Use of Bayesian Adaptive Design in Medical Device Trials

Weihua Cao *† Nelson Lu* Rajesh Nair*

Abstract

Bayesian adaptive design has seen its increase use in medical device trials, especially with the recent release of the FDA guidance document. In general, Bayesian adaptive trials are challenging to design, conduct and analyze. In this article we will address some practical issues with Bayesian adaptive device trials, involving prior selection, predictive probability calculation, and control of type I error rate among others. Some examples of Bayesian adaptive device trials will be discussed.

Key Words: Medical device, Bayesian, adaptive design

1. Introduction

With the release of FDA guidance for the use of Bayesian statistics in medical device clinical trials in 2010, we have seen an increasing number of Bayesian clinical trials, and majority of them employ Bayesian adaptive design. Bayesian adaptive design, if properly designed and conducted, may reduce the size and/or length of a trial and enables faster study decisions.

A Bayesian adaptive design can be quite flexible. It could be a sample size adaptation. That is, if the probability of trial success given current data is sufficiently high, then stop the trial enrollment and wait all enrolled patients to complete follow-up, then do the final analysis. The probability of trial success could be either posterior probability of trial success given current complete data, or predictive probability of trial success once all enrolled subjects complete their follow-ups, which will be discussed in details in Section 3. However, for a trial with fast accrual rate or short enrollment period, sample size adaption may not be needed, as by the time of interim looks, majority of patients may have been enrolled and doing sample size adaptation may not save rescources/time.

Another common Bayesian adaptive design is stopping early for success. If the probability of trial success at an interim look is sufficiently high (exceeding pres-pecified boundaries at this interim look), then stop the trial and make success claim. As with sample size adaptation, this probability of trial success could be either posterior probability or predictive probability.

An adaptation could also be stopping early for futility. If the probability of trial success given current data is lower than pre-specified futility boundaries, trial may be stopped early for futility. A cautionary note is that use of futility boundary yields larger effectiveness thresholds at interim looks. Therefore, deviation of the futility stopping rules may result an inflation of the Type I error rate. Therefore we recommend use of non-binding futility boundaries, where futility boundaries do not have to be followed.

A trial could be adapted as stoping early for harm, e.g., high stroke rate for a cardiovascular device may trigger an early stopping for harm. Another adaption is adaptive randomization, that is, modify the randomization rate during a trial to increase the probability that a patient be allocated to the best treatment. It is attractive when enrollment

*Division of Biostatistics, Center for Devices and Radiological Health, U.S. Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD 20993

†No official support or endorsement by the Food and Drug Administration of this presentation is intended or should be inferred

for one arm is problematic. One practical problem with adaptive randomization is that it requires rapid assessment of patient outcomes. Another concern is that the characteristics of patients enrolled in the trial may change systematically over time, and this may cause an adaptive randomization procedure to function poorly (Thall and Wathen, 2006).

However, Bayesian adaptive trials are challenging to design, conduct and analyze. This article is written with the aim of providing some practical perspective with regard to Bayesian adaptive clinical trials in regulatory setting. This article is organized as follows. In Section 2, we briefly discuss prior distribution for a Bayesian trial. In Section 3, we will demonstrate the calculation of predictive probability through a hypothetical example. In Section 4, we then conduct Monte Carlo simulations with regard to Type I error rate control in checking operating characteristics of a trial. In Section 4, we summarize with some discussion points.

2. Prior Distribution

The first important thing for a Bayesian clinical trial is determination of prior distribution, which represents a priori belief on quantity of interest. A non-informative prior can represent lack of information. For example, we may use uniform(0,1) for the probability of a binary endpoint.

