with the same a_2, sometimes a larger value of a_1 yields a larger maximal sample size n. For example when $(p_0, p_1) = (0.1, 0.3)$, the maximal sample size n in the BETEC design with $(a_1, a_2) = (0.02, 0.08)$ is 69, while that with $(a_1, a_2) = (0.01, 0.08)$ is 56. It is because the value of r_1 affects the posterior distribution of p given $\mathbb{R}$, which is used to determine the parameters (r, n). The posterior probabilities of H_1 when the drug is rejected in the entire trial and those when the trial reaches the minimally required level (ie, $\Pr(H_1 | \mathbb{R})$ and $\Pr(H_1 | r, n)$) are also presented in Table 1. Furthermore, we show the probabilities of early termination in the first stage (denoted as PET_0 and PET_1) and the expected sample sizes (denoted as ESS_0 and ESS_1) when response rate is p_0 and p_1, respectively.

It is clear to observe that stage 1 of the BETEC design tends to be conservative as PET_0 is not high enough under H_0. Due to the limited information at stage 1, early stopping for futility is typically not desirable and drugs tend to be allowed to proceed to stage 2. Under H_1, values of PET_1 are small under different trial parameters, which is consistent with our principle in the design. Furthermore, because the small PETs can help to reduce the risk of rejecting a promising drug which is the main concern of the first stage, the relatively small n_1 is acceptable at stage 1. At stage 2, the BETEC design ensures a low posterior probability of H_1 when rejecting drug (denoted as $\mathbb{R}$) and a high posterior probability of H_1 when

TABLE 1 Characteristics of the BETEC and δ-BETEC designs for binary endpoint with $(\pi_1, \pi_2) = (0.5, 0.55)$ under a Beta(1,1) prior distribution and various specifications of (p_0, p_1), (a_1, a_2) for BETEC and (b_1, b_2) for δ-BETEC in terms of the posterior probability of H_1, probability of early termination (PET), expected sample size (ESS), and frequentist type I and II error rates $(\alpha, \beta, \beta_\Delta)$, where β_Δ represents the type II error rate calculated at $p_1 + 0.05$

Design	p_0	p_1	a_1 or b_1	a_2 or b_2	r_1/n_1	r/n	$\Pr(H_1\|\mathbb{R})$	$\Pr(H_1\|r,n)$	PET_0	PET_1	ESS_0	ESS_1	α	β	β_Δ
BETEC	0.05	0.25	0.010	0.080	1/16	19/75	0.080	0.561	0.440	0.010	49.0	74.4	0.000	0.484	0.158
			0.020	0.080	1/13	19/75	0.080	0.561	0.513	0.024	43.2	73.5	0.000	0.487	0.162
			0.020	0.100	1/13	13/51	0.096	0.574	0.513	0.024	31.5	50.1	0.000	0.482	0.201
δ-BETEC			0.010	0.050	1/16	13/51	0.097	0.574	0.440	0.010	35.6	50.6	0.000	0.479	0.198
			0.025	0.050	1/12	13/51	0.099	0.574	0.540	0.032	29.9	49.8	0.000	0.484	0.203
			0.025	0.080	1/12	7/28	0.115	0.557	0.540	0.032	19.4	27.5	0.000	0.430	0.223
BETEC	0.10	0.30	0.010	0.080	1/12	17/56	0.078	0.554	0.282	0.014	43.6	55.4	0.000	0.475	0.196
			0.020	0.080	1/10	21/69	0.076	0.559	0.349	0.028	48.4	67.3	0.000	0.492	0.186
			0.020	0.100	1/10	10/33	0.098	0.554	0.349	0.028	25.0	32.4	0.001	0.453	0.233
δ-BETEC			0.010	0.050	1/12	17/56	0.077	0.553	0.282	0.014	43.6	55.4	0.000	0.475	0.196
			0.025	0.050	1/10	17/56	0.081	0.554	0.349	0.028	40.0	54.7	0.000	0.478	0.200
			0.025	0.080	1/10	10/33	0.095	0.554	0.349	0.028	25.0	32.4	0.001	0.453	0.233
BETEC	0.20	0.40	0.010	0.080	2/12	16/39	0.078	0.568	0.275	0.020	31.6	38.5	0.002	0.493	0.257
			0.020	0.080	1/7	16/39	0.079	0.568	0.210	0.028	32.3	38.1	0.002	0.496	0.261
			0.020	0.100	1/7	9/22	0.093	0.556	0.210	0.028	18.9	21.6	0.020	0.457	0.279
δ-BETEC			0.010	0.050	3/17	24/59	0.062	0.556	0.310	0.012	46.0	58.5	0.000	0.494	0.214
			0.025	0.050	1/7	24/59	0.065	0.556	0.210	0.028	48.1	57.5	0.000	0.499	0.220
			0.025	0.080	1/7	16/39	0.078	0.568	0.210	0.028	32.3	38.1	0.002	0.496	0.261
BETEC	0.30	0.50	0.010	0.080	4/13	13/25	0.077	0.577	0.421	0.046	20.0	24.4	0.017	0.501	0.307
			0.020	0.080	2/7	13/25	0.080	0.577	0.329	0.063	19.1	23.9	0.017	0.507	0.314
			0.020	0.100	2/7	8/15	0.099	0.598	0.329	0.063	12.4	14.5	0.050	0.502	0.349
δ-BETEC			0.010	0.050	7/24	28/55	0.052	0.553	0.388	0.011	43.0	54.6	0.001	0.500	0.228
			0.025	0.050	4/14	28/55	0.054	0.553	0.355	0.029	40.4	53.8	0.001	0.502	0.230
			0.025	0.080	4/14	19/37	0.064	0.564	0.355	0.029	28.8	36.3	0.005	0.501	0.271

declaring the drug as promising at the boundary values, which shows that BETEC design inherits the merits of the BET design.

