U.S. Food and Drug Administration
Protecting and Promoting *Your* Health

A to Z Index | Follow FDA | FDA Voice Blog

Search the FDA Archive

SEARCH

Home | Food | Drugs | Medical Devices | Radiation-Emitting Products | Vaccines, Blood & Biologics | Animal & Veterinary | Cosmetics | Tobacco Products

Medical Devices

Device Advice: Comprehensive Regulatory Assistance

- **Guidance Documents (Medical Devices and Radiation-Emitting Products)**
- Cross-Center Guidance Documents
- OC Guidance
- OCD Guidance
- OCER Guidance
- ODE Guidance 2010 - 2013
- ODE Guidance 1998 - 2009
- ODE Guidance 1976 - 1997
- OIR Guidance
- OSB Guidance
- OSEL Guidance
- Radiation-Emitting Products Guidance

Statistical Guidance for Clinical Trials of Non Diagnostic Medical Devices

The Division of Biostatistics
Office of Surveillance and Biometrics
Center for Devices and Radiological Health
U.S. Food and Drug Administration
January 1996

PREFACE

The Office of Surveillance and Biometrics (OSB) of FDA's Center for Devices and Radiological Health (CDRH) was established in July, 1993 to consolidate and focus CDRH postmarket surveillance programs. A major portion of the OSB mandate is to employ significant clinical, technical and scientific skills to identify and resolve public health problems. Towards this goal, the Office provides statistical, epidemiological, and biometrics services in support of the major operating programs of the Center. Reviewing premarket approval applications (PMA) to assure the safety and effectiveness of marketed medical devices is a particularly vital part of that support.

The controlled clinical trial is the primary vehicle used to advance new medical device technology through the PMA approval process. These investigations provide the basis of valid scientific evidence that FDA requires to evaluate new medical device technology. As such, it is critical that a sponsor correctly plan, conduct and analyze these trials.

The following guidance has been prepared by OSB's Division of Biostatistics with help from the Center's Office of Device Evaluation (ODE), academia, and the medical device industry. The primary purpose of this document is to assist medical device manufacturers in advancing their product through the premarket approval process. The guidance is based on expertise and experience in reviewing data from medical device clinical trials, and a major FDA workshop on Medical Device Clinical Trials held in September, 1993.

It is our hope that this document, along with the additional information and references that have been cited will help manufacturers save time, money, and human resources in the planning, conduct, and analysis of medical device clinical trials.

```
                    Larry G. Kessler, Sc.D.

                    Director,
                    Office of Surveillance and Biometrics
```

ii

Your comments and suggestions are welcome. Please address any correspondence regarding this guidance to:

Division of Biostatistics
Office of Surveillance and Biometrics
FDA/CDRH

10903 New Hampshire Avenue

Silver Spring, MD 20993

Tel: 301-796-5750
FAX: 301-847-8123
(Note: address and phone updated Nov. 2009)

iii

STATISTICAL GUIDANCE for CLINICAL TRIALS of NON-DIAGNOSTIC MEDICAL DEVICES

Table of Contents

I. INTRODUCTION

The collection and evaluation of sound clinical data are the basis of the approval process for many medical devices. The determination of the need for clinical data is made by the Center for Devices and Radiological Health (CDRH) based on requirements described elsewhere (DHHS, 1987; DHHS, 1990; DHHS, 1992). This guidance document assumes that the need for a clinical trial has been determined and describes procedures to assure that data from such studies can be interpreted in both a scientific and regulatory manner by the Food and Drug Administration (FDA, or the Agency).

This document is consistent with previously published clinical study guidance (DHHS, 1987; DHHS, 1990; DHHS, 1992) but provides a more comprehensive treatment of the clinical trial process from a statistical perspective. An accompanying guidance covers clinical aspects of device trials. This guidance describes how a sponsor should proceed to properly design and conduct a clinical trial in order to provide a meaningful evaluation and interpretation of clinical data in support of medical device Premarket Approval Applications (PMA).

The development of this clinical trial guidance resulted from a concern about the quality of clinical trials submitted to the Agency in support of medical device applications. This concern applied to many critical elements of clinical trial design, conduct, and analysis and was supported by the findings of the Committee for Clinical Review chaired by Dr. Robert Temple, Ann Witt served as co-chair, whose report became publicly available in March 1993. The CDRH recognized the need for a separate guidance document to address these concerns, and to clearly document those elements needed for a well designed, conducted, and analyzed device clinical trial.

The purpose of this document is to discuss important clinical trial issues and not to describe the contents of a medical device submission. It provides an explanation of each particular trial element and discusses why it should be incorporated into the clinical trial and what problems may be encountered if it is not included in the investigation.

The goal of a good clinical trial is to provide the most objective evaluation of the safety and effectiveness of the medical device based on its intended claims. Anything in the design, conduct, and analysis which impairs that objective assessment lessens the ability of the Agency staff and their advisory committees to make an informed decision concerning a "reasonable assurance of safety and effectiveness" for a device.

The cost of any decision in the design, conduct, and analysis of device clinical trials which may interfere with this objectivity must be weighed against the cost of delays or disapprovals in the review process encountered as a result of those decisions.

While this guidance serves as a road map and provides the key elements of good clinical trial design, conduct, and analysis, it is by no means exhaustive. Numerous books, only a few of which have been referenced here, exist on the topic of clinical trial design and the scientific literature is rich with papers on the topic.

II. VALID SCIENTIFIC EVIDENCE

While the manufacturer may submit any evidence to convince the Agency of the safety and effectiveness of its device, the Agency may rely only on valid scientific evidence as defined in the PMA regulation section entitled, "Determination of Safety and Effectiveness" (21 CFR 860.7). A thorough reading of that section is strongly recommended. It should be noted that while the Agency does not prescribe specific statistical analyses for given devices and/or situations, all statistical analyses used in an investigation should be appropriate to the analytical purpose, and thoroughly documented.

