We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,100
Open access books available

167,000
International authors and editors

185M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Chapter

Introduction to Bayesian Group Sequential Design

Chen Li, Ping Huang and Haitao Pan

Abstract

In classical group sequential designs, a clinical trial is considered as a success if the experimental treatment is statistically significantly better than placebo. The criteria for stopping or continuing the trial are chosen to control the false-positive rate (type I error). Bayesian group sequential design has an advantage of allowing inclusion of prior information in the analysis. The decision criteria can be based on the posterior or predictive distribution of the treatment effect to stop for success or futility, or to continue for each interim analysis and the final analysis. This chapter introduces Bayesian group sequential designs with examples in a confirmatory setting, including how to calibrate the tuning parameters to set up decision criteria for the interim and final analyses, how to derive the sample size, and how to evaluate the operating characteristics via simulations.

1. Introduction

In confirmatory trials, randomized controlled trials (RCTs) are the gold standard for treatment evaluation, which directly compare the investigational drug with the standard treatment or a placebo (if there is no standard of care). The essential component for a trial design is to find the sample size that is necessary to detect a clinically important treatment difference with sufficient power and controlled type I error rate. Once all observations have been collected, final analyses will be conducted. However, due to lack of information on both the magnitude and the sampling variability of the new treatment effect at the design stage, realized sample size may be different from what the design gives use. To that end, the fixed designs can be inefficient since they cannot accommodate this discrepancy.

There has been an increasing interest in group sequential designs that can adapt to the information collected during the process of the trial. In contrast to fixed designs, group sequential methods are flexible and adaptive to regularly examine the efficacy over administratively convenient time intervals [1]. During the process of a trial, strong evidence in favor of the benefit of the novel treatment may emerge early. If so, the extra study participants required to provide this protection against a false-negative result may not be necessary. Stopping the trial before the maximum planned sample size can save resources and accelerate the trial process. Of course, this advantage must

be balanced against the potential for the overestimation of the treatment effect and other limitations of smaller trials (e.g. limited safety data and less information about treatment effects in subgroups). Conversely, if strong evidence accumulating against the benefit of the new treatment, it would be unethical for patients continuing to be exposed to the futility therapy. Interim analysis is a useful tool to stop trials early for futility. In classical frequentist group sequential designs, the criteria for stopping or continuing the trial are chosen to control the type I error and *p*-values are used to make decisions.

Rather than making inference by using *p*-values, criteria for success and futility stopping, Go/No-go decisions in Bayesian design are based on the posterior probability (PoP) or posterior predictive probability (PreP) at the interim and final analyses [2, 3]. Based on these statistical tools, use of the cumulating data through interim analyses allows the trial design adapted to improve design efficiency. For example, ineffective treatment arms could be dropped; further treatment arms could be introduced; the trial could be stopped early (due to futility/efficacy); or randomization to treatment could be altered to favor the more effective treatment. Such adaptations are attractive to both researchers and patients, by making more efficient use of patient resource and potentially treating patients more effectively. In general, adaptive clinical trial designs are easier to implement within the Bayesian framework. Frequentist designs may not always work. While, Bayesian methods have particular advantage in rare disease scenarios where traditional methods can be difficult, if not impossible, to achieve due to limited sample size. To that end, the Bayesian approach is that they allow inclusion of external information, which can be historical, nonconcurrent information. By applying dynamic borrowing methods or matching approaches to create a synthetical control arm or augment a control arm, sample size may be saved.

In this chapter, we provide an introduction of Bayesian group sequential trials and discuss some commonly used design features with an example.

