CHAPTER TEN

STUDY QUALITY: INTERPRETING THE LITERATURE

The first nine chapters of this book have focused on the proper use of statistical methods in data analysis. This chapter will apply these concepts to interpretation of the medical literature with an emphasis on assessing study quality. As discussed in the INTRODUCTION, an understanding of basic statistical concepts and theory is essential to being able to interpret and use the medical literature. Several reviews of articles published in such journals as the New England Journal of Medicine, the American Journal of Surgery, and the British Journal of Surgery have documented that a significant number of published articles contain statistical errors or faults in study design. These errors are not intentional, but stem from the author's lack of knowledge of statistical theory and study design.To prevent these errors from being introduced into the literature, many journals now have statisticians on their editorial review boards to review manuscripts for the appropriate use of statistics. This is not yet universal, however, and it remains the reader's responsibility to actively assess and critique the medical literature to ensure that the conclusions proposed are statistically valid and applicable to patient care. By accepting a study's conclusions, without critically evaluating their veracity, we may be just as guilty of committing a Type I error as were the authors in proposing them. Thus, *caveat emptor* - "let the buyer beware", appropriately applies to the medical literature.

The following are questions which the reader should always ask when reviewing a clinical trial or research study. Based on the answers to these questions, one can determine whether the study and its conclusions are appropriate, statistically valid, and applicable to patient care.

STUDY DESIGN

- **Are the study patients comparable to my own patients?**

  Perhaps the first question which must be answered is whether the results of a study are applicable to an individual physician's patient population. This requires that the study patients be described in sufficient detail to allow such comparisons to be made. If the study patients are not adequately described, one can never be certain as to which patient population the study conclusions are applicable. An obvious example of this issue would be a Japanese study on the treatment of gastric cancer. Such a study would likely not be applicable to caucasian patients in the United States due to the differences in diet and genetic predisposition which separate Japanese and American patients. Further, earlier screening programs and differences in operative resection techniques might also make the results not applicable to Western patient populations. Thus, the nature of the sample population tested is crucial to determining whether the study conclusions are applicable to a clinical patient population.

- **What is the type of study and were patients collected appropriately?**

  The type of study performed, as discussed in Chapter Nine, may have a significant impact on the validity of the study conclusions. Randomized clinical trials are frequently the strongest study design and the most likely to demonstrate a true difference if one exists. Details of the randomization process should always be given to demonstrate that the groups are truly random. If a blinded study design is used, details of the blinding process should be presented. Non-randomized and retrospective study designs have a role in scientific research and their use should not necessarily cause the reader to ignore the results of such studies. When a non-randomized study design is used, study group comparability and attempts to minimize the effect of statistical bias (such as studying consecutive patients) must be demonstrated. Careful analysis of the Methods section should provide the reader with the information necessary to determine whether patients in the study groups were assigned appropriately and whether there is evidence of statistical bias.

- **Are the various patient groups being compared sufficiently similar to each other to allow accurate comparisons?**

  As we have previously discussed, in order to reach accurate and appropriate conclusions, the study and control groups must be as similar as possible to increase the likelihood that differences which are detected are due to true differences between the groups and not due to inequality of the patient samples. The distribution of age, sex, coexisting disease processes, and other pertinent variables among the patient groups should therefore be presented in the Results section to aid the reader in determining whether the study groups are comparable. In a randomized study, this information can be utilized to confirm that the randomization was appropriately performed.

- **Are the study hypotheses explicitly stated?**

  The study hypotheses should be clearly stated such that the purpose of the study is apparent to any reader. There should be evidence that the authors proposed their research hypotheses before beginning data collection. Conclusions based on *post hoc* hypotheses (i.e., those that were not proposed before data collection) should raise concerns regarding their validity. Statistical tests based on probability assume that hypotheses are proposed prior to data collection; creation of hypotheses after review of the raw data introduces bias and alters the probability value of the data. This is not to say that data should not be reviewed and new hypotheses proposed. These new hypotheses, however, should then be tested "from scratch" and not tested on the existing data.

- **Are the methods presented such that an investigator could reproduce the study if desired?**

  The methods by which the study was performed should be provided in sufficient detail to allow the reader to reproduce the study, and hopefully the same results, if desired. The Methods section should therefore contain details of the research protocol, the data acquisition techniques used, medications involved (including dosages), the diagnostic, monitoring, and laboratory equipment utilized (including model numbers and manufacturer), and any other aspects of the study which are not widely accepted in clinical or research practice and therefore familiar to the reader. The conclusions of a study which fails to present its methods and presents only a "black box" methodology should be considered suspect.

STATISTICAL ANALYSIS

- **Are the data presented such that the statistical analysis can be reproduced?**

  Sufficient information should be presented in the article to allow the reader to reproduce and confirm the author's statistical analysis. The data upon which the study conclusions are based should be provided either in graphical or tabular form. This allows the reader to evaluate the raw data for the presence of outliers and trends as well as to confirm that the appropriate statistical tests were performed (i.e., normally vs non-normally distributed data, discrete vs continuous data variables). This also provides the reader with the opportunity to apply other statistical methods to the data to determine whether the conclusions are strong enough to hold up to more than one method of analysis. Inclusion of graphs and other illustrations in the manuscript greatly facilitates this review of the data and avoids the need for tedious tables of raw numbers.