"Valid scientific evidence is evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, from which it can fairly and responsibly be concluded by qualified experts that there is a reasonable assurance of safety and effectiveness of a device under its conditions of use "(GPO, 1993).

The regulation further states, "The valid scientific evidence used to determine the effectiveness of a device shall consist principally of well-controlled investigations as defined in paragraph (f) of this section (860.7) unless the Commissioner authorizes the reliance upon other valid scientific evidence which the Commissioner has determined is sufficient evidence from which to determine the effectiveness of the device even in the absence of well-controlled investigations" (GPO, 1993). From these passages it is clear the Agency intends to require well-controlled clinical trials to provide the required reasonable assurance of safety and effectiveness for medical devices.

DEFINITION OF CLINICAL TRIAL

"A clinical trial is defined as a prospective study comparing the effect and value of intervention(s) against a control in human subjects" (Friedman et al., 1985). In this definition, intervention is used in the broadest sense to include "prophylactic, diagnostic, or therapeutic agents, device regimens, procedures etc." (Friedman et al, 1985).

Additional insight into clinical trials is given in a definition by Hill (1967), "The clinical trial is a carefully, and ethically, designed experiment with the aim of answering some precisely framed question." So, the clinical trial is an ethical experiment in humans and as such requires informed consent and Institutional Review Board (IRB) approval. Such considerations require careful deliberation in the design and conduct of trials. (This will be further addressed in the accompanying section on clinical aspects of trials.)

III. DESIGN OF THE CLINICAL TRIAL

A good clinical trial design controls or minimizes known or suspected sources of bias and other errors so that clinical device performance may be assessed clearly and objectively. Error is the result of our inability to accurately measure a variable. Bias results when any characteristic of the investigator, study population, or study conduct interferes in a systematic way with the ability to measure a variable accurately.

A. The Trial Objective (The Research Question)

An effective and efficient design of a clinical investigation cannot be accomplished without a clear and concise objective. Usually the study objective is posed as a research question, involving the medical claims for the device. This research question should be formulated with extreme care and specificity. A question such as "Is my device safe and effective?" is far too general to be meaningful.

The question must be refined to effectively evaluate a particular type of intervention. What is the proper way to evaluate effectiveness in the target condition and population? What are the unique safety concerns of the device intervention? Is the device as effective or more effective than another intervention? If so, is it as safe or safer? Is the evaluation of safety and effectiveness limited to a particular subgroup of patients? What is the best clinical measure of safety and effectiveness?

The attempt to answer these and similar questions will provide an essential focus to the trial and should provide the basis for labeling indications. For example, if a new device has been developed to treat a progressive, degenerative ophthalmic disorder for which there currently exists an alternative therapy using an approved device, how should effectiveness be determined? Does the new device slow or halt degeneration? If so, does it restore functions that had previously been lost? Does it reduce pain or discomfort? Is it to be compared with the approved device and is it thought to be as good as or better than the old device for some purpose? Does it have fewer adverse reactions?

One can see that asking these questions will lead not only to a focused study objective, but also will require the sponsor to consider a number of other issues, such as a suitable endpoint or outcome variable, a control population, the type of hypothesis that might be tested and others.

These issues must be addressed prior to protocol development, because one must determine if the stated research question can be adequately addressed by designing a sound clinical trial. That is, can we obtain specific and objective answer(s) to the research question(s) by the collection, analysis, and interpretation of data from the clinical trial.

B. Pilot or Feasibility Study

If a sponsor cannot answer the key questions necessary to focus the trial because of insufficient experience with the device in human populations, then the sponsor should design a limited human study to gather essential information. The purpose of this limited study (frequently called a pilot or feasibility study) is to identify possible medical claims for the device, monitor potential study variables for a suitable outcome variable, test study procedures, refine the prototype device, and determine the precision of those potential response variables. It may also allow a limited evaluation of factors that may introduce bias. A protocol for a pilot study should be submitted to the Agency, usually as an Investigational Device Exemption (IDE) application.

Pilot studies are often used to field test the device. That is, the sponsor has a good idea of the utility of the device and may need a limited trial to test a theory or new technique, but the pilot study should not be too broad, i.e., a "fishing expedition". A number of issues related to the clinical trial can be refined including device use, patient processing and monitoring, data gathering and validation, and physician capabilities and concerns. Care should be taken to refine the measurements of critical variables, including potential outcome variables and influencing variables including potential sources of bias. However, it should be noted that in situations where long-term endpoints are needed, these are usually not part of the pilot study.

Pilot studies allow for limited hypothesis testing and are the ideal place for exploratory data analyses, i.e., looking for meaningful relationships between the device and outcome variables since exploratory methods will often yield research questions that can be evaluated during the clinical trial.

C. Identification and Selection of Variables

The observations in a clinical study involve two types of variables: outcome variables and influencing variables. Outcome variables define and answer the research question and should have direct impact on the claims for the device. These variables, also known as response, endpoint, or dependent variables, should be directly observable, objectively determined measures subject to minimal bias and error. They should be directly related to biological effects of the clinical condition and this relationship itself may need validation. For example, it may be necessary to perform preliminary laboratory, animal, or limited human studies to determine that reducing a particular blood value is in fact clinically meaningful before attempting to study a device that claims to be safe and effective in decreasing this value to specific levels.

Influencing variables, also known as baseline variables, prognostic factors, confounding factors, or independent variables, are any aspect of the study that can affect the outcome variables (increase or decrease), or can affect the relationship between treatment and outcome. Imbalances in comparison or treatment groups in influencing variables at baseline can lead to false conclusions by improperly attributing an effect observed in the outcome variable to an intervention when it was merely due to the imbalance.

