BIOSTATISTICAL CONSULTATION FOR DENTAL RESEARCH

Jonathan Clive, PhD

What do dentists need to know about statistics, and why do they need to know it?

This article suggests some reasonable and convincing answers to these questions. To focus the discussion, dental health care providers are considered as either practitioners and specialists who see patients daily but who do not perform scientific research (PR), or as dental researchers (DR) who may see patients or students but are also actively engaged in research.

In this article, it is generally assumed that dental researchers' activities involve the acquisition and evaluation of some kind of data. The term *data* simply refers to a description, numeric or otherwise, of the attributes of the experimental units (patients, laboratory animals, teeth, periodontal tissue, and so forth.) being considered. These may be as basic as the number of decayed, missing, and filled teeth (DMFT), or the number of decayed, missing, and filled surfaces (DMFS), gingival index, or some more specialized and exotic measure, such as the number of cells of a particular type per unit volume.

This article does not provide a *crash course* in statistics; in fact, no specific statistics lessons are offered here (although some specific examples are provided). It would be counterproductive to attempt to cover in this limited space subject matter that most introductory statistics texts require several hundred pages to present. Instead, the author discusses related general concepts that he believes are crucial for both dental researchers and dental practitioners to understand before beginning the statistical consultation process.

From the Office of Biostatistical Consultation, University of Connecticut Health Center, Farmington, Connecticut

STATISTICAL NEEDS OF DENTAL PRACTITIONERS

Although dental practitioners may not possess formal statistical experience or training, they may frequently use several terms that comprise a basic statistics vocabulary. These terms may carry associations or interpretations that are intuitively understood. Perhaps the terms most often used in this fashion are *mean* and *average*. Given a set of numbers (which can be assumed to represent data acquired during the course of some research endeavor), the *mean* is often interpreted as the *most typical* or *most representative* single value describing all the numbers. Although there is some justification for this view, the term has a more rigorous definition. To determine the mean of a set of data, one sums the data values and divides by the number of observations. The mean represents the center of gravity of a set of numbers, the value around which all other numbers are distributed.

The *standard deviation* (or, equivalently, the *variance*, which is the square of the standard deviation) is another basic summary attribute of data that has a relatively straightforward meaning. It describes the spread or dispersion of the numbers in the dataset around the mean. The larger the standard deviation, the greater the variation, or heterogeneity, among observations. These two measures, the mean and the standard deviation, arise naturally in the logical study of data. Outlining this development is helpful in understanding means and standard deviations and in obtaining an overview of statistical procedures.

To do so, it is useful to consider a common graphic portrayal of data, as illustrated in Figure 1*A*. This figure shows a histogram, which illustrates the distribution of values for some variable, in a sample assessed from some hypothetic population. The *distribution of values* encompasses the values that occur in the data and the frequency with which they occur. To generate a histogram, the range of the data (smallest and largest values among the observations) is partitioned into successive intervals.

A set of DMFS scores, for example, might range from 0 to 40. Intervals of DMFS could be designated as from 0–4, 5–10, 11–15, up to 36–40. The number of observations falling in each of these intervals is then tallied, and a bar graph is generated. The height of the bar for each interval is proportional to the frequency or the number of observations falling within that interval. The resulting plot is called a histogram, and it, too, is an intuitively and easily understood representation of data. (Note that the width of the intervals must be constant so that the comparison of frequencies is meaningful.)

BASICS 1: HISTOGRAMS TO BELL-SHAPED CURVES

Although useful and informative, the histogram represents little more than a basic tool for exploratory data analysis and is purely descriptive. Statisticians and mathematicians who were not satisfied such heuristic descriptions, carried the idea further, asking what happens when a histogram presents all the information possibly available. This