  A critical review of the data in any study begins by verifying the arithmetic. It is not uncommon to find that the data rows and columns do not add up correctly due either to errors in transcription or the constant revisions that go with wiriting any research manuscript. Usually, the correct numbers can be ascertained from the Results section of the paper. Occasionally, however, such errors may point to the presence of larger errors in the statistical analysis which should alert the reader to skeptically review the study conclusions.

- **Are descriptive statistics presented to summarize the data?**

  The mean and standard deviation for any data set should be included to demonstrate the central tendency and variability of the data. Where appropriate, the median, mode, and/or range should be included to demonstrate the presence of non-normally distributed data. The standard error of the mean should not be used as a measure of the variability of a data set.

- **Were appropriate statistical methods used to analyze the data?**

  Critical evaluation of the data allows the reader to determine whether appropriate statistical methods were used in the data analysis. This involves a description of the statistical tests used and the computer statistics package utilized (if any). Data which are clearly non-normally distributed, for example, are inappropriately analyzed using methods designed for normally distributed data (such as a t-test). Conclusions which are based on the use of multiple comparisons (i.e., multiple t-tests) should also be considered suspect for the reasons discussed in Chapter Seven. If p-values are used, the actual significance level should presented (i.e., $p=0.02$, not $p<0.05$) to allow the reader to identify the exact significance attached to the study results. Confidence intervals are of more value than p-values alone as they provide an idea of how likely the results are due to chance alone.

- **Were patients excluded from the data analysis? If so, was their outcome reported separately?**

  The inclusion and exclusion criteria should be presented in sufficient detail to allow the reader to identify the patient population being evaluated and to determine whether patients were included or excluded appropriately. If patients were excluded, it should be clearly stated whether they were excluded before or after randomization due to the potential for introducing bias into the data analysis. As we saw in Chapters Two, Three, and Nine, patients should not, in general, be excluded from data analysis as this introduces statistical bias and makes the results of a study less reliable. When patients must be excluded from the analysis, their outcome should be reported such that the reader can determine for himself whether inclusion of the excluded patients would have had a significant impact on the study conclusions. The reason for their exclusion should be presented and any measures taken to minimize the effect of bias on the results explained. The impact of patients who were "lost to follow up" should also be considered. In analyzing studies with patients who were "lost to follow up", the missing patients should be considered to have had the least favorable outcome and the study conclusions reconsidered. This is known as **sensitivity analysis**. If the study conclusions are still valid, those patients lost to follow up can likely be safely ignored.

- **Were sample size calculations performed before data collection?**

  Before beginning any clinical trial, an author should always calculate the sample size necessary to identify a predefined significant difference based on the significance level and power desired. Failure to do so may result in insufficient patients being studied leading to a Type II error (if a difference truly exists) or may result in more patients being studied than is necessary (resulting in a waste of research time and economic resources). Thus, a research study should describe in its Methods section the number of patients that were calculated to be necessary before beginning the trial, the chosen significance level and power, and the treatment difference which was considered to be significant. Unfortunately, few research studies provide this information.

STUDY RESULTS AND CONCLUSIONS

- **Are the dates of the study described?**

  The dates of data collection for each study group should be presented. This is especially important for retrospective studies as therapies and treatment techniques change over time. Patient groups treated at varying periods in time may therefore not be comparable.

- **Are the study endpoints and data presented clearly?**

  The manuscript should include a table which lists the study endpoints and the summary data associated with these variables (including the statistical significance if appropriate). This allows the reader to quickly summarize the results of the study and determine whether the conclusions are appropriate. Tables such as these are very useful in comparing the results of one study with another as well as in combining the results of several studies as in a meta-analysis.

- **Are the study conclusions supported by the data?**

  Upon reading the study conclusions, one should always review the Results section to confirm that data are presented to support each of the conclusions proposed. This is not always the case. It is not uncommon to find conclusions which, upon review of the data, are not supported by the study

results. Thus, there is more to interpreting the medical literature than reading the abstract and conclusions. The reader must critically review the Methods and Results of each study to determine whether the authors are correct in reaching the conclusions they have proposed, and whether the study contributes useful information to the clinical treatment of patients.

Effective use and accurate interpretation of the medical literature requires a knowledge of basic statistical theory and study design. Review of the past ten chapters in this book should provide the reader with the fundamental skills necessary to critically read the majority of the articles published in today's journals. The next two chapters deal with the other side of the medical literature: the writing and publishing of research abstracts and manuscripts.

SUGGESTED READING

1. Cohn KH, Weimar JA, Feldman JG. Critically evaluating surgical literature. Surgical Rounds 1992; 371-378.
2. Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, et al. A method for assessing the quality of a randomized control trial. Control Clin Trial 1981; 2:31-49.
3. Bailar JC. Science, Statistics, and Deception. Ann Intern Med 1986; 104:259-260.
4. Yancey JM. Ten rules for reading clinical research reports. Am J Surgery 1990; 159:533-539.
5. Lawrence W. Some problems with clinical trials. Arch Surg 1991; 370-378.
6. van der Linden W. Pitfalls in randomized surgical trials. Surgery 1980; 87:258-262.