Draft Guidance
for Industry and FDA Staff

Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

DRAFT GUIDANCE
**This guidance document is being distributed for
comment purposes only.
Draft released for comment on May 23, 2006**

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Alternatively, electronic comments may be submitted to http://www.fda.gov/dockets/ecomments. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this document, contact Dr. Greg Campbell at 240-276-3133 or greg.campbell@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Division of Biostatistics
Office of Surveillance and Biometrics

Preface

Additional Copies

Additional copies are available from the Internet at:
http://www.fda.gov/cdrh/osb/guidance/1601.pdf, or to receive this document via your fax machine, call the CDRH Facts-On-Demand system at 800-899-0381 or 301-827-0111 from a touch-tone telephone. Press 1 to enter the system. At the second voice prompt, press 1 to order a document. Enter the document number 1601 followed by the pound sign (#). Follow the remaining voice prompts to complete your request.

Table of Contents

Draft Guidance for Industry and FDA Staff

Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

This draft guidance, when finalized, will represent the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

1.   Introduction

This document provides guidance on statistical aspects of the design and analysis of clinical trials for medical devices that use Bayesian statistical methods.

The purpose of this guidance is to discuss important statistical issues in Bayesian clinical trials for medical devices and not to describe the content of a medical device submission. Further, while this document provides guidance on many of the statistical issues that arise in Bayesian clinical trials, it is not intended to be all-inclusive. The statistical literature is rich with books and papers on Bayesian theory and methods; a selected bibliography has been included for further discussion of specific topics.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

2.   The Least Burdensome Approach

This draft guidance document reflects our careful review of what we believe are the relevant issues related to the use of Bayesian statistics in medical device clinical trials and what we believe would be the least burdensome way of addressing these issues. If

you have comments on whether there is a less burdensome approach, however, please submit your comments as indicated on the cover of this document.

3.   Foreword

3.1 What is Bayesian statistics?

Bayesian statistics is a statistical theory and approach to data analysis that provides a coherent method for learning from evidence as it accumulates. Traditional (frequentist) statistical methods formally use prior information only in the design of a clinical trial. In the data analysis stage, prior information is considered only informally, as a complement to, but not part of the analysis. In contrast, the Bayesian approach uses a consistent, mathematically formal method called Bayes' Theorem for combining prior information with current information on a quantity of interest. This is done throughout both the design and analysis stages of a trial.

3.2 Why use Bayesian statistics for medical devices?

When good prior information on clinical use of a device exists, the Bayesian approach may enable FDA to reach the same decision on a device with a smaller-sized or shorter-duration pivotal trial.

The Bayesian approach may also be useful in the absence of informative prior information. First, the approach can provide flexible methods for handling interim analyses and other modifications to trials in midcourse (e.g., changes to the sample size or changes in the randomization scheme). Second, the Bayesian approach can be useful in complex modeling situations where a frequentist analysis is difficult to implement or does not exist.

Good prior information is often available for a medical device; for example, from earlier studies on previous generations of the device or from studies overseas. These studies can often be used as prior information because the mechanism of action of medical devices is typically physical, making the effects local and not systemic. Local effects are often predictable from prior information when modifications to a device are minor.

Bayesian methods may be controversial when the prior information is based mainly on personal opinion (often derived by elicitation methods). The methods are often not controversial when the prior information is based on empirical evidence such as prior clinical trials. Since sample sizes are typically small for device trials, good prior information can have greater impact on the analysis of the trial and thus on the FDA decision process.

The FDA Modernization Act of 1997 mandates that FDA shall consider the least burdensome means of demonstrating effectiveness or substantial equivalence of a device (Section 513(a)(3)(D)(ii) and Section 513(i)(1)(D)). The Bayesian

approach, when correctly employed, may be less burdensome than a frequentist approach.[1]

3.3 Why are Bayesian methods more commonly used now?

Bayesian analyses are often computationally intense. However, recent breakthroughs in computational algorithms and many-fold increases in computing speed have made it possible to carry out calculations for virtually any Bayesian analysis. These advances have resulted in a tremendous increase in the use of Bayesian methods over the last decade. See Malakoff (1999). The basic tool that enabled the advances is a method called Markov Chain Monte Carlo (MCMC). For a technical overview of MCMC methods, see Gamerman (1997).

3.4 When should FDA participate in the planning of a Bayesian trial?

For any Bayesian trial, we recommend you schedule meetings to discuss experimental design, models, and acceptable prior information with FDA before the study begins. If an investigational device exemption (IDE) is required, we recommend you meet with FDA before you submit the IDE.

3.5 What software programs are available that can perform Bayesian analyses?

Currently, the only commonly available computer program dedicated to making Bayesian calculations is called WinBUGS[2] (Bayesian Inference Using Gibbs Sampling). It is non-commercial software. FDA expects other software packages to become available in the future. FDA recommends that you consult with FDA statisticians regarding your choice of software prior to analyzing your data.

---

[1] Two examples of successful use of Bayesian methods in device trials are:

TRANSCAN (SSE: http://www.fda.gov/cdrh/pdf/p970033b.pdf). Prior information was used to incorporate results from previous studies, resulting in a reduced sample size for demonstration of effectiveness.

INTERFIX (SSE: http://www.fda.gov/cdrh/pdf/p970015b.pdf). An interim analysis was performed; based on Bayesian predictive modeling of the future success rate, the trial was stopped early. No prior information was used.

[2] The WinBUGS program can be downloaded from the website of the Medical Research Center, Cambridge: www.mrc-bsu.cam.ac.uk. Other open-source Bayesian software packages are also under development.

3.6 What resources are available to learn more about Bayesian statistics?

Non-technical introductory references to Bayesian statistics and their application to medicine include Malakoff (1999), Hively (1996), Kadane (1995), Brophy & Joseph (1995), Lilford & Braunholtz (1996), Lewis & Wears (1993), Bland & Altman (1998), and Goodman (1999a, 1999b). Berry (1997) has written for FDA an introduction specifically on Bayesian medical device trials.

A comprehensive summary on the use of Bayesian methods to design and analyze clinical trials or perform healthcare evaluations appears in Spiegelhalter, Abrams, & Myles (2004).

Introductions to Bayesian statistics that do not emphasize medical applications, in order of complexity, are Berry (1996), DeGroot (1986), Stern (1998), Lee (1997), and Gelman, et al. (2004).

References with technical details and statistical terminology are Spiegelhalter, et al. (2000), Spiegelhalter, et al. (1994), Berry & Stangl (1996), Breslow (1990), and Stangl & Berry (1998).

An overview of Markov Chain Monte Carlo for Bayesian inference may be found in Gamerman (1997). Practical applications appear in Gilks, et al. (1996) and in Congdon (2003).

Brophy & Joseph (1995) provide a well-known synthesis of three clinical trials using Bayesian methods.

A list of resources on the Web appears on the International Society for Bayesian Statistics website.[3]

3.7 The Bayesian approach should not push aside sound science.

Scientifically sound clinical trial planning and rigorous trial conduct are important regardless of whether you use the Bayesian or the frequentist approach. We recommend you remain vigilant regarding randomization, concurrent controls, prospective planning, blinding, bias, precision, and all other factors that go into a successful clinical trial. See **Section 5.1: Bayesian trials start with a sound clinical trial design**.

