Practice and Guidance for Conducting Bayesian Interim Analyses in Clinical Trials

Hal Li, Davis Gates

***Abstract*— The theory and methods for the Bayesian Interim Analysis for Clinical Trials are motivators for this presentation. Introductions to the Bayesian statistics, Bayesian Inference, and interim analysis for clinical trials are provided, along with detailed discussion of the theory. Practical examples of the Bayesian interim analysis for clinical trials are provided to illustrate the methods. The Bayesian predictive probability approach, Interval estimation approach and posterior distribution summary approach on the Normal and Binomial distributed data are also discussed. Bayesian inference simulations were performed. Discussions are provided to the FDA's guidance and opinion on Bayesian analysis, as well as FDA's approval history on Bayesian analyzed clinical trials.**

Index Bayesian, Interim Analysis, Clinical Trails, Simulation

I. An Overview of the Bayesian Approach

Supposing that we are interested in estimating θ from a data set: $X=\{x_1, \ldots, x_n\}$. Bayes theorem provides a solution by using a well-known rule about conditional probabilities:

$$P(\theta_i \mid X) = \frac{P(X, \theta_i)}{P(X)} = \frac{P(X \mid \theta_i)P(\theta_i)}{\sum_j P(X \mid \theta_j)P(\theta_j)} \quad (1)$$

Overall, Bayesian inference is based on the posterior distribution of the parameter P(θ|X). However, in order to derive the posterior distribution, we need to specify the prior distribution, P(θ) – the distribution of θ, we also need to determine the likelihood function P(X|θ) from the data observed. From formula (1), one can see that the P(θ|X) is proportional to (i.e. has the same shape as) the product of the likelihood function and the prior distribution of the data:

$$P(\theta|X) \sim P(X|\theta)P(\theta) \quad (2)$$

Having derived the posterior distribution P(θ|X), in Bayesian analysis all further inferences about θ will be derived from that distribution. This includes calculations of location parameters including the posterior mean, median, mode, or percentiles, among other parameters.

II. Sequential use of Bayes Theorem

Manuscript received June 2, 2013; revised July 25, 2013.

Hal Li, M.S., Principal Biostatistician, is with Biostatistics, Accenture, Life Sciences, 585 East Swedesford Road, Wayne, PA 19087, Phone : 610-535-6500 Ext 5829, Fax: 610- 535-6515, Email : hal.h.li@accenture.com

Davis Gates, Ph.D., Director, is with Late Development Statistics, Merck Research Laboratories, 2015 Galloping Hill Road, K-15-2-2445, Kenilworth, NJ-07033, Phono: 908-740-5094, Fax: 908-740-3353, Email: davis.gates@merck.com

Now suppose that we decide to observe data in two or more segments, X_m followed by X_n. After the first segment is observed, our posterior distribution is given by:

$$P(\theta| X_m) \sim P(X_m|\theta)\, P(\theta).$$

This posterior distribution then becomes the prior distribution for the next use of Bayes theorem. Then, after the next segment X_n is observed, the posterior conditioning on all the data is:

$$P(\theta| X_m, X_n) \sim P(X_n|\theta, X_m)\, P(\theta|X_m)$$
$$\sim P(X_n|\theta, X_m)\, P(X_m|\theta)P(\theta)$$

This result can also be derived by considering a single step with the datasets (X_m, X_n), by factoring

$$P(X_n, X_m|\theta) = P(X_n|\theta, X_m)\, P(X_m|\theta).$$

Usually, the distribution of $P(X_n|\theta, X_m)$ is equal to the distribution of $P(X_n|\theta)$, so this equivalence will lead to the same posterior distribution.

