THE JOURNAL OF
THORACIC AND
CARDIOVASCULAR
SURGERY

Statistical Methods

Because data and their analyses often constitute the supporting evidence for inferences drawn from studies submitted to the *Journal*, the Editors remind authors that all manuscripts with statistical analyses undergo statistical review by the *Journal*. The review includes study design, analysis, interpretation, and reporting. The manuscripts are not published without an acceptable rating by the statistical editorial staff. Therefore, to minimize revision and delays, authors should request review of such manuscripts by a statistician prior to submission. This is best done by involving a statistician as a collaborator from the inception to completion of the study.

The Editors subscribe to the statistical guidelines contained in the "Uniform Requirements for Manuscripts Submitted to Biomedical Journals" (*Ann Intern Med*. 1997;126:36-47). These are as follows:

> Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as the use of *P* values, which fails to convey important quantitative information. Discuss the eligibility of experimental subjects. Give details about randomization. Describe the methods for and success of any blinding observations. Report complications of treatment. Give numbers of observations. Report losses to observation (such as dropouts from a clinical trial). References for the design of the study and statistical methods should be to standard works when possible (with pages stated) rather than to papers in which the designs or methods were originally reported. Specify any general-use computer programs used.
>
> Put a general description of methods in the Methods section. When data are summarized in the Results section, specify the statistical methods used to analyze them. Restrict tables and figures to those needed to explain the argument of the paper and to assess its support. Use graphs as an alternative to tables with many entries; do not duplicate data in graphs and tables. Avoid non-technical uses of technical terms in statistics, such as "random" (which implies a randomizing device), "normal," "significant," "correlations," and "sample." Define statistical terms, abbreviations, and most symbols.

We recognize that there are a number of schools of differing statistical philosophy, and we take these differences into account. Nevertheless, over the years, a number of specific items have been raised repeatedly by statisticians when reviewing manuscripts. The Editors have compiled a list of these and present them in the form of guidelines, with the intent of being helpful to authors, not prescriptive. The guidelines have been formulated as a checklist and appear twice yearly, in the January and July issues.

Statistical Guidelines for Manuscript Submission

Authors should prepare manuscripts in light of the guidance provided in "Notes from the Editors" (*J Thorac Cardiovasc Surg*. 1996;112:209-20). Authors should also consult published checklists such as Gardner MJ, Machin D, Campbell MJ, "Use of Check Lists in Assessing the Statistical Content of Medical Studies" (*BMJ*. 1986;292:810-2) and Bailar JC, Mosteller F, "Guidelines for Statistical Reporting in Articles for Medical Journals" (*Ann Intern Med*. 1988;108:226-73). For papers reporting events after heart valve procedures, consult Edmunds LH Jr, Clark RE, Cohn LH, Grunkemeier GL, Miller DC, Weisel RD, "Guidelines for Reporting Morbidity and Mortality After Cardiac Valvular Operations. Ad Hoc Liaison Committee for Standardizing Definitions of Prosthetic Heart Valve Morbidity of The American Association for Thoracic Surgery and The Society of Thoracic Surgeons" (*J Thorac Cardiovasc Surg*. 1996;112:708-11). Manuscripts reporting the results of randomized controlled trials (RCTs) should include the CONSORT flow diagram showing the progress of patients throughout the trial (http://www.consort-statement.org). The CONSORT checklist also should be completed and submitted with the manuscript. See also Piantadosi S, Gail M, "Statistical Issues Arising in Thoracic Surgery Clinical Trials." In: Pearson FG, Deslauriers J, Ginsberg RJ, Hiebert CA, McKneally MF, Urschel HC Jr, editors. Thoracic Surgery. New York: Churchill Livingstone; 1995. p. 1652-70, and Kirklin JW, Barratt-Boyes BC, "The Generation of Knowledge From Information, Data and Analyses." In: Cardiac Surgery, New York: Churchill Livingstone; 1993. p. 249-82.

Checklist

Notation and terminology. Explain meaning of notations such as SE, SD, CL, or CI in abstract, methods, and tables when they first appear. **Distinguish between a variable**, an item that can take on different values for each subject or observation, such as temperature and blood pressure, **and a parameter**, a constant, such as the mean. **Distinguish between prevalence,** a proportion of subjects or observations, and **incidence or rate**, a quantity expressed per unit of time. **Distinguish between multivariable**, referring to several predictor or explanatory variables or risk factors, **and**

multivariate, referring to simultaneous analysis of several outcome variables. The latter is a relatively recent change in statistical definitions. An accessible source of statistical terms can be found in Piantadosi S, Kirklin J, Blackstone E. Statistical Terminology and Definitions. In: Pearson FG, Deslauriers J, Ginsberg RJ, Hiebert CA, McKneally MF, Urschel HC Jr, editors. Thoracic Surgery. New York: Churchill Livingstone; 1995. p. 1649-52.