3.8 What are the potential benefits of using Bayesian methods?

Sample size reduction or augmentation

The Bayesian methodology may reduce the sample size FDA needs to reach a regulatory decision. You may achieve this reduction by using prior information and interim looks during the course of the trial. When results of a trial are

---

[3] http://www.bayesian.org/

unexpectedly good (or unexpectedly bad) at an interim look, you may be able to stop early and declare success (or failure).

The Bayesian methodology can allow for augmentation of the sample in cases where more information helps FDA make a decision. This can happen if the observed variability of the sample is higher than that used to plan the trial.

Midcourse changes to the trial design

With appropriate planning, the Bayesian approach can also offer the flexibility of midcourse changes to a trial. Some possibilities include dropping an unfavorable treatment arm or modifications to the randomization scheme. Modifications to the randomization scheme are particularly relevant for an ethically sensitive device or when enrollment becomes problematic for a treatment arm. Bayesian methods can be especially flexible in allowing for changes in the treatment to control randomization ratio during the course of the trial. See Kadane (1996) for a discussion.

Exact analysis

The Bayesian approach can sometimes be used to obtain an exact analysis when the corresponding frequentist analysis is only approximate or is too difficult to implement.

3.9 What are the potential difficulties in the Bayesian approach?

Extensive preplanning

Planning the design, conduct, and analysis of any trial are always important from a regulatory perspective, but they are crucial for the Bayesian approach. This is because decisions are based on:

- prior information
- information obtained from the trial, and
- the mathematical model used to combine the two.

Different choices of prior information or different choices of model can produce different decisions. As a result, in the regulatory setting, the design of a Bayesian clinical trial involves pre-specification of (and agreement on) both the prior information and the model. This includes clinical agreement on the appropriateness of the prior information and statistical agreement on the mathematical model to be used. Since reaching this agreement is often an iterative process, we recommend you meet with FDA early on to discuss and agree upon the basic aspects of the trial design.

A change in the prior information or the model at a later stage of the trial may imperil the scientific validity of the trial results. For this reason, formal agreement meetings may be appropriate when using a Bayesian approach.

Specifically, the identification of the prior information may be an appropriate topic of an agreement meeting.[4]

Extensive model-building

The Bayesian approach often involves extensive mathematical modeling of a clinical trial, including:

- the probability distributions chosen to reflect the prior information

- the influence of covariates on patient outcomes or missing data

- the relationships between various sources of prior information used in the model.

We recommend you determine modeling choices through close collaboration and agreement with FDA's and your statistical and clinical experts.

Specific statistical and computational expertise

The Bayesian approach often involves specific statistical expertise. Computer-intensive calculations are often used to:

- analyze trial data

- check model assumptions

- assess prior probabilities at the design stage

- perform simulations to assess probabilities of various outcomes

- estimate sample size.

The technical and statistical costs for the above are often offset by the savings of a shorter trial or a more flexible analysis.

Choices regarding prior information

An FDA advisory panel may question prior information you and FDA agreed upon beforehand. We recommend you be prepared to clinically and statistically justify choices of prior information. In some cases, we recommend you perform sensitivity analyses to check robustness of models and priors.

Device labeling

Results from a Bayesian trial may be expressed differently from the way trial results are usually described in device labels. We recommend you ensure trial results reported on the device label are easy to understand.

---

[4] The FDA Modernization Act (FDAMA) provided for two types of early collaboration meetings: agreement meetings and determination meetings. For details, see the FDA Guidance on early collaboration meetings at http://www.fda.gov/cdrh/ode/guidance/310.html.

Checking calculation

The flexibility of Bayesian models and the complexity of the computational techniques for Bayesian analyses create greater possibility for errors and misunderstandings. FDA, therefore, will carry out a detailed statistical review of a Bayesian submission.

Since the software used in Bayesian analysis is relatively new, FDA will often verify results using alternate software. FDA recommends you submit your data and any programs used for Bayesian statistical analyses electronically.

Bayesian and traditional analyses approaches may differ

Two investigators, each with the same data and a different preplanned analysis (one frequentist and one Bayesian), could conceivably reach different conclusions that are both scientifically valid. While the Bayesian approach is often favorable to the investigator with good prior information, the approach can be more conservative than a frequentist approach (e.g., see **Section 5: Planning a Bayesian Clinical Trial**).

If the results from your pre-planned analysis are not as positive as expected, we recommend you do not switch from a frequentist to a Bayesian analysis (or vice versa). Such *post hoc* analyses are not scientifically sound and would tend to weaken the validity of the submission.

4. Bayesian Statistics

4.1 Introduction

The fundamental idea in Bayesian statistics is that one's uncertainty about an unknown quantity of interest is represented by probabilities for possible values of that quantity. For instance, unknown quantities of interest in device trials might be:

- clinical safety and effectiveness endpoints
- a patient's outcome to be observed in the future
- a missing observation on a patient.

Prior distribution and non-informative prior distribution

Before a trial begins and data are obtained, the investigator assigns prior probabilities to the possible values of the unknown quantity, known as the prior distribution. In principle, the prior can be based on the investigator's personal knowledge of the quantities of interest or on another expert's opinion, etc. If absolutely nothing is known about that quantity, something called a *non-informative* prior distribution may be specified. In trials undergoing regulatory

review, however, the prior distribution is usually based on data from relevant previous trials.

Bayes' theorem and posterior probabilities

After data are gathered and information becomes available, the prior probabilities are mathematically updated according to a statistical result called Bayes' theorem. The updated probabilities, known as posterior probabilities, are probabilities for values of the unknown quantity after data are observed. This approach is a scientifically valid way of combining previous information (the prior probabilities) with current data. The approach adjusts to changing levels of evidence: today's posterior probabilities become tomorrow's prior probabilities.

The Bayesian paradigm

The Bayesian paradigm states that probability is the only measure of one's uncertainty about an unknown quantity. In a Bayesian clinical trial, uncertainty about an endpoint (also called parameter) is quantified according to probabilities, which are updated as information is gathered from the trial.

Decision rules

The pre-market evaluation of medical devices aims to demonstrate the safety and effectiveness of a new device. This demonstration is most commonly achieved through statistical hypothesis testing. For Bayesian trials, hypotheses are tested with decision rules. One common type of decision rule considers that a hypothesis has been demonstrated (beyond a reasonable doubt) if its posterior probability is large enough (e.g., 95 or 99 percent).

The Bayesian approach encompasses a number of key concepts, some of which are not part of the traditional statistical approach. Below, we briefly discuss these concepts and contrast the Bayesian and frequentist approaches.

4.2 What is a prior distribution?

Suppose that **x** is an endpoint (parameter) of interest in a clinical trial. The initial uncertainty about **x** should be described by a probability distribution for **x**, called the prior distribution and denoted by P(**x**).

As an example, suppose **x** is the rate of a serious adverse event. Its possible values will lie between 0 and 1. One prior distribution is the uniform distribution indicating no preference for any value of **x**. So the probability that **x** lies between 0.1 and 0.2 is the same as the probability that **x** lies between 0.4 and 0.5, or between 0.65 and 0.75, or in any interval of length 0.1.