A. Normally Distributed Data

For a single-arm study design, we take a sufficient number of independent observations such that $x|\theta \sim N(\theta,\sigma^2)$ where σ is known. In these circumstances:

$$f(x \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$$

In this normally distributed data pool, the prior distribution of the data is $\theta \sim N(\mu_0, \sigma_0^2)$. That is:

$$f(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}_0} e^{-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}}$$

By formula (2), the posterior distribution $P(\theta|X) \sim$

$$(\prod_{i=1}^{n} e^{\frac{(x_i-\theta)^2}{2\sigma^2}}) e^{\frac{(\theta-\mu_0)^2}{2\sigma_0^2}} \sim \exp[-\frac{1}{2(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2})}(\theta - \frac{\frac{\mu_0}{\sigma_0^2}+\frac{n}{\sigma^2}\bar{x}}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}})^2] \quad (3)$$

It is apparent $P(\theta|x) \sim N(\mu_1, \sigma_1^2)$. There are several aspects about this analysis that should be noted. Most importantly, Bayesian analysis of normal distributions is an example of conjugate analysis. Conjugate models occur when the posterior distribution is of the same family as the prior distribution.

In this case, a normally distributed prior and likelihood of the data leads to a posterior distribution that is also normal. From formula (3), one can also see that the posterior mean is a weighted average of the prior mean and the data mean. Thus, posterior precision is greater than prior precision.

B. Binomial Distributed Data

For this type of endpoint, we again use a single armed

ISBN: 978-988-19253-1-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

study with independent observations as an example. Data is obtained from n subjects, where the number of responders X has a Binomial (n, θ) distribution. In this case, θ is the parameter of response rate and:

$$f(X \mid \theta) = \binom{n}{x} \theta^{x} (1-\theta)^{n-x}$$

In this case, we assume the prior distribution of θ is Beta(a, b), that means:

$$f(\theta \mid a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}$$

By formula (2), the posterior then becomes θ|X ~ Beta(a+x, b+n-x), where the P(θ|X) ~

$$\binom{n}{x} \theta^{x} (1-\theta)^{n-x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \sim \theta^{x+a-1} (1-\theta)^{n+b-x-1}$$

It is important to note that, as with the normal distribution, the binomial distribution in this analysis is another example of conjugate analysis.

As an example, let us take a study design where there are 20 subjects in Stage 1 of the study, and another 20 subjects in Stage 2.

In Stage 1, it is found that there are 4 responders out of 20 subjects (20%). In this case specified here, the prior distribution is Beta(1,1) for response rate θ. Thus, the posterior at Stage 1 becomes:

θ|x ~ Beta (1+4, 1+20-4) = Beta(5,17)

In Stage 2, it is then found that there are 5 responders in the additional 20 subjects (25%). We can take the prior distribution in this case to be the posterior we derived from the Stage 1 data, which was θ~Beta (5, 17). Now, incorporating the data from Stage 2, we find that the final posterior of θ:

θ|x ~ Beta (5+5, 17+20-5) = Beta(10,32).

This is the same result as would have been obtained from a full data analysis, rather than the separate stage calculations as has been used above. In that case, the prior distribution remains: θ ~ Beta (1,1). Then, taking the full data into account (Stage 1 & Stage 2), we find that there are 9 responses out of 40 subjects leading to a posterior of:

Beta(1+9, 1+40-9)=Beta(10,32)

C. Poisson Distributed Data

The probability of a given number of events occurring in a fixed interval of time can often be modeled by the Poisson distribution, such that:

$$f(x \mid \theta) = \frac{\theta^{x} e^{-\theta}}{x!}$$

In this case, the prior distribution is usually taken to be θ~ Gamma(a, b) such that:

$$f(\theta \mid a,b) = \frac{b^{a}}{\Gamma(a)} \theta^{a-1} e^{-b\theta}$$

Again by applying formula (2), the posterior distribution becomes

$$\prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x!} \frac{b^{a}}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \sim \theta^{t+a-1} e^{-(b+n)\theta}$$

That means θ|X ~ Gamma (a+t, b+n), where t=Σ(xi).