For example, blood pressure generally increases with age. If a group of individuals in the treatment group is significantly younger, and possess lower mean pressures than subjects in the control group, and are then compared using blood pressure as the outcome variable, the investigators may falsely conclude that an intervention was responsible for the observed "reduction" in blood pressure. Appropriate statistical testing of these baseline values should reveal any significant imbalances between the two comparison groups before the trial begins.

In the development of a clinical trial design, extreme care should be taken to identify those influencing variables that are likely to affect the outcome. By taking such known or suspected variables into consideration when designing the trial, the sponsor minimizes the chance that conclusions drawn at the end of the study will be spurious.

Once the variables or factors to be included in the trial have been identified, the selection of measurement methods becomes critical. The most informative and least subjective methods should be used. Quantitative (continuous) variables are measures of physical dimension (height, weight, circumference, area, etc.). Qualitative or categorical (discrete) variables are measures of distinct states usually represented by whole numbers (alive or dead, healthy or diseased, tumor classes, etc.).

Quantitative data can contain more information than qualitative data, and this generally allows for the use of more mathematically sophisticated and statistically powerful analytical methods. However, there may be situations where qualitative data is most appropriate or the only information available for a specific comparison, and there are many powerful non-parametric or distribution-free techniques available for these types of analyses. For example, quality of life evaluations generally utilize these types of qualitative analytical approaches.

D. Study Population

The study population should be a representative subset of the population targeted for the application of the medical device. The study population should be defined before the trial by the development of rigorous, unambiguous inclusion/exclusion criteria. Clinical experts in the field of the device under investigation should develop these criteria. These inclusion/exclusion criteria will characterize the study population and in this way help to define the intended use for the device.

It is possible to narrowly define a study population such that it is rather homogeneous in its composition. The advantage of using a restrictive population is that it allows for a smaller sample size in the clinical trial. That is, in homogeneous populations, the variability in responses in general will be smaller than in a more heterogeneous group, and this reduction in variability, (all other critical factors being held constant), will result in a corresponding decrease in the sample size required to observe a specified significant difference between two groups.

The disadvantage is that it may limit generalization of the approval to a narrow subset of the general population as defined by the criteria. Thus, a sponsor should discuss how they intend to define the study population with the reviewing division in the Office of Device Evaluation before beginning the clinical trial.

Inclusion/exclusion criteria should include an assessment of prognostic factors for the outcome variable(s), since one or more of these variables may influence the effectiveness of the device. For example, gender may be a prognostic factor for a particular disease process. It seems reasonable then to assess what role, if any, that gender might play in device assessment and then determine inclusion/exclusion criteria, other design, and analytical considerations accordingly. Consideration should also be given to: patient age; concomitant disease, therapy or condition (at both baseline and subsequent follow-up times); severity of disease; and others.

E. Control Population

Every clinical trial intended to evaluate an intervention is comparative, and a control exists either implicitly or explicitly. The safety and effectiveness of a device is evaluated through the comparison of differences in the outcomes (or diagnosis) between the treated patients (the group on whom the device was used) and the control patients (the group on whom another intervention, including no intervention, was used). A scientifically valid control population should be comparable to the study population in important patient characteristics and prognostic factors, i.e., it should be as alike as possible except for the application of the device.

There are many types of control groups. For the purposes of this document, four types are described:

1. Concurrent controls are those who are assigned an alternative intervention, including no intervention or a placebo intervention, and are under the direct care of the clinical study investigator. Any concurrent control can be a treatment control if it is assigned another intervention. If a placebo or sham is assigned, then it becomes a placebo or sham control. If the controls do not receive any intervention, then they are called a "no treatment" control.

2. In a passive concurrent control design, patients receive an alternative intervention, including no intervention, but are not under the direct care of the clinical study investigator.

3. Self-controls or crossover controls are patients who are assigned one intervention, (the order of treatment presentation should be specified in advance), for a prescribed period of time and then, following a washout period, receive the alternate intervention.

    A washout period refers to allowing a period of time to elapse between the end of one experimental condition and the beginning of the next condition. The period of time between the two interventions should be based on current knowledge of how the device may affect any anatomical or physiological processes, so that it may be demonstrated that no residual effects of the first treatment remain which may confound the results obtained from the next scheduled treatment.

    It should be noted that there will still be instances where a patient may serve as his/her own control even if a crossover design is not necessary or appropriate. For example, a crossover design would not be necessary when it can be clearly demonstrated that current clinical consensus has determined that there are no residual effects of a device beyond the immediate treatment of the patient.

4. An historical control is a nonconcurrent group of patients with the same disease or condition that have received an intervention, including no intervention, but are separated in time and usually place, from the population under study.

Concurrent controls and, where applicable, self-controls allow the largest degree of opportunity for comparability. Passive concurrent controls can provide comparability only if the selection criteria are the same, the study variables are measured in precisely the same way as those in the study sample, and assuming there are no hidden biases.

The use of historical controls is the most difficult way to assure comparability with the study population, especially if the separation in time or place is large. The practice of medicine and nutrition is dynamic - hygiene and other factors change as well. Subtle differences (secular trends) in patient identification, concurrent therapies, or other factors can lead to differences in outcomes from a standard therapy or diagnostic algorithm. Such differences in patient selection, therapy or other factors may not be easily or adequately documented. These differences in outcome may be mistakenly attributed to a new intervention when compared to a historical control observed at a significantly different time and/or place.

In addition, it is often difficult or impossible to ascertain whether the measurement of critical study variables was sufficiently similar to those used in the current trial to allow comparison. It should not be assumed that the measurement methods are equivalent. For these reasons, historical controls will usually require much more work to validate comparability with the study population than would concurrent controls.

F. Methods of Assigning Intervention

A method of assigning treatments or interventions to patients must minimize the potential for selection bias to enter the study. Selection bias occurs when patients possessing one or more important prognostic factors appear more frequently in one of the comparison groups than in the others. For example, if we know that the mortality from a condition is twice as likely in males than in females, and that one group had a two-to-one ratio of males to females, and a second group had a two-to-one ratio of females to males, then a difference in mortality will appear between these two groups with no intervention effect. If an intervention is assigned to one of these groups, its effect on mortality will be confounded, i.e., inseparably mixed, by the effect of gender.