Prior distribution may be based on information from previous comparable studies, for example, out of US clinical trials, company's own pilot studies, data registries, literature results, etc. For example, suppose the severe adverse event rates from two previous similar clinical trials are 21% and 22%. One may consider an informative prior Beta(25, 75) for the current study, representing a prior belief of 25% severe adverse event rate. For informative prior, one has to show comparability between the prior studies and the current study: whether they have similar protocols, whether the endpoint is defined the same and measured at the same time frame, etc. Informative priors borrows information from historical studies, however, we do not want the informative prior to borrow too much information, which may lead to an inflation of type I error rate. We can use prior probability of study claim to evaluate the appropriateness of the prior distribution, which is the probability of claiming trail success without seeing any data. If the prior probability of study claim is too high, we may conclude the informative prior is too informative!

3. Predictive Distribution

At interim looks, there are two methods to analyze data and make decisions: posterior probability and predictive probability. One has to pre-specify which method to use prior to initiation of a clinical trial. Posterior probability only utilizes subjects with complete follow-up at interim looks. Predictive distribution is the posterior distribution of unobserved outcome, given what we have observed so far.

We now use an hypothetical example to to illustrate the calculation of predictive distribution. Suppose we have a single arm study with a performance goal of 70%. The study will be a success if the lower bound of the credible interval is above 70%. We further assume that the primary endpoint is a Binary endpoint: whether a patient is disease free at 6 months. Patient's disease statuses are also measured at 3 months.

At interim looks, patient may be categorized into three(3) groups: patients with 6 months data, patients with only 3 months data, and patients yet to be followed at 3 months. With some distributional assumptions, we may use observed data to predict unmeasured 6 month data for patients in the last two groups. Beta priors used to model the transition probabilities are given in Table 1. The transition probabilities used to predict the 6 months

outcome are chosen according to the following assumptions: 80% of subjects are expected to be successes at 6 months; 20% of subjects with disease at 3 months will be successes at 6 months; and 90% of subjects who are disease free at 3 months will be successes at 6 months.

Suppose at an interim look, we have enrolled 100 subjects. Of those, there are 33 subjects with 6 months data, and 23 of them are successes. There are 40 subjects with 3 months data available and 29 of them are disease free, 11 subjects with disease. The remaining 27 subjects are yet to be followed to 3 months.

For the group of 27 subjects with no 3 months data, the updated 6 months success rate would be $Beta(4+23, 1+10)$, thus the number of disease-free subjects at 6 months x_1 out of these 27 subjects follows a beta-binomial distribution: $x_1 \sim Beta - Binomial(27, 4+23, 1+10)$.

Further suppose among 33 subjects with 6 months data, 20 of 21 subjects who are disease free at 3 months were successes at 6 months. Thus the Number of disease-free subjects at 6 months x_2 for the 29 subjects who are free of disease at 3 months follows a Beta-Binomial $x_2 \sim Beta - Binomial(29, 4.5+20, 0.5+1)$. Similarly the number of of disease-free subjects at 6 months x_3 for the 11 subjects with disease at 3 months follows a Beta-Binomial distribution: $x_3 \sim Beta - Binomial(11, 1+3, 4+9)$.

So the total number of disease-free subjects at 6 months are predicted by the random quantity $23 + x_1 + x_2 + x_3$. Given this predictive distribution, we can make a random draw from this distribution, and conditional this random draw, we then calculate the posterior probability, and compare it to the pre-defined success criterion to see whether the study is successful. We then repeat this process many times, say 10,000 times, and the proportion of study success out of these 10,000 times is the predictive probability of study success at this interim look.

4. Type I error rate control

Because of the inherent flexibility of Bayesian adaptive design, extensive simulations are needed to check the operating characteristics. To illustrate out point, we conduct some simulations to check the type I error rate for a hypothetical Bayesian sample size adaptation study.

The simulation is a single arm study with maximum sample size 110. Sample size adaptation takes place when 50, 70, 90 and 100 subjects are enrolled. At an interim look, if the predictive probability cross the stopping boundaries, we will stop enrollment, and complete follow-up on all enrolled subjects and conduct final analysis. Otherwise, we will continue accrual until maximum sample size has been achieved. The stopping boundaries for predictive probabilities of trial success at interim looks are chosen as 0.90, 0.85, 0.80 and 0.80 for the four interim looks sequentially.