III. Bayesian Analysis on Clinical Trial Data

A. Decision making in Bayesian Frameworks

Bayesian analysis can answer many questions that scientists may ask about drug candidates at the end of the first efficacy study in order to help them decide on how to develop the drug further. For example, it can determine the probability that the treatment effect observed was a false positive. It also provides an answer to perhaps the most important question in clinical studies: What is the probability that the drug works given the evidence that have been collected so far? The common Bayesian analytical methods applied in clinical trial data are posterior distribution summaries, interval estimation and predictive probability approach.

B. Frequentist vs. Bayesian Approach

Bayesian statistics is gaining more ground in the area of clinical trials. In clinical trial regulation the frequentist view continues to dominate, however there are areas (e.g. the regulation of medical devices) where the Bayesian approach is being applied. The following diagram illustrates the processes of analyzing the clinical data by these two approaches.

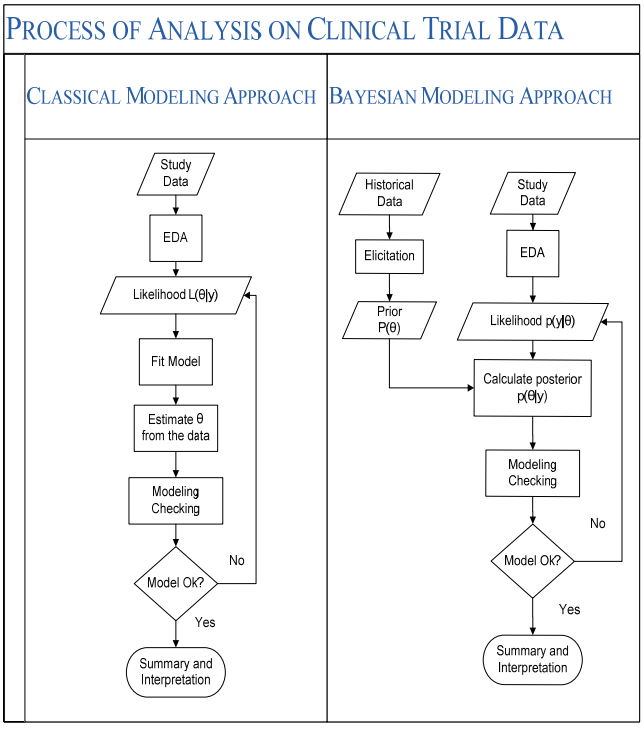

Figure 1 Diagram illustrates the processes of analyzing the clinical data by Frequentist and Bayesian approaches.

ISBN: 978-988-19253-1-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

(7)

C. Posterior Distribution Summaries

Both the posterior mean and the posterior mode can be determined through analysis and together they provide insight into the observed data.

Posterior mean: In general, the posterior mean is a compromise between the prior distribution and the observed data, and is calculated as:

$$E(\theta|x)=\int \theta P(\theta|x)d\theta \quad (4)$$

As in the binary data example that was shown presented, as we know that for Beta (a, b), the mean E(X) =a/(a+b). Hence, the posterior mean of Beta (10,32) is 10/42.

Posterior mode: In general, the posterior mode represents the 'most likely' posterior value given the prior distribution. As in the same binary data example, we know that for X~ Beta (a, b), if a, b > 1, the mode can be calculated as:

$$\frac{a-1}{a+b-2}$$

Hence, the posterior mode of Beta (10,32) is 9/40

D. Interval Estimation Approach

Credible intervals are used for purposes similar to those of confidence intervals in frequentist statistics. The difference between the two is shown below.

Confidence intervals: $P(a(x)< \theta <b(x)|\ \theta)=1-\alpha$,
x is the random variable and θ is unknown

Credible intervals: $P(a(x)< \theta <b(x)|\ x)=1-\alpha$,
θ is the random variable given the data x

The Bayesian Credible Interval represents a posterior probability interval that can be used for interval estimation.

For example, if the posterior probability that θ lies in an interval is 0.95, then the interval found is called a 95% credible interval. This interval would thus be the interval of θ between the 2.5th and 97.5th percentiles of the distribution.

