Title: INFERENTIAL STATISTICAL TESTS IN ANESTHESIOLOGY

Authors: M. J. Avram, Ph.D., C. A. Shanks, M.D., M. H. M. Dykes, M.D., M.Ed., A. K. Ronai, M.B., B.S., Ph.D., and W. M. Stiers, M.A.

Affiliation: Department of Anesthesia, Northwestern University, Chicago, Illinois 60611
Department of Counseling Psychology, Northwestern University, Evanston, Illinois 60201

INTRODUCTION. Critical evaluators of the biomedical literature report that statistical methods are used incorrectly in nearly half of the articles in which they are used.[1] The most important way to correct these errors is for the editors of the journals to insist on the proper use of appropriate statistical methods. ANESTHESIOLOGY has recently announced a new standard of statistical excellence.[2] The purpose of the present study was to compare the appropriateness of inferential statistical analyses in ANESTHESIOLOGY published in a six month period prior to setting the new standard with that in a later six month period.

METHODS. The inferential statistics of the original articles appearing in ANESTHESIOLOGY Volumes 55 and 59 were critically evaluated by each of the authors in an order determined by a random number generator. Data were classified as nominal, ordinal, or interval according to the criteria of Siegel.[3] The inferential procedures were judged appropriate or not using the "Statistical guidelines for contributors to medical journals" of Altman et al.[4] and the texts of Glantz[1] and Siegel[3]. The primary errors in TABLE 1 were those encountered in the chronologic order of the requirements of the test (e.g. if serial blood pressures were measured and multiple comparisons to the baseline pressure were made using unpaired t-tests the primary error was considered to be application to related samples of a test for independent samples even though the error of multiple applications of a test to the same data without correction was also made). The observed frequencies of nonparametric test errors and of parametric test errors in the two volumes were compared with the expected frequencies using the chi-square test statistic with the Yates correction for continuity; the null hypothesis was rejected when $P < 0.05$.

RESULTS. There were 90 inferential statistical tests (15 nonparametric and 75 parametric tests) in Volume 55 of ANESTHESIOLOGY and 106 such tests (23 nonparametric and 83 parametric) in Volume 59 (Table 2). The percentage of all inferential statistical tests that were erroneous was 80% in Volume 55 and 71% in Volume 59. The incidence of errors in nonparametric tests was 67% and 59% while that of parametric tests was 85% and 72% in Volumes 55 and 59, respectively. The frequencies of errors in nonparametric and in parametric testing in Volume 55 were not shown to differ from those in Volume 59. The principle sources of primary errors in both volumes were the application to related samples of a test for independent samples (and vice-versa) and multiple applications of a test to the same data without correcting for multiple applications; together they account for 83% of the primary errors in Volume 55 and 70% of the primary errors in Volume 59. There were no unidentified inferential statistical tests in Volume 55 but 8% of erroneous tests in Volume 59 were unidentified. Other origins of primary error include inadequate numbers in the various categories for the chi-square test (1% and 5% in Volumes 55 and 59), application of tests of interval data to ordinal data (7% and 9% in Volumes 55 and 59), and inappropriate post-hoc testing after variance analysis (8% and 7% in Volumes 55 and 59).

TABLE 1. ORIGINS OF PRIMARY STATISTICAL ERRORS

ANESTHESIOLOGY Volume/Year	55/1981	59/1983
Unidentified Test	0	6
Ordinal Data, Interval Test	5	7
Inadequate No.s for Chi Square	1	4
Related Data, Independent Test (and Vice-Versa)	24	22
Inappropriate Follow-up to Variance Analysis	6	5
Multiple Applications of an Uncorrected Test	38	31

TABLE 2. ERRORS IN APPLYING INFERENTIAL STATISTICS

ANESTHESIOLOGY Volume/Year	55/1981	59/1983
Number of Manuscripts	67	57
Nonparametric Tests	10 of 15	15 of 23
Both t-Tests	37 of 41	20 of 30
All ANOVA and ANCOVA	27 of 33	34 of 45
Other Parametric Tests	0 of 1	6 of 8
All Parametric Tests	64 of 75	60 of 83

DISCUSSION. The incidence of incorrect nonparametric and parametric inferential statistical analyses in ANESTHESIOLOGY did not differ in the two volumes. The shortest interval between the editorial calling for a new standard of excellence[2] and any acceptance data of the articles reviewed in either volume was five months. The Editorial stressed abuse of the t-test and, while it was not the purpose of the present study to assess this specifically, there appeared to be a change in its usage pattern. Table 2 shows a reduction in the number of t-tests (both unpaired and paired) with a concomitant increase in other tests of variance, both incorrect and total. Primary sources of error enumerated in Table 1 are the result of applications of tests inappropriate to the type of data presented or violation of simple concepts underlying the statistical models. The sources of error commonly encountered in the literature are seldom of a complex, controversial nature but, like those noted in the present study, are easily corrected.[1] Recognition of potential sources of error and increasing access to sophisticated computer programs should make it easier for investigators to use statistical analyses appropriate to their specific needs.

REFERENCES.

1. Glantz SA. Primer of Biostatistics. McGraw-Hill Book Co., St. Louis, 1981
2. Longnecker DE. Support versus illumination: Trends in medical statistics. Anesthesiology 57:73-74, 1982
3. Siegel S. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill, New York, 1956
4. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. Brit Med J 286:1489-1493, 1983