Influence of Prior on Bayesian Sample Size Calculation

Gongfu Zhou[1], Kevin Hou[1], Yufan Zhao[1]
[1]Incyte Corporation, 1801 Augustine Cut-Off, Wilmington, DE 19803

Abstract
Bayesian designs have been increasingly used in clinical development, especially in phases I and II trials. In this article, we focus on sample size calculation for single arm trials with response rate as the primary endpoint. We implement the predictive expectation criterion to account for all possible trial outcomes, therefore avoid the need to hypothesize the trial outcome. We evaluate the property of our approach under similar setting as Mayo and Gajewski (2004). Our result is comparable with that of Mayo and Gajewski (2004), and seems slightly more robust.

Key Words: Prior, Sample Size, Robust

1. Introduction

Sample size determination plays a critical role in clinical trial planning. Under the frequentist framework, sample size calculations are aimed at adequately maintaining statistical power at the pre-specified Type I error rate. The power depends on the values of the parameter in the hypothesis testing, which are rarely known in practice with high degree of precision; see M'Lan CE, Joseph L, Wolfson DB [1] for detail.

Since fixed parameters under frequentist frame work are considered as random in Bayesian approaches, this uncertainty can be easily dealt with. A large body of literature discusses various Bayesian sample size methodologies. Among them, the paper by Tan and Machin [2] proposes a Bayesian two-stage design for trials with binary outcome, the underlying response rate π is the parameter of interest. In order to calculate the sample size, it hypothesizes the observed response rate to be slightly higher than the target response rate, and then calculates the posterior probability for π to exceed the target response rate. The smallest n such that this posterior probability becomes higher than a threshold probability will be the calculated sample size. This approach is commonly called Single Threshold Design (STD). Using the same idea, the paper by Mayo and Gajewski [3] (we denote this as M&G 2004 throughout the rest of paper) extends the setting of Tan and Machine by introducing a class of informative prior for calculating the posterior distribution rather than restricting to non-informative priors. Using a prior for the binomial parameter can fully account for the uncertainty in the estimated parameter for planning the trial, thus offering appealing advantages over the frequentist approach. M&G 2004 demonstrates some desirable properties for the calculated sample size under a variety of priors. However the idea of hypothesizing a trial outcome before the experiment is challenged by researchers and a natural way of addressing this issue is to incorporate a predictive distribution to account for all possible trial outcomes. Following the idea of Gelfand and Wang [4], De Santis [5] formalizes the framework of using both an analysis prior and a design prior to calculate the sample size. Brutti, De Santis and

Gubbiotti [6] proposes a robust Bayesian sample size determination approach by using this framework and suggests two criteria for determining sample size: one criterion is called predictive expectation criterion (PEC) and the other is called predictive probability criterion (PPC). Sambucini [7] develops a predictive single threshold design (PSTD) for adaptive two- stage trials where the sample size determination for the second stage depends on the outcomes from the first stage. This method uses an analysis prior for calculating posterior probabilities and a design prior for obtaining prior predictive distributions.

Based on the framework proposed in De Santis [5], this paper utilizes both an analysis and a design prior to calculate the sample size for trials with binary outcome by using the PEC criterion to account for all possible trial outcomes, and investigates its performance under the same setting as M&G 2004. We show that our approach has similar properties as that of M&G 2004 and seems slightly more robust with respect to the choice of priors. The remainder of this paper is organized as follows: section 2.1 presents the Bayesian framework for calculating sample size for one-arm clinical trials, section 2.2 presents the result of M&G 2004 and section 2.3 provides our result and compares it with the result in M&G 2004. Section 3 concludes with a brief discussion.

2. Bayesian Sample Size Determination

Let the endpoint of a clinical trial be a binary outcome and we assume

$$X \sim Bin(n, \pi)$$

Where X is the number of response from a sample of n subjects and π is the true response rate. Let R_u be a target response rate and trial success is defined as response rate π exceedingR_u. With an adequately sized trial, we expect a minimum threshold probability for claiming trial success, which is denoted as γ. Based on R_u and γ, we will need to find the sample size for planning the trial, and Bayesian statistics offers a natural framework to address this problem.

2.1 Bayesian prior and sample size calculation

There are various ways to formally define the probability of success. In this section, we discuss how M&G 2004 defines this probability and the setting to perform sample size calculation; we then present how our approach defines the probability of success and also discuss its relationship to that defined by M&G 2004.

