computers which allow rapid computation of complex statistical equations. Continued development of the computer industry will probably stimulate further interest and development of "more appropriate" and complex medical statistical methods.

Statistical analysis of anesthesiology research data can be likened to the use of a cookbook. Level one users of statistical methods can read the recipes and have sufficient arithmetic and/or computer skills to insert numbers and crank out results. Level two users not only can use the recipes, but also understand, more or less, the assumptions and restrictions for each statistical tool. Level three users can do all of the above plus recreate old recipes and write new ones.

In our observation, the busy clinician reader is often able to read and use the simple recipes (*t* test and chi-square analysis); however, more advanced methods are out of reach because of mathematical unfamiliarity and absence of computational support. The typical research anesthesiologist usually has level two skills for simple statistics (*t* test and chi-square analysis); however, his understanding and use of the more advanced analyses mentioned by Dr. Longnecker (analysis of variance, multivariate analysis of variance, multiple comparisons, multiway contingency table analysis, etc.) is hindered by inadequate foundations in probability theory, calculus, matrix algebra, sampling theory, etc. As editorial reviewers and editorial board members are drawn from the more published and knowledgeable of these researchers, their skills at times may be slightly more advanced. Even academic anesthesiologists/editors are not at level three.

How should readers, researchers, and reviewers respond to this new standard of statistical excellence enunciated by the editorial? It may be appropriate that researchers like ourselves spend more than 30% of our research dollar and research time for statistical consultation, computer programming and computation, and other statistical self education. It also may be appropriate that the tools of reviewers include statistical glossaries, the most up-to-date statistical text books, and paid statistical consultation. But what will the clinician reader do when faced with a journal using incomprehensible statistical methods? One alternative for the non-expert statistical reader of ANESTHESIOLOGY is to simply accept the published statistical method as editorially approved and thus appropriate. Another alternative might be to simply ignore the statistics utilized, scan the data, and make an unsupported "gut feeling" judgement. These approaches can only lead to a more superficial appraisal of research reports.

Without greater efforts devoted to the statistical skills of the clinical reader, even greater estrangement of the clinician from the journal will occur. Some possible solutions include: 1) increased statistical teaching during residency; 2) increased attention to statistical topics during written and oral Board examinations; 3) journal review articles on statistics; 4) refresher course lectures on statistics; 5) frequent exhortations by prominent anesthesiologists for better statistical skills; and 6) statistical teaching in ASA Self Education and Evaluation Programs.

The solution to the posed problem is not easy. However, an appropriate and expedient solution to the problem seems crucial if medical statistics are to *support* rather than *strangulate* future anesthesiology research reporting to the clinician.

THEODORE H. STANLEY, M.D.
Professor of Anesthesiology/
Research Professor of Surgery
The University of Utah School of Medicine
Salt Lake City, Utah 84132

NATHAN L. PACE, M.D.
Associate Professor of Anesthesiology
The University of Utah School of Medicine
Salt Lake City, Utah 84132

REFERENCE

1. Longnecker DE: Support versus illumination: Trends in medical statistics. ANESTHESIOLOGY 57:73–74, 1982

(Accepted for publication September 28, 1982.)

Anesthesiology
58:298–300, 1983

Absence of Evidence Is not Evidence of Absence

To the Editor:—We applaud Longnecker's recent appeal for higher standards of statistical analysis,[1] but we disagree about Glantz's[2] finding that inappropriate use of *t* tests is the most frequently committed error. Using *t* tests to compare more than two means is probably the most frequently committed *Type One Error*[3] (errors that give false-positive results), but *Type Two Errors*[3] (false-negative results) are probably the most frequently committed, and least frequently detected, statistical errors in both social and hard science (including medical science).[4]

Type Two Errors result in the implication, if not the explicit conclusion, that a lack of statistical significance for an observed difference indicates a lack of actual dif-