Appropriate steps must be taken to assure that imbalances among known or suspected prognostic factors are minimized. The preferred method for protecting the trial against selection bias is randomization. The process of randomization assigns patients to intervention or control groups such that each patient has an equal chance of being selected for each group. If the trial is large with a limited number of comparison groups, randomization tends to guard against imbalances of prognostic factors.

It also protects the trial from conscious or subconscious actions on the part of the study investigators which could lead to non-comparability, e.g., assigning (or selecting) the most seriously ill patients to the therapy thought by the physician to be the more aggressive treatment.

Finally, randomization provides a fundamental basis on which most statistical procedures are founded. Generally, randomization methods utilize random number tables, computer generated programs, etc. Specific methods of randomization with examples are discussed in textbooks on clinical trials and medical statistics (Friedman et al, 1985; Fleiss, 1986; Hill, 1967; Pocock, 1983). The method of randomization used in a trial should be specified.

On occasion, when trial sizes are small and/or the number of comparison groups is large, simple randomization may not provide adequate balance among prognostic factors within comparison groups. In such situations it may be reasonable to form subgroups, called strata, by grouping subsets of selected prognostic variables.

Other methods of treatment assignment can be devised for active concurrent controls but, unless a true randomization scheme is used, it is difficult for the sponsor to assure that the resulting assignments are free from systematic or other possible biases. For example, assigning the intervention to patients in some systematic order, say every other or every third patient, seems random. However, such periodic assignments can sometimes coincide with cyclical patterns of patient presentation at the clinic such that imbalances can occur or can lead to selection bias because the intervention assignment is predictable. Thus, systematic or patterned intervention assignments are best avoided.

The intervention assignment process should be routinely monitored to assure crude balance in the important factors that are known or suspected to affect outcome. There are grouped randomization schemes which automatically preserve balance, while other methods require monitoring and adjustment. Caution must be exercised in adjusting randomization methods to assure that the random nature is preserved. For example, some imbalance between intervention and control group is tolerable because adjustment methods exist in analysis which can be applied to make the groups comparable. Large imbalances cannot be adequately adjusted by such techniques and should be avoided by employing appropriate randomized assignment.

G. Specific Trial Designs

There have a substantial impact upon the final analyses.

It should be clear, then, that deviations from the protocol by particular investigators for individual patients may create substantial problems for the trial analysis. Ultimately, it is the sponsor's responsibility to assure investigator compliance with the protocol. Potential investigators who for whatever reasons indicate that they may not be willing to strictly adhere to the protocol throughout the course of the investigation should not be asked to participate in the clinical trial.

J. Sample Size and Statistical Power

A discussion of sample size and statistical power requires knowledge of some elementary statistical principles which will be briefly reviewed here.

The object of the clinical trial is to collect data concerning the safety and effectiveness of a device in a sample of the target population. Statistical analysis is then used to infer relevant information concerning properties of the target population from the observations of those same properties in the trial sample. These inferences require that the research questions be translated into numerical statements of relationships of those population properties. Tests of the stated hypotheses should provide unequivocal answers to the research questions.

For example, if the research question is "For some disease A, is the mean valn to protect against hidden or unknown biases. The conduct of a crossover design is somewhat more complicated than parallel designs and requires closer monitoring.

Analyses for crossover designs are also more complicated because the patient's response to any particular intervention is usually correlated with the response to another intervention. This is because more than 1 interventions are applied to the same patient and the response is likely to be influenced heavily by that patient's individual characteristics. However, patient-to-patient variability is controlled by employing a crossover design.

A third design that is applicable in medical device clinical trials is the factorial design. In a simple version of a factorial design, patients in the study population are assigned to one of four groups: one of two interventions under study, a control intervention or both interventions. Such a trial may be used if a medical device was being tested against an alternate therapy, say a drug, and the research question is to determine if either intervention acting alone was effective, or if in combination they "interacted" to produce a stronger beneficial or detrimental effect.

The negative aspect of this design is that it is more complicated to conduct and the sponsor must assure that investigators are adhering to the study protocol.

A factorial design may require a larger sample size, but since this type of design is essentially two clinical trials in one, it offers an efficiency that should not be overlooked. If a drug intervention is proposed for a factorial design, the sponsor will have to adhere to the requirements of the Center for Drug Evaluation and Research if the drug is not already approved for the proposed claim.

Other aspects of experimental design, such as blocking or stratification, may further complicate the evaluation. The design chosen for a particular study must be the one that is most applicable to the sponsor's objectives. These objectives may appropriately result in complicated studies that need to be developed, monitored, and evaluated carefully. Sometimes, less complicated designs can be used by limiting the scope of the trial. Such a move, however, should be very carefully considered because it will nearly always result in a restriction on the claims for the device.

H. Masking (or Blinding)

Three of the more serious biases that may occur in a clinical trial are investigator bias, evaluator bias, and placebo or sham effect. An investigator bias occurs when an investigator either consciously or subconsciously favors one group at the expense of others. For example, if the investigator knows which group received the intervention, he/she may follow that group more closely and thereby treat them differently from the control group in a manner which could seriously affect the outcome of the trial.

Evaluator bias can be a type of investigator bias in which the person taking measurements of the outcome variable intentionally or unintentionally shades the measurements to favor one intervention over another. Studies that have subjective, or quality of life, endpoints are particularly susceptible to this form of bias.

The placebo or sham effect is a bias that occurs when a patient is exposed to an inactive therapy mode but believes that he/she is being treated with an intervention and subsequently shows or reports improvement.

