

GUIDELINES FOR STATISTICAL REPORTING IN ARTICLES FOR
MEDICAL JOURNALS
AMPLIFICATIONS AND EXPLANATIONS

John C. Bailar III, M.D., Ph.D.; and Frederick Mosteller, Ph.D.; Boston, Massachusetts
Annals of Internal Medicine, 15 February 1988. 108:266-273.

The 1988 edition of the *Uniform Requirements for Manuscripts Submitted to Biomedical Journals* includes guidelines for presenting statistical aspects of scientific research. The guidelines are intended to aid authors in reporting the statistical aspects of their work in ways that are clear and helpful to readers. We examine these guidelines for statistics using 15 numbered statements. Although the information presented relates to manuscript preparation, it will also help investigators in earlier stages make critical decisions about research approaches and protocols.

[MeSH terms: clinical protocols; clinical trials; eligibility determination; manuscripts, medical; probability; random allocation; statistics. Other indexing terms: blinding; blocking; confidence intervals; International Committee of Medical Journal Editors; matching; *P* values; software; statistical methods; stratification; study design; treatment complications; Uniform Requirements for Manuscripts]

From the Department of Health Policy and Management, School of Public Health Harvard University; Boston, Massachusetts; Office of Disease Prevention and Health Promotion, U.S. Dept. of Health and Human Services, Washington, D.C.; Department of Epidemiology and Biostatistics, McGill University; Montreal, Quebec, Canada.

In 1979, the group now known as the International Committee of Medical Journal Editors first published a set of uniform requirements for preparing manuscripts to be submitted to their own journals. These uniform requirements have been revised several times [\(1\)](#), and have been widely adopted by other biomedical journals. In the 1988 revision [\(2\)](#), the Committee added guidelines for presenting and writing about statistical aspects of

research. The purpose of these guidelines is to assist authors in reporting statistical aspects of their research in ways that will be responsive to the queries of editors and reviewers and helpful to readers.

We present the statistical guidelines as a sequence of 15 numbered statements, and amplify and explain some of the reasoning behind the guidelines. The material focuses on manuscript preparation, but it should also be helpful at earlier stages when critical decisions about research approaches and protocols are made. This article does not provide a short course in statistics because we can deal with only a few important aspects of what should be reported in publications about work already done, but we provide references to general statistical texts. The International Committee is not responsible for these amplifications; however, we have tried to present the spirit of the Committee's discussions as well as our own views.

The International Committee's statistical guidelines 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 sole reliance on statistical hypothesis testing, such as the use of P values, which fails to convey important quantitative information. Discuss eligibility of experimental subjects. Give details about randomization. Describe the methods for, and success of, any blinding of observations. Report treatment complications. Give numbers of observations. Report losses to observation (such as dropouts from a clinical trial). References for study design and statistical methods should be to standard works (with pages stated) when possible, rather than to papers where designs or methods were originally reported. Specify any general use computer programs used.

Put general descriptions 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 nontechnical uses of technical terms in statistics, such as "random" (which implies a randomizing device), "normal," "significant," "correlation," and "sample." Define statistical terms, abbreviations, and most symbols.

Our general approach is that scientific and technical writing should be comprehensible at the first reading for the average reader who is knowledgeable about the general area but not a subspecialist in the specific topic of investigation.

1. Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results.

Authors should report which statistical methods they used, and why. In many instances they should also report why other methods were not used, although this is rarely done.

Readers must be told about weaknesses in study design and about study strengths in enough detail to form a clear and accurate impression of the reliability of the data, as well as any threats to the validity of findings and interpretations. Such details are often omitted, although investigators probably know them [\(3, 4\)](#).

The researcher must decide which statistical measures and methods are appropriate, given that a statistical goal has been defined. Investigators often have a choice: Mean or median? Nonparametric test or normal approximation? Adjustment, matching, or stratification to deal with confounding factors? Choosing statistical methods generally requires an appreciation of both the problem and the data, and an experienced biostatistician, statistician, or epidemiologist can often provide substantial help. This help ideally begins before the study, because the foundation for reporting one's findings is laid before the study even begins.