2. The decision rule in classical frequentist framework

Classical group sequential trials rely on null hypothesis testing involving calculation of test statistics, along with *p*-values and confidence intervals. The critical statistical issue with early stopping, particularly for success, is accounting for multiple "looks" and repeatedly testing the null hypothesis over time. However, the more frequently the data analyzed, the greater the chance of observing one of these fluctuations. Therefore, the decision criteria for early stopping or continuing the trial are chosen to control the overall type I error rate (e.g. 0.05) [4, 5]. For example, the null ($H_0 : \delta \leq 0$) and alternative ($H_0 : \delta > 0$) hypotheses are formulated for the true treatment difference δ (large values of δ correspond to a positive effect) in a two-arm trials, and the type I error (α) is set to a specified value. The null hypothesis will be rejected if the observed one-sided *p*-value is less than . The interim analysis can be performed on a specific calendar date within the planning period of times, which is called calendar time. It can also be performed at the information time, which is a predefined proportion of maximum subjects or events/outcomes already observed, especially in time-to-event trials. To account for multiple testing on several interim analyses, interim hypothesis testing always based on α-spending functions, such as Pocock method and O'Brien-Fleming method [6]. The stopping rules require a very small *p*-value smaller than the false-positive rate boundaries in the interims. The more

conservative the early stopping criteria, the more assurance there is that an early stop for success is not a false-positive result.

3. The decision rule in Bayesian perspectives

Unlike frequentist approaches where parameters of interest are considered deterministic, the parameters in the Bayesian paradigm are treated as random, while the data collected in the trial have been observed and thus considered fixed. The prior distributions that we assign to the unknown interest parameters (e.g. treatment effect) can be viewed as our uncertainty initial belief about them. Once data of the current trial collected, new information becomes available and is summarized by another distribution—the likelihood. Using Bayes' theorem, the prior can be combined with the likelihood and updated to become a posterior distribution. Accordingly, various posterior probabilities and inferences can be drawn. Thus, the decision can be made by the posterior probabilities which summarize all information available at that point in time. We can find the empirical frequentist error rates for a Bayesian testing procedure by fixing certain parameter boundaries at prespecified values. Bayesian approach provides an alternative statistical framework and uses probability distributions to represent uncertainty of the parameter estimation. By carefully calibrating design parameters, not only do the methods enhance flexibility of trial conduct and monitoring, but they can also maintain the frequentist type I and type II error rates at the nominal levels.

From a Bayesian perspective, the decision-making is based on the PoP for the treatment effect given the trial data. If the PoP for the interest parameter δ beyond the effect threshold s is sufficiently high, i.e. above a prespecified boundary θ_s, denoted by $Pr(\delta > s|D) \geq \theta_s$, the trial could allow for early stopping for efficacy. By the same token, early stopping for futility may be permitted if the PoP is below the futility boundary θ_f, expressed as $Pr(\delta > s|D) \leq \theta_f$. If the probability falls between these two values θ_s and θ_f, then the trial may continue recruiting. Since the Bayesian methods are not required for multiple looks corrections, the decision can be made at any time with the updating PoP. Also, the stopping rules in each interim can be constructed independently and multiple criteria can be required based on several treatment effect thresholds. For example, the success criteria of a trial can be quantified as:

$$Pr(\delta > s_1|D) \geq \theta_{s1} \text{ and } Pr(\delta > s_2|D) \geq \theta_{s2.} \tag{1}$$

where s_1 and s_2 are specified effect thresholds, and θ_{s1} and θ_{s2} are specified or tuned probability boundaries. The multiple quantitative criteria based on the PoP may greatly help to achieve a clinically meaningful decision-making. This is appealing to clinicians and statisticians who will often want to know how a given design will conclude in favor of some particular treatment effects.

Although the type I error and power are frequentist concepts, the Bayesian approach can calculate something analogous to these quantities for any prespecified decision rules. It can also consider multiple "looks" as the frequentist approach to control the type I error. The multiple corrections, such as based on α-spending functions, may be conducted as the Bayesian early stopping boundaries on the interim analysis to maintain the total false-positive rate. The PoP can also be considered as the power for a specific treatment effect when the target treatment effect is assumed to be the true value. It assists with decision-making to demonstrate that the Bayesian design has good frequentist operating characteristics.