To protect the trial against these potential biases, masking should be used. The degree of masking needed depends on the strength and seriousness of the potential bias. Single mask designs shield the patient from knowing what intervention has been assigned. Double mask trials shield both the patient and the study investigator.

Third party mask trials allow the patient and investigator to know the intervention assignment but restrict the evaluator, i.e., the third party, from knowing, such as in the reading of imaging films or laboratory tests.

Masking is accomplished by coding the interventions and having an individual who is not on the patient care team control the key to breaking the code. The bias introduced by breaches in masking can be very difficult to assess in the analysis, therefore it is important not to break the code until the analysis is completed.

The evolution of medical device evaluation has demonstrated that it is often difficult or impossible to mask the

patient or investigator because a placebo or convincing sham treatment may not be feasible. In such cases extra care must be exercised by the study staff to assure that these biases are minimized by assuring that the evaluator is blinded to the assignment of patients to a particular intervention or control group.

I. Study Site and Investigator

Because pooling of data across study sites and investigators is almost always necessary in order to attain the required sample size, the selection of study sites and investigators is critical in planning a clinical trial.

The sites that have been selected must have sufficient numbers of eligible patients who are representative of the target population for the device. Each site must have facilities that are capable of processing patients in the manner prescribed by the protocol, and must have staff who are qualified to conduct the trial. It should be noted, however, that despite a common protocol and the best efforts of the study monitor, site effects may be present which can invalidate pooling the data. A careful analysis to rule out potential bias due to site effects is an important part of the investigational protocol.

The principal investigator at each site must be able to recruit eligible patients to the trial and must be willing to abide by the procedures established by the protocol. Potential investigators may overestimate their capabilities to recruit and process study patients, so a review of the demographics and records of patients for a recent calendar period is advisable. If the investigator consistently violates the protocol, the data from that site cannot be used to establish the safety and effectiveness of the sponsor's device.

Participating physicians have a primary responsibility to their patients and must provide for individual patients what they consider to be the best medical care. While there is no question a physician must do what is best for the patient, if a specific treatment regimen happens to violate the protocol, a patient enrolled in the study becomes disqualified from the trial and that patient's data cannot be used in the analysis.

The clinical trial is basically an experiment in a human population and as such differs from the routine practice of medicine. It should be noted that in many investigations, the Center may require an intention to treat analysis, which would record data of disqualified patients as a failure. Clearly, a relatively small number of patients that are disqualified in an intention to treat model could have a substantial impact upon the final analyses.

It should be clear, then, that deviations from the protocol by particular investigators for individual patients may create substantial problems for the trial analysis. Ultimately, it is the sponsor's responsibility to assure investigator compliance with the protocol. Potential investigators who for whatever reasons indicate that they may not be willing to strictly adhere to the protocol throughout the course of the investigation should not be asked to participate in the clinical trial.

J. Sample Size and Statistical Power

A discussion of sample size and statistical power requires knowledge of some elementary statistical principles which will be briefly reviewed here.

The object of the clinical trial is to collect data concerning the safety and effectiveness of a device in a sample of the target population. Statistical analysis is then used to infer relevant information concerning properties of the target population from the observations of those same properties in the trial sample. These inferences require that the research questions be translated into numerical statements of relationships of those population properties. Tests of the stated hypotheses should provide unequivocal answers to the research questions.

For example, if the research question is "For some disease A, is the mean value of a critical outcome variable after prescribed treatment, greater for the device-treated group than for the control group?" Two hypotheses would be formed: a null hypothesis that states that the mean value of patients post treatment in the treatment group is equal to (or worse than) that in the controls; and an alternative (or research) hypothesis that states that the mean value post treatment in the treatment group is greater than that in the controls. There are two types of decision errors that can be made by inferring results from a sample to the population. If the sample indicates that the mean is greater in the device treated group than in the controls (i.e., rejecting the null hypothesis) when in the population there is no difference between means, a Type I error (also called an alpha error) is made. If, on the other hand, the sample indicates no difference between means, (i.e., accepting the null hypothesis), when the device mean is actually greater, then a Type II error is made. The probability of making a Type II error is also known as Beta error, and statistical power is defined as 1 - Beta.

The probabilities of these two types of errors factor heavily into all sample size calculations for hypothesis tests (see Section VIII Appendix on Sample Size for a more thorough discussion). Usually these probabilities are fixed in advance, giving more weight to the error with more serious consequences.

For example, a Type I error occurs if the aim of the trial is to show that the test device is "better than" the control, and we falsely reject the null hypothesis, and conclude that the device may be better than the comparison device, when in fact it is equivalent or even worse than the control. Conversely, if the object of the trial is to show that the device mean survival is "as good as" (really, "no worse than") that of the control, then it would be more serious to accept a false null hypothesis (a Type II error).

Additionally, clinical trial hypothesis tests should involve clinically meaningful differences, that is, those differences in the outcome variable(s) determined by experts in the medical community to be clinically significant. The most common sample size formulas include an estimate of the variability of the clinically meaningful difference in the numerator and an estimate of the clinically meaningful difference to be detected in the denominator. Thus, for a given outcome variable, the larger the variability, the larger the sample size that will be required. Similarly, for a given variability, the smaller the clinical difference to be detected, the larger the sample size.

Meinert (1986) provides an excellent discussion of these computations for both sample size and power.

IV. THE PROTOCOL

Each well-designed clinical trial should have a detailed protocol, i.e., the comprehensive plan that precisely describes how the trial is to be conducted and how the clinical data are to be collected and analyzed.

The protocol may be submitted to the Agency as part of an IDE or as an IDE supplement, but those study protocols not submitted as part of an IDE must be included in the submission of the PMA.