Trying several reasonable statistical methods is often appropriate, but this strategy must be disclosed so that readers can make their own adjustments for the authors' industriousness or skill in fishing through the data for a favorable result. Whatever statistical task is defined, it is inappropriate, and indeed unethical, to try several methods and report only those results that suit the investigator. Results of overlapping methods need not be presented separately when they largely agree, but authors should state what additional approaches were tried, and that they did agree. Of course, results that do not agree also should be given, and investigators may sometimes find that such disagreements arise from important and unexpected aspects of the data.

Units should always be specified in text, tables, and figures, although not necessarily every time a number appears if the unit is clear to the reader. Often, careful choice of units of measurement can help clarify and unify the study question, biological hypothesis, and statistical analysis. Careful reporting of units can also help to avoid

serious misunderstanding. Are quantities in milligrams or millimoles? Are rates per 10 000 or per 100 000? Does a figure show number of different patients, or number of myocardial infarcts among those patients (including second infarcts), or number of admissions to a given hospital (including readmissions)? Research investigators often use an abbreviated language that is clear to their colleagues, but they may have to make a special effort to assure that such usage will not confuse nonspecialists, or even other experts.

2. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals).

Investigators have to choose a way to report their findings. The most useful ways give information about the actual outcomes, such as means and standard deviations as well as confidence intervals. The tendency to report a test of significance alone—rather than with this additional information—should be resisted, although a significance test in the context of other information may be helpful.

Readers have many reasons for studying a research report. One reason is to find out how a particular treatment does in its own right, not just in comparison with another treatment. At a minimum, readers should be offered the mean and standard deviation for every appropriate outcome variable. Significance levels (P values), such as $P = 0.03$, are often reported to show that the difference seen or some other departure from a standard (a null hypothesis) had little probability of occurring if chance alone was the cause. Merely reporting a P value from a significance test of differences loses the information about both the average level of performance and the variability of individual outcomes for the separate treatments.

Exact P values rather than statements like " $P < 0.05$ " or " P not significant" should be reported where possible so that readers can compare the calculated value of P with their own choice of critical values. In addition, other investigators may need exact values of P if they are to combine results of several separate studies.

In independent samples, information about means, standard deviations, and sample sizes can often be readily converted to a significance test and thus into a P value. From the P value alone, none of the others can be reconstructed, so that important information is lost when only a P value is reported (5, 6).

Make clear whether a reported standard deviation is for the distribution of single observations, or for the distribution of means (standard errors), or for the distribution of some other statistic such as the difference between two means. If the standard deviation for single observations is given, together with sample sizes, then in independent samples the reader can compute the other standard deviations.

Each statistical test of data implies both a specific null hypothesis about those data (such as "The 60day survival rate in Group A equals that in Group B," so that the difference is zero) and a specific set of alternative hypotheses (such as "The survival rate is different in Group B, which allows for a range of values for the difference). It is critical that both the null hypothesis and the alternatives be clearly stated, although many authors fail to do so. Clear reporting will not only help readers, it is also likely to reduce the frequency of abuse of P values.

It is critical also that authors specify how and when they developed each null hypothesis in relation to their consideration of the data. Statistical theory requires that null hypotheses be fully developed before the data are examined—indeed, before even the briefest view of preliminary results. Otherwise, P values cannot be interpreted as meaningful probabilities.

Authors should always specify whether they are using two tail or one tail tests.

3. Avoid sole reliance on statistical hypothesis testing, such as the use of P values, which fails to convey important quantitative information.

Confidence intervals offer a more informative way to deal with the significance test than does a simple P value. Confidence intervals for a single mean or a proportion provide information about both level and variability. Confidence intervals on a difference of means or proportions provide information about the size of difference and its uncertainty, but not about component means, and these should be given.

A significance test of observed data, generally to determine whether the (unknown) means of two populations are different, usually winds up with a score that is referred to a table, such as a t , normal, or F table. The table then presents the P value.

Although confidence limits offer appraisals of variability and uncertainty, in some studies, such as certain large epidemiologic and demographic studies, biases are often greater threats to the validity of inferences than ordinary random variability (expressed in the standard deviation) Coding or typing errors may exaggerate the number of deaths from a cause, nonresponse to treatment may be selective (those patients more ill being

less likely to respond), and so on. Although the potential sources of bias are many, books on applied statistics, epidemiology, and demography alert the research worker to common difficulties, and often to steps that may be taken toward their amelioration.