E. Predictive Probability Approach

This approach can be used to predict a future i.i.d. data point $\tilde{x}$ on the basis of a currently observed i.i.d. data x.

In this approach, the prior predictive distribution of x is:

$$p(x)=\int p(x|\theta)p(\theta)d\theta \quad (5)$$

Therefore, the posterior predictive distribution of $\tilde{x}$ will be:

$$p(\tilde{x}|x)=\int p(\tilde{x}|x,\theta)p(\theta|x)d\theta \quad (6)$$

Usually $\tilde{x}$ and *x* are conditionally independent given θ. That means

$$p(\tilde{x}|x,\theta)=p(\tilde{x}|\theta)$$

Combining (6) and (7) we have

$$p(\tilde{x}|x)=\int p(\tilde{x}|\theta)p(\theta|x)d\theta \quad (8)$$

Provided we can do this integration in (8), prediction becomes straightforward.

IV. Interim Analysis for Clinical Trials

A. Normally Distributed Data

Assuming the analysis will be performed on the normally distributed endpoint, to perform these studies, n pairs of subjects are enrolled with two possible treatments. We call the estimated treatment difference $\bar{x}$ while the true treatment difference is θ.

Independent observations are then taken such that

$$\bar{x}\ |\theta \sim N(\theta,\sigma^2/n),$$

where σ is known. In this case, the prior distribution is taken to be:

$$\theta \sim N(\mu_0, \sigma_0^2),$$

where $\sigma_0^2=\sigma^2/n_0$ and n_0 reflects the precision of the prior information about the difference.

The posterior distribution is thus found to be:

$$\theta \mid \bar{x} \sim N\left[\frac{n_0\theta + n\bar{x}}{n_0+n}, \frac{\sigma^2}{n_0+n}\right]$$

For these studies, we use the decision rule that the experimental treatment is clinically superior if:

$$\theta > \Delta_E$$

and the control treatment is clinically superior if:

$$\theta < \Delta_C$$

The trial will thus be stopped if either:

$$P_E = \int_{\Delta_E}^{t} p(\theta \mid \bar{x})d\theta < \varepsilon_E$$

Or

$$P_C = \int_{t}^{\Delta_C} p(\theta \mid \bar{x})d\theta < \varepsilon_C$$

The stopping rules is illustrated in the following Figure 2.

ISBN: 978-988-19253-1-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

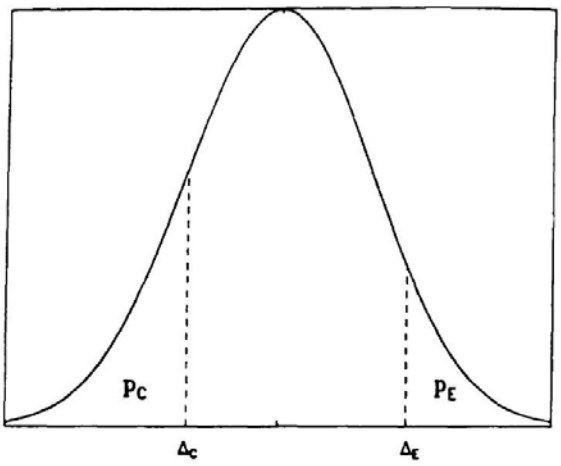

Figure 2 Stop Criterion for two-treatment trial based on the current posterior distribution: stop the trial if $P_C < \varepsilon_c$ or $P_E < \varepsilon_E$

In 1989, L. Freedman[1] showed that depending on the precision of the prior information (n_0), Bayes' boundaries tended to converge at a rate somewhat between the Pocock[2] scheme and O'Brien and Fleming (OBF)[3] scheme as is shown below Figure 3.

