Randomized Phase II Trials: a Bayesian Two-Stage Design

Matteo Cellamare[1,*], Valeria Sambucini[1], Federica Siena[1]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 7

[1] Sapienza Università di Roma, Dipartimento di Scienze Statistiche, Roma, Italia
[*] matteo.cellamare@uniroma1.it

Abstract

Single-arm two-stage designs are commonly used in phase II of clinical trials. However, the use of randomization in phase II trials is currently increasing. We propose a randomized version of a Bayesian two-stage design due to Tan and Machin [2]. The idea is to select the two-stage sample sizes by ensuring a large posterior probability that the true response rate of the experimental treatment exceeds that of the standard agent, assuming that the experimental treatment is actually more effective. This optimistic assumption is realized by fixing virtual outcomes.

Keywords: Bayesian approach; phase II clinical trials; randomization; two-stage design.

1 Introduction

Phase II trials are typically conducted as single-arm studies based on a binary endpoint, where the patients are recruited in two stages to let the trial stop if the observed response rate is unacceptably low. In this context the most popular two-stage designs developed under a frequentist framework are due to Simon [1].

Among the Bayesian two-stage designs proposed in the literature, we in particular focus on the Single Threshold Design (STD) presented by Tan and Machin [2].

Let us denote by p_X the unknown response probability of an experimental treatment and define the treatment promising if p_X exceeds a target of clinical interest, p^*. The STD selects the two-stage sample sizes by ensuring a large posterior probability that $p_X > p^*$, under the assumption that the observed

response rate is slightly larger than the target. The results are strongly affected by the choice of p^*, that is typically defined *a priori* from historical data on the expected efficacy of the best available treatment.

The use of historical response rates is one of the main criticisms moved to single-arm studies and the introduction of randomization in phase II of clinical trials is widely debated in the recent literature [3, 4].

A general scheme to conduct a randomized two-stage design is provided by Jung [5]. Let $\mathcal{X}$ and $\mathcal{Y}$ be the experimental and the standard arm, respectively. At the first stage, n_1 patients are enrolled in each arm. Let us denote by x_1 and y_1 the observed number of responders for $\mathcal{X}$ and $\mathcal{Y}$, respectively. If $x_1 - y_1 \geq a_1$, where $a_1 \in [-n_1, n_1]$, the trial continues to the second stage, otherwise it stops. At the second stage we accrue n_2 additional patients to each arm and observe the number of responders, x and y, out of the total of $n = n_1 + n_2$ patients. Then if $x - y \geq a$, where $a \in [a_1 - n_2, n]$, we proceed to phase III; otherwise the trial terminates. In particular, Jung [5] suggests to select the values (n_1, a_1, n, a) by minimizing either the maximum sample size or the expected sample size under the null hypothesis of no treatment difference, subject to pre-specified restrictions on Type I and Type II error probabilities. These proposals represent randomized versions of the single-arm "minimax" and "optimal" designs due to Simon [1].

2 A Bayesian two-stage design

To avoid the use of a historical control, we propose a randomized version of the STD. Let p_Y be the efficacy probability of the standard therapy. The criterion we suggest to select n_1 and n is based on the control of the posterior probability that $p_X > p_Y$, under the assumption that the observed response rate for the standard treatment is equal to the target p^*, while the one for the experimental treatment is equal to the target plus a small quantity $\varepsilon > 0$.

More formally, let us denote by $Pr(p_X > p_Y | X_1 = x_1, Y_1 = y_1)$ and $Pr(p_X > p_Y | X = x, Y = y)$ the posterior probabilities that $p_X > p_Y$ at the end of the first and the second stage, respectively. We select the smallest sample size n_1, such that

$$Pr(p_X > p_Y | X_1 = n_1(p^* + \varepsilon), Y_1 = n_1 p^*) \geq \lambda_1, \tag{1}$$

where $\lambda_1 \in (0, 1)$ is a pre-specified threshold. Since the data arise from a binomial distribution, we introduce independent beta prior densities for the parameters, i.e. $\pi(p_j) = \text{Beta}(\alpha_j, \beta_j)$, for $j = X, Y$, where

$$\alpha_j = n_j^0 p_j^0 + 1 \qquad \text{and} \qquad \beta_j = n_j^0 (1 - p_j^0) + 1.$$

With this choice of the hyperparameters, the beta prior $\pi(p_j)$, for $j = X, Y$, has mode at p_j^0 and is based on an implicit *prior sample size*, n_j^0, such that the larger

its value, the more concentrated is the prior distribution (see Sambucini [6]). As it is well known, the corresponding independent posterior densities for p_X and p_Y are still beta with updated parameters. Then, the posterior probability in (1) can be easily computed using, for instance, Monte Carlo simulation techniques.

Analogously, at the second stage we choose the smallest n that satisfies

$$Pr(p_X > p_Y | X = n(p^* + \varepsilon), Y = np^*) \geq \lambda_2, \tag{2}$$

for a suitable $\lambda_2 \in (0, 1)$. Once the optimal sample sizes have been determined and the trial started, following Tan and Machin [2], at the end of each stage we compute the posterior probability of interest corresponding to the observed outcome and check whether it exceeds the pre-specified threshold (λ_1 or λ_2), in order to make a go/no-go decision.

Finally, it is important to point out that statistical considerations about the irrelevance of stopping rules in Bayesian inference let us conclude that the posterior probability in (2) is not affected by the first stage results.

Moreover the behaviour of $Pr(p_X > p_Y | X_1 = n_1(p^* + \varepsilon), Y_1 = n_1 p^*)$ as a function of n_1 is the same as that of $Pr(p_X > p_Y | X = n(p^* + \varepsilon), Y = np^*)$ as a function of n and we need to set $\lambda_2 > \lambda_1$ in order to obtain $n > n_1$. Then, since the posterior distributions involved in both the criteria (1) and (2) are actually the same, in the following we will use the first stage notation in describing the numerical results related to both the stage.