4. Discuss eligibility of experimental subjects.

Reasons for and methods of selecting patients or other study units should always be reported, and if the selection is likely to matter, the reasons should be reported in detail. The full range of potentially eligible subjects, or the scope of the study, should be precisely stated in terms that readers can interpret. It is not enough to say that the natural history of a condition has been seen in "100 consecutive patients." How do these patients compare with what is already known about the condition in terms of age, sex, and other factors? Are patients from an area or population that might be special? Are patients from an "unselected" series with an initial diagnosis, or do they include referral patients (weighted with less serious or more serious problems)? In comparing outcomes for patients who underwent surgery to outcomes for patients treated medically, were the groups in similar physical condition initially? What about probable cases not proved? Many other questions will arise in specific instances. Sometimes information is obvious (for example, if the investigator studied patients from one hospital because that is where he or she practices). Other questions about scope need answers. (Why begin on 1 January 1983? Why include only patients admitted through the emergency room?) Authors should try to imagine themselves as readers who know nothing about the study.

Although every statistically sound study has such "scope" criteria to determine the population sampled by the investigator, many also have more detailed "eligibility" criteria. Medical examples include the possible exclusion of patients outside a specified age range, those previously treated, those who refuse randomization or are too ill to answer questions, and other groups.

Which criteria are used to establish scope and which are used to establish eligibility may be uncertain, although both must be reported. Scope pushes study boundaries outward, toward the full range of patients or other study units that might be considered as subjects, whereas eligibility rules narrow the scope by removing units that cannot be studied, that may give unreliable results, that are likely to be atypical (for example, the extremes of age), that cannot be studied for ethical reasons (for example, pregnant women in some drug studies), or that are otherwise not appropriate for individual study.

The first goal is to state both scope and eligibility so that another knowledgeable investigator, facing the same group of patients or other study units, would make nearly the same decisions about including patients in the study.

The second goal is to provide readers with a solid link between the patients or cases studied and the population for which inferences will be made. Both scope and eligibility constraints can introduce substantial bias when results are generalized to other subjects, and readers need enough information to make their own assessment of this potential. Thus, reasons for each eligibility criterion should be stated. The two critical elements in setting the base for generalization are first to document each exclusion under the eligibility criteria with the reasons for that exclusion; and second, to present an accounting (often in a table) of the difference between patients falling within the scope of the study and those actually studied. The article should also say how patients excluded for more than one reason are handled; common approaches are to show specific combinations or to use a priority sequence. Such information helps the reader better understand how the study group is related to the population it came from, and also helps to assure that all omissions are accounted for. It should be so stated if no subject was ineligible for more than one reason.

Another critical element in reporting is to say how and when the scope and eligibility criteria were devised. Were scope and eligibility criteria set forth in a written protocol before work was started? Did they evolve during the course of the study? Were some eligibility criteria added at the end to deal with some problems not foreseen? For example, a written protocol might call for the study of "all" patients, but if only 5% of patients were female, they might be set aside at this point—especially if they are thought to differ from male patients in ways relevant to the subject of the study.

5. Give details about randomization.

The reporting of randomization needs special attention for two reasons. First, some authors incorrectly use "random" as a synonym for "haphazard." To prevent misunderstanding, simply tell readers how the randomization was done (coin toss, table of random numbers, cards in sealed envelopes, or some other method). Readers will then know that a random mechanism was in fact applied and they can also judge the likelihood that it was subject to bias or abuse (such as peeking at cards). Second, randomization can enter in many ways. For example, a sample may be selected from a larger population at random, or study patients may be randomly allocated to treatments, or treated patients

may be randomly given one or another test. Thus, it is not enough just to say that a study was "randomized." The many possible roles of randomization can be dealt with by careful reporting to assure there is no ambiguity.

Even with randomization, imbalances occur, with their predicted frequency, and these may need attention even if they do not call for special steps in the analysis. Stratification or matching may be used in combination with randomization to increase the similarity between the treated and control groups, and should be reported. Sometimes an assessment of the efficacy of stratification or matching in overcoming the imbalance is feasible; if so, it should be done and reported.