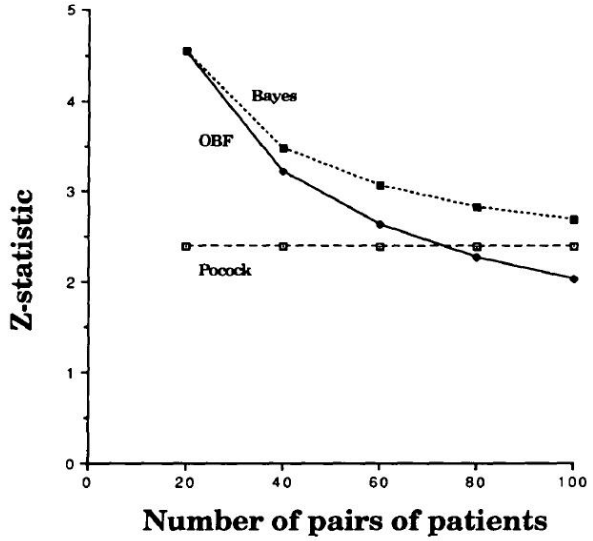

Figure 3 Stopping boundaries from Bayes, Pocock and O'Brien and Fleming (OBF) schemes. Stopping rule: |z|>Z for Pocock and OBF scheme and Bayes scheme with n_0=22 in a trial with 200 patients

B. Binomial Distributed Data – By Example

For this example, the study design will be a two period cross over design where patients are randomized to one of two treatment sequences, either Treatment x followed by Treatment y (xy), or Treatment y followed by Treatment x (yx).

The endpoint will be the proportion of patients achieving clinical response (CR/PR), which is also known as the Response Rate. Our H_0 is therefore that there is no difference between the study medication and placebo in response rate.

The procedure to be followed in conducting this Bayesian interim analysis is shown in Figure 4:

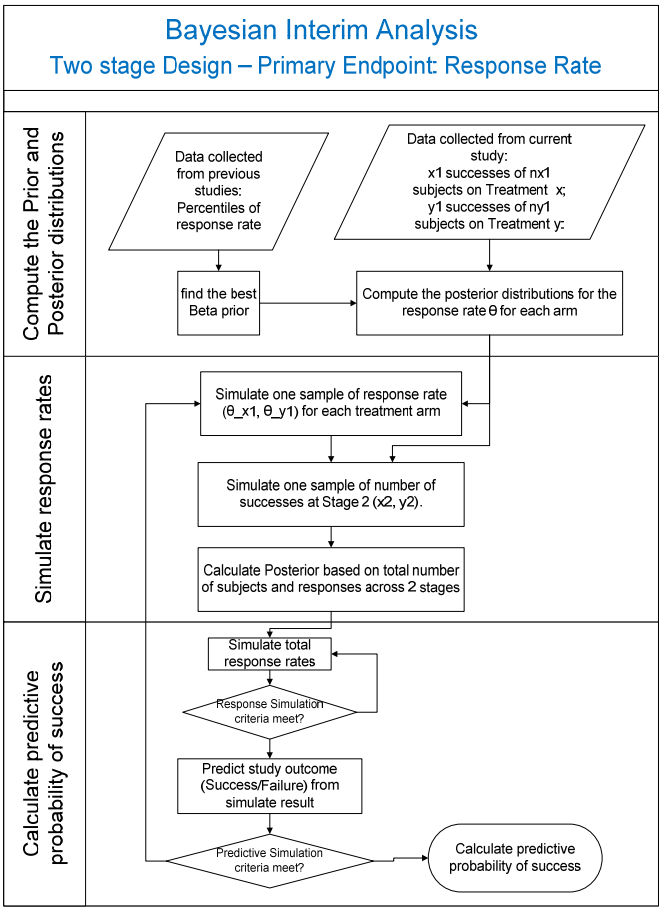

Figure 4 The procedure in conducting Bayesian interim analysis (example)

Let us suppose that that in Stage 1, we observed x1 successes out of the n_{x1} subjects on Treatment x such that $x1 \sim Binomial(\theta\_x1, n_{x1})$ and y1 successes out of n_{y1} subjects on Treatment y such that $y1 \sim Binomial(\theta\_y1, n_{y1})$

The procedure for a 2-stage design is as follows:

1) Find the Beta Prior: We can find the best Beta prior from previous studies by specifying the most likely value and extremes (quintiles) for the response rate.