To better demonstrate the characteristics of the BETEC design, we further calculate the frequentist type I and type II error rates α and β. As the focus of our design is on controlling the false positive rate, the type II errors are inevitably inflated. Note that the type II error rate is calculated at $p = p_1$, the boundary value of H_1. If we recalculate the type II error rate at $p_1 + 0.05$ (denoted as β_Δ) which is marginally larger than p_1, then the frequentist false negative errors are dramatically reduced to a decent level (around 0.25). Meanwhile, as the type II errors are inflated, the BETEC design yields extremely small type I errors which are typically around the scale of 10^{-3} or even lower. Such results are consistent with our design principle that is to control the false positive rates. From a Bayesian perspective, the posterior probability of H_1 when rejecting the drug, $\Pr(H_1|\mathbb{R})$, is well controlled which indicates that if the drug is rejected, the probability of overlooking a promising drug is small under the BETEC design. Such a conflict between β and $\Pr(H_1|\mathbb{R})$ is due to the differences between frequentist and Bayesian perspectives.

In the left panels of Figure 1, we exhibit the density plots of the posterior distributions when the total response number reaches the minimum required level (panel A) and when rejecting the drug in the entire trial (panel C) over a range of specified (p_0, p_1) for the BETEC design. Clearly, the majority of the posterior distribution of p under the minimally required level lies beyond the desirable target response rate p_1, which shows that the BETEC design guarantees the effectiveness