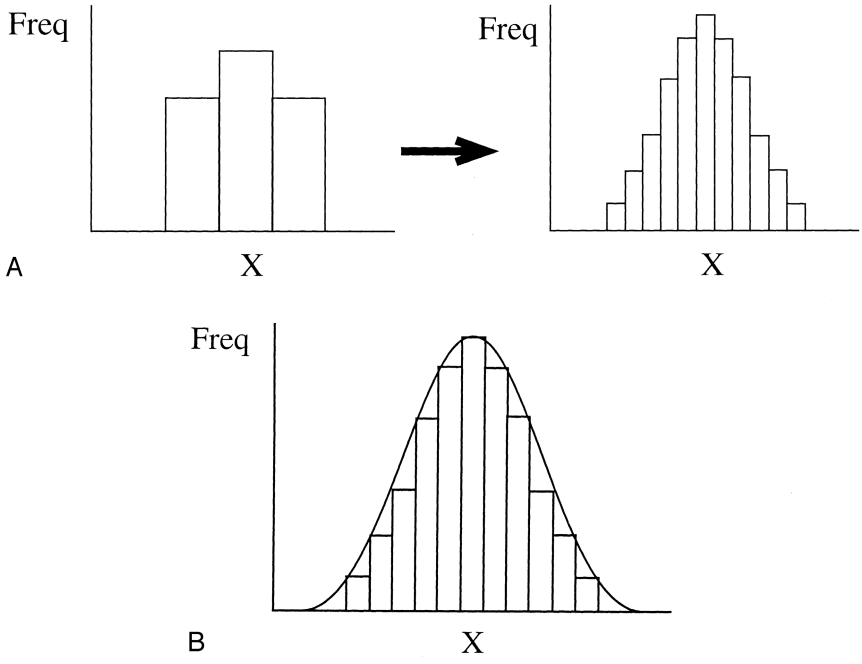

Figure 1. *A*, A histogram can be used to graphically represent the distribution of values contained in a sample of observations. Each bar corresponds to a range of values of the measure being considered (X), and the height of the corresponding bar (Freq) is proportional to the count or percentage of all observations falling into that category. Histograms can be refined by designating more narrow intervals and increasing the number of observations. *B*, A histogram may be modeled, or approximated, by an appropriate mathematical function (see text). FREQ = frequency.

limit can be approached, it was suggested, by making the intervals more and more narrow as the sample size (the number of observations or data points) becomes larger and larger. As successive histograms are generated under these circumstances, their appearance is increasingly seen to resemble a relatively smooth or continuous curve, in contrast to a single histogram defined over a few intervals, which resembles a conventional bar chart with adjacent bars.

The generation of such a curve suggests that it might be possible to use some type of mathematical function to characterize it. A mathematical representation would have several desirable features. First, it would provide a succinct way to describing a distribution, namely, the mathematical function describing the curve. Second, it would serve as a basis for comparing distributions across different populations or groups. Making such comparisons is a basic activity of statistical inference.

Using a well-defined mathematical function to describe or model a histogram is illustrated in Figure 1*B*. Here, a curve has been superimposed around the histogram shown in Figure 1*A*. The particular curve

used here is called a *normal curve*, or *normal distribution*. It is also well known as the *bell-shaped curve* familiar to most scientists. The term *normal* in normal distribution is used as a name, and derives from the suggestion that the distribution of most attributes in the normal (here used as an adjective) population is well represented by the type of curve shown in Figure 1*B*.

In fact, the normal curve is in many cases an acceptable representation of a distribution, even when the observed histogram is skewed or not symmetric or, in general, somewhat poorly behaved. Furthermore, one of the most remarkable results from mathematical statistics shows that in almost all situations, no matter what the underlying form of the histogram of observations, the distribution of means from the population under study does tend to follow a normal distribution. The normal distribution forms the core of much of statistical theory and practice.

It is important to understand the distinction between a histogram (Fig. 1*A*) and the normal curve (Fig. 1*B*). The histogram represents the data, or the real world, whereas the normal curve is strictly a model, an approximation generated by statisticians. A normal distribution is uniquely characterized by its mean and standard deviation. These are the same parameters so intuitively familiar to nonstatisticians. The mean of a normal distribution reflects the center of gravity of the values or observations and is often referred to as a *measure of location*, or a *measure of central tendency*. The normal distribution peaks at the mean value; that is, the maximum value of the curve along the *y*-axis occurs at the location of the mean on the *x*-axis. The standard deviation indicates the spread of the distribution around the mean; the larger the standard deviation, the more flat or less spiked the normal curve.

BASICS 2: CONFIDENCE-INTERVAL ESTIMATES

The mean and standard deviation of a normal curve (or any set of data, for that matter) are often presented in the scientific literature as mean ± standard deviation. For example, an author might write that "the mean and standard deviation DMFS for the test group was 8.3 ± 4.5, whereas for the control group the mean and standard deviation DMFS was 11.6 ± 5.2." The symbol "±" means *plus or minus* and is technically not correctly used in this context. The standard deviation is a positive quantity. By convention, however, the use of this notation has been almost universally adopted.

The best interpretation of a statement of the form mean ± standard deviation is that approximately 68% of all observations lie within one standard deviation of the mean. Approximately 95% of all observations lie within two standard deviations of the mean. (These statements apply to observations that are well modeled by a normal distribution.) The use of this notation does provide the reader with the notion of an interval in which most of the data are contained. A confidence interval represents

the natural extension of this notion of an interval estimate and is another statistical concept commonly encountered in the scientific literature.