The performance goal for this study is assumed to be 70%, and the study will be a success if the lower bound of the credible interval is above 70%. The primary endpoint is a binary endpoint: : whether a patient is disease free at 6 months. Patient's disease statuses are also measured at 3 months. The study success criterion to which the posterior probability is compared was initially chosen as 0.98 with the aim to control the Type I error rate at 2.5%. The accrual rate is assumed to be 5 patients per month. We will use simulation studies to see whether the type I error rate has been controlled or not for this particular design.

The data were generated as follows. We first generate 3 months data with a disease-free rate of 80%, and then generate 6 months data conditional on these 3 months data:

$$pr(\text{disease-free at 6 months}|\text{disease-free at 3 months}) = 80\%,$$

$$pr(\text{disease-free at 6 months}|\text{disease at 3 months}) = 30\%.$$

According to this setting, the 6 months success rate is 70%. Same transition probabilities as in Table 1 are used.

The results based on 5000 simulations are provided in Table 2, where the second column is the proportion of trials which are stopped at the current sample size, and the fourth column shows how the type I error rate is allocated across different looks. It seems that the type I error rate is 2.48%. Now we conduct some sensitivity analyses to see whether the type I error rate is still under control or not.

The original beta priors amount to borrowing 5 observations from historical studies. As a first sensitivity analysis, we vary the beta priors for the transition probabilities and triple the borrowing evidence. The new Beta priors are given in Table 3, and the corresponding simulation results are given in Table 4

A second sensitivity analysis is that we modify the transition probabilities as in Table 5 , and the corresponding simulations results are given by Table 6.

Another interesting scenario is to change the 3 month success rate, but keep the 6 month rate at 70%.

$$pr(\text{disease-free at 3 months}) = 60\%,$$

$$pr(\text{disease-free at 6 months}|\text{disease-free at 3 months}) = 90\%,$$

$$pr(\text{disease-free at 6 months}|\text{disease at 3 months}) = 40\%.$$

All other parameters are set to the same, and the simulation results are provided in Table 7 We can see from this table is inflated from 2.48% to 2.82%, thus not robust to the data generating process. From these sensitivity analyses, we see the type I error rate is not controlled at 2.5%. Thus we need to go back to the study design to vary some of the parameter settings to get the Type I error rate under control.

There are other interesting scenarios which may need to be checked, for example, we can vary the accrual rate from 5 pts/month to 10 pts/month; we can also change the constant accrual rate to varying accrual rate, etc. We can also change the effectiveness boundaries at interim looks.

In addition, We would also like to see how the type I error rate is allocated across looks. It is not ideal if the type I error are spent too much at the first couple looks for several reasons. The first is that there are only limited data at first few interim looks, and we feel comfortable only when the results are extreme, just like what the OBrien-Flemming boundaries do group sequential design. Also for the sponsors interest, it is better to be conservative upfront, otherwise later data may cross back the boundary. For example, in a sample size adaptation study, suppose a trial is stopped at an interim look, but after all enrolled patients complete their follow-up, it is found out that the primary analysis failed. This dilemma could have been avoided that if we were more conservative at the beginning and enroll more patients before we stop the enrollment. The third reason is from clinical perspective that whether there are enough safety information captured at first few looks.

In summary, Type I error rate depends heavily on various parameter settings, and extensive simulation results and sensitivity analyses are desired for clinical reasonable scenarios. In addition, in this simulation, we only present scenarios for controlling Type I error rate in a regulatory setting. Among other scenarios, power, sample size distribution should be checked as well as part of the operating characteristics. Again, because of inherent flexibility associated with a Bayesian adaptive design, a thorough evaluation of operating characteristics is needed. Therefore Upfront simulation burden could be quite high for Bayesian adaptive studies.