Bayesian statistics models the uncertainty for π by assigning a prior distribution for π. A commonly used prior distribution for π is the conjugate beta distribution due to its flexibility and computational tractability. Let $\pi \sim Beta(\alpha, \beta)$, where α and β are hyper-parameters specified before trial is conducted. After the trial is concluded, the distribution for π is updated to be the posterior distribution, which is influenced by both the prior distribution and the data.

2.1.1 Approach by M&G 2004

G&M 2004 defines the probability of success as

$$p_n = Prob[(\pi \mid X=x, \alpha, \beta) > R_u]$$

$(\pi \mid X=x, \alpha, \beta)$ is the posterior distribution of π after assuming $x=(R_u+\varepsilon)n$, and $\varepsilon>0$ is a constant, therefore making the hypothesized trial outcome favorable for claiming trial success. This definition for probability of success is called single threshold design (STD). The smallest n such that p_n exceeds γ is the calculated sample size. M&G 2004 considers sample size calculation for a variety of priors for π, which covers both non-informative

and informative priors. We will describe their setup and report the sample size calculation result in section 2.2.

Hypothesizing a trial outcome before conducting the trial will encounter difficulty in extending the approach to multi-arm trials. This paper will present alternative framework to define the probability of success, this framework avoid the need to hypothesize the trial outcome and we subsequently investigate its property by comparing the sample size calculation result with that of M&G 2004.

2.1.2 Approach by using predictive probability

This paper utilizes the idea of De Santis [5] to incorporate an extra design prior π^d to obtain predictive distribution for the trial outcome, and treat the prior Beta(α, β) used in M&G 2004 as the analysis prior π^a, which is used for calculating the posterior distribution.

We define the probability of success as

$$e_n = E_{m^d}\{\text{Prob}[(\pi^a \mid X = x, \alpha, \beta) > R_u]\}$$

Where X will follow m^d where $m^d = E_{\pi^d}[\text{Bin}(n, \pi)]$ is the predictive density for the trial outcome under the design prior π^d. This definition for probability of success is called predictive expectation criterion (PEC) in Brutti, De Santis and Gubbiotti [6]. By taking expectation over every possible realization of trial outcome, we avoid the need to hypothesize a particular trial outcome.

The smallest n such that e_n exceeds γ is the desired sample size and π^d is usually chosen to center around a value solicited from clinician with various degree of diffusiveness. To offer a fair comparison with M&G, we set π^d to have a point mass at R_u +ε, therefore $m^d \sim \text{Bin}(n, (R_u + \varepsilon))$ in this article. The setting for analysis prior π^a remains the same as π in M&G 2004.

2.2. Bayesian sample size from M&G 2004

M&G 2004 calculates sample size for a range of target response rate, they let R_u running from 0.20 to 0.80 with a 0.05 increment, and for each R_u, three settings are investigated:

- very optimistic: $\pi_{\text{prior}} = R_u + 0.05$
- optimistic: $\pi_{\text{prior}} = R_u$

and

- pessimistic: $\pi_{\text{prior}} = R_u - 0.20$

where π_{prior} is the center for the prior distribution. They provide four ways to interpret the center and therefore provides a variety of prior distributions.

Case 1: mode prior, Beta(α, β) has $\alpha + \beta$ =3 and has mode equals to π_{prior}.
Case 2: mode informative prior, Beta(α, β) has $\alpha + \beta$ =13 and has mode equals to π_{prior}.
Case 3: median informative prior, Beta(α, β) has median equals to π_{prior} and 90% percentile range W_{90} equals to 0.30.
Case 4: mean informative prior, Beta(α, β) has mean equals to π_{prior} and 90% percentile range W_{90} equals to 0.30.

In practice, most studies would not use a very optimistic prior for calculating posterior distribution. Therefore, we only list the result for optimistic priors and pessimistic priors.

The parameter ε is set to be 0.05 and γ is set to be 0.80, and the following table is obtained according to M&G 2004.