The use of a mean ± standard deviation attempts to indicate both the value of a mean and the precision with which the value was determined. Thus, one might simply note that "the estimated mean DMFS was 8.3." This statement can be extended to include a 95% confidence interval: "The mean DMFS was 8.3, with a 95% confidence interval given by (6.8, 9.8)." This statement means that one is 95% confident that the interval from 6.8 to 9.8 inclusive contains the true mean DMFS for the population under study. A 95% confidence interval given by (6.8, 9.8) is much more precise than a 95% confidence interval given, for example, by (4.8, 11.8).

Most scientific journals currently insist on the use of confidence intervals beyond simple point estimates when discussing numerical data. The difference between a simple point estimate and a confidence interval is illustrated in the following statements: (1) "It is likely that the restaurant is on 7th Avenue and 55th Street"; and (2) "I am 95% confident that the restaurant is on 7th Avenue between 52nd and 58th streets."

BASICS 3: COMPARING MEANS

If two distributions have identical means, one can assume that the distribution of values of the variable being measured are identical in the two groups from which the data were drawn. If two distributions have similar means, the distributions are similar; finally, if two distributions have different means, the distributions are different. Statistics provides a way of estimating how probable it is that two or more means originated from the same underlying population. This probability is called a "*P*-value" and is the last of the routinely invoked statistical terms considered here.

P-values arise when two means are compared statistically. Thus, one may report that "when the two means were compared using a *t*-test of independent group means, it was observed that t on 44 degrees of freedom was 4.55, $P < 0.001$." The interpretation of this statement is as follows: if there is really no difference between the groups being observed, then the probability of observing a mean difference as extreme or more extreme than that observed is less than 1 in 1000.

Because the *P*-value in the example is less than 0.05, the difference is said to be *statistically significant*. This cutoff point for statistical significance (0.05) is rather arbitrary but has developed into a standard in the scientific literature over time. The *P*-value indicates the strength of the evidence against the hypothesis of no difference in means. (The perspective of no difference is used because doing so reflects how the theory of statistical hypothesis testing developed.) Small *P*-values indicate that the hypothesis of no difference is unlikely. Unlikely does not mean *impossible*, however; therefore the researcher (or reader) must choose between rejecting the hypothesis of no difference or accepting the conclusion

that the data represent the unlikely instance of a large difference in sample means.

STATISTICAL NEEDS OF DENTAL PRACTITIONERS—GENERAL CONSIDERATIONS

Although dental practitioners generally do not need assistance in designing and executing an experiment, there are circumstances in the activities of daily practice for which some statistical insight is required. For example, dental practitioners need to be able to keep up with the scientific literature, studying and evaluating scientific reports appearing in professional journals. Also, patients may ask questions such as, "Am I at risk for losing my teeth?" or, "Am I at risk for periodontal disease?" or, "Am I at risk for oral cancer?"

These are statistical questions. *Risk* is a statistical concept in epidemiology, and the appropriate estimation of risk factors for major diseases and other clinical conditions is a major topic in biomedical research. The dental practitioner may not need to estimate the risk but rather may need to be able to explain it and discuss it intelligently with a concerned patient. In other circumstances, practitioners may need to evaluate an article in the scientific literature and ponder the implications for their practice. For example, a practitioner may need to decide whether an article presents convincing evidence for switching to a different type of material for restorations.

The study by Kilburn and Asmundsson[10] serves as an example of the importance of reviewing the scientific literature with a certain degree of skepticism. These authors claimed to disprove the long-held clinical maxim that the anteroposterior (AP) diameter of the chest is increased in patients with advanced pulmonary emphysema (who were compared with a group of nondiseased controls and a second group of patients with non-emphysema diagnoses). In fact, the authors were not at all reluctant to assert that "it is contended that measurement has destroyed an apparently long-established and often repeated maxim that an increased AP diameter is a common and useful sign of emphysema."

The experimental approach used in this study was suspect; furthermore, the authors cited no statistical evidence to support their claim. They did, however, present sufficient tabulated summary measures to enable the reader to carry out the basic statistical test that would have been appropriate for the clinical question. When the reader carries out the test (the level of statistical knowledge necessary to do so would be acquired during the first third of an introductory biostatistics course), the difference in mean AP diameter between the emphysema group and the nondiseased controls is found to be statistically significant, with $P = 0.04$! Thus, not only does this result contradict the main conclusion; the authors themselves have provided the reader with the resources to refute the paper.

In this case, the statistical test is a *t*-test of independent group means,

two-tailed, carried out at the 0.05 level of significance. Note that is also possible to perform a one-way analysis of variance given the tabulated summary measures. A one-way analysis of variance permits comparison of all three group means simultaneously, as well as the appropriate multiple-comparison procedures to isolate group differences.

Is it reasonable to expect practitioners to be familiar with these terms and to be able to duplicate statistical procedures of the type discussed here in the course of evaluating a journal article? Probably not. Any statistical background practitioners acquired in dental school may be long forgotten, and the practitioner probably faces more pressing concerns involving office management and patient treatment. In addition, the practical import of a study in the literature may need time to propagate to the office of a practitioner.

