Perspectives on Clinical Studies for Medical Device Submissions

by

Richard P. Chiacchierini, Ph.D. and Melvin M. Seidman

Introduction

Statistics play a vital role in the evaluation of the safety and effectiveness of medical devices submitted for premarket approval. Two previous papers, (1,2), presented the statistical aspects of device submissions and the design, conduct, and interpretation of clinical studies. While each paper focused on a different aspect of the statistical quality of submissions to the Food and Drug Administration (FDA), Center for Devices and Radiological Health (CDRH), both described solid statistical methods and principles which, if used, can avoid lengthy delays in the approval process. The sponsor must choose the appropriate submission vehicle and must provide valid scientific evidence of safety and effectiveness for the product under evaluation.

The best means to provide valid scientific evidence is to follow good clinical study concepts (2).

- Pose the appropriate research questions associated with the claims for the device.

- Design the clinical study with adequate statistical "power" and "size" to answer all appropriate research questions.

- Exercise maximum control of biasing factors.

- Provide complete patient follow-up with the appropriate duration to allow a reasonable assurance of safety and effectiveness.

- Use the most objective and most informative measurement variables for both outcome variables and influencing factors.

- Finally, apply appropriate statistical analyses consistent with the study design and conduct to demonstrate adequately all safety and effectiveness claims.

In this paper, the evaluation of clinical data is presented from the statistical reviewers' perspective. The CDRH has compiled a checklist of statistical deficiencies that the statistical staff consider to be fatal flaws in any submission. The checklist is discussed briefly and a copy can be found in Appendix I. The remainder of the paper discusses the statistical fatal errors most frequently found by the reviewing statisticians during the past few years.

THE STATISTICAL CHECKLIST

The statistical checklist is comprised of those omissions or errors

which would prevent an orderly in-depth review of the statistical aspects of the device submission. The checklist applies primarily to premarket approval applications (PMA's) but will also be used for supplements and amendments to PMA's where appropriate. Included in the checklist are three major areas of focus: organizational and administrative elements, summary of safety and effectiveness, and the clinical investigation.

The organizational and administrative review determines if all major elements of the document are in place and are intelligible. Is there a table of contents? Are the volumes, pages and tables numbered? Are the tables adequately labeled?

The summary of safety and effectiveness is reviewed for completeness with special emphasis on the following concerns. Are the indications for use stated? Are the claims for the device clearly presented? Is there a summary of all pertinent scientific studies relevant to the sponsor's demonstration of safety and effectiveness?

The most intense review is conducted on the clinical investigation. Minimally, the concerns which must be addressed are the following. Is there a protocol provided, was it followed, and are all protocol deviations described? Has the sponsor accounted for all patients entered into the study? Is the patient follow-up fully described? Is there a full discussion of the safety and effectiveness variables and parameters? Is there full documentation of the statistical analysis and results including comparison group selection (control), sample size justification, stated hypothesis test(s), complete demographics, study site pooling justification, description of statistical tests applied, clear presentation of data and clear discussion of the statistical results and conclusions?

The absence of any of these items is sufficient to make an in-depth statistical review either extremely cumbersome or even impossible. It is essential, therefore, that the sponsor quality control the submission "before" it goes to CDRH by checking to determine that the submission is at least complete relative to the checklist. After this is done, the in-depth statistical review can be properly focused on the quality of the clinical study instead of the quality of the mechanics of the submission.

MOST COMMON STATISTICAL FATAL ERRORS

1. Patient Accountability

    It is not unusual for a sponsor to fail to account for some part of the population of patients originally enrolled in the protocol. There may be numerous valid reasons why patients are dropped from the study, but one of the weakest reasons is "patients lost to follow-up." It is possible that a small number of patients truly cannot be found, but the sponsor must demonstrate that all

appropriate steps, including extraordinary measures, were exercised in a thorough search for all missing patients.

If a sufficiently large number of study patients remain lost to follow-up, the resulting analysis may be biased. Frequently, the sponsor assumes that patients who do not complete the protocol are continuing to do well because they were doing well at the last known examination. The FDA has found, however, that it is just as probable that the patient either could not or would not return for clinical evaluation because of a worsening medical condition, death, loss of confidence in the study or investigator, etc.

By way of example, a sponsor had 54 patients entered into their original study. The PMA submission reported on 26 patients. No explanation was offered concerning the absence of the remaining 28 subjects. Fourteen of the twenty-six were considered improved. Some questions come to mind. Did the other 28 people die as a result of the device? Did all or a large proportion of the 28 subjects improve? Is the sponsor trying to hide something by "losing" the 28 patients? It is certain that CDRH can not simply assume that the 28 patients lost to follow-up are all doing fine. Therefore, complete accountability is always required.