If the randomization was "blocked" (for example, by arranging that within each successive group of six patients, three are assigned to one treatment and three to another), reasons for blocking and the blocking factors should be given. Blocking should ordinarily affect statistical analysis, and authors should say how they used blocking in their analysis or why they did not.

6. Describe the methods for, and success of, any blinding of observations.

"Blinding," sometimes called "masking," is the concealment of certain information from patients or members of the research team during phases of a study. Blinding can be used to good effect to reduce bias, but because it can be applied in different ways, a research report should be explicit about who was blinded to what. An unadorned statement that a study was "blind" or "double blind" is rarely enough.

Patients may be blinded to treatment, or to the time that certain observations are made, or to preliminary findings regarding their progress. A decision to admit a patient to a study may be made blind to that patient's specific circumstances, and a decision that a patient randomized to treatment was not eligible may be made blind to the assigned treatment. The observer who classifies clinical outcomes may be blinded to the treatment, as may be the pathologist who interprets specimens or the technician who measures a chemical substance. These and other efforts to prevent bias by blinding should be reported in enough detail for readers to understand what was done.

The effectiveness of blinding should also be discussed in any situation where the person who is blinded may learn or guess the concealed information, such as by side effects that may accompany one treatment but not another. Such discoveries are particularly important for observations reported by patients themselves and for third party observations of endpoints with a subjective component, such as level of patient activity.

A particularly critical aspect of blinding is whether the decision to admit a patient to a study was made before (or otherwise entirely and demonstrably independent of) any decision about choice of treatment to be used or offered. Where random allocation to treatments is used, the timing of randomization in relation to the decision to admit a patient should always be stated.

7. Report treatment complications.

Any intervention, or treatment, has some likelihood of causing unintended effects, whether the study is of a cell culture, a person, an ecologic community, or a hospital management system. Side effects may be good (quitting smoking reduces the risk of heart disease as well as the risk of lung cancer) or bad (drug toxicity). Side effects may be foreseen or unexpected. In most studies side effects will be of substantial interest to readers. Does a drug cause so much nausea that patients will not take it? If we stock an ecologic area with one species, what will happen to a predator? Does a new system for scheduling the purchase of hospital supplies at lower overall cost change the likelihood that some item will be exhausted before the replacement stock arrives?

Nearly every medical treatment carries some risk of complications—that is, of unintended adverse effects. Such effects should be sought at least as assiduously as beneficial effects, and they should be reported objectively and in detail. Treatment failure often gives the most useful information from a study. If no adverse effects can be found, the report should say so, with an explanation of what was done to find them.

8. Give numbers of observations.

The basic observational units should be clearly specified, along with any study features that might cause basic observations to be correlated. A study of acid rain might take samples of water from five different depths in each of seven different lakes—35 measurements in all. But the relevant sample size for one or another purpose may be five (depths), or seven (lakes), or 35 (depths in different lakes). In a metaanalysis of such work [\(7\)](#) the whole study may count as only a single observation. Lake water may tend to mix, so that five samples from different depths tell little more about acidity than a single sample; or lake to lake differences may be small within a geographic region, so that the study of one lake effectively studies them all.

Similarly, a study in several institutions of rates of infection after surgery may be considered to have a sample size of three hospitals, 15 surgeons, 600 patients, or 3000

days of observation after surgery. But infection rates may differ so much by hospital or surgeon that it is more important to include many hospitals or surgeons, perhaps with only a few patients from each, than to have large samples per surgeon.

Reporting decisions about the basic unit of observation and about sample size, as well as proper method of analysis, may require an informed understanding of statistics as well as the subject matter. The analysis and reporting of correlated observations, such as the water samples and the infection rates described above, raise difficult issues of statistical analysis that often require expert statistical help.

A different kind of problem arises from ambiguity in reporting ratios, proportions, and percents, where the denominator is often not specified and may be unclear to readers. Authors should be meticulous about specifying which study units are included in denominators (which then specifies the group examined) each time there may be any uncertainty.

Whatever the investigators adopt as their basic unit of observation, relationships to and possible correlations with other units must be discussed. Such internal relationships can sometimes be used to strengthen an analysis (when a major source of difference is balanced or held constant), and sometimes they weaken the analysis (by obscuring a critical limitation on effective sample size). Complicated data structures require special attention in study reporting, not just in study design, performance, and analysis.