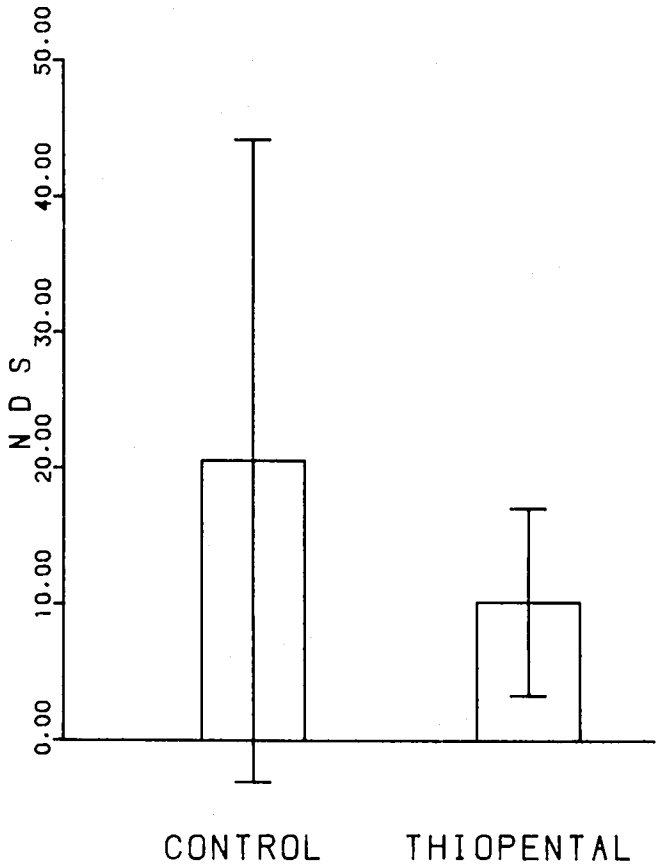

FIG. 1. 95% confidence intervals for the means of two samples. Control: N = 8, mean = 20.6 ± 10 (SE). Thiopental: N = 9, mean = 10.2 ± 3.

ference. If you sent a diffident colleague to look for a piece of equipment and he returned in three minutes with the pronouncement, "I didn't see it, I don't think we have one," you might well ask: "How hard did you look?" Likewise, if a colleague's experimental results cause him to conclude that a drug has no effect, you should ask: "How hard did you look?" In more formal terms, the question is: "what was the probability of falsely failing to reject the null hypothesis?" That is a calculable probability called "beta." When Type Two Errors are committed, authors almost invariably report the attending "alpha"—the probability of falsely rejecting the null hypothesis (that is, the significance level, or *P* value). Unfortunately, beta cannot be deduced from alpha.When you wish to check the veracity of a lack of difference between means (or correlation coefficients, or any other parameter) you need *Power* (power = 1 − beta), *not* significance (significance = 1 − alpha).

To illustrate this point, we decided to find an example of a Type Two Error. We did not have to look far. The first article in the issue containing Longnecker's editorial is a case-in-point.[5] One of the paper's two major conclusions (explicitly stated) is based on a Type Two Error.

The authors report that high-dose thiopental has "no effect on the neurologic function of survivors" after a 12-min period of electrically induced ventricular fibrillation. In a control group of eight cats that survived seven days, the mean neurologic deficit score (NDS) was 20.6 ± 10 (SE). In the test group, it was 10.2 ± 3. This is not a statistically significant difference at the author's chosen significance level of $P < 0.01$ (or even at $P < 0.05$). However, this does not stand as evidence that the two means are the same, or even very similar. Drawing the 95% confidence interval (CI) around each of the reported means gives an indication of their similarity: the 99% CI would show an even greater potential disparity (see Fig. 1).

Clearly, the confidence intervals overlap—so the observed difference between the means may be the result of sampling error. Just as clearly, within the lowest conventionally accepted CI, the mean for the thiopental sample may have come from a population whose true mean is as low as 3.3 while the true control mean could be as high as 44. In short, the standard error of the difference between these means is 10.44, so one can be 95% certain that their actual difference is somewhere between 0 and 33, or 99% certain that their difference is between 0 and 41. Since the NDS only goes from 0 to 100, the data indicate plenty of room for a highly significant effect (clinically significant, as distinct from statistically significant).

The best guess about the true difference between the means is 10.4—the observed difference. Our point is that the conclusion that the drug had no effect is not only not justified by the data, it contradicts the data. Our point is also that the paper referred to is not an exception when it comes to statements of "no effect." Indeed, we agree with Michenfelder[6] that, otherwise, the paper "should stand as model for future investigators."