4. Decision boundaries

The decision boundaries for the PoP can be specified based on a clinically meaningful treatment effect threshold by the investigator. For example, one can stop for efficacy if the PoP of having a hazard ratio (HR) <1 above 90%, i.e. $Pr(HR < 1|D) > 0.9$, and one can stop for futility if $Pr(HR < 1|D) < 0.2$. In practice, however, Bayesian designs usually rely on simulations to determine the decision boundaries and parameter calibration. This is achieved by determining how frequently the Bayesian design incorrectly declares a treatment to be effective or superior when it is assumed that there is truly no difference. It has often been used to tune stopping boundaries to ensure acceptable type I error, e.g. 2.5% one-sided type I error or 5% two-sided type I error. The power for a specific treatment effect can be calculated as the proportion of simulations that declare the trial to be "successful" when the target treatment effect is assumed to be the true value. This approach has been recommended by the FDA [4] and has been used in practice for Bayesian adaptive designs [7, 8]. These simulations should be performed in the planning stage of a Bayesian group sequential trail. In the analysis stage, no further adjustments are required to account for the previous interim analyses that have been performed.

We use a simulations study to introduce how to obtain the decision boundaries. Consider a two-arm RCT with two interim analyses, and a final analysis is planned with time-to-event outcomes, such as progression-free survival (PFS) times. Let T denote the underlying failure time, which may be right-censored, and let C denote the censoring time. We denote the observed time as $X = \min(T, C)$ with a censoring indicator $\Delta = I(T \leq C)$, i.e. if $\Delta = 1$ then $X = T$, which is the failure time, and if $\Delta = 0$, then $X = C$, which is the censoring time. We assume the survival times for both the treatment and placebo arms followed exponential distributions with means of μ_T and μ_C, respectively. The null hypothesis is equivalence of the two treatments in terms of the efficacy, and the alternative hypothesis is the treatment better than the control. Under the exponential survival model, the mean survival time is the reciprocal of the hazard, that is, the hazard ratio (HR) $= \mu_C/\mu_T$ and a lower value means better treatment. It could be claimed success if HR between two groups given the observed data D satisfies

$$Pr\left(\log\left(\frac{\mu_C}{\mu_T} \right) < \delta | D \right) > \theta_T \tag{2}$$

where δ is an effect threshold for a clinically meaningful treatment difference, and θ_T is a probability boundary for decision-making.

In the Bayesian framework, we specify a prior distribution for the mean survival time μ following an inverse-gamma (IG) prior distribution for μ:

$$\mu \sim IG(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \mu^{-\alpha-1} \exp\left(-\beta/\mu\right) \tag{3}$$

where $\alpha > 0$ and $\beta > 0$. Since the IG distribution is conjugate with the exponential likelihood function, the posterior distribution of μ also follows an IG distribution:

$$p(\mu|D) \propto \mu^{-\sum_{i=1}^{n} \Delta_i - \alpha - 1} \exp\left(-\frac{\sum_{i=1}^{n} x_i + \beta}{\mu} \right) \tag{4}$$

That is, $\mu|D \propto IG\left(\sum_{i=1}^{n}\Delta_i + \alpha, \sum_{i=1}^{n}x_i + \beta\right)$. Thus, we can compute the PoP of the treatment better than the control as shown in Eq.(1). If $PoP > \theta_T$, we claim the treatment superior to the control. We should also specify the lower and upper probability boundaries θ_L and θ_U for decision-making in interim analysis. The decision rules in the interim analysis are given as follows:

1. *Success stopping* If $Pr\left(\log\left(\frac{\mu_C}{\mu_T}\right) < \delta|D\right) > \theta_U$, we stop the trial to claim a superior treatment.

2. *Futility stopping* If $Pr\left(\log\left(\frac{\mu_C}{\mu_T}\right) < \delta|D\right) < \theta_L$, we stop the trial to claim a futility treatment.

Then, the design parameters ($\theta_T,\theta_U,\theta_L$) can be calibrated via simulation to achieve desirable trial operating characteristics.