Unfortunately, in the presentation of scientific studies, the situation is often "let the reader beware."[18] It is true that since the appearance of the Kilburn, et al article, most journals have increased their requirements for statistical rigor in submitted manuscripts. Many journals retain statistical consultants for special reviews and will use statistically sophisticated referees where necessary. Nonetheless, a practitioner may need to know how to evaluate such issues as the suitability of the experimental design, the appropriateness of the statistical tests used, and whether the results of the test have been interpreted correctly.

STATISTICAL TRAINING FOR DENTAL PRACTITIONERS

The practitioner must determine what level of statistical insight is appropriate and how to acquire it. A suggested knowledgebase is shown in Table 1. Although there is no one-size-fits-all statistics curriculum, this table lists the basic topics an introductory statistics student should

Table 1. SUGGESTED TOPICS FOR A BASIC STATISTICAL EDUCATION, WITH ASSOCIATED ENHANCEMENTS

Topic	Contents	Enhancements
Descriptive statistics	Summary measures/graphics	
Hypothesis testing	Paired, 2-sample *t*-tests	
	Compare independent proportions (analysis of 2×2 tables)	Higher order contingency tables
	Power/sample size	
	Type I and type II errors	
	One-way analysis of variance	Linear models/higher order ANOVA designs
Bivariate analysis	Univariate regression/ correlation	Multiple linear regression Logistic regression

ANOVA = analysis of variance

master. Beyond these basic topics, certain enhancements or intermediate topics are given.

In the absence of a working knowledgebase of this type, a practitioner may feel it sufficient to "ask around," at professional meetings, for example, concerning a particular topic. In special circumstances, it may be necessary to secure the services of a statistical consultant from a local university, college, or school of public health. (The department secretary of the statistics or biostatistics department will, in most cases, direct a PR to an available consultant or the director of a consulting facility.) Sometimes these two approaches can be combined, as when a local dental society, for example, invites a statistician to address a meeting and discuss a reference of special interest to the members.

The practitioner may also want to audit a course in introductory biostatistics. The course should be presented from a research perspective. Most institutions of higher learning from the community college level to the college, university, or academic health center (including schools of public health) level offer such courses, and auditing privileges can often be obtained.

Alternatively, practitioners may want to follow a self-study regimen. There are a number of excellent references for this purpose, well written and especially suited for individual use. For biostatistics, the author recommends Glantz, *Primer of Biostatistics*[8]; for clinical epidemiology, Sackett, Haynes, and Tugwell, *Clinical Epidemiology*[17] and Friedman *Primer of Epidemiology*[7]; and for general evaluation of the scientific literature, Riegelman, *Studying a Study and Testing a Test*.[16] The truly ambitious practitioner can acquire some statistical software. Statistical package for the social sciences (SPSS) is a Windows-based, user-friendly program, and inexpensive student versions are available. A number of self-instruction texts are available for this package to complement the useful in-program documentation.[2]

STATISTICAL NEEDS OF DENTAL RESEARCHERS

At this point, it is appropriate to consider the statistical needs of the dental researcher, although readers who consider themselves practitioners are encouraged to keep reading. The statistical needs of dental researchers are generally more pressing (e.g., there is generally less time for a self-instruction approach) and may be more technically advanced than the statistical needs of a practitioner.

Because the researcher is actively involved in some form of dental research, the need for statistical input during all phases of the research is important, from the initial formulation of the scientific or clinical hypotheses being considered, through the execution of the project, to the preparation of research papers and presentations. The extent and amount of this input depend, of course, on the magnitude of the project. Clearly, a survey of several thousand patients involving many study

variables will require more statistical resources than a study involving 20 or 30 laboratory animals and few study variables. Nonetheless, the planning issues are similar in both cases.

THE ROLE OF THE STATISTICAL CONSULTANT IN DENTAL RESEARCH

Certain primary tasks are the biostatistician's responsibility when interacting with dental researcher. For the present discussion, it is assumed that the consultation goes beyond a simple drop-in visit during which the biostatistician can respond to a few simple questions such as responding to comments in a manuscript review. Ideally, the researcher and the biostatistician meet early in the planning process (at the researcher's initiative, of course), to discuss the nature of the research, whether it is an intricate experiment with many outcome measures or a large-scale clinical trial in which observations are collected at multiple points over time on a large group of patients or experimental animals.

It is advisable for the biostatistician to become as familiar as possible with the purpose of the research and the underlying scientific or clinical considerations. This familiarity can accrue over time, as the biostatistician and the researcher meet repeatedly and interact regularly over the course of planning and executing the project. It is not reasonable to expect the statistical consultant to manifest the same level of understanding of the scientific issues as the researcher. It is also not reasonable, however, to expect the researcher to be able to handle the technical mathematical issues of the statistics involved in the research.