Table 1: Sample size calculation for $W_{90} = 0.3$ and $\gamma = 0.80$ by using STD

RU	optimistic: $\pi_{prior} = R_u$				pessimistic:$\pi_{prior} = R_u - 0.20$			
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
0.25	44	54	72	77	53	124	103	81
0.3	53	62	80	84	61	133	135	126
0.35	60	69	86	89	68	140	166	163
0.4	66	75	90	92	74	146	192	193
0.45	70	79	92	93	78	150	213	215
0.5	73	82	93	93	81	152	229	231
0.55	75	83	92	91	82	153	240	242
0.6	75	83	88	86	82	152	245	247
0.65	73	81	83	80	81	150	245	246
0.7	70	78	76	72	77	146	240	240
0.75	66	73	68	62	73	140	229	228
0.8	60	67	57	50	67	132	212	210

2.3. Bayesian sample size using PEC

We use the same setup as that of M&G 2004: R_u running from 0.20 to 0.80 with a 0.05 increment, and for each R_u, we consider optimistic and pessimistic setting with four cases for interpreting the center π_{prior}.

Since we define e_n as the probability of success, which incorporates all possible trial outcome, the threshold probability γ we use does not have the same meaning and value as that of M&G 2004, therefore we set γ to be from 0.70 to 0.80 with a 0.01 increment and compare with the result in section 2.2, and we found that γ=0.73 gives sample size that are most comparable with that of M&G 2004. By setting γ=0.73 for our method, the resulting sample size has almost the same mean and median as that of M&G 2004, therefore we presented results and discuss its properties based on γ=0.73.

Table 2: Sample size calculation for $W_{90} = 0.3$ and $\gamma = 0.73$ by using PEC

RU	optimistic: $\pi_{prior} = R_u$				pessimistic:$\pi_{prior} = R_u - 0.20$			
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
0.25	51	55	71	77	59	129	109	88
0.3	59	64	78	83	67	137	140	131
0.35	65	70	83	86	73	144	169	167
0.4	71	75	86	88	78	149	195	196
0.45	74	79	88	89	82	152	215	217
0.5	76	81	87	87	84	153	230	232
0.55	77	81	85	84	84	153	239	241
0.6	75	80	81	79	83	151	243	245
0.65	73	77	76	72	80	148	242	243
0.7	69	73	68	64	76	143	235	235
0.75	63	67	59	53	70	135	223	222
0.8	56	60	48	41	63	126	205	203

As indicated in table 2, our sample size has these properties similar to that of table 1:

- For all cases, the sample size gets larger as target value increases, the trend reverses as target value reaches certain point and the sample size gets smaller as target value increases, the point at which the trend changes is different for each cases, but the change point for table 1 and table 2 are about the same for each case.
- For each case, pessimistic setting always incurs a larger sample size that of optimistic setting, the case with mode non-informative prior involves the least amount of increase. For optimistic setting, mode non-informative prior incurs a smaller sample size than that of mode informative prior across all target rate; while for pessimistic setting, mode non-informative prior incurs a much smaller sample size than that of mode informative prior across all target rate.
- For optimistic setting, median informative prior has smaller sample size than mean informative prior for small target response rate, the relationship reverses when target response rate is approaching 0.50, and median informative prior has bigger sample size than mean informative prior for large target response rate; For pessimistic setting, median informative prior has bigger sample size than mean informative prior for both small and big target response rate, while median informative prior has smaller sample size than mean informative prior for target response rate between 0.40 and 0.65.

Plot offers more insight on the relative performance between these two approaches, and we plot the sample size as a function of target response rate. Plots for optimistic setting (figure 1.a) and pessimistic setting (figure 1.b) are presented separately. Red curve is based on result from STD, and black curve is based on result from PEC.

Figure 1.a: comparison between sample sizes based on $STD\ and\ PEC$.