3 Numerical results

The proposed design has been implemented and applied to some reasonable prior scenarios. Table 1 provides the optimal sample sizes for different values of p^* and λ_1, when $\varepsilon = 0.05$ and we consider informative prior distributions that express skepticism or enthusiasm about the efficacy of the experimental treatment. In particular we obtain a skeptical prior by specifying the prior modes $p_0^X = p^* - 0.05$ and $p_0^Y = p^* + 0.05$, while an enthusiastic prior is obtained by setting $p_0^X = p^* + 0.05$ and $p_0^Y = p^* - 0.05$. Different values of the prior sample sizes are also considered, in order to take into account different levels of skepticism or enthusiasm expressed by the prior densities. As expected, larger values of λ_1 determine higher values for the optimal sample size. We can also note that, when we adopt skeptical prior densities about the effectiveness of the new treatment, we need larger sample sizes with respect to those obtained when we use enthusiastic priors. Of course, the differences are more relevant as we increase the values of n_X^0 and n_Y^0.

Figure 1 represents the behaviour of the optimal sample size as a function of ε, when $p^* = 0.3$ and $\lambda_1 = 0.75$, under the skeptical and the enthusiastic scenarios considered in Table 1 for $n_X^0 = n_Y^0 = 5$ (see left panel) and $n_X^0 = n_Y^0 = 15$ (see right panel). The case of non-informative priors is also considered by setting $n_X^0 = n_Y^0 = 0$, so that $\pi(p_j) = \text{Beta}(1, 1)$, for $j = X, Y$. As ε increases, the fixed

virtual results used in the criteria (1) and (2) express a larger level of optimism about the efficacy of the experimental treatment and the design requires smaller sample sizes.

Moreover, since the larger the *prior sample size* the higher the weight assigned to the prior opinions, the difference in the optimal sample sizes under the skeptical and the enthusiastic scenarios are more evident in the right panel of Figure 1.

Table 1: Optimal sample sizes for different values of the prior sample sizes, p^* and λ_1, when $\varepsilon = 0.05$ and we elicit skeptical and enthusiastic prior distributions.

		$n_X^0 = n_Y^0 = 1$			$n_X^0 = n_Y^0 = 5$			$n_X^0 = n_Y^0 = 10$		
		λ_1			λ_1			λ_1		
p^*	*prior*	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
0.2	*skeptical*	15	45	106	28	61	124	42	79	144
	enthusiastic	8	38	98	1	25	86	1	6	69
0.3	*skeptical*	17	55	131	31	72	149	46	90	170
	enthusiastic	10	47	123	1	35	111	1	17	95
0.4	*skeptical*	18	60	145	32	77	163	47	96	185
	enthusiastic	11	52	137	1	40	125	1	23	109

Prior modes of *skeptical* prior distributions: $p_0^X = p^* - 0.05$ and $p_0^Y = p^* + 0.05$
Prior modes of *enthusiastic* prior distributions: $p_0^X = p^* + 0.05$ and $p_0^Y = p^* - 0.05$

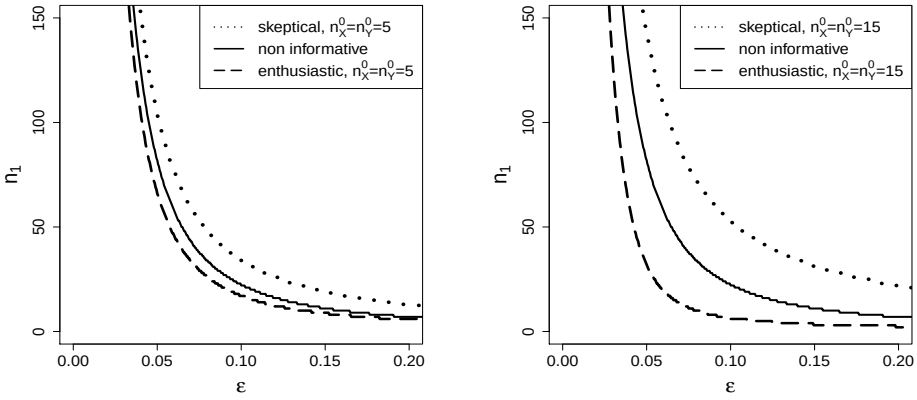

Figure 1: Optimal sample size as a function of ε, when $\lambda_1 = 0.75$ and $p^* = 0.3$, using skeptical, enthusiastic and non-informative prior distributions.

References

[1] R. Simon. **Optimal two-stage designs for phase II clinical trials**. *Controlled Clin Trials*; 1989; 10; pp. 1-10.

[2] S.B. Tan, D. Machin. **Bayesian two-stage designs for phase II clinical trials**. *Stat Med*; 2002; 21; pp. 1991-2012.

[3] M.J. Ratain, D.J. Sargent. **Optimising the design of phase II oncology trials: the importance of randomisation**. *Eur J Cancer*; 2009; 45(2); pp. 275-280.

[4] L. Rubinstein, M. LeBlanc, M.A. Smith. **More randomization in phase II trials: necessary but not sufficient**. *Natl Cancer Inst*; 2011; 103(14); pp. 1075-1077.

[5] S.H. Jung. **Randomized phase II trials with a prospective control**. *Stat Med*; 2008; 27(4); pp. 568-583.

[6] V. Sambucini. **A Bayesian predictive two-stage design for phase II clinical trials**. *Stat Med*; 2008; 27; pp. 11991224