Thus, in the interaction between researcher and statistician, what the researcher needs to know about statistics mirrors in many ways what the statistical consultant or dental patient needs to know about clinical dentistry. A patient does not need to know the intricate and minute clinical and scientific details of how a therapy works to benefit from it. It is likely that over time a patient will become somewhat familiar with aspects of the dental procedures received. For example, a patient receiving implants would be able to advise other patients on the nature of the process but would not be qualified to apply it.

Similarly, the researcher does not need to know whether maximum likelihood or least squares algorithms were used to generate estimates of model parameters. The model needs to target the specific aims of the research and to address the fundamental hypotheses. It is the interpretation of the results of the analysis or modeling process that is crucial. (It is assumed here that the biostatistician has appropriately diagnosed the adequacy of the model—a process referred to as evaluating the fit of the model. That is, this discussion assumes good statistical practice on the part of the statistician, just as it assumes high standards of clinical and scientific conduct on the part of the researcher.)

STATISTICAL HYPOTHESES AND METHODS

The biostatistician and the researcher need to carefully establish a one-to-one relationship between a set of clinical or experimental hypotheses and corresponding statistical hypotheses. These hypotheses are often expressed in opposite ways. For example, the clinical hypothesis, "This new treatment, together with proper oral hygiene, will greatly reduce the rate of increase in DMFS compared with proper oral hygiene alone," might be expressed in a statistical context as, "There is no difference in mean change in DMFS between the drug-and-oral-hygiene and oral-hygiene-only groups." This transcription will help to specify the appropriate statistical procedures to be used in analyzing the data.

In some cases, the relationship may be more subtle and may require extensive interaction and question-and-answer sessions between the researcher and the biostatistician. It is quite possible that the type of statistical methods selected will change as new understanding arises on the part of both the researcher and the statistical consultant. New issues may arise that require additional planning. In any case, it is important to establish statistical hypotheses for all primary and secondary clinical and scientific hypotheses and to take the nature of the corresponding outcome variables into account. The statistical procedures should be coordinated with the research-specific aims when a research proposal is being jointly prepared.

COMPLICATING FACTORS

A number of characteristics of dental research need to be considered when planning statistical analyses. The first characteristic is structural; quite simply, one is dealing with multiple units if one considers individual teeth or even individual tooth surfaces. Statisticians refer to multiple-unit data as high-dimensional data, and such data can seriously complicate both the research planning and the data analysis. The usual approach is to attempt a type of statistical analysis known as a *dimension reduction procedure* and then study the reduced number of data units. Alternatively, researchers can confine attention to a specified subset of the original units. These issues are discussed by Clive and Woodbury.[3]

An example of such a situation is the paper by Löe et al, which analyzes data from the well-known study of the progression of periodontal disease among Sri Lankan tea laborers.[12] The authors seek to identify and characterize subtypes of disease development based on analysis of loss-of-attachment measurements. Three disease subtypes were identified. Although the clinical utility of these types remains to be established, their descriptive value is clear.

Many experiments in dental research involve the acquisition and analysis of longitudinal, or repeated measurements, data, because the development and manifestation of dental disease is often a gradual

process. Longitudinal research concerns the assessment of experimental units at several points in time over the entire chronologic course of the research. Such observations are referred to as *clustered*, or *correlated,* data because the outcomes for individual subjects may be related over time.

It is generally inadvisable to analyze such data in orthodox ways, as if the observations were independent. For example, if a subject's DMFS is above average at time 1, it is likely to be so again at subsequent readings; this degree of association can influence the analytic results and needs to be accounted for. Developments in both applied and theoretical statistics and computer science over the past two decades have made it possible to deal with these analyses on a fairly routine level; these approaches are discussed in Diggle, Liang, and Zeger's *The Analysis of Longitudinal Data*[4] and in Littell, Milliken, and Stroup's *SAS® System for Mixed Models*.[11]

Armitage et al[1] provide an interesting illustration of such research in assessing the use of elastase as a marker for the progression of periodontal disease. This paper, which appeared in the dental literature, is paired with a technical paper from the statistical literature[20] that assesses the advanced longitudinal data analytic techniques in the specific context of periodontal disease. These articles are noteworthy in that together they present research evaluating the appropriateness of a new class of data analytic models as well as clinical and scientific applications of the new analyses. The software for implementing these procedures is now routinely available; this was not the case when the papers were published.

Other interesting examples of longitudinal data analysis for dental research are given by Neely[14] and Chugal et al.[2] Neely identifies key risk variables for tooth loss based on analysis of the Sri Lanka data cited previously.[12] Subjects were seen between one and seven times, and loss of attachment for two surfaces for each tooth were measured at each point. Tooth loss was also assessed at each point and used as the outcome variable for the analysis.