9. Report losses to observation (such as dropouts from a clinical trial).

When the sample size for a table, graph, or text statement differs from that for a study as a whole, the difference should be explained. If some study units are omitted (for example, patients who did not return for 6month followup), the reduced number should be reconciled with the number eligible or expected by readers. Reporting of losses is often easiest in tables, where entries such as "patients lost," "samples contaminated," "not eligible," or "not available" (for example, no 15meter sample from a lake with a maximum depth of 10 meters) can account for each study unit.

Loss of patients to follow up, including losses or exclusions for non compliance, should generally be discussed in depth because of the likelihood that patients lost are atypical in critical ways. Have patients not returned for examination because they are well? Because they are still sick and have sought other medical care? Because they are dead? Because they do not wish to burden a physician with a bad outcome? Failure to

discuss both reasons for loss (or other termination of followup) and efforts to trace lost patients are common and serious. Similarly, issues of noncompliance (reasons, as well as numbers) are often slighted by authors.

10. References for study design and statistical methods should be to standard works (with pages stated) when possible rather than to papers where designs or methods were originally reported.

An original paper can have great value for the methodologist, but often does little to explain the method and its implications or the byways of calculation or meaning that may have emerged since the method was first reported. Standard works such as textbooks or review papers will usually give a clearer exposition, put the method in a larger context, and give helpful examples. The notation will be the current standard, and the explanation will orient readers to the general use of the method rather than the specific and sometimes peculiar use first reported. For example, it would be hard to recognize Student's t distribution in his original paper; indeed, " t " was not even mentioned. Exceptions to the general advice about using textbooks, review papers, or other standard works occur where the original exposition is best for communication and where it is the only one available.

11. Specify any general use computer programs used.

General purpose computer programs should be specified, with the computer that ran them, because such programs are sometimes found to have errors (8). Readers may also wish to know about these programs for their own use. In contrast, programs written for a specific task need not be documented, because readers should already be alert to the likelihood of errors in ad hoc or "private" programs, and because they will not be able to use the same, programs for their own work.

12. Put general descriptions of statistical methods in the Methods section. When data are summarized in the Results section, specify the statistical methods used to analyze them.

Where should statistical methods be described? There are good arguments for putting such material in one place, usually in the Methods section of a paper, but our preference (9) is generally to specify statistical methods at the places where their uses are first presented. Methods may differ slightly from one to another application within a given paper; and decisions about which results to report in full, or which methods to use in

exploring critical or unexpected findings, generally depend on the data and earlier steps in the analysis. Keeping the specification of statistical methods close to their point of application will sometimes lead to more thought about choices and to better discussion of why a particular method was used in a particular way. Some editors, as well as some of our statistical colleagues, disagree, and authors should follow the instructions of the journal to which they submit their work.

Statements such as "statistical methods included analysis of variance, factor analysis, and regression, as well as tests of significance," when divorced from the outcomes or reasons for their use, give the reader little help. On the other hand, if the only method was the use of chisquared tests for 2×2 contingency tables, that fact might be sufficiently informative.

Some general suggestions about reporting clinical trials have been discussed by Mosteller and associates [\(10\)](#).

13. 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.

Authors have an understandable wish to tell readers everything they have learned or surmised from their data, but economy is much prized by scientific readers as well as editors. A basic point is that economy in writing and exposition gives an article its best chance of being read. Although many tables may help support the same basic point, and might be appropriate in a monograph, an article generally requires only enough information to make its point—the mathematician's concept of "necessary and sufficient."

There are occasional exceptions. Sometimes the study generates data that have consequences beyond the article. For example, if information about certain biological or physical constants is obtained, it should be retained in the article. An author should inform the editor of this situation in a cover letter. Sometimes such data need to be preserved, but not in the article itself, many journals have some plan for the preservation and documentation of unpublished supporting material. Such plans are often mentioned in a journal's instructions to authors.