The following points should be included in the protocol and determined before initiating the trial:

1. The background of the trial that completely describes and summarizes all previous scientific studies that are pertinent to the subject matter.
2. A clear statement of the trial objective(s), specifying any medical claim and indication that is related to the research question(s), a clinically meaningful effect, and associated outcome variables.
3. A complete description of the trial design including design type, method of data collection, type of control, method and level of masking, justification of sample size, and method of treatment assignment (randomization, stratification, other).
4. A complete description of the study population, including study site(s), method of selecting subjects (inclusion and exclusion criteria), and type of patients (e.g., inpatient or outpatient). Pertinent clinical and

demographic characteristics of study subjects should be discussed in relation to the characteristics of the target population and the intended use of the device (clinical utility).

5. A complete description of the intervention including frequency and duration of application, and measures of physician and patient compliance.

6. A complete description of the procedure for each follow-up visit and a schedule of required follow-up. Include identification of all measurements to be made and information collected at each visit. Also include how patient withdrawal is to be handled and those steps the sponsor will take to determine the health status of individuals who fail to return for follow-up visits or who withdraw from the study.

7. A detailed description of the data gathering and analysis, including data collection and validation methods, data monitoring, methods of statistical analysis, and specific rules as to how and why the clinical trial would be terminated early - i.e., for statistically significant un-expected positive or negative results.

8. A thorough description, including Curriculum Vitae of the participating investigators, monitoring methods, and trial administration techniques (trial monitor, policy and data monitoring committee, etc.) including methods to identify and make necessary adjustments to the protocol.

9. A list of precisely defined clinical terminology and other relevant terms to be used during the trial. This should include detailed descriptions of trial entrance criteria and all criteria for observing either an outcome or influencing variable.

10. All informed consent forms and a list of provisions not already discussed above that may be required by the Institutional Review Board (IRB).

V. CLINICAL TRIAL CONDUCT

If a detailed protocol is established that completely describes the trial design, relevant methodologies, and the proposed analysis, then conducting the trial should be straightforward. However, it will not be simple or routine. It is imperative that those charged with the oversight of the clinical trial have contingency plans available for unforeseen problems that may occur during the trial and have means to rapidly implement those plans.

Contingency plans should be carefully crafted with the goal of preserving the integrity of the established design. Any modification of the protocol may reduce the efficiency of the design. It is difficult to envision, however, any clinical trial conducted precisely as it was designed. Therefore, it is wise to anticipate possible problems and have plans to address them if they occur.

A. Trial Monitoring

The primary concerns in conducting the clinical trial lie in assuring that the study subjects are entered, the interventions assigned, the relevant variables measured (at the appropriate times), and the data accurately and completely recorded as specified in the protocol. This requires extreme care by the trial sponsor to closely monitor the conduct of the trial. A designated trial monitor should assure compliance with the protocol and identify potential weaknesses that may require modification of the protocol.

Clinical trials generally incorporate multiple study sites with one or more investigators at each location. It is critical to the integrity of the trial that the monitor assure that each site and investigator is executing the protocol just as it was planned.

For example, if a modification of the protocol is thought to be necessary by one or more investigators and the trial is not closely monitored, it is possible that each site or investigator will modify the protocol in his/her own way. This could result in as many distinct protocol changes as there are sites or investigators, thus jeopardizing the ability to pool the trial results.

If the investigator consistently violates the protocol, the data from that site cannot be used to establish the safety and effectiveness of the sponsor's device. To avoid this possibility, the sponsor should establish a mechanism to consider protocol modification, and appoint a monitor or gatekeeper to ensure that all sites and investigators make the same modification at the appropriate time.

B. Baseline Evaluation

Whether or not the clinical trial will use randomization, the baseline observations should be made on all prospective study patients before assigning or applying an intervention. The accurate determination of baseline information on all study subjects is critical for a number of reasons. It allows:

- the investigator to evaluate a subject's eligibility
- subclassification for stratification (if necessary)
- the descriptive characterization of the target population or an assessment that the trial sample represents the target population
- baseline physical/laboratory measurements prior to intervention

The assessment of baseline data is instrumental in the identification of prognostic factors which must be balanced among intervention groups. That is, the patient's current disease status; concomitant medication, therapy, or condition; age; gender; socioeconomic status; prior disease history; and other factors may affect the outcome variable. The assessment of baseline data allows for the selection and implementation of methods that minimize the impact of any potential bias on the comparison of outcome measures. For example, for those prognostic factors known to affect outcome, stratification or balanced allocation can be used at the time interventions are assigned.

If a prognostic factor is discovered during the course of the trial and adequate baseline measurements exist, then adjustment or standardization methods can be employed during data analysis to minimize the effect of imbalance on comparisons.

C. Intervention

The assignment and application of the intervention should be done with strict adherence to the protocol. A pre-specified regimen should be followed on every subject. In so far as it is possible, every procedure scheduled for the treatment group should also be scheduled for the control group except for the active application of the device. If the individual administering the treatments is masked to the intervention group assignment, it is more likely that all groups will be treated the same way.

D. Follow-Up

The follow-up of subjects after intervention extends beyond the simple scheduling of follow-up appointments for the study subjects. Mechanisms should be in place to assure a high degree of subject compliance with the follow-up schedule. Even moderate deviations in follow-up between comparison groups can lead to substantial biases in the analysis.

Two characteristics of follow-up are critical: completeness and duration. Completeness is defined as the proportion of patients entering the trial who come back for each and every follow-up appointment. It is extremely important that this proportion be as close to 100% as possible, because statistical power will decrease as patients are lost to follow-up. Follow-up percentages of less than 80% are generally considered poor and these trials are labeled incomplete. It is also important for the follow-up percentages

to be similar across comparison groups and across study sites.

Incomplete follow-up is a major concern in analysis. The trial must have procedures available to trace subjects who fail to appear for scheduled follow-up. Accounting for patients lost to follow-up is a critical analytical issue because those patients may provide the most important information from the clinical trial, particularly if the outcome in such patients is poor. So, it is essential to determine the health status of all patients entered into the trial even for those who do not return to the clinic for all follow-up appointments.