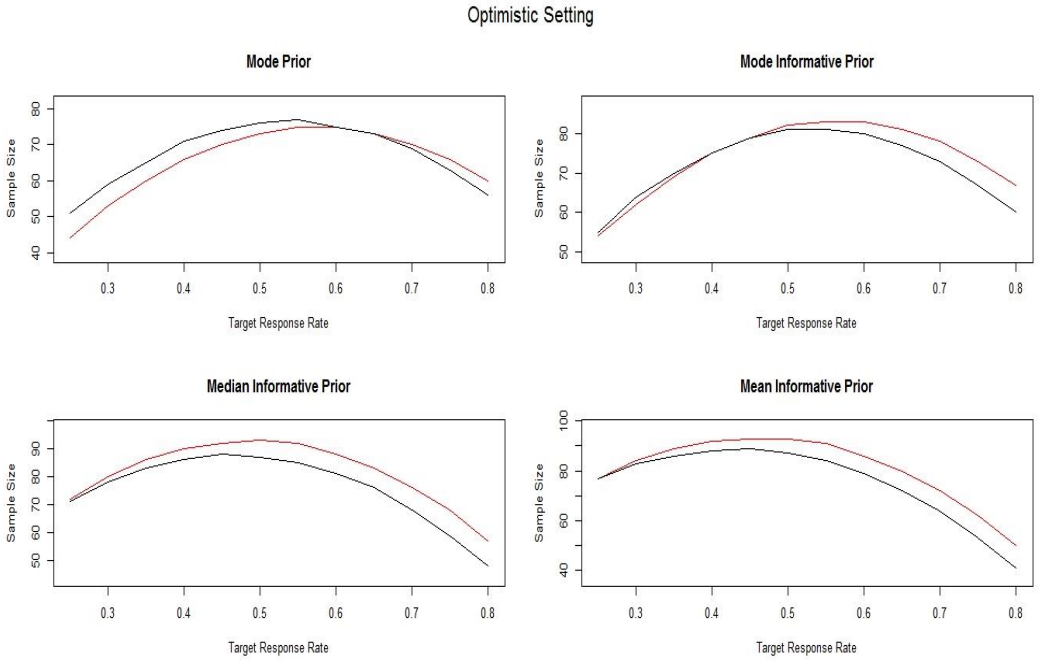
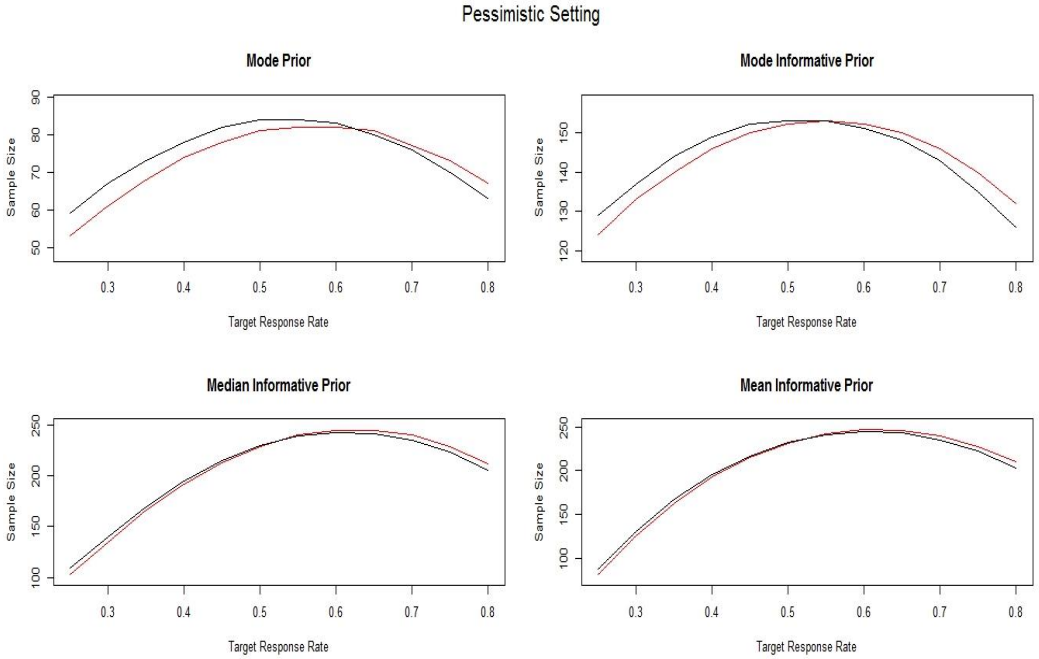


Figure 1.b: comparison between sample sizes based on $STD\ and\ PEC$.



Both figure 1.a and figure 1.b suggest that the sample sizes calculated based on STD and PEC have very similar trends with respect to target response rate. Another plot will demonstrate the difference between these two approaches.

Figure 2: comparison between sample sizes based on $STD\ and\ PEC$.