5. Parameter calibration and setup

The parameters calibration needs two stages.

In stage 1, we focus on choosing an appropriate probability boundary θ_T for the specific treatment effect δ. We simulate the data under a "null" scenario of HR = 1 to calibrate the parameter that can control the false-positive rate. In this step, we firstly set $\theta_L = 1$ and $\theta_U = 0$ such that the trial will not be terminate early. Considering null hypothesis $H_0 : \mu_T = \mu_C$ and the alternative hypothesis $H_1 : \mu_T = \mu_C \times e^{\delta}$, an initial sample size can be obtained by the frequentist method. Let the prior distribution of μ for each arms be non-informative prior such as IG (0.01,0.01) and simulate the data under the null hypothesis. We simulate millions of trials and count the number of trials that were declared to be successful in which the decision rule was shown in Eq.1. The proportion of trials that were successful when assuming a HR = 1 provides the simulated type I error rate. Then we vary the value of θ_T (e.g. from 0.6 to 0.95) and calculate the PoP in Eq.1 $Pr(\log HR < \delta|D) > \theta_T$. If the final PoP is higher than the given type I error, it means that the probability boundary is too loose to control the false-positive rate, and we should increase θ_T and use more stringent stopping boundaries. On the contrary, if PoP is lower than the given type I error, θ_T can be decreased to loosen the efficacy stopping boundaries. Until PoP is close to the type I error, the corresponding θ_T can be chosen as the final efficacy success boundary.

In the second stage, we fixed the chosen θ_T in stage 1 and varied the value of θ_U and θ_L, such as $\theta_U = 0.90, ... 0.99$ and $\theta_L = 0.01, ... 0.10$, to calibrate early stopping boundaries in the interim analysis. Similar to the calibration procedures in stage 1, we select the appropriate combination of θ_U and θ_L to control the given type I error with simulations presented above.

6. Sample size estimation

In this step, based on the above calibrated parameters ($\theta_T,\theta_U,\theta_L$), we simulate the data under the alternative hypothesis using the initial sample size estimated by the frequentist method. The proportion of the simulation trials declared to be successful

based on the given decision criteria can be interpreted as the conditional power. If it is lower than the given power, it means the current sample sizes cannot provide adequate power, and we should increase sample size and vice versa. Until the PoP reaches to the given power, the corresponding sample sizes can be finally determined.

7. Prior distribution and ESS

Although noninformative prior has been widely applied in the design stage, the prior can also be constructed using domain knowledge based on expert clinical opinion or information from previous studies. Borrowing data on the control arm may result in a more efficiently design and more favorable operating characteristics by way of a smaller trial overall or patients on the control arm. Although the idea of using historical data is not new, proper application is critical. Challenges exist in quantifying the level of relevance of historical data. When incorporating prior information, it is significant to choose prior beliefs into trials. For example, we can use skeptical or enthusiastic priors to decide partial or fully borrowing of the external information. In practice, to utilize historical data as enthusiastic prior data for the current trials, it must be assumed that the historical data are fully relevant to this trial patient population. If unsure of the relevance of prior information, a probability of relevance can be incorporated as part of the prior distribution, such as

$$prior = (1 - \alpha) * f(D) + \alpha * g(D) \qquad (5)$$

where $f(D)$ is skeptical prior distribution (the external information or historical data are completely different from current data), $g(D)$ is the enthusiastic prior distribution (historical data reflects current data) and relevance factor, and α is the applicability probability of current data [9]. A number between 0 and 1 for the relevance factor α corresponds to the amount of information borrowed from the historical data, that is, the interpretation of applicability of the historical data. Some other common methods for discounting are weighted average of the means for the randomized and historical controls to control bias, such as the power prior approach [10], the commensurate prior approach [11], meta-analytic predictive (MAP) approaches [12], and so on. Modeling and simulation are useful tools to explore and set expectations on the relevance of the historical data. Even if prior information seems very relevant, sufficient skepticism about potential efficacy exists; therefore, requiring that prior information should be discounted.