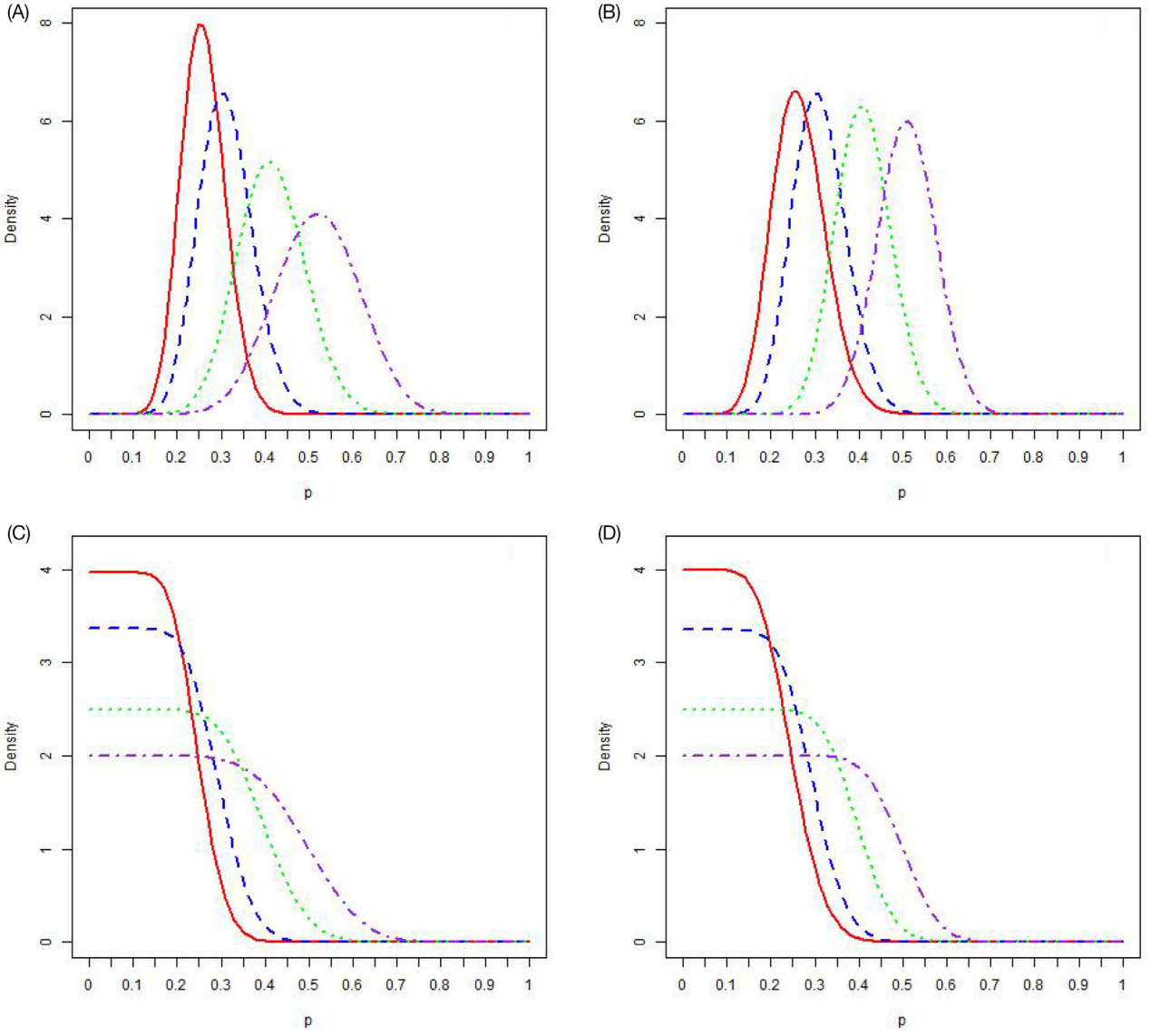

FIGURE 1 Comparisons of posterior distributions of p when the total response number reaches the minimum required level (top panels) and when rejecting the drug in the entire trial (bottom panels) over a range of specified (p_0, p_1) under the BETEC design (left panels) and δ-BETEC design (right panels). The red solid line is for $(p_0, p_1) = (0.05, 0.25)$, the blue dashed line for $(p_0, p_1) = (0.1, 0.3)$, the green dotted line for $(p_0, p_1) = (0.2, 0.4)$, and the purple dash-dotted line for $(p_0, p_1) = (0.3, 0.5)$ [Colour figure can be viewed at wileyonlinelibrary.com]

of the drug when declaring it promising. Meanwhile, most of the posterior distribution of p when rejecting the drug concentrates below p_1, which demonstrates a good control of the posterior error rate when rejecting the drug under the BETEC design. These results indicate that the BETEC design strikes a good balance between the posterior false positive and false negative error rates.

In a summary, the BETEC design provides better interpretability with more transparent design parameters. In the BET design,[31] the HPD intervals are used for restricting the variance of the posterior distribution of p. However, the lengths of HPD intervals are difficult to interpret and thus may cause ambiguity in specifying their values in practice. That is, (ℓ_1, ℓ_2) do not have a clear range to choose from, so selection of proper values for (ℓ_1, ℓ_2) is not an easy task. Yet, in the BETEC design, the meaning of (a_1, a_2) is straightforward and intuitive. They represent the maximal posterior error rates that can be tolerated when rejecting the drug at stage 1 and at stage 2, respectively. The design parameters (a_1, a_2) can be chosen in a natural way: if the high risk of rejecting a promising drug is unacceptable, small values of (a_1, a_2) should be chosen, while if enrolling an adequate number of subjects is a difficult task, values of (a_1, a_2) can be increased to reduce the sample size.

Similar discussions can be applied to the δ-BETEC design based on Table 1 and the right panels of Figure 1. It is worth noting that compared with the BETEC design, the maximal sample size of the δ-BETEC design is more stable under the same $(b_1,\ b_2)$ when varying the values of (p_0, p_1). The δ-BETEC design is easy to implement and has the main merits of the BET design (ie, ensuring the efficacy of the drug when claim it to be promising), while the introduction of δ in the second stage helps to confine most of the posterior distribution of p between p_1 and $p_1 + \delta$. The design parameters (b_1, b_2) are meaningful, which are used to restrict the posterior error rates when rejecting the drug at the maximal rejecting numbers of responses at stage 1 and stage 2, respectively.

Finally, we compare our methods with three Bayesian phase II single-arm designs, including the BET design,[31] the dual threshold design[22] (DTD) and the predictive probability design[24] (PPD). We also include Simon's optimal two-stage design[2] and the modified Simon (m-Simon) design in the comparisons. The DTD is a two-stage design with decision boundaries based on posterior probabilities. At the end of the first stage, the DTD would stop for futility if $\Pr(p < p_0|y_1, n_1) > \gamma_1$, otherwise the trial continues to the second stage. At stage 2, the drug is declared as promising if $\Pr(p > p_1|y, n) > \gamma_2$. The PPD, based on the predictive probabilities under the Beta-Binomial model, can monitor the trial continuously, which focuses on the posterior probabilities of $p > p_0$. By contrast, the three BET-based designs and DTD pay more attention to posterior probabilities of $p \geq p_1$. The m-Simon design refers to Simon's optimal design when controlling the size at $p_{0m} = (p_0 + p_1)/2$ and limiting power at $p_{1m} = p_{0m} + (p_1 - p_0)$. Such modifications are based upon the observation that the original Simon's design always yields the minimal response rate r/n between p_0 and p_1 to declare the drug as promising, while our designs generally yield r/n close to p_1. For a direct comparison, we set $(\pi_1, \pi_2) = (0.5, 0.55)$ for all BET, BETEC, and δ-BETEC designs. To ensure the maximal sample sizes under different designs comparable, we calibrate the maximal sample sizes by tuning the trial parameters for different methods under the corresponding recommended scales. For the PPD, we fix the maximal sample size at $n = 45$ and monitor the trial at $n_1 = 15$.

Table 2 shows the results of the comparisons among the seven phase II trial designs, where the frequentist error rates $(\alpha, \beta, \beta_\Delta)$ under the m-Simon design are calculated based on $(p_0, p_1, p_1 + 0.05)$ rather than $(p_{0m}, p_{1m}, p_{1m} + 0.05)$. The Bayesian posterior probabilities for all methods are calculated based on a Beta(1, 1) prior. Among the seven designs, the DTD yields the largest $\Pr(H_1|r, n)$ as well as the largest frequentist type II error rate. All the BET-related methods and DTD aim to guarantee high posterior probabilities of H_1 when declaring the drug as promising, and thus they unanimously show reasonable posterior probabilities from the Bayesian perspective while being over-restrictive in terms of the frequentist error rates. The m-Simon design, which is a frequentist method, shows similar features to the four Bayesian designs, as it controls the size at p_{0m} the middle point between p_0 and p_1. The PPD focuses on the posterior probability of $p > p_0$, and it possesses similar properties to the original Simon design, that is, low posterior probabilities of H_1 when declaring the drug as promising and low frequentist type II error rates. The minimally required response rates r/n under the BET-related designs are slightly above p_1, while those of PPD and Simon's design fall in the middle of p_0 and p_1 which are not high enough to declare $p > p_1$ intuitively. The above phenomenon indicates that due to the existence of the gap between p_0 and p_1, to reduce the false positive rate, more attention should be paid on p_1 rather than p_0 in the trial design. At the same time, the Bayesian operating characteristics are substantially different from the frequentist counterparts. Due to the limited maximal sample sizes, it is very difficult to control both the Bayesian and frequentist error rates under some decent level simultaneously. Among the seven designs, the PPD has significantly smaller PET_0 compared with the rest methods, while the m-Simon design yields notably larger PET_0 and PET_1 which helps to reduce the corresponding ESS_0 and ESS_1.