Whether tables or graphs better present material is sometimes a vexing question. Some readers go blind when faced with a table of numbers; others have no idea how to read graphs; unfortunately, these groups are not mutually exclusive, and some users of statistical data need to see quantitative findings in text. Overall there is a general failure

to tolerate or understand the problems of any group that does not include oneself. Most of what we know about tables and graphs comes from the personal experiences of a few scholars, and little scientific information has been gathered on these subjects. Cleveland (11) has begun some scientific studies of what information can be communicated with graphs (for example, many people read bar charts better than pie charts). Tufte (12) has a beautiful book on the art of graphics.

In the field of tabular presentation, even less scientific investigation has been done, but there seems to be much value in some rules proposed by Ehrenberg (13): Give marginal (row and column) averages to provide a visual focus. Order the rows and columns of the table by the marginal averages or some other measure of size or other logical order (keeping to the same order if there are many similar tables). Put figures to be compared into columns rather than rows (with larger numbers on top if possible). Round to two effective (significant) digits. Use layout to guide the eye and facilitate comparisons. In the text give brief summaries to lead the reader in the main patterns and exceptions.

To show the effect of Ehrenberg's rules, we devised [Table 1](#) showing data on infant mortality, and we used Ehrenberg's rules to produce [Table 2](#). Our primary interest is in the association of the father's education with infant mortality, with a secondary interest in region.

[Table 1](#) is obviously "busy" with four digit numbers, and we have reduced them to two digits. [Table 2](#), with fewer digits, is easier to read although it has more numbers.

Because our primary interest is in the father's education, we put years of education in the rows. We want the big numbers at the top of the table, so in arranging the rows we started with the lowest level of education rather than the highest. We did not reorder the rows because years of education already provided an order. The regions were reordered according to their average values. The issue of whether to put northeast or north central first depends on whether we want to emphasize what is best or what is poorest. Some people like to have numbers rising as the eye goes from left to right.

We have added averages for the rows and for the columns, and given the grand mean without additional decimals to keep the table simple. The text might read as follows: "The table shows that the infant death rate has a grand mean of 23 per 1000 live births. Lower education of the father is associated with higher infant mortality, but education beyond the completion of high school (12 years) seems to have no further beneficial effect on the infant mortality rate. The northeast and west have the lowest rates, and the

south did slightly better than the north central region. Father's education seems to matter more than region of the country, a variation of 13 deaths per 1000 births for education (range, 30 to 17) compared with 5 for regions (range, 20 to 25). The highest rate seen was in Southern families whose father had no more than a grammar school education (no more than 8 years). The lowest rate was 16, the highest 39, a ratio of nearly two and a half."

14. Avoid nontechnical uses of technical terms in statistics, such as "random" (which implies a randomizing device), "normal," "significant," "correlation," and "ample."

Many words in statistics, and in mathematics more generally, come from everyday language and yet have specialized meanings. Thus, when statistical reporting is an important part of a paper, the author should not use statistical terms in their everyday meanings.

The family of *normal* (or Gaussian) distributions refers to a collection of probability distributions described by a specific formula. The distribution of usual or average values of some quantity found in practice is rarely "normal" in the statistical sense, even when the data have a generally bellshaped distribution. Normal also has many other mathematical meanings, such as a line perpendicular to a plane. When we mix these meanings with the meaning of "normal" for a patient without disease, we have the makings of considerable confusion.

Significance and related words are used in statistics, and in scientific writing generally, to refer to the outcome of a formal test of a statistical hypothesis or test of significance (essentially the same thing).

Significant means that the outcome of such a test fell outside a chosen, predetermined region. Careful statisticians and other scientists often distinguish between statistical and medical or social significance. For example, a large enough sample might show statistically significant differences in averages on the order of one tenth of a degree in average body temperature of groups of humans. Such a difference might be regarded as of no biological or medical significance. In the other direction, a dietary program that reduces weight by an average of 5 kg might be regarded as important to health, and yet this finding may not be well established, as expressed by statistical significance.

Although the 5 kg is important, the data do not support a firm conclusion that a difference has actually been achieved.

Association is a usefully vague word to express a relation between two or more variables. Correlation, a more technical term, refers to a specific way to measure association, and should not be used in writing about statistical findings except in referring to that measure. *Sample* usually refers to an observation or a collection of observations gathered in a well-defined way. To describe a sample as having been drawn at random means that a randomizing device has been used to make the choice, not that some haphazard event has created the sample, such as the use of an unstructured set of patient referrals to create the investigator's control group.