2. Statistical Documentation - Protocol Not Followed

The sponsor must provide CDRH with the protocol and must describe all protocol deviations or changes. In many instances, the sponsor supplies the study protocol and on review, substantial deviations are found between the protocol and what was actually done. There may be legitimate reasons to modify or deviate from the protocol, but failure to describe all changes made is certain to cause major problems in the review. Protocols may not be followed for a variety of reasons, some of which may be considered reasonable. Physicians, when trying to implement a protocol, may find that procedural problems make following the protocol impossible. On the other hand, a monitoring mechanism may determine that physicians at different study sites are arbitrarily "doing their own thing" and are, in effect, doing as many different study protocols as there are study sites.

The impact of these deviations range from minor to extremely severe. If each primary investigator is not controlling their patient population for possible influencing factors such as major medical procedure changes, patient age, sex, concomitant medication, etc., the results may not be comparable across study sites which may prevent pooling.

Consider, for example, a sponsor who required their physician-investigators to take a given essential measurement (concerning a safety parameter) on every patient. Out of 200 patients, measurements were taken on only 141 patients with no discussion

concerning the remaining 59 patients. The CDRH is likely to ask why the measurements were not taken. Did the patients fail to follow some instruction, say for medication? Did the attending physician's measurement instrument fail for some period of time? Or, were the patients on which the measurements were missing considered by the evaluating physician to be doing poorly?

3. Statistical Documentation - Hypothesis Test Not Stated or Hypothesis Test is Incorrect

    Frequently, sponsors present their data in tabular form, perform statistical tests, and interpret their results without clearly specifying what was being tested. Ideally, the sponsor should have developed appropriate research questions before writing the protocol. These questions generally can be readily translated into numerical sentences called hypotheses. All such hypotheses must be clearly stated so that any test procedures used to test these hypotheses can be evaluated with respect to use of the appropriate statistical test procedures. Furthermore, specific hypotheses concerning safety and effectiveness must be stated so that an appropriate statistical determination can be made on the validity of each claim for the device.

    For example, in one instance, a sponsor submitted data on the lowering of the temperature of a certain bodily function by application of their medical device. The sponsor proceeded to show a statistically significant decrease in temperature, but failed to demonstrate a link between the reduced temperature and a physiological benefit. Furthermore, the reviewing statistician had to infer from the context of the document that a test of average temperature was the object of the investigation because it was never clearly stated.

    Several questions come to mind. Was there a particular level of temperature reduction which would be considered to be effective? What is the biological link between temperature and the physiologic activity? Are these two variables correlated? Just what is the sponsor trying to demonstrate anyway? (If the sponsor doesn't argue its case for the safety and effectiveness of the device, how is CDRH to infer safety and effectiveness?)

4. Statistical Documentation - Unjustified Pooling

    In most submissions, sponsors choose to pool data from several study sites in order to satisfy minimal sample size requirements for their clinical investigations. Pooling requires a clear demonstration that the study sites are sufficiently similar in protocol adherence, population demographics, and other factors to assure that the study sites are truly representative of a single clinical study on a relatively homogenous population.

The most important issue in data pooling is the minimization of bias. Some study sites or principal investigators may have executed the protocol differently in important ways or may have a selected different study populations. In either circumstance, the data from such clinical sites may not be sufficiently similar to that from the other sites and may not be pooled.

For example, a sponsor conducted a clinical study at four study sites and reported their results in a single combined table. The sponsor stated that the effectiveness of the device was highly dependent on surgical procedure. Under scrutiny, it was observed that patients at two of the sites had markedly different survival rates than patients at the other two sites.

The CDRH is required to ask the following questions. Was the medical technique clearly specified in the protocol and did all investigators use the same medical technique as specified? Were all study sites routinely monitored for protocol adherence by investigators and patients? Were treated patients demographically similar (age, sex, etc.) across study sites? Were there sufficient data from each site to analyze demographics?

5.   Statistical Documentation - Potential Bias Not Evaluated

If an imbalance appears between a treated group and control group of patients in factors which could change the response of either group in the variables being evaluated, a bias could occur which can falsely increase or decrease the response attributed to the device under investigation. If there are more patients in one group who are on prescribed medication affecting the condition under study than in the other group, the group having the excess may show an improvement due solely to the medication. If a patient or physician evaluator knows the treatment assignment that the patient received, a more or less favorable response may be indicated by the patient or recorded by the physician due to such knowledge, not solely due to the device under investigation.

Several methods exist to reduce bias and preserve balance in clinical studies. Randomization, stratification, development of patient inclusion/exclusion criteria, and blinding can all be used to minimize bias. (2)

To demonstrate a situation where a serious bias could exist, consider a sponsor whose primary investigator, responsible for a large proportion of the clinical population, has a financial interest in the sponsoring company. The sponsor did not discuss any procedures such as blinding, randomization, or other bias minimizing methods, and the results showed an overwhelming effectiveness for the sponsor's device based largely on the results from the principal investigator's study.

Again, CDRH will ask some difficult questions. Were patient inclusion/exclusion criteria developed and followed? Were there any methods used to make certain that neither the patient nor the evaluating physician knew the treatment status of the patient (blinding)? How were treatments assigned to patients? What was the possible impact on the outcome variable of any differences in patient demographics?