Chugal and colleagues[2] investigated factors influencing the success or failure of endodontic therapy. This research modeled the success or failure for each canal. The number of canals within teeth varied, as did the number of treated teeth across patients. In this case, there were, in fact, two levels of clustered data: canals within teeth, and treated teeth within patients.

Still another attribute of dental data that complicates analytic considerations concerns the intrinsic lack of precision of some basic measurements; this imprecision is sometimes referred to as *noise*. A good example of noise is the limit on precision in measuring loss of attachment. Although loss of attachment is a crucial outcome measure in many studies, investigators need to be aware of the extent to which experimental conclusions can be compromised by measurement deficiencies. Certain experimental design considerations devised to deal with the problem of noise are discussed by Imrey and Chilton.[9]

SAMPLE SIZE AND POWER ANALYSIS

The estimation of sample size is a crucial aspect of research development. Having too many experimental subjects wastes time and money, a circumstance especially frowned upon by funding agencies. Having too few subjects increases the risk of failing to detect a real effect or a statistically significant difference. In statistical terms, the smaller the sample size, the smaller the power of the experiment, which is the chance of detecting a difference or effect that is really present. In the chronology of planning, it could be argued that sample size estimation precedes the specification of statistical techniques described previously. The selection of methods of analysis often dictates the approach to sample size estimation, however.

The design of the study is another factor influencing the estimation of sample size. There are several well-established experimental designs to consider, especially when the research is in the form of a clinical trial. A randomized clinical trial is a design with at least two study groups (test and control) to which eligible patients are assigned randomly. Other designs include case-control, prospective cohort, crossover, and the less scientificly rigorous pilot and observational studies. The selection of an appropriate study design is an important aspect of researcher and biostatistician interaction.

Sample size estimation must include a statistical justification in terms of testing the primary research hypotheses and specification of what a clinically or scientifically significant difference is for each of the main outcome variables in the study. The term *difference* is used here to denote the magnitude of the difference between summary outcome measures across experimental groups.

Note that clinical or scientific significance may be different from statistical significance. As Feinstein, citing Gertrude Stein, has noted, a difference has to make a difference to be a difference.[6] Thus, rational experiment planning requires the researcher to estimate what difference is noteworthy and to specify sample size accordingly. This planning is in contrast to the haphazard selection of a sample size with the hope that something statistically significant and worth reporting will *turn up*.

Although the biostatistician actually provides sample size estimates, these estimates are based on extensive input from the researcher. In addition to the concepts cited previously, the researcher needs to have a good working understanding of type I and type II errors. In statistical terms, a type I error involves rejecting a true hypothesis of no difference, and a type II error involves accepting a false hypothesis of no difference. The probability of these events are denoted by α and β; the power of a test is the complement of a type II error, with probability $1-\beta$.

In research terms, type I and type II errors correspond to concluding falsely that an effect is present or concluding falsely that an effect is absent, respectively. Many researchers consider the latter more serious than the former, because failing to detect an experimental effect might

lead to loss of interest or motivation in the particular type of research being performed. On the other hand, it is likely that a false effect will be exposed sooner or later in the course of further research.

The researcher also needs to provide estimates of the magnitude of the effect of the intervention on each of the outcome variables in one of the research groups, together with an estimate of the variability; these estimates are called *pilot data*. Making these estimates may seem counter-intuitive, because it may reasonably be asked what purpose the research serves if some concept of the size and variance of the intervention is available a priori. In fact, researchers are not presuming to estimate the effect of the experimental intervention but rather to make a reasonable speculation on the response that could be expected in the control or nonintervention group. The pilot data form the basis for estimating the sample size required to observe a given difference.

Assume, for example, that an investigator is planning to test the effect of an intervention hypothesized to reduce the rate of loss of attachment. Suppose further that it is known that over some time period untreated individuals will lose an average of 4 mm attachment, with a standard deviation of 3.5, and that these estimates apply to the type of patient population being studied. The biostatistical consultant can use this information to estimate the number of patients needed to detect a specified difference based on given values of α and β and the type of analysis to be used.

SOURCES OF PILOT DATA

The division of labor is straightforward in sample size estimation. The researcher needs to supply estimates of the appropriate summary measures for the important outcome variables for at least one of the groups in the study (probably the control or non-intervention group). The researcher also needs to provide an idea of the magnitude of a clinically significant effect. The statistician uses these data, together with specifications of α and β and the statistical method to be used, to estimate a sample size. Table 2 outlines the main steps in performing a power analysis. It is permissible to estimate sample sizes for a range of values of α and β, as illustrated in Table 3, which shows a sample size table for the hypothetical experiment concerning loss of attachment discussed previously.