5. Discussion

Another interesting issue related to Bayesian adaptive design is the Number of interim looks. How many interim looks are appropriate? Is 10 or 20 interim looks too many? Clinical justification for number of interim looks is needed. Another potential problem with too many interim looks is that it may yield higher bar at interim looks, thus the study is more difficult to stop.

And like any trial involving interim look, access to unmasked interim results could introduce operational bias in future conduct of the trial. So we recommend only DMC has access to interim results, and have external independent statisticians conduct interim analyses. For a single-arm adaptive trial, as it is difficult, if not possible, to mask study results, there might be more severe operational bias

If the study involves primary safety and primary effectiveness endpoint, the Type I error rate should be controlled independently at a desired level for both endpoints. That is, simulations for operating characteristics should be conducted separately for these two endpoints. Another issue is related to analyses of secondary endpoints and subgroup analysis. If Bayesian analysis is conducted for primary analyses, should we use Bayesian or Frequentist approach for secondary analyses?

In summary, a Bayesian adaptive clinical trial should be planned in advance and the operating characteristics of the design be assessed in a variety of scenarios and come to talk with FDA early to reach an agreement on study design.

REFERENCES

Thall, Peter F. and Wathen, Jay K. (2006), "Practical Bayesian Adaptive Randomization in Clinical Trials" , UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series, Working Paper 31. http://biostats.bepress.com/mdandersonbiostat/paper31

FDA guidance (2010), "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials," http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071072.htm.

Irony , T. (2009), "Bayesian Statistics for Designing and Analyzing Medical Device Clinical Trials in the Regulatory Setting", 2nd Annual FDA/MTLI Medical Device and IVD Statistical Issues Workshop.

Table 1: Prior distributions for transition probabilities

Group	α	β
No 3-month follow-up	4	1
Disease at 3months	1	4
Disease free at 3 months	4.5	0.5

Table 2: Operating characteristics based on 5000 simulations

Sample size	Proportion of Trials	Stop & Lose	Stop & Win
50	0.0154	0.0114	0.004
70	0.0106	0.006	0.0046
90	0.0102	0.0058	0.0044
100	0.0026	0.001	0.0016
110	0.9612	0.951	0.0102
Total	1.00	0.9752	0.0248

Table 3: Prior distributions for transition probabilities

Group	α	β
No 3-month follow-up	12	3
Disease at 3months	3	12
Disease free at 3 months	13.5	1.5

Table 4: The first sensitivity analysis for the operating characteristics based on 5000 simulations

Sample size	Proportion of Trials	Stop & Lose	Stop & Win
50	0.0108	0.0054	0.0054
70	0.0146	0.0078	0.0068
90	0.0112	0.0056	0.0056
100	0.0024	0.0016	0.0006
110	0.9610	0.9480	0.0130
Total	1.00	0.9684	0.0316

Table 5: Prior distributions for transition probabilities

Group	α	β
No 3-month follow-up	6	4
Disease at 3months	5	5
Disease free at 3 months	7	3

Table 6: The second sensitivity analysis for the operating characteristics based on 5000 simulations

Sample size	Proportion of Trials	Stop & Lose	Stop & Win
50	0.0026	0.0018	0.0008
70	0.0084	0.0044	0.0004
90	0.0102	0.0054	0.0048
100	0.0026	0.0010	0.0016
110	0.9762	0.9620	0.0142
Total	1.00	0.9746	0.0254

Table 7: The third sensitivity analysis for the operating characteristics based on 5000 simulations

Sample size	Proportion of Trials	Stop & Lose	Stop & Win
50	0.0098	0.0007	0.0028
70	0.0138	0.0007	0.0068
90	0.0104	0.0058	0.0046
100	0.0006	0.0058	0.0012
110	0.9640	0.9512	0.0128
Total	1.00	0.9718	0.0282