Alternatively, the prior distribution might give preference to lower values of **x**. For example, the probability that **x** lies between 0.2 and 0.3 can be larger than the probability that **x** lies between 0.7 and 0.8.

4.3 What is the likelihood of the observed data?

Now suppose data have been obtained from a clinical trial. The likelihood of these data being observed can be formally expressed in terms of a likelihood function, P(data |$\mathbf{x}$), which is the conditional probability of observing the data, given a specific value of $\mathbf{x}$, for each possible value of $\mathbf{x}$ (the parameter). The likelihood is the mathematical and statistical model that reflects the relationships between the observed outcomes in the trial, the covariates, and the endpoint $\mathbf{x}$ of interest.

4.4 What is the posterior distribution?

The final objective is to obtain the probability of each possible value of the endpoint $\mathbf{x}$ conditional on the observed data, denoted P($\mathbf{x}$| data). Using exclusively the laws of probability, Bayes' theorem combines the prior distribution for $\mathbf{x}$, P($\mathbf{x}$), with the likelihood, P(data|$\mathbf{x}$), in order to obtain the posterior distribution for $\mathbf{x}$, P($\mathbf{x}$|data). In the Bayesian approach, all available information about $\mathbf{x}$ is summarized by the posterior distribution, P($\mathbf{x}$ |data), and all inferences about the endpoint are based on it.

As more data are obtained, more updating can be performed. Consequently, the posterior distribution that has been obtained today may serve as a prior distribution later, when more data are gathered. The more information that is accrued, the less uncertainty there will be about the posterior distribution for $\mathbf{x}$, and as more and more information is collected from the trial, the influence of the prior will become less and less. If enough data are collected, the relative importance of the prior distribution will be negligible compared to the likelihood.

For more introductory material on conditional probability, Bayes' theorem, and Bayesian statistics, see DeGroot (1986), Lee (1997), Berry (1996), Lindley (1985). For an introduction specific to medical devices, see Berry (1997).

4.5 What is a predictive distribution?

The Bayesian approach allows for the derivation of a special type of posterior probability; namely, the probability of future events given outcomes that have already been observed. This probability is called the *predictive probability*. Collectively, the probabilities for all possible values of future outcomes are called the predictive distribution. Predictive distributions have many uses, including:

- determining when to stop a trial
- helping a physician and patient make decisions by predicting the clinical outcome of the current patient, given the observed outcomes of past patients in a clinical trial
- predicting a clinical outcome from a validated surrogate
- adjusting trial results for missing data (imputation)

- model checking.

These uses are discussed in more detail in **Section 6. Analyzing a Bayesian Clinical Trial**.

4.6 What is exchangeability?

Exchangeability is a key idea in statistical inference in general, but it is particularly important in the Bayesian approach. Two observations are exchangeable if they provide equivalent statistical information. So, two patients randomly selected from a particular population of patients can be considered exchangeable. If the patients in a clinical trial are exchangeable with the patients in the population for which the device is intended, then the clinical trial can be used to make inferences about the entire population. Otherwise, the trial tells us very little about the larger population. The concept of a *representative sample* can thus be expressed in terms of exchangeability.

Exchangeability may depend on the statistical model used. If we know, for example, that the adverse event rate for a particular device depends on the patient's body mass index (BMI), then we say that patients are exchangeable *conditional on* BMI. That is, two patients will provide equivalent statistical information, but only after we account for differences in BMI. Therefore, any discussion of exchangeability should also include a discussion of the statistical models used.

We can also think of exchangeability in terms of clinical trials. Two trials are exchangeable if they provide equivalent statistical information about some *super-population* of clinical trials. Again, the trials may be exchangeable, but only after we account for (that is, condition on) other factors with the appropriate statistical model.

The use of Bayesian hierarchical models enables us to combine information from different sources that may be exchangeable on some levels but not on others (see **Section 5: Planning a Bayesian Clinical Trial**). If trials are exchangeable, then Bayesian hierarchical models enable us to make full use of the information from all the trials. For technical definitions of exchangeability, see Bernardo & Smith (1995).

4.7 What is the likelihood principle?

The likelihood principle is important in all of statistics, but it is especially central to the Bayesian approach. The principle states that *all* information about the endpoint of interest, **x**, obtained from a clinical trial, is contained in the likelihood function. In the Bayesian approach, the prior distribution for **x** is updated using the information provided by the trial through the likelihood function, and nothing else. Bayesian analysts base all inferences about **x** solely on the posterior distribution produced in this manner.

A trial can be altered in many ways without changing the likelihood function. As long as the modification schemes are pre-specified in the trial design, adherence to the likelihood principle allows for flexibility in conducting Bayesian clinical trials, in particular with respect to:

- modifications to sample size

- interim looks for the purpose of possibly stopping the trial early

- adaptive designs in general.

For more on the topic, see Berger & Wolpert (1988), Berger & Berry (1988), Irony (1993), and Barlow, et al., (1989).

4.8 How do the Bayesian and frequentist approaches differ?

As outlined above, Bayesian analysts base all inferences on the posterior distribution, which (in adherence to the likelihood principle) is the product only of the prior and the likelihood function. Although the frequentist approach makes extensive use of the likelihood function, it does not always strictly adhere to the likelihood principle. For example, the interpretation of a frequentist p-value is based on outcomes that might have occurred but were not actually observed in the trial; that is, on something external to the likelihood.

Another way of saying this is that Bayesian inferences are based on the "parameter space" (the posterior distribution), while frequentist inferences are based on the "sample space" (the set of possible outcomes of a trial).

5. Planning a Bayesian Clinical Trial

5.1 Bayesian trials start with a sound clinical trial design

The basic tenets of good trial design are the same for both Bayesian and frequentist trials. Parts of a comprehensive trial protocol include:

- objectives of the trial

- endpoints to be evaluated

- conditions under which the trial will be conducted

- population that will be investigated

- planned statistical analysis.

We recommend you follow the principles of good clinical trial design and execution, including minimizing bias. Randomization minimizes bias that can be introduced in the selection of which patients get which treatment. Randomization allows concrete statements about the probability of imbalances in covariates due

to chance alone. For reasonably large sample sizes, randomization ensures some degree of balance for *all* covariates, including those not measured in the study.

Masking (also known as blinding) of physicians avoids bias that can be introduced by intended or unintended differences in patient care or evaluation of patient outcomes based on the treatment received during the course of the trial. Masking of patients minimizes biases due to the placebo effect.

We recommend you choose beforehand the type of analysis to be used (Bayesian or frequentist). Switching to a more favorable analysis method introduces a bias. It is difficult to justify a switch or account for such a bias in the analysis stage. In some cases, a Bayesian analysis of a new trial may salvage some information obtained in a previous non-Bayesian clinical trial that deviated from the original protocol. The information provided by such a trial may be represented by a prior distribution to be used in a prospective Bayesian clinical trial.