Quantification of the amount of information induced by the prior is important to avoid domination of the prior information on posterior inference. The effective sample size (ESS) reflects the amount of borrowing by incorporating prior information, which equates prior information to a certain number of observations. Since the historical information may not be commensurate with the information collected during the current trial, there may be a prior-data conflict observed. The ESS can quantify the strength of prior information and its contributions to the inference.

Prior effective sample sizes are well understood for conjugate of one-parameter exponential families. It can be motivated in the updating rule from prior to posterior parameters. For example, for Poisson data with a $Gamma(a, b)$ prior, the second parameter of the posterior Gamma distribution is $b + n$, implying b as the prior ESS. In another way, the posterior mean is a weighted average of the prior mean and the

standard parameter estimate, with weights proportional to the prior ESS and the sample size n. For Poisson data, the prior mean and parameter estimate are a/b and $\sum Y_j/n$, and the posterior mean $(a + \sum Y_j)/(b + n)$ is the weighted average of the two, with weights proportional to b and n. The ESS under conjugacy can be concluded with different distribution as follow:

Distribution	Prior	ESS
Normal	$Normal(\mu, s^2/n_0)$	n_0
Binary	$Beta(a, b)$	$a + b$
Poisson	$Gamma(a, b)$	b

Another more involved information-based ESS has been suggested in the seminal paper by Morita [13], which has been denoted as Morita method for short. In addition to the Fisher information, it uses the information of the prior distribution $p(\theta)$:

$$i(p(\theta)) = -\frac{d^2 \log p(\theta)}{d\theta^2} \tag{6}$$

and the information of and ϵ-information (large variance) prior $p_0(\theta)$ with the same mean $(\overline{\theta})$ as $p(\theta)$:

$$i(p_0(\theta)) = -\frac{d^2 \log p_0(\theta)}{d\theta^2} \tag{7}$$

The ESS can be defined as an interger m that minimizes the distance (evaluated at the prior mean $\overline{\theta}$) between the expected posterior information for a sample of size m based on the same mean large variance prior $p_0(\theta)$ and the information of the actual prior:

$$|i(p_0(\overline{\theta})) + E_{Y_m}\{i_F(Y_m; \overline{\theta})\} - i(p_0(\theta))| \tag{8}$$

where the expectation of Y_m is taken over the prior-predictive distribution under $p(\theta)$. This approach is noteworthy because it appears to be the first formal, metric-based approach to ESS that complies with the standard one-parameter exponential family ESS.

There is also another information-based ESS, which is described as expected local-information-ratio (ELIR) method [14]. It also uses the prior and Fisher information, but instead of locally evaluating the respective information ratio at the mean (or mode), and it is defined as the mean of the prior information to Fisher information ratio $r(\theta)$:

$$\text{ESS}_{\text{ELIS}} = E_\theta\{r(\theta)\} = E_\theta\left\{\frac{i(p(\theta))}{i_F(\theta)}\right\} \tag{9}$$

ESS_{ELIS} gives the well-known effective sample sizes for some standard one parameter exponential families. For the natural parameter η, it is the standard ESS without any boundary restriction on the parameters. The information ratio $i(\eta) = i(p(\eta))/i_F(\eta)$ does not depend on the parameter. For the natural parameter, the sampling and prior distribution can be written as:

$$f(y|\theta) = \exp\{y\eta - M(\eta)\}, p|(\eta) = \exp\{n_0 m_0 \eta - n_0 M(\eta)\} \tag{10}$$

Since $i_F(\theta) = d^2 M(\eta)/d\eta^2$, it follows that $ESS_{ELIS} = n_0$. Take Poisson data for example, with a Gamma prior for the mean μ, $\eta = \log(\mu)$, $M(\eta) = exp(\eta)$ and $n_0 = b$. Therefore, the *ELIS* method seems to be simple and superior to current versions.