15. Define statistical terms, abbreviations, and most symbols.

Although many statistical terms such as mean, median, and standard deviation of the observations have clear, widely adopted definitions, different fields of endeavor often use the same symbols for different entities. Authors have extra difficulty when they need to distinguish between the true value of a quantity (a parameter such as a population mean, often symbolized by the Greek letter μ) and a sample mean.

We usually take for granted the mathematical symbols =, +, , and /, as well as the usual symbols for inequalities (greater than or less than); we do the same for powers such as x^3 , and for the trigonometric and logarithmic abbreviations such as sin, cos, tan, and log, although it is well to report what base the logarithms are using. Typography for ordinary multiplication differs, but is rarely a problem. Generally, symbols such as r for the correlation coefficient should be defined, as should n or N for the sample size, even though these are widely used.

Terms like *reliability* and *validity* are much more difficult, and they should always be defined when they are used in a statistical sense.

One difficulty with an expression such as $a \pm b$, even when a is a sample mean, is that b has many possibilities. (Some journals prefer the notation $a(b)$, but the ambiguities remain unchanged.) The author may use b for the observed sample standard deviation of individual measurements, or the standard error of the mean, or twice the standard error of the mean, or even the interquartile range, depending on the situation. The commonest ambiguity is not knowing whether b represents the standard deviation of individual observations or the standard error of the statistic designated by a . And no single choice is

best in all situations. If the measure of variability is used only to test the size of its associated statistic, as for example in a *P* value to test whether a correlation coefficient differs from zero, then use the standard error. If the measure of variability needs to be combined with other such measures, the standard deviation of single observations is often more useful.

The same difficulty occurs with technical terms. A danger is that a special local language will become so ingrained in a particular research organization that its practitioners find it difficult to understand that their use of words is not widespread. Nearly every laboratory has special words that need to be defined or eliminated in reports of findings.

When one or two observations, terms, or symbols are not defined, readers may be able to struggle along. When several remain uncertain, readers may have to give up because the possibilities are too numerous.

A well established convention is that mathematical symbols should be printed in italics (1517). This practice has many advantages, including the reduction of ambiguity when the same character is commonly used to designate both a physical quantity and a mathematical or statistical quantity. In typescripts, an underline is generally used to indicate that a character is to be printed in italics, and authors may need to give special instructions to editors or printers if underlines are used for other purposes, such as to designate a mathematical vector (which might be printed both underlined and in italics).

ACKNOWLEDGMENTS: The authors thank Marcia Angell, Alexia Anctzak, M. Anthony Ashworth, John J. Bartko, Thomas Chalmers, Eli Chernin, Victor Cohn, David Hoaglin, Susan Horn, Edward J. Huth, Deborah A. Lambert, Kathleen N. Lohr, Thomas Louis, Thomas R. O'Brien, Kenneth Rothman, Stephen B. Thacker, Anne A. Scitovsky, Barbara Starfield, Wallace K. Waterfall, Jonathan Weiner, Alfred Yankauer, Cleo Youtz, and two reviewers, for their thoughtful and helpful advice.

Grant support: in part by grant GAHS8414 from the Rockefeller Foundation, grant S8406311 from the Josiah Macy Jr. Foundation, and the Methods Panel of the Institute of Medicine's Council for Health Care Technology.

Requests for reprints should be directed to John C. Bailar III, M.D., Ph.D.; Department of Epidemiology and Biostatistics, McGill University School of Medicine, 1020 Pine Avenue West; Montreal, PQ, H3A IA2, Canada.