For further information on planning a trial, see FDA's Statistical Guidance for Non-Diagnostic Medical Devices.[5]

5.2 Selecting the relevant endpoints or parameters of interest

Endpoints (also called parameters in this document) are the measures of safety and effectiveness used to support a certain claim. Ideally, endpoints are:

- clinically relevant
- directly observable
- related to the claims for the device
- important to the patient.

For example, an endpoint may be a measure of the size of a treatment effect, or the difference between the effects in the treatment and control groups. The objective of a clinical trial is to gather information from the patients in the trial to make inferences about these unknown parameters.

5.3 Collecting other important information: covariates

Covariates, also known as confounding factors, are characteristics of the study patients that can affect their outcome. There are many statistical techniques (Bayesian and frequentist) to adjust for covariates. Covariate adjustment is especially important in any situation where some degree of covariate balance is not assured through randomization, such as a Bayesian trial in which other trials are used as prior information. If adjustments are not made for differences in the covariate distribution between trials, the analysis can be biased. Covariate

---

[5] http://www.fda.gov/cdrh/ode/ot476.html

adjustment is also often used to reduce variation, which leads to a more powerful analysis.

5.4 Choosing a comparison: controls

To facilitate evaluation of clinical trial results, we recommend you use a comparator, or control group, as a reference. Types of control groups you may use are:

- Concurrent controls
- self controls
- historical controls.

We believe that self controls and historical controls are scientifically less rigorous than concurrent controls because of:

- potential problems with covariate adjustment
- placebo effect
- regression to the mean.

Another way to characterize the type of control is to distinguish between controls that are treated with an effective therapy (active controls) vs. controls that either receive no treatment (inactive controls) or are treated with a sham device (placebo controls). Bayesian methods are especially useful with active controlled trials seeking to demonstrate a new device is not only non-inferior to the active control but is also superior to no treatment or a sham control. A Bayesian trial can investigate this question by using previous studies comparing the active control to the inactive control. Bayesian methods for active control trials are discussed in Gould (1991) and Simon (1999).

5.5 Initial information about the endpoints: prior distributions

The initial uncertainty about the endpoints or parameters of interest, both in the control and treatment groups, is quantified through probability distributions, called *prior distributions*. See Gelman et al. (2004) for background on different types of prior distributions. You should select the appropriate prior information and incorporate it into the analysis correctly. Discussions regarding study design with FDA will include an evaluation of the model to be used to incorporate the prior information into the analysis. Irony & Pennello (2001) discuss prior distributions for trials under regulatory review.

Informative priors

Prior knowledge is described by an *informative prior* distribution. Because using prior information may decrease the sample size in a trial, we recommend you identify as many sources of good prior information as possible when planning a

trial. FDA should agree with your choice of prior distributions. Possible sources of prior information include:

- clinical trials conducted overseas

- patient registries

- clinical data on very similar products

- pilot studies.

We recommend the proposed prior information be submitted as part of the IDE (when an IDE is required). In some cases, existing valid prior information may be unavailable for legal or other reasons (e.g., the data may belong to someone else who is unwilling to allow access).

We recommend you hold a pre-IDE meeting with FDA to come to agreement on what prior information is scientifically valid and how it will be used in the analysis. Quantitative priors (i.e., those based on data from other studies) are the easiest to evaluate. We recommend the prior studies be similar to the current study in as many as possible of the following aspects:

- protocol

- endpoints

- target population

- sites

- physicians

- time frame.

Priors based on expert opinion rather than data are problematic. Approval of a device could be delayed or jeopardized if FDA advisory panel members or other clinical evaluators or do not agree with the opinions used to generate the prior.

To avoid bias, we recommend you avoid using studies that are not representative (e.g., if non-favorable studies are systematically excluded). We recommend you check for selection bias by examining:

- the population of all studies that might have been considered

- the reasons for including or excluding each study.

A Bayesian analysis of a current study of a new device may include prior information from:

- the new device

- the control device

- both devices.

Most commonly, prior information based on historical controls is "borrowed," which can significantly decrease the sample size in a concurrent control group. As a result, a greater proportion of patients can be allocated to the experimental device, increasing the experience with that device at a faster pace. However, if differences between the historical control studies and the current study are large, the use of a historical control as prior information for a concurrent control may not be advantageous.

For example, consider a study with the objective of demonstrating the experimental device is non-inferior to the control regarding rate of complication. If patients in the study are more likely to have complications than those in the historical control studies because they are sicker, for example, indiscriminate use of the historical control data will bias downward the concurrent control complication rate. This bias will make it more difficult to demonstrate the experimental device is non-inferior to the control than if the historical control data were ignored. This phenomenon is less likely to occur if the historical control data are properly calibrated to the current study by, for example, adjusting for important covariates.

If the prior information for a study is based on many more patients than are to be enrolled in the study, the prior distribution may be too informative. In this case, the prior probability that the pivotal study is a success (i.e., demonstrates the proposed claims) will be excessively high. If the prior probability of a successful study is too high, it can be lowered in various ways, including:

- modifying or discounting the prior distribution in some way
- increasing the stringency of the decision rule.

Non-informative priors

Lack of any prior knowledge may be reflected by a *non-informative prior* distribution. Usually, it is easy to define maximum and minimum values for the parameters of interest, and in this case, a possible "non-informative" prior distribution is a uniform distribution ranging from the minimum to the maximum value.

Non-informative priors are reviewed in Kass & Wasserman (1996). Standard and related *improper priors* are explained and used extensively in Box & Tiao (1973). *Reference priors* are extensively discussed in Bernardo & Smith (1993).

5.6 Borrowing strength from other studies: hierarchical models

Bayesian hierarchical modeling is a specific methodology you may use to combine prior results with a current study to obtain estimates of safety and effectiveness parameters. The name hierarchical model derives from the hierarchy in which observations and parameters are structured. The Bayesian analyst refers to this approach as "borrowing strength." For device trials, strength

can be translated into sample size, and the extent of borrowing depends on how closely results from the new study reflect the prior experience.

If results are very similar, the current study can borrow great strength. As current results vary from the previous information, the current study borrows less and less. Very different results borrow no strength at all, or even potentially "borrow negatively". In a regulatory setting, hierarchical models can be very appealing: They reward having good prior information on device performance by lessening the burden in demonstrating safety and effectiveness. At the same time, the approach protects against over-reliance on previous studies that turn out to be overly optimistic for the pivotal study parameter.

An example hierarchical model

Suppose you want to combine information from a treatment registry of an approved device with results from a new study. You may decide to use two hierarchical levels: the patient level and the study level.

The first (patient) level of the hierarchy assumes that (1) within the current study, patients are exchangeable; and (2) within the registry, patients are exchangeable. Registry patients are not, however, exchangeable with patients in the current study, so patient data from the registry and the current study may not be simply pooled.

The second (study) level of the hierarchy applies a model that assumes the success probabilities from the registry and the current study are exchangeable, but the rates may differ (e.g., they may depend on covariates). This assumption is prudent any time you are not sure if patients from the prior experience (i.e., the registry) are directly exchangeable with the patients from the current study. However, the two success probabilities are related in that they are assumed exchangeable. As a result, the registry provides some information about the success probability in the current study, although not as much information as if the patients in the two groups were directly poolable.