4 | TRIAL APPLICATIONS

4.1 | Gemcitabine-eribulin combination trial

As an illustration, we apply the BETEC and δ-BETEC designs to the gemcitabine-eribulin trial.[35] This single-arm phase II trial aimed to assess the efficacy of a gemcitabine-eribulin combination for csiplatin-ineligible patients with metastatic urothelial carcinoma. The primary endpoint was treatment efficacy evaluated by the overall objective response, which included overall confirmed complete response (CR) and partial response (PR). In Sadeghi et al,[35] Simon's optimal two-stage design was used with $(\alpha, \beta) = (0.09, 0.09)$ for the hypothesis test $H_0 : p \leq 0.2$ vs $H_1 : p \geq 0.5$. Under such trial specifications, the design parameters were $(r_1, n_1, r, n) = (2, 7, 7, 21)$. In the first stage, the trial observed 4 responses among 7 patients. Based on Simon's design, the trial proceeded to the second stage. In the end, the trial enrolled three additional subjects beyond the originally planned 21 patients, and observed 12 responses out of a total of 24 patients.

TABLE 2 Comparisons of the BETEC, δ-BETEC, BET, DTD, PPD, Simon's optimal two-stage design, and the modified Simon (m-Simon) design in terms of the posterior probability of H_1, probability of early termination (PET), expected sample size (ESS), frequentist type I and II error rates (α, β, β_Δ), and the minimally required response rate $\hat{p}$, where β_Δ represents the type II error rate calculated at $p_1 + 0.05$

Design	p_0	p_1	r_1/n_1	r/n	$\Pr(H_1\|\mathbb{R})$	$\Pr(H_1\|r,n)$	PET_0	PET_1	ESS_0	ESS_1	α	β	β_Δ	$\hat{p}$
BETEC	0.05	0.25	1/9	12/47	0.108	0.577	0.630	0.075	23.1	44.1	0.000	0.494	0.225	0.255
δ-BETEC			1/10	12/47	0.105	0.577	0.599	0.056	24.8	44.9	0.000	0.489	0.218	0.255
BET			1/20	11/43	0.103	0.580	0.358	0.003	34.8	42.9	0.000	0.477	0.215	0.256
DTD			1/36	14/45	0.204	0.847	0.158	0.000	43.6	45.0	0.000	0.784	0.509	0.311
PPD			0/15	4/45	0.000	0.005	0.000	0.000	45.0	45.0	0.186	0.002	0.000	0.089
Simon			2/19	6/47	0.012	0.027	0.755	0.031	25.9	46.1	0.024	0.039	0.011	0.128
m-Simon[a]			4/20	11/46	0.098	0.478	0.984	0.225	20.4	40.1	0.000	0.433	0.193	0.239
BETEC	0.10	0.30	1/7	14/46	0.098	0.559	0.478	0.082	27.3	43.8	0.000	0.492	0.238	0.304
δ-BETEC			1/8	14/46	0.092	0.559	0.430	0.058	29.6	43.8	0.000	0.483	0.227	0.304
BET			2/20	14/44	0.108	0.635	0.392	0.008	34.6	43.8	0.000	0.548	0.278	0.318
DTD			2/36	17/46	0.187	0.860	0.113	0.000	44.9	46.0	0.000	0.809	0.555	0.370
PPD			1/15	7/45	0.002	0.017	0.206	0.005	38.8	44.9	0.153	0.012	0.002	0.156
Simon			3/21	8/43	0.009	0.056	0.648	0.027	28.7	42.4	0.054	0.049	0.013	0.186
m-Simon			5/20	13/45	0.088	0.470	0.957	0.238	21.1	39.1	0.000	0.439	0.209	0.289
BETEC	0.20	0.40	2/12	18/44	0.073	0.564	0.275	0.020	35.2	43.4	0.001	0.494	0.246	0.409
δ-BETEC			2/12	18/44	0.073	0.564	0.275	0.020	35.2	43.4	0.001	0.494	0.246	0.409
BET			5/24	18/44	0.072	0.564	0.460	0.013	34.8	43.7	0.001	0.492	0.244	0.409
DTD			6/35	22/47	0.146	0.835	0.272	0.001	43.7	47.0	0.001	0.790	0.543	0.468
PPD			2/15	13/45	0.005	0.068	0.167	0.005	40.0	44.8	0.098	0.048	0.010	0.289
Simon			5/20	13/45	0.011	0.068	0.630	0.051	29.0	43.7	0.085	0.079	0.025	0.289
m-Simon			7/21	18/45	0.072	0.516	0.891	0.200	23.6	40.2	0.001	0.474	0.236	0.400
BETEC	0.30	0.50	4/13	23/45	0.060	0.559	0.421	0.046	31.5	43.5	0.002	0.504	0.254	0.511
δ-BETEC			4/14	23/45	0.059	0.559	0.355	0.029	34.0	44.1	0.002	0.502	0.251	0.511
BET			7/22	24/46	0.067	0.615	0.494	0.026	34.1	45.4	0.001	0.559	0.296	0.522
DTD			8/31	25/45	0.102	0.769	0.245	0.002	41.6	45.0	0.000	0.724	0.468	0.556
PPD			4/15	18/45	0.007	0.092	0.297	0.018	36.1	44.5	0.095	0.077	0.020	0.400
Simon			7/21	18/45	0.008	0.092	0.551	0.039	31.8	44.1	0.090	0.088	0.024	0.400
m-Simon			8/18	23/46	0.062	0.500	0.859	0.240	21.9	39.3	0.003	0.492	0.255	0.500
BETEC	0.40	0.60	4/10	27/44	0.051	0.556	0.382	0.055	31.0	42.1	0.003	0.515	0.258	0.614
δ-BETEC			6/15	27/44	0.058	0.556	0.403	0.034	32.3	43.0	0.003	0.510	0.252	0.614
BET			10/25	27/43	0.060	0.629	0.425	0.013	35.4	42.8	0.002	0.582	0.317	0.628
DTD			10/26	30/45	0.095	0.808	0.364	0.008	38.1	44.9	0.000	0.775	0.525	0.667
PPD			5/15	23/45	0.006	0.109	0.217	0.009	38.5	44.7	0.085	0.090	0.021	0.511
Simon			8/18	23/46	0.008	0.082	0.563	0.058	30.2	44.4	0.095	0.100	0.030	0.500
m-Simon			10/18	27/45	0.050	0.484	0.865	0.263	21.6	37.9	0.004	0.494	0.253	0.600
BETEC	0.50	0.70	7/13	33/46	0.040	0.567	0.500	0.062	29.5	43.9	0.002	0.535	0.249	0.717
δ-BETEC			7/14	33/46	0.039	0.567	0.395	0.031	33.4	45.0	0.002	0.531	0.245	0.717