8. Example

In the following, we use the above example to illustrate the design of Bayesian group sequential trial. It will be used in a randomized, double-blinded, placebo-controlled study on the efficacy of new treatment to improve the survival in advanced triple-negative breast cancer patients. The primary time-to-event end point is PFS within 30 months. It is allowed for 90% power to detect an improvement in median PFS from 6 months in the control arm to 10 months in the new treatment arm, that is, the target HR = 0.6 with 2.5% level of significance (one-sided). Accrual is projected to occur over 15 months, and the final PFS analysis is expected 30 months after the first patient enrolls. Assuming normal distribution for logHRs, it could be claimed success if the posterior distribution of logHR satisfies $Pr(\log HR < \delta \,|D) > \theta_T$ given the observed data D, and no futility criterion is required at the end of the trial. Considering two interim analyses at approximately 50% and 80% of information fraction, the success and futility early stopping criteria are $Pr(\log HR < \delta \,|D) > \theta_U$ and $Pr(\log HR < \delta|D) < \theta_L$, respectively. Now we need to calibrate the design parameters $(\theta_T, \theta_U, \theta_L)$ for the decision-making.

Firstly, we calculate an initial sample size by classical frequentist method for the simulation. With an average 5% dropout rate per year, there will be approximately 100 patients required in each arm and totally 164 events occurred. Then, we simulate the data under null hypothesis (HR = 1) and did not allow early stopping in the interim analyses. Assuming a non-informative prior distribution of mean survival μ for each arm, e.g., IG (0.01,0.01), we took 5000 posterior samples of μ for the PoP calculation. Varying θ_T from 0.6 to 0.99, we performed simulations to calibrate the cutoff probability values to satisfy 2.5% type I error. For each configuration, we carried out 5000 simulated trials to summarize the operating characteristics. The resulting type I error are presented in **Table 1**. With the decreasing of the target

δ	θ_T	Success probability	Sample size	Total time
log(0.8)	0.7	0.0247	218	23.1796
log(0.7)	0.6	0.0083	208	22.0439
log(0.7)	0.5	0.0157	208	22.0022
log(0.7)	0.46	0.0248	208	22.1248
log(0.7)	0.45	0.0210	209	22.2274
log(0.7)	0.43	0.0200	209	22.1462
log(0.6)	0.25	0.0118	190	20.1316
log(0.6)	0.2	0.0123	189	20.0138
log(0.6)	0.18	0.0227	188	19.8901

Table 1.
Stage 1 parameter calibration with no early stopping by varying the design parameter θ_T at different values of δ under null hypothesis with HR = 1.

δ	θ_T	θ_L	θ_U	Success probability	Sample size	Total time
$\log(0.6)$	0.18	0.01	0.91	0.0245	219	17.0610
		0.01	0.93	0.0228	218	16.8673
$\log(0.7)$	0.46	0.01	0.89	0.0236	200	19.8338
		0.01	0.91	0.0328	214	20.2497

Table 2.
Stage 2 parameter calibration with the tuned θ_T by varying the design parameter θ_L and θ_U under null hypothesis.

logHR, the success probability boundary also decreased to yield 2.5% for the type I error. For example, when δ is expected to be $\log(0.7)$, we should select the cutoff value of $\theta_T = 0.46$. When the target HR decreased to 0.6, the boundary decreased sharply to 0.18 to maintain the type I error. So the investigator can choose a clinical meaningful treatment as the target value δ, and multiple criteria can be also required based on several probabilities.

In the second stage of parameter calibration, given a specific δ and the tuned θ_T, we varied the value of θ_U and θ_L: $\theta_U = 0.90, \ldots 0.99$ and $\theta_L = 0.01, \ldots 0.10$. To control the type I error, we finally selected $\theta_U = 0.89$ and $\theta_L = 0.01$ to calibrate early stopping in the interim analysis under the null to detect the effect size of log (HR = 0.7). Similarly, if we want to detect $\log(\text{HR} = 0.6)$, we should select $\theta_U = 0.91$ and $\theta_L = 0.01$ under the null to control the type I error rate (**Table 2**).