Similarity of previous studies to current study

The key clinical question in using hierarchical modeling to borrow strength from previous studies is whether the previous studies are sufficiently similar to the current study in covariates such as:

- devices used
- objectives
- endpoints
- protocol
- patient population
- sites

- physician training

- patient management

- time frame.

Statistical adjustments for certain differences in covariates such as demographic and prognostic variables may be appropriate, using patient-level data. Generally, proper calibration of your study depends on using the same covariate information at the patient level as in previous studies.

Calibration based only on covariate *summaries* (such as from the literature) may be inadequate because the relationship of the covariate level to the outcome can be determined in the current study but not in the previous studies. This forces the untestable assumption that covariate effects in your study and previous studies are the same; that is, that study and covariate effects do not interact.

When you use more than one study as prior information in a hierarchical model, the prior distribution can be very informative. As discussed previously, if the prior probability of a successful trial is too high, we recommend the study design and analysis plan be modified.

You may also use hierarchical models to combine data across centers in a multi-center trial. For an example, see the Summary of Safety and Effectiveness for PMA P980048, BAK/Cervical Interbody Fusion System by Sulzer Spine-Tech.[6]

Outcomes for devices can vary substantially by site due to such differences as:

- physician training

- technique

- experience with the device

- patient management

- patient population.

A hierarchical model on centers assumes that the parameters of interest vary from center to center but are related via exchangeability. This kind of model adjusts for center-to-center variability when estimating parameters across studies.

Non-technical discussion of hierarchical models and technical details on their implementation appear in Gelman et al. (2004). Other, more complex approaches are described in Ibrahim & Chen (2000) and Dey et al., 1998.

---

[6] www.fda.gov/cdrh/pdf/p980048b.pdf.

5.7 Determining the sample size

The sample size in a clinical trial depends on:

- prior information

- mathematical model used in analysis

- distributions of parameters in the analytic model

- specific decision criteria

- variability of the sample.

If the population of patients is highly variable, the sample size increases. If there is no variability (i.e., everyone in the population has the same value for the measurement of interest), a single observation is sufficient. The purpose of sizing a trial is to gather enough information to make a decision while not wasting resources or putting patients at unnecessary risk.

In traditional frequentist clinical trial design, the sample size is determined in advance. Instead of specifying a particular sample size, the Bayesian approach (and some modern frequentist methods) may specify a particular criterion to stop the trial. Appropriate stopping criteria may be based on a specific amount of information about the parameter (e.g., a sufficiently narrow credible interval, defined in **Section 6: Analyzing a Bayesian Clinical Trial**) or an appropriately high probability for a pre-specified hypothesis.

At any point before or during a Bayesian clinical trial, you can obtain the posterior distribution for the sample size. Therefore, at any point in the trial, you can compute the expected additional number of observations needed to meet the stopping criterion. In other words, the sample size distribution is continuously updated as the trial goes on. Because the sample size is not explicitly part of the stopping criterion, the trial can be ended at the precise point where enough information has been gathered to answer the important questions.

Special considerations when sizing a Bayesian trial

When sizing a Bayesian trial, FDA recommends you decide in advance on the minimum sample size according to safety and effectiveness endpoints because safety endpoints may lead to a larger sample size. FDA also recommends you include a minimum level of information from the current trial to enable verification of model assumptions and appropriateness of prior information used. This practice also enables the clinical community to gain experience with the device.

When hierarchical models are used, we recommend you provide a minimum sample size for determining the amount of information that will be "borrowed" from other studies.

We recommend the maximum sample size be defined according to economical, ethical, and regulatory considerations.

Various approaches to sizing a Bayesian trial are described in Inoue et al. (2005), Katsis & Toman (1999), Rubin & Stern (1998), Lindley (1997), and Joseph et al. (1995a,b).

5.8 Assessing the operating characteristics of a Bayesian design

Because of the inherent flexibility in the design of a Bayesian clinical trial, a thorough evaluation of the operating characteristics should be part of the trial design. This includes evaluation of:

- probability of erroneously approving an ineffective or unsafe device (type I error)

- probability of erroneously disapproving a safe and effective device (type II error)

- power (the converse of type II error: the probability of appropriately approving a safe and effective device)

- sample size distribution (and expected sample size)

- prior probability of claims for the device

- if applicable, probability of stopping at each interim look.

A more thorough discussion appears in the **Appendix**.

6.   Analyzing a Bayesian Clinical Trial

6.1 Summaries of the posterior distribution

The results, conclusions, and interpretation of a Bayesian analysis all rely on the posterior distribution, which contains all information from the prior distribution, combined with the results from the trial via the likelihood. Consequently, results and conclusions for a Bayesian trial are based only on the posterior distribution. FDA recommends you summarize the posterior distribution with a few numbers (e.g., posterior mean and standard deviation), especially when there are numerous endpoints to consider. FDA also recommends you include graphic representations of the appropriate distributions.

6.2 Hypothesis testing

Statistical inference may include hypothesis testing or interval estimation, or both. FDA often bases approval on demonstrating claims via hypothesis tests. For Bayesian hypothesis testing, you may use the posterior distribution to calculate the probability that a particular hypothesis, either null or alternative, is true, given the observed data.

Although probabilities of type I and II errors are frequentist notions, and not formally part of Bayesian hypothesis testing, Bayesian hypothesis tests often enjoy good frequentist properties. FDA recommends you provide the type I and II error rates of your proposed hypothesis test.

6.3 Interval estimation

Bayesian interval estimates are based on the posterior distribution and are called *credible intervals*. If the posterior probability that an endpoint lies in an interval is 0.95, then this interval is called a 95 percent *credible interval*.

For construction of credible intervals, see Chen & Shao (1999) and Irony (1992). Other types of Bayesian statistical intervals include *highest posterior density* (HPD) intervals (Lee, 1997) and *central posterior intervals*.

6.4 Predictive probabilities

You may use predictive probabilities, a special type of posterior probabilities, in a number of ways:

Deciding when to stop a trial

If it is part of the clinical trial plan, you may use a predictive probability at an interim point as the rule for stopping your trial. If the predictive probability that the trial will be successful is sufficiently high (based on results thus far), you may be able to stop the trial and declare success. If the predictive probability that the trial will be successful is small, you may stop the trial for futility and cut losses.

Exchangeability is a key issue here: these predictions are reasonable only if you can assume the patients who have not been observed are exchangeable with the patients who have. This assumption is difficult to formally evaluate but may be more credible in some instances (e.g., administrative censoring) than others (e.g., high patient drop-out).

Predicting outcomes for future patients

You may also calculate the predictive probability of the outcome of a future patient, given the observed outcomes of the patients in a clinical trial, provided the current patient is exchangeable with the patients in the trial. In fact, that probability answers the following questions:

- Given the results of the clinical trial, what is the probability that a new patient receiving the experimental treatment will be successful?

- What would that probability be if the patient were treated in the control group?

After device approval, these probabilities could be very useful in helping physicians and patients make decisions regarding treatment options.