There are several sources of assistance for the researcher in determining what constitutes a clinically significant difference, as well as providing pilot data for use by the statistical consultant in estimating sample size or power. The scientific literature is a valuable source of background data for this planning. It is quite likely that the researcher has exhaustively reviewed the literature in the course of formally developing the research plan. The literature review may provide multiple sources of pilot data as well as indications of the variation in response across different classes of patients or potential research subjects.

Table 2. PRIMARY STEPS IN CARRYING OUT A POWER ANALYSIS

Step	Activity	Responsibility
1	Specify clinical hypotheses	DR
2	Determine primary outcome measures	DR, B
3	Transcribe clinical hypotheses to statistical hypotheses	DR, B
4	Specify range of clinically meaningful differences for outcome measures	DR
5	Specify α, β	DR
6	Obtain pilot data for calculations	DR
7	Obtain power/sample size estimates	B
8	Evaluate results for final sample size/power specification	DR, B

DR = dental researcher; B = biostatistician; α = probability of a type I error; β = probability of a type II error

Input from colleagues is also a useful source of data for research planning, especially when inquiries can be focused, so that matters of pilot data and significant effect can be addressed directly. A researcher who is also a practicing dentist may have a set of patient records worth examining. Many providers use the readily available software (such as spreadsheet packages) to construct their own databases, which may be useful for planning; however, it is important to verify the standards under which the data were assessed.

Some commercially available software packages for power analysis provide interactive prompts for assisting users through the steps of a sample size determination.[6] These steps include a variety of techniques for generating pilot data summary measures from limited input (e.g., estimating the standard deviation for an outcome variable based on estimation of the range or percentile values). These techniques are useful when prior knowledge or data are limited.

In some cases, suitable pilot data are lacking altogether. Such a situation may arise, for example, early in the history of a line of scientific inquiry or when first testing a new drug or intervention. Here, the researcher may consider implementing a pilot study. A pilot study is a

Table 3. SAMPLE SIZE ESTIMATES FOR HYPOTHETICAL LOSS OF ATTACHMENT STUDY*

Percent Mean Difference	β Value	
	0.10	***0.20***
10	40	35
15	33	28
20	29	24
25	25	19

*The table is designed to show the number of subjects needed in each of two groups, assuming a repeated measures design with hypothesis testing carried out at $\alpha = 0.05$. Sample size estimates are shown for each of two values of β, and each of several effect sizes.

α = probability of a type I error; β = probability of a type II error

small-scale, preliminary study designed to assess the feasibility of the proposed research and entails evaluating all aspects of the research (including administrative matters if a clinical trial is being contemplated). One of the primary objectives is to acquire a database for use in formal calculations of sample size needed for a larger experiment or clinical trial to be carried out at a later date.

Although relatively small in scale, pilot studies often require as much interaction between researcher and statistician as more formal research, particularly in determining which variables will be assessed. Stopping rules need to be established because, by definition, detailed power analyses are not possible in a pilot study.

DATA MANAGEMENT

Data management is a deliberately broad term, incorporating a variety of tasks concerned with data acquisition, storage, confidentiality, editing, and retrieval. These aspects of research execution are of vital importance in assuring the quality of the research. Data management is especially vital in large-scale projects involving the determination of many variables from many research subjects and possibly at multiple time points.

Data acquisition begins with the design and planning of data-gathering forms. Completed forms need to be machine entered, although machine-readable forms can facilitate this activity. Data should be checked thoroughly. One procedure for checking is dual entry, in which data forms are entered twice, by independent data technicians. The final data files for the two operators are compared data point by data point. With dual entry, the only way an incorrect value can enter the final file is for each operator to make the identical mistake in the identical location.

Although dual entry is an effective mechanism for data-entry quality control, it is not always feasible. In large studies or clinical trials, the researcher should plan on printing a randomly selected subset of the entire data file for verification against the data-gathering forms. It may also be prudent to check all values of any variables that are particularly significant. This checking will provide an estimate of the overall data-entry error rate and may suggest variables or areas in the data file that need further attention in the editing process.

Once a data file has been checked, edited, and found satisfactory, further examination should involve the search for potential outlying values. This search can be done for all variables in the file and is a basic procedure in the exploratory data analysis phase of the research. All values above the 95th percentile and below the 5th percentile (or some other specified cutoff points) are listed, together with identifying information indicating which record in the data file contains the value. The researcher can then consider this output and flag blatantly out-of-range

values for further checking. This procedure should be followed for all major study variables.

Data confidentiality is important in clinical trials involving human subjects. Assuring confidentiality usually entails an intermediate step in the data-entry process in which identifying information is replaced with some numeric patient identifier. This is the responsibility of the researcher, who establishes and maintains the key relating the two data fields. Access to the key is limited by and is under the direction of the researcher. The file supplied to the statistical consultant should contain no unique patient data that could be used for identification or to breach subject anonymity.