Then, we obtained the tuned parameters ($\theta_T, \theta_U, \theta_L$) for a specific effect size of . To maintain 90% power, we simulate data from the H_1, e.g. $\log\text{HR} = \delta = 0.6$, and the expected number of events was found to be 172.

9. Conclusion

In this chapter, we introduce the Bayesian group sequential framework with an example with details for planning and executing interim analyses. The concepts of PoP and predictive probability are intuitive and efficient tools for making decisions about continuation or early stopping and can be used at interim analyses even if the final planned analysis is to be performed in the classical frequentist hypothesis testing framework. Simulations can help assess the performance of different decision rules and assist in the determination of the sample size and are needed to tune desirable design's parameters. Bayesian approaches are often simpler to interpret than frequentist methods and allow teams to consider the evidence in support of different effects. Using these methods in clinical drug development can result in efficient studies that make the best use of resources while ensuring good chances of success.

Author details

Chen Li[1*†], Ping Huang[2†] and Haitao Pan[3]

1 Department of Health Statistics, School of Preventive Medicine, Fourth Military Medical University, Xi'an, Shaanxi, P. R. China

2 State Key Laboratory of Military Stomotology, National Clinical Research Center for Oral Diseases, Shaanxi Clinical Research Center for Oral Diseases, Department of Clinical Laboratory of Stomatology, Fourth Military Medical University, Xi'an, Shaanxi, China

3 Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA

*Address all correspondence to: lc.biosta@qq.com

† These authors contributed equally.

IntechOpen

References

[1] Lai TL, Lavori PW, Tsang KW. Adaptive design of confirmatory trials: Advances and challenges. Contemporary Clinical Trials. 2015;**45**(Pt A):93-102

[2] Gsponer T, Gerber F, Bornkamp B, Ohlssen D, Vandemeulebroecke M, Schmidli H. A practical guide to Bayesian group sequential designs. Pharmaceutical Statistics. 2014;**13**(1):71-80

[3] Yin G, Lam CK, Shi H. Bayesian randomized clinical trials: From fixed to adaptive design. Contemporary Clinical Trials. 2017;**59**:77-86

[4] FDA. Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry. FDA; 2019. Available from: https://www.fda.gov/regulatory-inf ormation/search-fda-guidance-docume nts/adaptive-design-clinical-trials-drug s-and-biologics-guidance-industry

[5] Vandemeulebroecke M. Group sequential and adaptive designs - a review of basic concepts and points of discussion. Biometrical Journal Biometrische Zeitschrift. 2008;**50**(4): 541-557

[6] Kittelson JM, Emerson SS. A unifying family of group sequential test designs. Biometrics. 1999;**55**(3):874-882

[7] Connor JT, Elm JJ, Broglio KR. Bayesian adaptive trials offer advantages in comparative effectiveness trials: An example in status epilepticus. Journal of Clinical Epidemiology. 2013;**66**(8 Suppl):S130-S137

[8] Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. The New England Journal of Medicine. 2018; **378**(1):11-21

[9] Greenhouse JB, Wasserman L. Robust Bayesian methods for monitoring clinical trials. Statistics in Medicine. 1995;**14**(12): 1379-1391

[10] Ibrahim JG, Chen MH, Gwon Y, Chen F. The power prior: Theory and applications. Statistics in Medicine. 2015; **34**(28):3724-3749

[11] Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics. 2011;**67**(3): 1047-1056

[12] Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics. 2014;**70**(4): 1023-1032

[13] Morita S, Thall PF, Müller P. Determining the effective sample size of a parametric prior. Biometrics. 2008; **64**(2):595-602

[14] Neuenschwander B, Weber S, Schmidli H, O'Hagan A. Predictively consistent prior effective sample sizes. Biometrics. 2020;**76**(2):578-587