(Continues)

TABLE 2 (Continued)

Design	p_0	p_1	r_1/n_1	r/n	$\Pr(H_1\|\mathbb{R})$	$\Pr(H_1\|r,n)$	PET_0	PET_1	ESS_0	ESS_1	α	β	β_Δ	$\hat{p}$
BET			13/26	32/44	0.046	0.620	0.423	0.009	36.4	43.8	0.002	0.583	0.294	0.727
DTD			10/20	35/46	0.073	0.794	0.412	0.017	35.3	45.6	0.000	0.767	0.489	0.761
PPD			7/15	27/45	0.003	0.068	0.304	0.015	35.9	44.5	0.112	0.063	0.012	0.600
Simon			10/18	27/45	0.006	0.068	0.593	0.060	29.0	43.4	0.100	0.094	0.025	0.600
m-Simon			10/15	31/44	0.044	0.491	0.849	0.278	19.4	35.9	0.004	0.518	0.260	0.705
BETEC	0.60	0.80	7/11	36/44	0.030	0.559	0.467	0.050	28.6	42.3	0.002	0.533	0.209	0.818
δ-BETEC			8/13	36/44	0.030	0.559	0.426	0.030	30.8	43.1	0.002	0.531	0.207	0.818
BET			15/24	36/43	0.042	0.677	0.511	0.013	33.3	42.8	0.000	0.650	0.313	0.837
DTD			8/14	40/46	0.067	0.856	0.308	0.012	36.2	45.6	0.000	0.841	0.546	0.870
PPD			9/15	32/45	0.002	0.062	0.390	0.018	33.3	44.5	0.080	0.062	0.008	0.711
Simon			10/15	31/44	0.004	0.052	0.597	0.061	26.7	42.2	0.085	0.088	0.019	0.705
m-Simon			14/18	36/44	0.034	0.559	0.906	0.284	20.4	36.6	0.001	0.569	0.248	0.818

[a]The frequentist error rates $(\alpha, \beta, \beta_\Delta)$ under the m-Simon design are calculated based on $(p_0, p_1, p_1 + 0.05)$.

TABLE 3 Comparisons of the BETEC, δ-BETEC, BET designs and Simon's design based on the gemcitabine-eribulin combination trial in terms of design parameters, posterior probabilities of H_0 and H_1

Designs	r_1/n_1	r/n	$\Pr(H_1\|y_1 < r_1, n_1)$	$\Pr(H_0\|r_1, n_1)$	$\Pr(H_1\|\mathbb{R})$	$\Pr(H_1\|r,n)$
BETEC	1/7	13/24	0.004	0.497	0.096	0.655
δ-BETEC	1/7	14/26	0.004	0.497	0.092	0.649
BET	2/9	14/26	0.006	0.322	0.094	0.649
Simon	2/7	7/21	0.019	0.203	0.017	0.067

Under Simon's design, even if we assume all three additional patients are responders and remove them from the total number of responses, we can still draw a conclusion that the gemcitabine-eribulin combination is promising (as we observed $12 - 3 = 9$ responses among 21 patients). However, if we adopt Beta(1, 1) as the prior distribution of the response rate p and compute the posterior probability of H_1 based on the trial results $(y, n) = (12, 24)$, $\Pr(H_1|y, n) = 0.5$, which is still not high enough.