Distinguish between descriptive statistics and expression of uncertainty of parameter estimates. When describing the values for a variable (eg, baseline information), descriptive statistics such as median and quartiles or the mean and standard deviation are appropriate. In particular, when the distribution of values is skewed, nonparametric descriptors such as quartiles are appropri-ate, not mean and standard deviation. In contrast, the uncertainty of parameter estimates is expressed commonly in terms of confidence limits (intervals) or, when these are symmetric, the standard error.

P *values.* Although it is not possible with all statistical tests, and although it is contrary to some statistical philosophies, we recommend use of exact *P* values unless $P < .001$ or $P > .2$ as measures of evidence. Thus, we recommend against use of "$P < .05$ was considered significant" or the abbreviation NS, or symbols representing various levels of statistical significance.

Authors sometimes interpret large *P* values to mean, "There is no difference between groups." This is generally contrary to the facts because differences are evident. We prefer the use of the phrase, "The differences could be due to chance ($P > .2$)."

The term "significant" is ambiguous, because it fails to distinguish so-called statistical from clinical significance. We recommend against the use of the term "significant," suggesting that a synonym such as "important" be used to signify "clinical significance." Statistical significance often can go unstated when accompanied by a *P* value.

P values alone do not convey the magnitude of the effect or difference, nor its precision. Therefore, we will recommend the use of estimates of strength (eg, coefficients, odds ratios, hazard ratios) and confidence limits (intervals), tolerance intervals, or credibility intervals to convey this information. Use of these intervals is particularly important when the conclusion is that no effect or association was observed (equivalence).

Other specific expressions of uncertainty. In many settings, particularly when the number of patients or subjects is small, proportions should be accompanied by confidence limits (intervals). We do not prescribe a specific confidence interval, such as 95%, or intervals equivalent to ± 1 standard deviation, since the appropriate confidence limits may vary with the situation. A consistent schema for expressive variability would include ± 1 standard deviation for normally distributed continuous variables, 15 and 85 percentiles for skewed distributions, and 70% confidence limits for proportions. Increasingly, approximations to parameter estimation and measures of uncertainty are being supplemented by computer-intensive resampling (bootstrap) methods.

Presentation of time-related events. In most circumstances, we recommend that the following information accompany presentations of time-related events: point estimates, preferably at the time of each event using a product limit method; asymmetric confidence limits at periodic intervals; and the number of patients at risk at periodic intervals.

Nonrandomized comparisons. Unlike experimental comparison studies that nearly always should be randomized, randomization in the clinical setting is often neither feasible nor ethical. Increasingly, multivariable matching methods for adjusting for ascertainable selection bias are becoming prevalent, well understood, and accessible (eg, use of propensity scores).

Multiple group comparisons and repeated measurements. In comparing three or more groups, statistical methods appropriate for multiple group comparisons and contrasts should be employed. If these groups have a natural ordinal relationship one with another, then methods that account for trend should be employed. When multiple measurements are obtained across time in the same patient or subject, methods of longitudinal data analyses (a relatively new field of statistics that has supplanted traditional repeated measures methodology) are recommended.

Multivariable analyses. Many studies lend themselves to methods that take into account simultaneously multiple variables (risk factors, predictor variables, independent variables, co-variables). Reports of multivariable analyses must state the model used, all variables that were examined, how the variables were coded in the final models, the extent of testing for interactions, the degree to which conformity to a linear gradient (for continuous or ordinal variables) was examined and accounted for, the degree to which the assumption of proportional hazards was tested when using such models, colinearity of variables, possibility of overfitting, and methods used for model validation.

Statistical Collaboration/Review Release Statement

I am a person with a master's or doctoral degree (or equivalent) in biostatistics or related field. I have experience in the design, analysis and interpretation of biomedical data of the type used in this paper. I take scientific responsibility for the analysis of the data and interpretation of those analyses as presented in the final version of this manuscript.

Statistician's signature: _____ Date: _____

There are no statistical methods presented in this paper.

Corresponding author's signature: _____ Date: _____