The duration of follow-up is that period of time after the intervention during which the study subjects are scheduled to be observed and evaluated. Follow-up duration must be consistent with safety and effectiveness claims, i.e., it must equal the duration of claimed effectiveness and must also be long enough to accurately estimate the rate of known or suspected adverse events. The duration of follow-up should also be the same across comparison groups and study sites.

E. Collection and Validation of Data

Methods for obtaining and verifying the accuracy of all measured variables in the trial must be in place before the trial begins and must be monitored for compliance. Each study site must have sufficient staff with suitable expertise to assure the collection of valid data. Attention to detail is critical because it is impossible to retrospectively assess data not taken at the scheduled time or data taken without adequate precision.

These methods must include quality-control techniques for data measurement, recording, transfer to electronic media, and verification. The measurement of trial variables begins with an unequivocal definition of each variable, condition, or characteristic to be observed in the trial. Trial staff should completely understand all defined terms, and care must be taken to assure consistency across investigators and study sites. Consistency of trial terminology is also essential for comparisons with other trials or research studies in the literature, and for use of historical controls, where appropriate.

VI. CLINICAL TRIAL ANALYSIS

When the clinical trial reaches the analysis stage, except for deviations that may have unexpectedly occurred during the trial, the analysis should have been previously determined in the protocol. The protocol, revised by any alteration made during the trial, dictates what can or cannot be done with statistical analysis. In most cases, large biases that have been introduced by any element of trial conduct and that affect the observations of the outcome variables cannot be satisfactorily rectified by statistical adjustment procedures.

A. Validations of Assumptions

Before beginning a detailed statistical analysis it is necessary to validate the assumptions to be used in the proposed analysis. Such assumptions include underlying characteristics of the probability distribution used for hypothesis tests or estimation, similarity of distribution of prognostic factors among study sites and comparison groups, and validation of suspected relationships (dependence) or lack of relationship (independence) among variables.

It is quite important to validate the distribution and variance assumptions of the statistical test to be used. A test statistic possesses the properties of the test only if all assumptions are valid. For example, if the normal (Gaussian) distribution is assumed, the data should be tested by appropriate statistical techniques to be certain that the sample does not deviate substantially from that which would be predicted by the normal distribution. If it does, then other more appropriate tests such as non-parametric (distribution-free) procedures should be used.

Likewise if the test requires equal variance among comparison groups, an appropriate procedure to detect unequal variances should be used. If unequal variances are detected, either the data will have to be adjusted or transformed to account for the unequal variances, or the statistical test will have to be modified.

An evaluation of the balance of prognostic factors across comparison groups and study sites is also necessary. Any observed imbalances must be adjusted so that the ultimate comparison is made between comparable samples. Analysis of covariance is a powerful statistical adjustment tool if the number of variables that require adjustment is small and the variables are highly correlated to the response variable. If the number of variables requiring adjustment is large, it is more difficult to adequately account for all of them. It is critical that extreme care be exercised in the conduct of the trial because in the words of Hill (1967) "to start out without thought and with all and sundry included, with the hope that the results can somehow be sorted out statistically in the end, is to court disaster."

If the analysis assumes that certain prognostic or response variables are unrelated to outcome, appropriate statistical tests should be performed to confirm these assumptions. Performing tests on variables that are assumed to be independent, but are in fact related, or dependent, can lead to significant errors in tests of hypotheses.

B. Hypotheses and Statistical Tests

In essence, all comparative analyses result in a hypothesis test. The report of the analysis should clearly state the hypotheses to be tested, the statistical tests to be used, and the assumptions behind the tests. All procedures should be referenced so that the Agency can validate the procedure. References should be provided even for common procedures. If any innovative analytical procedures are developed by the sponsor, complete documentation of those procedures must accompany the analysis.

In some instances it may be appropriate to use available (historical) data to develop a mathematical model of the progression or other characteristic of a disease or condition. Data gathered in a clinical trial could be used to "validate" the model by comparing the projected characteristics of the model with results obtained during the investigation. These types of comparisons can be used to form a hypothesis test of the model characteristics.

C. Pooling

It is almost always necessary for the sponsor to pool study subjects across investigational sites in order to obtain adequate sample sizes. Pooling must be justified by testing balance among prognostic factors and verifying that all clinical procedures were conducted in the manner prescribed in the protocol. On occasion, data from a given study site will exhibit characteristics that make it stand out from the others locations. The sponsor must investigate all relevant effects due to investigational site and report on these instances to determine why that particular site had results that differed.

D. Accountability for Patients

The sponsor should be prepared to use extensive measures to document the post-trial health status of every patient who was enrolled in the trial. While it is often not possible to find all patients, the sponsor must demonstrate that everything possible was done to attempt to find patients lost to follow-up. It is not appropriate to coerce the patient against their will to keep follow-up appointments, but, at the very least, a reasonable assessment of the morbidity or mortality of the patient should be made.

Sometimes a determination of safety and effectiveness will hinge on the differences of a small subset of patients in the comparison groups. If the number of patients lost-to-follow-up is large relative to the subset that has been observed to be different, then our ability to document safety and effectiveness is substantially weakened.

The Agency will require an analysis of the data by "intention-to-treat." This is an analysis method in which "the primary tabulations and summaries of outcome data are by assigned treatment" (Meinert, 1986). In such analyses, patients lost-to-follow-up in the intervention and control groups must be counted as though they actually completed the study in their assigned group. Since there is no observation of outcome variable after the time the patient is lost-to-follow-up, the observation cannot be counted as a success (and is considered failure).

The impact of intention-to-treat analyses on interventions that may be effective but for which there is a large number of patients lost-to-follow-up can be devastating. An observation of effectiveness in the intervention trial patients who are followed can be eclipsed entirely by a large number of patients lost to follow-up whose outcomes are recorded as ineffective. It is crucial, therefore, to keep the number of patients lost to follow-up as small as possible.