We then apply the BET, BETEC, and δ-BETEC designs to this trial. For all the three designs, we adopt a non-informative prior Beta(1,1) and set $\pi_1 = \pi_2 = (0.5, 0.6)$. We use $(\ell_1, \ell_2) = (0.18, 0.16)$ for the BET design, $(a_1, a_2) = (0.005, 0.1)$ for the BETEC design and $(b_1, b_2) = (0.005, 0.15)$ for the δ-BETEC design. The results in Table 3 show that the posterior probability of H_1 under the minimally required level of Simon's design at the end of stage 2 is 0.067, which is rather low. Thus based on Simon's design, even if we declare the combination as promising, it is still very likely that the subsequent phase III trial may end up with a failure. The BETEC and δ-BETEC designs require one response among seven subjects in the first stage, while the BET design requires two responses out of nine subjects. Nevertheless, this trial would proceed to the second stage under all the three Bayesian designs. At stage 2, the BETEC design requires 13 responses out of 24 patients and this trial fails to reach this critical value. Under the BETEC design, the gemcitabine-eribulin combination is not promising for csiplatin-ineligible patients with metastatic urothelial carcinoma and should not proceed into a phase III study. Under the δ-BETEC and BET designs, 14 responses are required out of 26 subjects, which is also more stringent than Simon's design.

4.2 | Recurrent glioblastoma trial

Silvani et al[36] reported a phase II study on the efficacy of ortataxel in recurrent glioblastoma (GBM). The primary objective of the study was to evaluate the activity of ortataxel in terms of progression free survival (PFS) at 6 months after the

TABLE 4 Comparisons of the BETEC, δ-BETEC, BET designs, and Simon's design based on the recurrent glioblastoma trial in terms of design parameters, posterior probabilities of H_0 and H_1 and the minimally required response rates $\hat{p}$

Designs	r_1/n_1	r/n	$\Pr(H_1 \mid y_1 < r_1, n_1)$	$\Pr(H_0 \mid r_1, n_1)$	$\hat{p}$
BETEC	7/37	22/62	0.002	0.500	0.355
δ-BETEC	7/36	20/56	0.002	0.467	0.357
BET	7/35	21/58	0.003	0.434	0.362
Simon	7/33	16/58	0.005	0.367	0.276

enrollment. The trial adopted Simon's minimax two-stage design under which the type I and type II error rates were both fixed as 10%. The trial was built upon the hypothesis test $H_0 : p \leq 0.2$ versus $H_1 : p \geq 0.35$. Under these design parameters, at the first stage, at least seven responses out of 33 subjects are needed to carry the trial to the second stage. At stage 2, a total number of patients would be increased to 58 and to declare the drug as promising, at least 16 responders are needed. This trial was terminated early for futility because there were only four patients alive and free of progression among 35 subjects at the end of the first stage (two additional subjects were for administrative and logistic issues).

As another illustration, we also apply the BETEC, δ-BETEC, and BET designs to this trial. Intuitively, with similar sample sizes in the first stage, these three designs would all reject the drug for futility as they are intrinsically more stringent than Simon's design. In all the three designs, we use Beta(1, 1) as the prior distribution for the response rate p and specify $(\pi_1, \pi_2) = (0.5, 0.55)$. We set $(a_1, a_2) = (0.03, 0.19)$ for the BETEC design, $(b_1, b_2) = (0.01, 0.05)$ for the δ-BETEC design, and $(\ell_1, \ell_2) = (0.09, 0.095)$ for the BET design. The results in Table 4 show that as expected under the three designs, the trial would be terminated early for futility. With comparable sample sizes at stage 1, Simon's design only needs seven responses to continue the trial, while the three Bayesian designs require 10 responses. We also investigate the minimal response rates to declare the drug as promising at the end of stage 2 under the four designs. Simon's design only needs $\hat{p} = r/n = 0.276$ to claim the drug to be promising, which is not high enough because the target response rate is $p_1 = 0.35$. As for the Bayesian designs, they all require similar minimal response rates above 0.35.

5 | DISCUSSION

As an extension of the BET design, we propose the BETEC and δ-BETEC designs by replacing the HPD interval length constraints with two posterior probability constraints, which focus on controlling posterior error rates when rejecting the drug. Our designs greatly enhance the transparency and interpretability of the design parameters. They inherit the virtues of the original BET design, which pays more attention to the target response level rather than the uninteresting response level. It helps to reduce the failure rate of the subsequent phase III clinical trial when the drug is declared as promising. In the meanwhile, the BETEC design is easier to implement compared to the BET design as all the constraints for searching boundary values and sample size parameters (r_1, n_1, r, n) are based on posterior probabilities, and the pre-determined design parameters (π_1, π_2, a_1, a_2) have intuitive interpretations. In particular, (π_1, π_2) are the minimal posterior probabilities of $p > p_0$ and $p > p_1$ we aim to reach when the trial achieves the minimally required level at stage 1 and stage 2, respectively, while (a_1, a_2) are the maximal posterior error rates we can tolerate when rejecting the drug at stage 1 and stage 2, respectively. In this way, we can balance the posterior false positive rate and false negative rate directly via choosing proper design parameters. To ease the computation, we also propose an alternative δ-BETEC design which has similar properties to the BETEC design while immensely reducing the computational burden. From a Bayesian perspective, the BETEC and δ-BETEC designs can incorporate the information of the historical data by selecting a suitable prior. Furthermore, the extension to other endpoints such as the survival endpoint is straightforward.