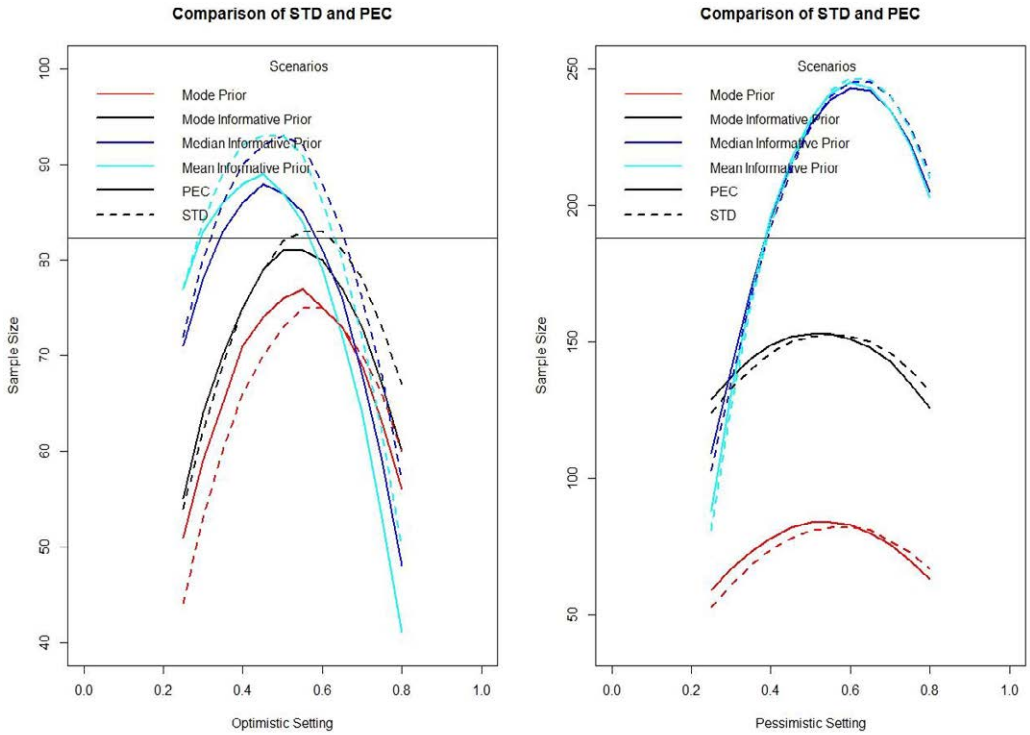


Figure 2 overlays the plots of sample size for all four cases and compares how different priors affect the sample size. For pessimistic priors, the solid lines and dotted lines are very close throughout different scenarios. For optimistic priors, the solid lines always have a tighter or similar range than dotted lines, especially for target response rate less than 0.6. Therefore figure 2 shows that sample size calculation based on PEC is slightly more robust to the choice of priors than sample size based on STD in M&G 2004.

3. Discussion

In this paper, we discuss a general framework to study Bayesian sample size estimation based on using both a design prior and an analysis prior. We use PEC criterion to account for all possible trial outcomes to avoid the need to hypothesize a trial outcome. We focus on studying the influence of analysis priors on sample size for trials with binary outcome. We compare our result with that of M&G 2004 and demonstrate that our approach has similar properties and is slightly more robust. A degenerate design prior is used in our study, however sample size calculation using non-degenerate design priors can be similarly performed and will be studied in the future. One main advantage of our approach is the avoidance to hypothesize a trial outcome, therefore offers a natural framework to study sample size problem for two-arm trials with binary outcomes. With the good properties we demonstrated by using this approach, we hope this paper will stimulate further interest in this field.

References

1. M'Lan CE, Joseph L, Wolfson DB. Bayesian sample size determination for binomial proportions. Bayesian Analysis 2008; 3: 269–296
2. Tan SB, Machin D. Bayesian two-stage designs for phase II clinical trials. Statistics in Medicine 2002; 21: 1991–2012.
3. Mayo M, Gajewski B. Bayesian sample size calculations in phase II clinical trials using informative conjugate priors. Controlled Clinical Trials 2004; 25: 157–167.
4. Gelfand AE, Wang F. A simulation based approach to Bayeisan sample size determination for performance under a given model and for separating models. Statistical Science 2002; 17:193–208.
5. De Santis F. Sample size determination for robust Bayesian analysis. Journal of the American Statistical Association 2006; 101:278–291.
6. Brutti P, De Santis F, Gubbiotti S. Robust Bayesian sample size determination in clinical trials. Statistics in Medicine 2008; 27: 2290–2306.
7. Sambucini V. A Bayesian predictive strategy for an adaptive two-stage desin in phase II clinical trials. Statistics in Medicine 2010; 29: 1430–1442.