Statistical procedures have been designed intentionally to assess evidence for effects. Unfortunately, little attention has been paid to procedures that assess evidence for a lack of effect. Cohen[4] has devoted much of his professional career to Power Analysis, but the applied literature suggests that his crusade is yet to be appreciated. This is partially because showing a lack of effect requires much larger samples than detecting an effect. Also, figuring beta is much more involved than figuring alpha. A good rule of thumb for assessing the conclusion "little or no effect" is to ask whether a *smaller* sample size would have made the conclusion *more* likely (according to the procedure used). When that is the case, as it would be in the example cited here, you can be sure that something is wrong.

Another reason that Type Two Errors have been all but ignored is because they are often unimportant. In the social sciences there is usually considerable debate over the meaning of true-positive results (valid statistically significant differences). Social scientists tend to be at even more of a loss over true-negative results. So it is rare that a false-negative result impedes progress. However, when the conclusion at hand is that two drugs have similar effects, or that one drug has no effect because its observed effect is not statistically different from control, the most important question that can be asked is: "How hard did you look?"

Glantz[2] and others have made suggestions for improving the quality of published analyses, and all of the suggestions we have seen should be taken. The problem is that they seldom are taken. The suggestion, "submit all analyses to expert reviewers," is editorially impractical. We would like to make a suggestion in this regard: require that an appendix containing *all* raw data be submitted with any article that contains a statistical inference—with the understanding that a copy of that appendix will be sent by the journal to any reader willing to cover copying and handling costs. The logic here is that raw data usually belie unjustified inferences. Authors would be more careful if they knew that anyone would be able to perform a truly independent analysis of their results. ANESTHESIOLOGY could be the vanguard of such a revolution.

JOHN HARTUNG, PH.D.
Instructor

JAMES E. COTTRELL, M.D.
Professor and Chairman

JOSEPH P. GIFFIN, M.D.
Assistant Professor

Department of Anesthesiology
State University of New York
Downstate School of Medicine
Brooklyn, New York 11203

REFERENCES

1. Longnecker DE: Support versus illumination: Trends in medical statistics. ANESTHESIOLOGY 57:73–74, 1982
2. Glantz SA: Biostatistics: How to detect, correct, and prevent errors in the medical literature. Circulation 61:1–6, 1980
3. Hays WL: Statistics. New York, Holt, Rinehart, and Winston, 1963, pp 280–281
4. Cohen J: Statistical Power Analysis for the Behavioral Sciences. New York, Academic Press, 1977
5. Todd MM, Chadwick HS, Shapiro HM, Dunlop BJ, Marshall LF, Dueck R: The neurologic effects of thiopental therapy following experimental cardiac arrest in cats. ANESTHESIOLOGY 57:76–86, 1982
6. Michenfelder JD: Barbiturates for brain resuscitation: Yes and no. ANESTHESIOLOGY 57:74–75, 1982

(Accepted for publication September 17, 1982.)

Anesthesiology
58:300, 1983

Use of the Left-entry Laryngoscope Blade in Patients with Right-Sided Oro-Facial Lesions

To the Editor:—Malformations and tumors that deform the right side of the face and oropharynx usually render the visualization of the larynx with a regular laryngoscope more difficult. One frequently has to resort to fiberoptic instrumentation or blind nasal intubation. We have found the "left-entry" laryngoscope blade* also clinically useful in these patients with right hemifacial lesions.

This blade was originally designed for the left-handed anesthesiologist. It actually is a mirror image of the standard Macintosh or Miller blade, in that the ridge or curve is placed on the right. Thus, when the laryngoscope blade is introduced on the left side of the mouth, the tongue is pushed to the right, towards the lesion. A better exposure of the larynx is obtained and an endotracheal tube, introduced from the left, is positioned more easily.

MARIA R. G. LAGADE, M.D.
Assistant Professor of Anesthesiology

PAUL J. POPPERS, M.D.
Professor and Chairman
Department of Anesthesiology
State University of New York
Stony Brook, New York 11794

(Accepted for publication September 17, 1982.)

* The "left-entry" laryngoscope blade is made in different sizes and configurations by the Foregger Company, Hauppauge, New York 11788.