Predicting (imputing) missing data

You may use predictive probabilities to predict (or *impute*) missing data, and trial results can be adjusted accordingly. There are also frequentist methods for missing data imputation.

Regardless of the method, the adjustment depends on the assumption that patients with missing outcomes follow the same statistical model as patients with observed outcomes. This means the missing patients are exchangeable with the non-missing patients, or that data are *missing at random*. If this assumption is questionable, FDA recommends you conduct a sensitivity analysis using the prediction model. For examples of missing data adjustments and sensitivity analysis, see the Summary and Safety Effectiveness for PMA P980048, BAK/Cervical Interbody Fusion System, by Sulzer Spine-Tech.[7]

Predicting a clinical outcome from a surrogate

If patients have two different measurements at earlier and later follow-up visits, you may make predictions for the later follow-up visit (even before the follow-up time has elapsed). Basing predictions on measures at the earlier visit requires that:

- some patients have results from both follow-up visits

- there is sufficiently high correlation between the early and the later measurement.

In this example, the outcome at the first time point is being used as a surrogate for the outcome at the second. This type of prediction was used to justify stopping the clinical trial of the INTERFIX Intervertebral Body Fusion Device.[8]

The surrogate may also be a different outcome; for example, for breast implants, rupture may be predictive of an adverse health outcome later.

Surrogate endpoints for predictive distributions should be validated. However, validation of a surrogate endpoint is a complex scientific and statistical issue that is outside the scope of this document.

Model checking

FDA recommends you verify all assumptions important to your analysis. For example, an analysis of a contraceptive device might assume the monthly pregnancy rate is constant across the first year of use. To assess this assumption, the observed month-specific rates may be compared to their predictive distribution. You may summarize this comparison using a *Bayesian p-value.* For more information, refer to Gelman et al. (1996; 2004).

---

[7] www.fda.gov/cdrh/pdf/p980048b.pdf.

[8] See Summary of Safety and Effectiveness for PMA P970015 at www.fda.gov/cdrh/pdf/p970015b.pdf.

You may also assess model checking and fit by Bayesian deviance measures as described in Spiegelhalter et al. (2002). Any predictive analysis assumes that patients for whom outcomes are being predicted are exchangeable with patients on whom the prediction is based. This assumption may not be valid in some cases. For example, patients enrolled later in a trial may have a different success rate with the device than those enrolled earlier if:

- physicians have to overcome a learning curve in using the device
- physicians form an opinion on whom the device treats best, and then enroll more or less favorable patients in the trial
- an effective adjunctive therapy becomes available to the patients in the trial
- an alternative treatment becomes available during the course of the trial, altering the characteristics of the patients who choose to enroll in the trial.

6.5 Interim analyses

There is more than one method for analyzing interim results in a Bayesian trial. FDA recommends you specify the method in the trial design and ensure FDA agrees in advance of the trial. FDA may ask you to calculate the probability of a type I error through simulations before accepting a method. Although this is a frequentist calculation, it can help in evaluating the application of a method to a trial.

The following describes three specific Bayesian interim analysis methods:

Applying posterior probability

One method stops the trial early if the posterior probability of a hypothesis at the interim look is large enough. In other words, the same Bayesian hypothesis test is repeated during the course of the trial.

Applying predictive distribution

Another method calculates at interim stages the probability that the hypothesis test will be successful. This method uses the Bayesian predictive distribution for patients yet to be measured. If the predictive probability of success is sufficiently high, the trial may stop early. If the predictive probability is very low, the trial may stop early for futility. This method was used in the submission of the INTERFIX Intervertebral Body Fusion Device.[9]

Applying formal decision analysis

A decision analysis method considers the cost of decision errors and experimentation in deciding whether to stop early. Carlin, Kadane, & Gelfand (1998) propose a method to approximate a decision analysis approach.

---

[9] See the Summary of Safety and Effectiveness for PMA P970015 at http://www.fda.gov/cdrh/pdf/p970015b.pdf).

7. Post-Market Surveillance

FDA believes the Bayesian approach is well suited for surveillance purposes. The key concept: "Today's posterior is tomorrow's prior" allows you to use the posterior distribution from a pre-market study to serve as a prior distribution for surveillance purposes, to the extent that data from the clinical study reflect how the device is used after approval. In other words, you may readily update information provided by a pre-market clinical trial with post-market data via Bayes' theorem *if you can justify exchangeability between pre- and post-market data.* You may continue to update post-market information via Bayes' theorem as more data are gathered. You may also use Bayesian models to mine large databases of post-market medical reports.

DuMouchel (1999) discusses Bayesian models for analyzing a very large frequency table that cross-classifies adverse events by type of drug used. DuMouchel uses a hierarchical model to smooth estimates of relative frequencies of adverse events associated with drugs to reduce the number of falsely significant associations. It is unclear at this time if this approach is as useful for medical device reports as it is with drug reports.

8. References

Barlow, R. E., Irony, T. Z., & Shor, S. W. W. (1989). *Informative sampling methods: The influence of experimental design on decision, in influence diagrams, beliefs nets and decision analysis*. Oliver and Smith (Eds.), John Wiley & Sons.

Berger, J. O., & Berry, D. A. (1988). The relevance of stopping rules in statistical inference. *Statistics decision theory and related topics, IV 1*. S. S. Gupta and J. O. Berger (Eds.). Berlin: Springer, 29–72 (with discussion).

Berger, J. O., & Wolpert, R. L. (1988). *The likelihood principle*, Second Ed. CA: Hayward: IMS.

Bernardo & Smith. (1993). *Bayesian theory*, John Wiley & Sons.

Berry, D. A. (1996). *Statistics, a Bayesian perspective*. Duxbury Press.

Berry, D. A. (1997). *Using a Bayesian approach in medical device development*. Technical report available from Division of Biostatistics, Center for Devices and Radiological Health, FDA.

Berry, D. A., & Stangl, D. K. (Eds). (1996). *Bayesian biostatistics*. New York: Marcel Dekker.

Bland, J. M., & Altman, D. G. (1998). Bayesians and frequentists. *BMJ, 317(24)*, 1151.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. MA: Reading: Addison Wesley.

Breslow, N. (1990). Biostatistics and Bayes. *Stat Sci 5(3)*, 269–284.

Brophy, J. M., & Joseph, L. (1995). Placing trials in context using Bayesian analysis: GUSTO, revisited by Reverend Bayes. *JAMA, 273*, 871–875.

Carlin, B. P., Kadane, J. B., & Gelfand, A. E. (1998). Approaches for optimal sequential decision analysis in clinical trials. *Biometrics, 54*, 964–975.

Chen, M. H., & Shao, Q. M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computation and Graphical Statistics, Vol. 8, N.1*, 69–92.

Congdon, P. (2003). *Applied Bayesian modeling*, Wiley.

De Groot, M. H. (1986). *Probability and statistics*, Addison Wesley.

Dey, D., Muller, P., & Sinha, D. (Eds.) (1998). *Practical nonparametric and semiparameteric Bayesian statistics.* New York: Springer-Verlag.

DuMouchel, W. (1999). Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *American Statistician, 53* (3), 177–202.