References

1. **International Committee of Medical Journal Editors.** Uniform requirements for manuscripts submitted to biomedical journals. *Ann Intern Med.* 1982;**96**(6 Pt 1):76671.
2. **International Committee of Medical Journal Editors.** Uniform requirements for manuscripts submitted to biomedical journals. *Ann Intern Med.* 1988;**108**:25865.
3. **DerSimonian R, Charette LJ, Mcpeck B, Mosteller F.** Reporting on methods in clinical trials. *N Eng J Med.* 1982;**306**:13327.
4. **Emerson JD, Colditz GA.** Use of statistical analysis in the *New England Journal of Medicine.* *N Engl J Med.* 1983;**309**:70913.
5. **Simon R.** Confidence intervals for reporting results of clinical trials. *Ann Intern Med* 1986;**105**:42935.
6. **Rothman KJ.** Significance questing. *Ann Intern Med* 1986;**105**:4457.
7. **Louis TA, Fineberg H, Mosteller F.** Findings for public health from metaanalyses. *Annu Rev Public Health.* 1985;**6**:120.
8. **Emerson JD, Moses L.** A note on the WilcoxonMannWhitney test for $2 \times k$ ordered tables. *Biometrics.* 1985;**41**:3039.
9. **Bailar JC, Mosteller F, eds.** *Medical Uses of Statistics.* Waltham, Massachusetts: NEJM Books; 1986.
10. **Mosteller F, Gilbert JP, Mcpeck B.** Reporting standards and research strategies for controlled trials: agenda for the editor. *Controlled Clinical Trials.* 1980;**1**:3758
11. **Cleveland WS.** *The Elements of Graphing Data.* Monterey, California: Wadsworth Advanced Books and Software; 1985.
12. **Tufte ER.** *The Visual Display of Quantitative Information.* Cheshire, Connecticut: Graphic Press; 1983.
13. **Ehrenberg AC.** The problem of numeracy. *Am Statistician.* 1981;**35**:6771.
14. **U.S. Department Of Health, Education, And Welfare.** *Infant Mortality Rates: Socioeconomic Factors, United States.* Rockville, Maryland: National Center for Health Statistics; 1972; DHEW publication no. (HSM) 721045. (*Vital and Health Statistics*; series 22; no. 14).

15. Huth EJ. Mathematics and statistics. In: Huth EJ. *Medical Style & Format, An International Manual for Authors, Editors and Publishers*. Philadelphia: ISI Press; 1987:1706.

16. International Organization for Standardization. *Units of Measurement*. 2nd ed. (ISO standards handbook 2). Geneva: International Organization for Standardization; 1982.

17. International Organization for Standardization. *Statistical Methods: Handbook on International Standards for Statistical Methods*. (ISO standards handbook 3). Geneva: International Organization for Standardization; 1979:2878.

Additional References

The following references may be helpful to readers who want to pursue these topics further.

1. Cochran WG. *Sampling Techniques*. 3rd ed. New York: Wiley and Sons; 1977.

2. Cohn V. *News and Numbers: A Guide to Reporting Statistical Claims and Controversies in Health and Other Fields*. Ames, Iowa: Iowa State University Press; 1988. (In press).

3. Colton T. *Statistics in Medicine*. Boston: Little, Brown and Co.; 1974.

4. Committee for Evaluating Medical Technologies In Clinical Use. *Assessing Medical Technologies*. Washington, DC: National Academy Press; 1985.

5. Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: John Wiley and Sons; 1981.

6. Gardner MJ, Maclure M, Campbell MI. Use of check lists in assessing the statistical content of medical studies. *Br Med J*. 1986;**292**:8102.

7. Gore S, Altman DG. *Statistics in Practice*. London: Taylor and Francis; 1982.

8. Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH. *Biostatistics in Clinical Medicine*. 2nd ed. New York: Macmillan; 1987.

9. Meinert CL. *Clinical Trials: Designs, Conduct, and Analysis*, New York: Oxford University Press; 1986.

10. Mike V, Stanley KE, eds. *Statistics in Medical Research: Methods and Issues, with Applications in Cancer Research*. New York: John Wiley and Sons; 1982.

11. Moses LE, Mosteller F, eds. *Planning and Analysis of Observational Studies*. New York: Wiley and Sons; 1983.

12. Shapiro SH, Louis TA, eds. *Clinical Trials: Issues and Approaches*. New York: Marcel Dekker; 1983.

13. Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Boston: Little, Brown & Co.; 1985.

14. Snedecor GW, Cochran WG. *Statistical Methods*. 7th ed. Ames, Iowa: Iowa State University Press; 1980.

Note: This document was obtained from

<http://www.acponline.org/journals/resource/guidelines.htm> on 26 March 2002