2) Compute the posterior distributions: This is determined for each treatment group (x & y), based on the observed data and the prior distribution: $\theta \sim Beta(Best\_a, Best\_b)$

Thus, for treatment x:

$$\theta\_x1 \sim Beta(Best\_a+x1, Best\_b+n_{x1}-x1) \quad (9)$$

Similarly, for treatment y:

$$\theta\_y1 \sim Beta(Best\_a+y1, Best\_b+n_{y1}-y1) \quad (10)$$

3) Take a random sample of response rate based on the previously calculated posterior distributions: For each treatment arm response rate (θ_x1, θ_y1), a sample is generated based on the previously calculated posterior distributions illustrated in formula (9) and (10).

ISBN: 978-988-19253-1-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

As an example the following Figure 5 illustrates the best Beta prior, the likelihood and the posterior.

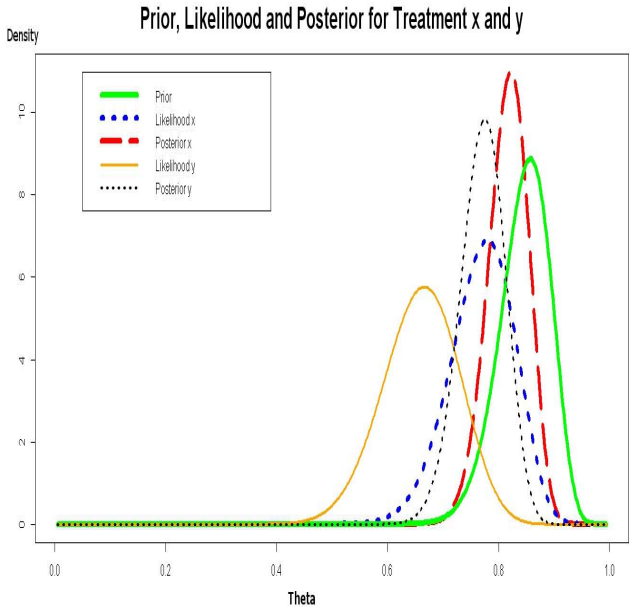

Figure 5 Prior, Likelihood and Posterior for Treatment X and Y: Best Prior Beta(52.2,9.5), Data in treatment X~Binom(50, 39), Posterior X~Beta(91.2, 20.5), Data in treatment Y~Binom(45, 30), Posterior Y~Beta(82.2, 24.5)

4) Simulate the number of responders in stage 2 for each treatment group (x2, y2): Suppose that there are n_{x2} and n_{y2} subjects on treatment groups x and y at this stage. Here we assume that the response rate will be the same as in stage 1, i.e. (θ_x1, θ_y1). In this case, it can be simulated that:

$$x2 \sim BINOM(n_{x2}, \theta\_x1)$$
$$y2 \sim BINOM(n_{y2}, \theta\_y1)$$

5) Calculate totals across both stages: Both the number of responders and the total number of subjects in each group across the 2 stages are determined. Respectively, these are:

$$x = x1+x2$$
$$y = y1+y2$$
$$Nx = n_{x1}+n_{x2}$$
$$Ny = n_{y1}+n_{y2}$$

6) Simulate total response rates: This will be performed on each arm of from the posterior distributions (θ|X) and (θ|Y) with all subjects.

In this case, the experimental Treatment x can be considered superior to reference Treatment y if ≥ 95% of the predictive distributions lead to a relative risk> 1.

Here, we are assuming that the posterior distributions in both stages are the same. A sample R software code could be read as:

```
postRX <- rbeta(10000, a+x, b+Nx-x)
postRY <- rbeta(10000, a+y, b+Ny-y)
Relative_risk=postRX/postRY
```

7) Repeat steps (3) to (6) (e.g. 1,000 times): This is done in to obtain the predictive probability of success (i.e. how often we can declare that the trial was a 'success').