According to the extensive simulation results, it is also worth noting that our methods tend to be conservative at stage 1, as PET_0 under H_0 is not high enough and PET_1 under H_1 is very low. Such features can help to reduce the risk of rejecting a promising drug in the first stage and thus reduce the required sample size n_1 at stage 1. However, the expected sample size is not decreased due to the low PETs. Compared with Simon's design, the low type I error and relatively high type II error indicate that our methods are more strict to declare the investigational drug as promising. In term of the minimally required response rate r/n, the BETEC and δ-BETEC designs are not over-strict, as the values of r/n in the simulation are all marginally larger than the target response rate p_1. In this sense, the BETEC and δ-BETEC designs are more desirable when the positive conclusion of the current trial will directly lead to a large-scale phase III trial, because

the low false positive rate can help to reduce the risk of the failure of the phase III trial. However, if the current trial will be followed by another small-scale randomized trial, then Simon's design is more preferable due to its higher power. Finally, although our methods are developed from a Bayesian perspective, the commonly used frequentist inference methods (ie, point estimation and confidence intervals) can still be applied to the data from the trial under the BETEC design.

ACKNOWLEDGMENTS

We would like to thank two anonymous referees and associate editor for their insightful suggestions in greatly improving the quality of this article. The research was supported by a grant No. 17307218 for Guosheng Yin from the Research Grants Council of Hong Kong.

DATA AVAILABILITY STATEMENT

The corresponding R codes are available at https://github.com/JINhuaqing/BETEC.

ORCID


Guosheng Yin https://orcid.org/0000-0003-3276-1392

REFERENCES

1. Brown SR, Gregory WM, Twelves CJ, Brown JM. *A Practical Guide to Designing Phase II Trials in Oncology*. Hoboken, NJ: John Wiley & Sons; 2014.
2. Simon R. Optimal two-stage designs for phase II clinical trials. *Controll Clin Trials*. 1989;10(1):1-10.
3. Verger E, Gil M, Yaya R, et al. Temozolomide and concomitant whole brain radiotherapy in patients with brain metastases: a phase II randomized trial. *Int J Radiation Oncol Biol Phys*. 2005;61(1):185-191.
4. Cristofanilli M, Valero V, Mangalik A, et al. Phase II, randomized trial to compare anastrozole combined with gefitinib or placebo in postmenopausal women with hormone receptor–positive metastatic breast cancer. *Clin Cancer Res*. 2010;16(6):1904-1914.
5. Shi H, Yin G. Bayesian two-stage design for phase II clinical trials with switching hypothesis tests. *Bayesian Anal*. 2017;12(1):31-51.
6. Shi H, Yin G. Two-stage seamless transition design from open-label single-arm to randomized double-arm clinical trials. *Stat Methods Med Res*. 2018;27(1):158-171.
7. Shi H, Zhang T, Yin G. START: single-to-double arm transition design for phase II clinical trials. *Pharmaceut Stat*. 2020;19(4):454-467.
8. Grayling MJ, Mander AP. Do single-arm trials have a role in drug development plans incorporating randomised trials? *Pharmaceut Stat*. 2016;15(2):143-151.
9. Sambucini V. Comparison of single-arm vs. randomized phase II clinical trials: a Bayesian approach. *J Biopharmaceut Stat*. 2015;25(3):474-489.
10. Gehan EA. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *J Chronic Diseas*. 1961;13(4):346-353.
11. Fleming TR. One-sample multiple testing procedure for phase II clinical trials. *Biometrics*. 1982;38(1):143-151.
12. Ensign LG, Gehan EA, Kamen DS, Thall PF. An optimal three-stage design for phase II clinical trials. *Statistics in Medicine*. 1994;13(17):1727-1736.
13. Chen TT. Optimal three-stage designs for phase II cancer clinical trials. *Stat Med*. 1997;16(23):2701-2711.
14. Shuster J. Optimal two-stage designs for single arm phase II cancer trials. *J Biopharmaceut Stat*. 2002;12(1):39-51.
15. Lin Y, Shih WJ. Adaptive two-stage designs for single-arm phase IIA cancer clinical trials. *Biometrics*. 2004;60(2):482-490.
16. Chen K, Shan M. Optimal and minimax three-stage designs for phase II oncology clinical trials. *Contemp Clin Trials*. 2008;29(1):32-41.
17. Mander A, Thompson S. Two-stage designs optimal under the alternative hypothesis for phase II cancer clinical trials. *Contemp Clin Trials*. 2010;31(6):572-578.
18. Englert S, Kieser M. Optimal adaptive two-stage designs for phase II cancer clinical trials. *Biometr J*. 2013;55(6):955-968.
19. Shan G, Wilding GE, Hutson AD, Gerstenberger S. Optimal adaptive two-stage designs for early phase II clinical trials. *Stat Med*. 2016;35(8):1257-1266.
20. Yin G. *Clinical Trial Design: Bayesian and Frequentist Adaptive Methods*. Vol 876. Hoboken, NJ: John Wiley & Sons; 2012.
21. Thall PF, Simon R. Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics*. 1994;50(2):337-349.
22. Tan SB, Machin D. Bayesian two-stage designs for phase II clinical trials. *Stat Med*. 2002;21(14):1991-2012.
23. Wang YG, Leung DHY, Li M, Tan SB. Bayesian designs with frequentist and Bayesian error rate considerations. *Stat Methods Med Res*. 2005;14(5):445-456.
24. Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. *Clinical Trials*. 2008;5(2):93-106.
25. Dong G, Shih WJ, Moore D, Quan H, Marcella S. A Bayesian–frequentist two-stage single-arm phase II clinical trial design. *Stat Med*. 2012;31(19):2055-2067.
26. Van Norman GA. Phase II trials in drug development and adaptive trial design. *JACC Basic Trans Sci*. 2019;4(3):428-437.
27. Gan HK, You B, Pond GR, Chen EX. Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. *J Nat Cancer Inst*. 2012;104(8):590-598.