Gamerman, D. (1997). *Markiv chain Monte Carlo: Stochastic simulation for Bayesian inference*. Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. Second Ed., London: Chapman and Hall.

Gelman, A., Meng, X., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6*, 733–760, discussion 760–807.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain monte carlo in practice*. London: Chapman & Hall.

Goodman, S. (1999a). Toward evidence-based medical statistics, 1: the p value fallacy. *Annals of Internal Medicine, 130*, 995–1004.

Goodman, S. (1999b). Toward evidence-based medical statistics, 2: the Bayes factor. *Annals of Internal Medicine, 130,* 1005–1013.

Gould, A. L. (1991). Using prior findings to augment active-controlled trials and trials with small placebo groups. *Drug Information Journal, 25*, 369–380.

Hively, W. (1996, May). The mathematics of making up your mind. *Discover*, 90–97.

Ibrahim, J. G., & Chen, M. H. (2000). Power distributions for regression models. *Statistical Science*, 46–60.

Inoue, L. Y. T., Berry, D.A., & Parmigiani, G. (2005). Relationship between Bayesian and frequentist sample size determination. *The American Statistician, 59*, 79–87.

Irony, T. Z. (1992). Bayesian estimation for discrete distributions. *Journal of Applied Statistics, 19*, 533–549.

Irony, T. Z. (1993). Information in Sampling Rules. *Journal of Statistical Planning and Inference, 36*, 27–38.

Irony, T. Z., & Pennello, G. A. (2001). Choosing an appropriate prior for Bayesian medical device trials in the regulatory setting. In *American Statistical Association 2001 Proceedings of the Biopharmaceutical Section*. VA: Alexandria: American Statistical Association.

Johnson, V. E., and Albert, J. H. (1999). *Ordinal data modeling*. New York: Springer-Verlag.

Joseph, L., Wolfson, D. B., & Berger, R., du. (1995a). Sample size calculations for binomial proportions via highest posterior density intervals. *The Statistician: Journal of the Institute of Statisticians, 44*, 143–154

Joseph, L., Wolfson, D. B., & Berger, R., du. (1995b). Some comments on Bayesian sample size determination. *The Statistician: Journal of the Institute of Statisticians, 44*, 167–171

Kadane, J. B. (1995). Prime time for Bayes. *Controlled Clinical Trials, 16*, 313–318.

Kadane, J. B. (1996). *Bayesian methods and ethics in a clinical trial design*. John Wiley & Sons.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of American Statistical Association, 91 (435)*, 1343–1370.

Katsis, A., & Toman, B. (1999). Bayesian sample size calculations for binomial experiments. *Journal of Statistical Planning and Inference, 81*, 349–362

Lee, P. M. (1997). *Bayesian statistics: an introduction*. New York: John Wiley & Sons.

Lewis, R. J., & Wears, R. L. (1993). An introduction to the Bayesian analysis of clinical trials. *Ann. Emerg. Med.*, *22(8)*, 1328–1336.

Lilford, R. J., & Braunholtz, D. (1996). The statistical basis of public policy: A paradigm shift is overdue. *BMJ, 313*, 603–607.

Lindley, D. V. (1985). *Making decisions*. John Wiley & Sons.

Lindley, D. V. (1997). The choice of sample size. *The Statistician, 46, N. 2*, 129–138.

Malakoff, D. (1999, Nov 19). Bayes offers a "new" way to make sense of numbers. *Science, 286*, 1460–1464.

Rubin, D. B., & Stern, H. S. (1998). Sample size determination using posterior predictive distributions. *Sankhyā, Series B, 60*, 161–175

Simon, R. (1999). Bayesian design and analysis of active control clinical trials. *Biometrics, 55*, 484–487.

Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. New York: Wiley.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64, 583-616

Spiegelhalter, D. J., Freedman, L. S., and Parmar, M. K. B. (1994). Bayesian approaches to randomized trials. *Journal of Royal Statistical Society, Series A, 157*, 356–416.

Spiegelhalter, D. J., Myles, J. P., Jones, D. R., & Abrams, K. R. (2000). Bayesian method in health technology assessment: A review. *Health Technology Assessment, 4*, 38.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (1996). *BUGS: Bayesian inference using Gibbs sampling*, version 0.5 (version ii). MRC Biostatistics Unit. Retrieved February, 2002, from http://www.mrc-bsu.cam.ac.uk.

Stangl, D.K., & Berry, D.A. (Eds.) (1998). Bayesian statistics in medicine: Where are we and where should we be going? *Sankhya Ser B, 60*, 176–195.

Stern, H. S. (1998). A primer on the Bayesian approach to statistical inference. *Stats, 23*, 3–9.

Tanner, M.A. (1996). *Tools for statistical inference*. New York: Springer-Verlag.

9. Appendix

9.1 Suggested Information to Submit to FDA

In addition to the standard clinical trial protocol, FDA believes there are statistical issues unique to Bayesian trial designs requiring additional information in your submission. The following suggestions (not an exhaustive listing) will facilitate a smoother review process and serve as a starting point when writing your protocol. Not all points apply to all studies.

Prior information

FDA recommends you indicate all prior information you will use, including:

- relevant studies

- expert opinions

- assumptions.

Criterion for success

FDA recommends you provide a criterion for success of your study (related to safety and effectiveness).

Method for choosing sample size

FDA also recommends you state your method for choosing a sample size. To assist in choosing an appropriate sample size for the trial, you may simulate data assuming a range of different true parameter values and different sample sizes. For each simulated data set, we recommend you determine the posterior distribution of the parameter. This posterior distribution is used in calculating the posterior probability of the study claim for the chosen sample size and true parameter value.

Frequentist power tables

FDA recommends you provide frequentist power tables of the probability of satisfying the study claim, given various "true" parameter values (e.g., event rates) and various sample sizes for the new trial. This table provides probabilities of observing data that allow the study claim to be met, given the indicated true parameter value. This table will also provide an estimate of the type I error rate in the case where the true parameter values are consistent with the null hypothesis.

For example, for an adverse event rate, you can generate power tables by:

- choosing a true adverse event rate

- choosing a sample size, and finding the largest count of events that satisfies the study claim

- calculating the probability of observing a count less than or equal to this largest count (e.g., assuming a binomial distribution with the true event rate).

This probability is the power of the criterion for the chosen sample size and true event rate.

If the study design is complex it may be necessary to use simulation to compute power. Some suggestions on simulation are outlined in **Appendix 9.4: Simulations to Obtain Operating Characteristics**.

Interim looks (monitoring)

If your design contains interim looks (monitoring), we recommend you also simulate those.

Predictive probability

FDA recommends you evaluate the prior predictive probability of your study claim. This is the predictive probability of the study claim prior to seeing any new data, and it should not be too high. In particular, we recommend the prior predictive probability not be as high as the simulated posterior probability of the claim identified in the sample size section above.

FDA makes this recommendation to ensure the prior information does not overwhelm the current data, potentially creating a situation where unfavorable results from the proposed study get masked by favorable prior results. In an evaluation of the prior probability of the claim, FDA will balance the informativeness of the prior against the gain in efficiency from using prior information as opposed to using noninformative priors.

