

EDITORIAL VIEWS

Anesthesiology
57:73-74, 1982

Support versus Illumination: Trends in Medical Statistics

THERE is a curious attitude among clinicians and biomedical scientists regarding statistical analysis of data. When simple statistical tests are applied, the tendency is to believe that the study was well-designed or that the differences are especially real, since they can be demonstrated with elementary statistical techniques. When more complex or newer statistical approaches are used, there is a tendency to view the interpretations with skepticism and distrust. Some even view the use of more advanced statistical approaches as manipulation of the data.

Most clinicians and biomedical scientists are not skilled mathematicians and it is not uncommon for us to mistrust something which we do not understand completely. Biomedical statistics were commonly viewed as unimportant by most of us as medical students, and biostatistics courses were characterized by declining attendance, indifference, and outright hostility by some. New investigators learned the methods of their seniors, and these methods were usually elementary and simple approaches to data analysis. Some have even expressed the opinion that well-designed experiments do not require statistical analysis, since the results will be obvious. Perhaps these attitudes are best characterized by the familiar quip: "He uses statistics like a drunk uses a lamp post: for support and not for illumination."

Unfortunately, these attitudes have resulted in significant errors in the medical literature. These errors are costly, both in terms of research dollars spent for erroneous results, and in terms of patient welfare, since therapeutic decisions are often based upon experimental results.

Glantz has summarized both the historical and current aspects of the problem.¹ The essence of the problem in-

volves the misuse and abuse of *t* tests (either independent or paired *t*), the most commonly used tests in biomedical research.² Glantz critically analyzed the statistical methods used in articles appearing in *Circulation* and *Circulation Research* during 1977. Of those articles which used statistical methods, about 50 per cent (61 per cent for *Circulation Research* and 44 per cent in *Circulation*) applied *t* tests inappropriately when they should have used analysis of variance or a multiple group comparison technique. Inappropriate statistical analysis is not a new problem nor is it unique to the American medical literature. The problem was just as common in articles which appeared in the American medical literature in 1964,³ and it occurs commonly in British journals as well.⁴

While a variety of errors have been identified, the most common one is the use of the *t* tests to analyze more than two groups. The *t* test is a specific application of analysis of variance and it is intended to be used for two groups only. When used to compare multiple groups, it may identify "significant" differences which are not real. For example, if three groups are compared by *t* tests then there are three possible comparisons (A versus B, A versus C, and B versus C). If we accept $P \leq 0.05$ as the minimal level of significance, then this value is applied for each *t* test and the overall probability is no longer 0.05 per cent but 0.05 for each application, or approximately 15 per cent (0.05×3 ; the actual value is 13 per cent, but this is a close approximation). Similarly, if four groups are compared by *t* tests, then there are six possible comparisons and the overall probability is increased from 0.05 per cent to 30 per cent; there is approximately a 30 per cent chance of declaring differences where none exists. There are appropriate statistical methods which ensure that a defined level of significance (for example, $P \leq 0.05$) applies for *all* comparisons among groups.

Accepted for publication February 23, 1982.

These methods are well-summarized in common statistical texts. They include techniques for normally distributed data (Bonferroni's method, Tukey's method, Dunnett's method and Scheffe's method, for example) and nonparametric techniques as well (the Kruskal-Wallis method). The reader who seeks more information on these methods is referred to a recent review of several possible approaches to these and other common problems in biostatistics.⁵ An introductory text, which was stimulated by an awareness of these problems, was published recently and it explains many of these principles in a simple and understandable manner, even for those with limited mathematical skills.⁶

Over the past years we have seen the development of more rigorous scientific approaches to biomedical research. Studies which formerly involved only a single dose of a drug now commonly include dose-response data. Blood levels of drugs are reported with increasing frequency. Randomized allocation to various treatment groups, blind or double-blind experimental protocols, and the administration of a placebo in a control group are common examples of the more rigorous approaches which are employed in biomedical research. Despite these advances in the scientific method, biomedical scientists have been slow to adopt sensitive, reliable, and appropriate statistical techniques which allow proper interpretation of results. A product is no better than its weakest link. If valid data are analyzed improperly, then the results become invalid and the conclusions may well be inappropriate. At best, the net effect is to waste time, effort, and money for the project. At worst, therapeutic decisions may be based upon invalid conclusions and patient's wellbeing may be jeopardized.

Anesthesiology
57:74-75, 1982

Barbiturates for Brain Resuscitation: Yes and No

PREVIOUS STUDIES concerned with the question of whether barbiturates can favorably alter brain outcome following a period of complete global ischemia have yielded either a yes or no answer and have created controversy and confusion. Dr. Shapiro's group in San Diego has now addressed that question and reports in this issue that the answer is neither yes nor no; rather it is yes *and* no.¹ It is hoped that this study represents the first step toward clarifying what has been a near-hopeless muddle of contradictory reports.

The study was done in cats. Complete ischemia was induced by electrically fibrillating the heart. The periods of circulatory arrest ranged from 12 to 16 min. Resus-

While the journal cannot and will not endorse any single method or approach for statistical analysis, the readers will note an increasing level of sophistication in the techniques used to analyze the data which appear in *Anesthesiology*. We owe that to our patients, whose medical care is based on the statistical analysis of experimental data. We owe it to our readers, who expect valid results to be presented in the journal. Finally, we owe it to the authors, many of whom are competing for research support in an environment which requires that all aspects of the research, including the statistical methods, receive proper attention in order to ensure valid conclusions.

DAVID E. LONGNECKER, M.D.
Professor of Anesthesiology
Department of Anesthesiology
Box 238
University of Virginia Medical Center
Charlottesville, Virginia 22908

References

1. Glantz SA: Biostatistics: How to detect, correct and prevent errors in the medical literature. *Circulation* 61:1-6, 1980
2. Feinstein AR: Clinical biostatistics. XXV. A survey of the statistical procedures in general medical journals. *Clin Pharmacol Ther* 15:97-107, 1974
3. Schor S, Karten I: Statistical evaluation of medical journal manuscripts. *JAMA* 195:1123-1128, 1966
4. Gore SM, Jones IG, Rytter EC: Misuses of statistical methods: Critical assessment of articles in *BMJ* from January to March, 1976. *Br Med J* 1:85-87, 1977
5. Wallenstein S, Zucker CL, Fleiss JL: Some statistical methods useful in circulation research. *Circ Res* 47:1-9, 1980
6. Glantz SA: *Primer Of Biostatistics*. New York, McGraw-Hill, 1981

citation was always accomplished within four minutes. The barbiturate was thiopental; the dose was 60 mg/kg initiated five minutes after resuscitation. And no, among the survivors there were no differences in the neurologic deficits of treated animals compared to controls. But yes, there *were* significantly fewer deaths due to neurologic dysfunction in the treated group. And yes, there was a significantly lesser incidence of a seizure-like pattern appearing in the EEG of treated animals following resuscitation.

The design of this study and the close control that was achieved of the many difficult variables encountered in this type of protocol should stand as a model for future investigators interested in studying the chronic effects of a period of complete global ischemia. Todd *et al.*¹ should

Accepted for publication February 3, 1982.