

Regarding the Misuse of *t* Tests.

To the Editor:—Reduction of Type 1 errors (alpha, false-positive results) seems to be the emphasis of new editorial policies on medical statistics.¹ This will necessarily increase the incidence of Type 2 errors (beta, false-negative results) in science, published or not, because of the mutual dependence of alpha and beta errors.² Hartung *et al.*³ recently have suggested that beta errors already are the most frequent type of statistical error in the literature. These authors³ focused their attention on what they believe is the most harmful aspect of beta error, namely, false acceptance of the null hypothesis. In light of the adoption of more stringent editorial policies, I believe the other aspect of beta error, namely, false rejection of the alternate hypothesis, deserves further consideration.

One of the targets singled out in the campaign to improve medical statistics is the analysis of data by multiple *t* tests.^{1,2} Even when global analysis of variance (ANOVA) is used in these experiments, multiple *t* tests almost invariably are justified and even required following the ANOVA, depending on the results of that analysis and the nature of the hypotheses. In order to perform multiple comparisons while maintaining a given level of alpha for the entire set of tests, a variety of multiple comparison procedures (MCP) have been developed to determine the appropriate critical values for these *t* tests under various conditions. Thus, the MCP range in complexity from the reasonably simple Bonferroni method² to much more elaborate methods.⁴ According to these more or less complicated formulas, the criterion of significance (alpha) for any single test is made much more conservative than if only one test were made. It must be remembered that in reducing alpha to some overall level (arbitrary to begin with and traditionally set at 0.05 or 0.01) the MCP simultaneously increases beta error.

A much simpler (therefore better?) alternative, which would not increase beta error, would be to use the usual *t* test (or the modified *t* test using the mean squared error from the ANOVA²) together with the usual alpha level and simply alert the reader that having performed *n* simultaneous tests of significance, there is an increased chance $[1 - (1 - \alpha)^n]$ of finding one or more of these tests "significant" due to sampling error. Regardless of

how the statistical tests turn out, we all need to judge the significance of effects noted in an article in terms of their importance, not just their validity.

It seems clear that greater editorial attention to statistical methods such as the appropriate use of control groups, randomization, *etc.*, will improve the quality of science. However, rather than being an improvement, the new emphasis on the MCP simply may be a subtle shift to a more conservative alpha with perhaps overlooked or underestimated losses in statistical power. Statistical power is typically low in science because of the small size of the treatment effects (statistical support is unnecessary for the large, obvious effects) and the small numbers of samples due to time and budgetary constraints. If my perception of the MCP situation is correct, the net result is that larger sample sizes (investment of added time and resources) now will be needed to achieve the traditional level of statistical significance needed for demonstration of an effect. Although more rigorous and conservative treatment of scientific data gives the appearance of an improvement in science, I wonder whether the type of increased statistical precision exemplified by the MCP is worth either the added resources needed to achieve it or the increased complexity it introduces to the conduct and understanding of science.

CARL F. SCHAEFER, PH.D.

Associate Professor of Anesthesiology
University of Oklahoma Health Sciences Center
Oklahoma City, Oklahoma 73190

REFERENCES

1. Longnecker DE: Support versus illumination: Trends in medical statistics. *ANESTHESIOLOGY* 57:73-74, 1982
2. Glantz SA: *Primer Of Biostatistics*. New York, McGraw-Hill, 1981
3. Hartung J, Cottrell JE, Giffin JP: Absence of evidence is not evidence of absence. *ANESTHESIOLOGY* 58:298-300, 1983
4. Games PA, Keselman HJ, Rogan JC: Simultaneous pairwise multiple comparison procedures for means when sample sizes are unequal. *Psych Bull* 90:594-598, 1981

(Accepted for publication September 30, 1983.)