To calculate this prior probability, you can simulate data using only the prior information. For example, if you are using a computer program that performs Markov Chain Monte Carlo (MCMC) simulation, you can leave blank the slot where you normally insert current data and have the program simulate these values instead. Simulations done in this manner provide the prior probability of the study claim.

The prior predictive probability of the study claim can be altered by inflating or deflating the variance between studies. Inflating the variance by modifying the parameters of its prior distribution might be difficult if there are few studies, resulting in an unstable variance estimate. FDA recommends first fixing the parameters of the prior distribution, and then experimenting with adding a constant to the study variance until the prior predictive probability of the claim is relatively low.

Program code

FDA recommends you submit the electronic program code you use to conduct simulations and any prior data with the IDE submission. We also recommend you

include an electronic copy of the data from the study and the computer code used in the analysis with the PMA submission.

Additional items

A useful summary that can be computed for any simulations using posterior variance information is the effective sample size in the new trial. That is,

Effective sample size (ESS) is given by:

$$ESS = n * V_1/V_2,$$

Where n = the sample size in the new trial
V_1 = the variance of the parameter of interest without borrowing
V_2 = the variance of the parameter of interest with borrowing.

Then, ESS – sample size from new trial = number of patients "borrowed" from the previous trial. This information can be useful for deciding how much efficiency you are gaining from using the prior information.

9.2 Model Selection

Some statistical analysis plans allow for comparison of several possible models of the data and parameters before a final model is chosen for analysis. For example, in a Bayesian analysis of a study outcome that borrows strength from other studies, the effects of a factor on the outcome might vary from study to study.

One method of comparing two models tests the null hypothesis that one model is true against the alternative that the other model is true. The result of such a test depends on the posterior probability of the alternative model, or the *posterior odds* for the alternative model. *Posterior odds* refer to the ratio of the posterior probability of the alternative model to the posterior probability of the null model.

9.3 Calculations

Almost all quantities of interest in a Bayesian analysis involve the calculation of a mathematical integral. All the following are expressed as an integral involving the posterior distribution:

- the posterior mean
- the posterior standard deviation
- the posterior probability of a hypothesis.

The following are some numerical integration techniques used to compute these integrals:

- Gaussian quadrature

- posterior distribution sampling

- Laplace approximation

- importance sampling

- *Markov Chain Monte Carlo* (MCMC) techniques (Gamerman 1997).

MCMC techniques are probably the most commonly used; the most used MCMC technique is the *Gibbs sampler*. The *Metropolis-Hastings algorithm* is a generalization that can be used in cases where the Gibbs sampler fails.

MCMC techniques are popular because in many analyses, the posterior distribution is too complex to write down, and therefore traditional numerical integration techniques like quadrature and Laplace approximation cannot be carried out. The Gibbs sampler draws samples from other, known distributions to create a (Markov) chain of values. A Markov chain is a set of samples where the value at each point depends only on the immediately preceding sample. Eventually, as the chain *converges*, the values sampled begin to resemble draws from the posterior distribution. The draws from the Markov chain can then be used to approximate the posterior distribution and compute the integrals.

Tanner (1996) provides a survey of computational techniques used in statistics, including numerical integration and MCMC techniques. Gilks et al. (1996) explains MCMC techniques and their application to a variety of scientific problems. Discussion of MCMC and other techniques also appear in Gelman et al. (2004).

When MCMC techniques are used, FDA recommends you check that the chain of values generated has indeed converged at some point so that subsequent draws are from the posterior distribution. If the chain has not converged, we recommend you sample more values.

Various techniques have been developed to diagnose nonconvergence. You may refer to diagnostic techniques discussed in Gelman et al. (2004), Gilks et al. (1996), Tanner (1996), and the manual for CODA (Convergence Diagnosis and Output Analysis), a set of SPlus functions that process the output from the program BUGS (Bayesian inference using Gibbs sampling).[10]

Convergence difficulties

Depending on how the Bayesian model is parameterized, the Markov chain might converge very slowly. Alternative parameterizations can help to speed up convergence. One possible explanation for a chain that does not seem to converge is that an improper prior distribution was used (see **Section 6: Analyzing a Bayesian Clinical Trial**). Thus, the chain does not have a (posterior) distribution to converge to. When improper prior distributions are

---

[10] Both programs may be downloaded from the Medical Research Center, Cambridge, at
http://www.mrc-bsu.cam.ac.uk

used, you should check that the posterior distribution is proper. Convergence difficulties can also occur when the prior distribution is nearly improper.

Data augmentation

The technique of *data augmentation* introduces auxiliary variables into a model to facilitate computation. The use of auxiliary variables can also aid in the interpretation of your analysis. For example, *latent variables* are now commonly introduced in analyses of ordinal outcomes (i.e., outcomes with only a few possible values that are ordered). Examples of such outcomes include answers to multiple-choice questions for a quality of life questionnaire. Johnson & Albert (1999) discuss Bayesian analysis of ordinal outcomes using latent variables. Tanner (1996) discusses data augmentation as a general technique.

Electronic submission of calculations

FDA routinely checks the calculations for a Bayesian analysis (e.g., for convergence of the Markov chain when using MCMC techniques). We recommend you submit data and any programs used for calculations to FDA electronically.

9.4 Simulations to Obtain Operating Characteristics

FDA recommends you provide trial simulations at the planning (or IDE) stage. This will facilitate FDA's assessment of the operating characteristics of the Bayesian trial; specifically, the type I and type II error rates. We recommend your simulated trials mimic the proposed trial by considering the same:

- prior information

- sample size

- interim analyses

- possible modifications of the trial in midcourse.

You can assess the type I error rate from simulated trials where the parameters are fixed at the borderline values for which the device should not be approved. The proportions of successful trials in these simulations provide estimates of the type I error rate. FDA recommends that several likely scenarios be simulated and that the expected sample size and estimated type I error be provided in each case.

You can assess power and the type II error rate from simulated trials where the parameters are fixed at plausible values for which the device should be approved. The proportions of unsuccessful trials in these simulations provide estimates of the type II error rate. The complement estimates the power provided by the experimental design. FDA recommends several likely scenarios be simulated and that the expected sample size and estimated type II error rate and power be provided in each case.

If FDA considers the type I error rate of a Bayesian experimental design to be too large, we recommend you modify the design to reduce that rate. Determination of "too large" is specific to a submission since some sources of type I error inflation (e.g., large amounts of valid prior information) may be more acceptable than others (e.g., inappropriate choice of prior studies or inappropriate criteria for study success).

There are several options for decreasing the type I error rate:

- increase the posterior (or predictive) probability that defines a successful trial (e.g., 97 percent, 98 percent, 99 percent)

- increase the number of patients before the first interim analysis

- discount the prior information*

- reduce the number of interim analyses

- increase the maximum sample size (again, to reduce the influence of the prior)

- any combination of the above options.

To discount prior information, FDA recommends (1) you increase the number of patients before the first interim analysis until the type I error rate reduces to an acceptable level; or (2) you iteratively increase the variance of the prior distribution by trial and error until the type I error rate reduces sufficiently.

In case the experimental design is modified, we recommend you carry out a new set of simulations to evaluate the new design.