If there is at least a 90% probability of a successful trial outcome for Treatment x being superior to Treatment y, then the trial is stopped at the interim analysis with the conclusion that Treatment x is superior in efficacy.

Conversely, if there is a lower than 10% probability of a successful trial, then the trial is terminated early with the conclusion that further trials would be futile.

If, however, the treatment's efficacy is not demonstrated but also not shown to be futile (i.e. the predictive probability is between 10% and 90%), then the study is continued in order to collect more data for next stage.

V. Simulation for Bayesian Inference

The Bayesian approach can, in theory, deal with realistically complex situations. Despite this, there are some non-standard posterior distributions that are mathematically intractable or computationally too intensive to be easily solved.

In these cases, posterior information can be obtained from looking at a set of values that are drawn from the posterior distribution, rather than ascertaining the precise mathematical equation that describes the distribution.

For example, a technique known as the Markov Chain Monte Carlo (MCMC) method can be used to draw samples from the posterior distribution. Therefore, results from this method can be used to find the posterior distribution empirically.

This method focuses on producing a Markov chain in which the distribution for the next simulated value $(\theta^{(j+1)},\psi^{(j+1)})$ depends only on the currently specified value $(\theta^{(j)},\psi^{(j)})$. Thus, under broad conditions, the samples will eventually converge into an 'equilibrium distribution'.

The MCMC method can be performed via WinBUGS; however, this program has a number of pros and cons. In its favor, it has both built-in graphics and convergence diagnostics as well as a flexible model specification. In contrast though, it is a 'stand-alone' program. This means that it is not very user friendly and also assumes that users are already skilled at Bayesian analyses (they must know how to choose the prior distribution and likelihood, check the fit of their model, and check convergence).

In addition to this specific program, though, the MCMC simulations can also be carried out using R or S+ software as well.

ISBN: 978-988-19253-1-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

VI. FDA Guidance on Bayesian Analysis

A. Guidance from the FDA

The FDA has laid out its guidance for how to use Bayesian analysis in the following article:

Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials – FDA CDRH Feb. 5 2010

In addition, Dr. Greg Campbell, who is the Director of Biostatistics at the CDRH has provided his opinion on these analysis[4]. In his view, Bayesian studies should be:

1) Prospectively designed

2) Restricted to studies that can utilize good data-based prior information.

3) Guided by an agreement between companies and the FDA with regard to the validity of its prior information.

4) Conducted such that the control group cannot be used as a source of prior information.

In addition, simulations are extremely important to any Bayesian study in order for it to be approved by the FDA.

More specifically, numerous simulations must be conducted to show that Type 1 errors (or some analog of them) are well-controlled. This is perhaps the key to FDA approval!

Also, simulations need to be conducted in order to help in estimating the approximate size of the trial and the strategy of interim looks. This is needed because usually Bayesian studies are not a fixed size.

B. FDA Approval History

Despite these challenges, there have been at least 15 original PMAs and PMA Supplements that the FDA has approved with a Bayesian analysis as its primary method. These supplements are varied and include numerous stent systems, a heart valve, and a number of spinal cage systems.

References

[1] Laurence S. Freedman, David J. Spiegelhalter, "Comparison of Bayesian with group sequential methods for monitoring clinical trials" Controlled Clinical Trials, Volume 10, Issue 4, December 1989, Pages 357–367.
[2] Pocock SJ (1977). Group sequential methods in the design and analysis of clinical trials. Biometrika, 64, 191-99.
[3] O'Brien PC, Fleming TR (1979). A multiple testing procedure for clinical trials. Biometrics, 35, 549-56.
[4] Gregory Campbell, "Bayesian and "Small n" Trials", Public Workshop on Pediatric Cardiovascular Devices September 30. 2010.

ISBN: 978-988-19253-1-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)