VII. BIBLIOGRAPHY

1. Premarket Approval (PMA) Manual. (1987). U.S. Department of Health and Human Services. FDA 87-4214.
2. Investigational Device Exemption Manual. (1990). U.S. Department of Health and Human Services. FDA 90-4159.
3. Regulatory Requirements for Medical Devices. (1992). U.S. Department of Health and Human Services. FDA 92-4165.
4. Food and Drug Administration. 21 CFR 800-1299 (Food and Drugs). Revised as of April 1993 pp 146-149. GPO 1993.
5. Friedman, L.M., C.D. Furberg, D.L. DeMets. (1985). Fundamentals of Clinical Trials (2nd ed). Mosby Year Book, (2nd ed), St. Louis.
6. Fleiss, J.L. The Design and Analysis of Clinical Experiments. (1986). John Wiley and Sons, New York.
7. Hill, A.B. Principles of Medical Statistics. (1967). Oxford University Press, New York.
8. Pocock, S.J. Clinical Trials: A Practical Approach (1983). John Wiley and Sons, New York.
9. Cox, D.R. Planning of Experiments. (1958). John Wiley and Sons, New York.
10. Cochran, W.G. and G.M. Cox. Experimental Designs. (1957). John Wiley and Sons, New York.
11. Meinert, C.L. Clinical Trials Design, Conduct, and Analysis. (1986). Oxford University Press, New York.

VIII. APPENDIX on SAMPLE SIZE

If we let pi be the proportion surviving two years in the intervention group and pc be the proportion surviving two years in the control group, then numerically the hypotheses are stated as:

$H_o: p_i = p_c$

$H_a: p_i > p_c$.

In the study population, one of these two conditions is true. If, based on the data, we reject H_o (and accept H_a) when H_o is true, we make a Type I statistical error. When we accept H_o (and reject H_a) based on the data when in fact H_a is true in the study population, then we make a Type II statistical error. The object of the sample size estimate is to minimize the chances of making either of these types of errors.

Probability statements are used to determine the chances of making Type I or Type II errors. The probability is based on the distribution of possible values for the outcome variable, or in the case of our example p_i or p_c.

In common statistical notation, α designates the probability of making a Type I error, i.e., the probability of rejecting the null hypothesis when it is true. The probability of the Type II error, i.e., the probability of accepting the null hypothesis when it is false, is denoted by β. The statistical power of a test method is the probability that the null hypothesis will be rejected when it is false. The power is denoted $1-\beta$

Consider the formula for sample size for the example given here:

$$n > (Z_\alpha \sqrt{2pq} + Z_\beta \sqrt{(p_i q_i + p_c q_c)})^2 / d^2$$

where n = the sample size for each comparison group. (intervention and control)

p_i = the probability of surviving two years in the intervention group
p_c = the probability of surviving two years in the control group.
p = the mean probability of surviving two years, i.e., $(p_i + p_c)/2$
Z_α = the standard normal (Gaussian) variate corresponding to the tail probability of size α
Z_β = the standard normal variate corresponding to the tail probability of size β
d = the difference between the intervention and control groups considered to be clinically meaningful.
Note $d = p_i - p_c$.
q_k = the probability of not surviving two years after treatment in the kth group. Note $q_k = 1-p_k$ for k = i or c, and q = 1-p.

In most situations α and β are pre-specified to account for the seriousness of making a Type I or Type II error, respectively. By convention, α is given the value of 0.05, i.e., the chance of rejecting the null hypothesis when it is true is 1 in 20. The determination β of depends on the claim for the device. If the claim is that the two year survival of the intervention is "better than" that of the control intervention, then β can be less restrictive (meaning larger). However, it should never be greater than 0.20.

If the claim is that the two-year survival of the intervention is "as good as" (or no worse than) the control intervention, then the object would be to make β as small as feasible. When an "as good as" hypothesis is being tested, the test is attempting to "prove" the null hypotheses. The failure to reject the null hypothesis can occur under two conditions: either the two probabilities are truly not different or they are different but the sample is too small (too little power to detect the observed difference). If β is small, the power $(1-\beta)$ to detect the specified difference d is large. Under the "as good as" hypothesis it is not unusual for β to be 0.1 or even 0.05.

The difference, d, is also dependent on the claim. If the hypothesis involves the claim of "better than," then d is that increase in two year survival considered by the medical community to be clinically meaningful. If the hypothesis involves the claim of "as good as," then d is that decrease in the two-year survival considered by the medical community to be clinically significant.

Whenever possible, the determination of d should be based on previous data. Where data are not available, it may be necessary to convene a panel of medical experts to provide a value for d which is considered by the panel to be reasonable. In either situation, the sponsor should provide a detailed justification for the choice of the d used in the calculation.

The final elements of the formula are estimates of the variability of p_i, p_c, and p. The term $\sqrt{2pq}$ involves the variability of the difference under the null hypothesis, i.e., $p_i = p_c$. The term $\sqrt{(p_i q_i + p_c q_c)}$ is the variability of the difference under the alternate hypothesis, i.e., $p_i > p_c$.

Page Last Updated: 11/20/2009
Note: If you need help accessing information in different file formats, see Instructions for Downloading Viewers and Players.

FDA

Accessibility | Contact FDA | Careers | FDA Basics | FOIA | No Fear Act | Site Map | Transparency | Website Policies

U.S. Food and Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993
Ph. 1-888-INFO-FDA (1-888-463-6332)
Email FDA

USA.gov

For Government | For Press

Combination Products
Advisory Committees
Science & Research
Regulatory Information
Safety
Emergency Preparedness
International Programs
News & Events
Training and Continuing Education
Inspections/Compliance
State & Local Officials
Consumers
Industry
Health Professionals
FDA Archive

U.S. Department of **Health & Human Services**