28. Mandel JJ, Yust-Katz S, Patel AJ, et al. Inability of positive phase II clinical trials of investigational treatments to subsequently predict positive phase III clinical trials in glioblastoma. *Neuro-oncology*. 2017;20(1):113-122.
29. Grayling MJ, Dimairo M, Mander AP, Jaki TF. A review of perspectives on the use of randomization in phase II oncology trials. *JNCI J Nat Cancer Inst*. 2019;111(12):1255-1262.
30. Ritchie G, Gasper H, Man J, et al. Defining the most appropriate primary end point in phase 2 trials of immune checkpoint inhibitors for advanced solid cancers: a systematic review and meta-analysis. *Jama Oncol*. 2018;4(4):522-528.
31. Shi H, Yin G. Bayesian enhancement two-stage design for single-arm phase II clinical trials with binary and time-to-event endpoints. *Biometrics*. 2018;74(3):1055-1064.
32. Englert S. *Flexible Designs for Single-Arm Phase II Trials in Oncology*. Berlin, Germany: Epubli; 2014.
33. Shan G, Zhang H, Jiang T, Peterson H, Young D, Ma C. Exact p-values for Simon's two-stage designs in clinical trials. *Stat Biosci*. 2016;8(2):351-357.
34. Lee SJ, Zelen M. Clinical trials and sample size considerations: another perspective. *Stat Sci*. 2000;15(2):95-110.
35. Sadeghi S, Groshen SG, Tsao-Wei DD, et al. Phase II California cancer consortium trial of gemcitabine-eribulin combination in cisplatin-ineligible patients with metastatic urothelial carcinoma: final report (NCI-9653). *J Clin Oncol*. 2019;37(29):2682-2688.
36. Silvani A, De Simone I, Fregoni V, et al. Multicenter, single arm, phase II trial on the efficacy of ortataxel in recurrent glioblastoma. *J Neuro-oncology*. 2019;142(3):455-462.

How to cite this article: Jin H, Yin G. Bayesian enhancement two-stage design with error control for phase II clinical trials. *Statistics in Medicine*. 2020;39:4452–4465. https://doi.org/10.1002/sim.8734

APPENDIX A. PROOF OF MONOTONICITY

We assume $p \sim \text{Beta}(a, b), y_1|n_1 \sim \text{Binomial}(p,n_1)$ and $p_0 \in (0,1)$. We aim to prove the monotonicity of $\Pr(p > p_0|y_1, n_1)$ with respect to y_1.

To prove $\Pr(p > p_0|y_1, n_1)$ is a non-decreasing function of y_1 when $p_0 \in (0, 1)$, it is sufficient to show $\Pr(p > p_0|y_1, n_1) - \Pr(p > p_0|y_1 - 1, n_1) \geq 0$ for $n_1 \geq y_1 \geq 1$. Let $G(p_0) = \Pr(p > p_0|y_1, n_1) - \Pr(p > p_0|y_1 - 1, n_1)$, and it is clear that $G(0) = 0$ as the support of p is [0,1]. By definition, we have

$$G(p_0) = \int_{p_0}^{1} \left\{ \frac{p^{a+y_1-1}(1-p)^{b+n_1-y_1-1}}{B(a+y_1, b+n_1-y_1)} - \frac{p^{a+y_1-2}(1-p)^{b+n_1-y_1}}{B(a+y_1-1, b+n_1-y_1+1)} \right\} \mathrm{d}p,$$

where $B(a,b)$ is the Beta function, satisfying

$$B(a+y_1-1, b+n_1-y_1+1) = \frac{b+n_1-y_1}{a+y_1-1} B(a+y_1, b+n_1-y_1).$$

Thus, we obtain

$$\begin{aligned} G(p_0) &= \int_{p_0}^{1} \frac{(b+n_1-y_1)p^{a+y_1-1}(1-p)^{b+n_1-y_1-1} - (a+y_1-1)p^{a+y_1-2}(1-p)^{b+n_1-y_1}}{(b+n_1-y_1)B(a+y_1, b+n_1-y_1)} \mathrm{d}p \\ &= \int_{p_0}^{1} g(p) \mathrm{d}p, \end{aligned}$$

where

$$g(p) = \frac{p^{a+y_1-2}(1-p)^{b+n_1-y_1-1}\{(a+b+n_1-1)p - (a+y_1-1)\}}{(b+n_1-y_1)B(a+y_1, b+n_1-y_1)}.$$

Note that for $p \in [0, (a+y_1-1)/(a+b+n_1-1)]$, $g(p) \leq 0$; otherwise $g(p) > 0$. Therefore, $G(p_0)$ is a non-decreasing function of p_0. For any $p_0 \in (0,1)$, we have $G(p_0) \geq G(0) = 0$, and thus $\Pr(p > p_0|y_1, n_1)$ is a non-decreasing function of y_1.