6. Statistical Documentation - Sample Size Not Justified

Rarely do sponsors provide any discussion on the number of patients in their clinical investigation. The absence of justification for the size of the samples presents a difficult problem for CDRH. In a good study, the sponsor should determine the sample size before beginning the clinical study. To accomplish this, the sponsor should determine the increase or decrease in the outcome variable that would be considered clinically significant and then derive the sample size which will give a high probability of detecting that increase or decrease at a specified significance level. Numerous formulas are available to assist in determining the proper sample size.

The determination of sufficient sample size is important to the sponsor for two reasons. It provides a reasonable assurance of detecting a clinically significant difference as described above. It also can prevent wasted resources on the part of the sponsor because, without determining the proper sample size, too many patients may be assigned to the study.

To demonstrate, consider a sponsor who used a device to treat 31 acutely injured and 43 chronically ill patients. The sponsor also included 24 acute and 36 chronically ill patients treated with a control device. The average pooled response for treated patients was 68.2 and for control patients was 67.8.

Several questions related to sample size are in order. What is the natural uncertainty or variance in the outcome variable, 0.1, 0.5, or what? What is considered a clinically significant increase, 0.4, 0.5 or what? What power (probability of detecting the above difference if it exists) is to be used? Eighty percent is nominal but higher levels can be employed. Does the given sample have a good chance of detecting the given difference at an appropriate significance level? If the variance is 0.1, then 10 patients in each group (treatment and control) would be required to have an 80 percent chance of detecting a difference of 0.4 at the 5 percent significance level. If the variance is 0.5, then 50 are needed in the treated group and 50 in the control group. If the variance is 1.0, then 99 patients would be needed in each group for an 80 percent chance of detecting a 0.4 unit change in the outcome variable at the 5 percent significance level. (See Chiacchierini and Bushar (2) for formulation.)

Conclusion

The design, conduct, and analysis of clinical studies are crucial elements in the submission process for premarket approval. Good common sense coupled with sound statistical and medical precepts should yield studies which are free from statistical flaws discussed in this paper. The process has few short cuts, and failure to adhere to sound clinical study design principles is costly. Time and resources saved by taking "short cuts" on the study design or conduct are lost many times over by delays in the product approval process when the deficiencies resulting from such short cuts induce one or more fatal flaws.

While only six statistical deficiencies are discussed here, many others exist which are equally as devastating to a PMA submission. To avoid such problems, use a proper design, conduct and monitor the study according to the appropriate protocol, use correct and properly applied statistical procedures, and interpret the results in the context of the intended use(s) or claim(s) for the device. While it is nearly impossible to foresee all possible problems that may arise in the conduct of a clinical investigation, the steps described in this and earlier papers (1,2) lay the groundwork to minimize the probability of adverse circumstances arising in the sponsor's study.

References:

1. Chiacchierini, R. and H. Lee. Statistical aspects of submissions to FDA: A medical device perspective. In the Proceedings of the Biopharmaceutical Section, American Statistical Association, January 1985.

2. Chiacchierini, R. and H. Bushar. Regulatory Submissions for New! Improved! Devices. In 43rd Annual Quality Congress Transactions, May 1989.

PMA REVIEW STATISTICAL CHECKLIST

	Yes	No	Comment

I. Organizational and Administrative
Elements including Table of Contents
with volume and page numbers   ___   ___   _____

II. Summary of Safety and Effectiveness

   A. Indications for use   ___   ___   _____
   B. Claims for the device   ___   ___   _____
   C. Summary of studies   ___   ___   _____

III. Clinical Investigations

   A. Protocol

      1. included   ___   ___   _____
      2. adhered to   ___   ___   _____
      3. deviations described   ___   ___   _____

   B. Patient Accountability

      1. patient inclusion/
         exclusion criteria   ___   ___   _____
      2. follow-up schedule   ___   ___   _____
      3. study period completed   ___   ___   _____
      4. all patients accounted for   ___   ___   _____

   C. Description of Safety and
      Effectiveness Parameters

      1. safety   ___   ___   _____
      2. effectiveness
         a. sensitivity   ___   ___   _____
         b. specificity   ___   ___   _____
         c. false positive   ___   ___   _____
         d. false negative   ___   ___   _____
         e. reproducibility   ___   ___   _____
         f. repeatability   ___   ___   _____
         g. stability   ___   ___   _____

   D. Documentation of Statistical
      Analysis and Results

      1. control (comparison)
         group   ___   ___   _____
      2. sample size justified   ___   ___   _____
      3. hypothesis test stated   ___   ___   _____

		Yes	No	Comment

4. potential of bias adequately
   evaluated
   a. randomization or
      blinding techniques — — ——————
   b. descriptive and stratified
      analyses
      (1) patient demographics — — ——————
      (2) investigator — — ——————
      (3) site — — ——————
      (4) surgical technique — — ——————
5. pooling of data justified — — ——————
6. statistical test given — — ——————
7. clear presentation of data — — ——————
8. statistical results stated — — ——————
9. statistical conclusions
   drawn from results — — ——————