RESEARCH ADMINISTRATION

The biostatistical consultant can help the researcher with other aspects of the actual administration and execution of the research project. Several such topics noted here arise in clinical trials and include the randomization of patients, protocol deviations, and the analysis of dropouts and missing data.

The randomization of patients refers to the assignment of patients to study groups, usually by some random mechanism. Randomization is generally a straightforward task, and the method of randomization depends largely on the study design. Both the researcher and the statistical consultant need to keep careful track of any protocol deviations. Protocol deviations involve changes in the study design or plan once subject intake has begun. These planning changes are sometimes unavoidable.

Subject dropout can be a major problem in longitudinal clinical trials. Patients can leave for a variety of reasons. Some may decide not to continue participating, especially if the experiment involves some unpleasant or invasive procedures. Others may leave the area. Some may become injured or ill. The researcher and the biostatistician will hope fervently that such dropout is random; that is, one is not primarily losing the treatment responders or non-responders or only the most compliant or non-compliant participants. Random dropout implies that subjects who leave a clinical trial before completion of the protocol do so at random, and that the remaining subject groups are still homogeneous with respect to potentially confounding factors.

The term *intent to treat* refers to the analysis of data from all subjects, including those who drop out. The rationale for this method of analysis is discussed in most clinical trial guides; see, for example, Spilker's *Guide to Clinical Trials*[19] and Piantadosi's *Clinical Trials: A Methodological Perspective*.[15] Most researchers will also analyze those subgroups of participants completing the protocol. A crucial phase of the analysis of data from a clinical trial is the analysis of dropouts and their comparison with subjects who remained.

Missing data can be a problem in surveys or retrospective studies. The difficulty, as with patient dropouts, arises when missing data occur in a nonrandom fashion. Although the treatment of missing data (involving, for example, multiple imputation procedures) is the responsibility of the biostatistician, the researcher needs to participate in planning the procedures for carrying out the study so that that the occurrence of missing data is minimized or the variables prone to absence are of relatively minor importance in the study.

DISCUSSION

The primary theme of this article is that it is not essential that the researcher have a strong working knowledge of elementary (or higher) statistics to perform valid scientific research. Rather, the researcher should be prepared to work with a biostatistical consultant on an extensive and ongoing basis to plan and execute a research project carefully. The need for this emphasis was recognized by Moses and Louis, who suggested that effective collaboration between clinician and statistician can help identify tractable scientific and statistical problems that need attention and can help avoid undertaking intractable ones.[13] Furthermore, they assert that the "central requirement for successful collaboration is clear, broad, specific, two-way communication on both scientific issues and research roles."[13]

The researcher will need to assist the biostatistician in estimating sample size, in understanding the basics of the science involved, and in relating scientific and statistical hypotheses. The biostatistician should come to appreciate the scientific and clinical issues and underlying principles; likewise, the researcher will come to appreciate how appropriately executed data analysis can extract valuable scientific knowledge from experimental data.

Over time, the researcher will acquire the statistical knowledge needed to interpret and present study results. The statistical understanding may be focused and restricted to the methods relevant to the particular study, but it will constitute a useful body of knowledge, appropriate for future studies or as a basis for using other statistical methods in different studies.

Most, if not all, scientists are convinced of the utility of mathematical models in representing and studying natural phenomena. Statistical models are mathematical models that incorporate probabilistic measures of uncertainty. In the study of oral health, two primary sources of variation impart this uncertainty. The first is the natural variation among patients in measures of oral health; the second is the variation resulting from sampling, or selecting a subgroup of patients for study, because the entire population of such patients is impossible to access.

The progression of data analyses intended to account for this variation, from simple independent group *t*-tests through complex multivari-

ate methods, is one of increasing technologic, mathematical, and statistical sophistication and advancement. It is also a progression that describes considerable theoretic and applied advances by researchers attempting to understand dental disease and how to deal with it. Often, theoretic clarification and understanding derive from the application of more detailed models, as new information about the processes being modeled derive from formalization and logical representation.

The influence of this trend of increasing detail and complexity in data analysis for dental research will become more profound in the immediate future, as new developments in dental science occur simultaneously with new advances in statistical theory and computer science. This progression will only increase the need for dental researchers to establish and develop lines of communication with data analysts.

References

1. Armitage GC, Jeffcoat MK, Chadwick DE, et al: Longitudinal evaluation of elastase as a marker for the progression of periodontitis. J Peridontol 65:120–128, 1994
2. Chugal N, Clive J, Spangberg L: A prognostic model for assessment of the outcome of endodontic treatment: Effect of biologic and diagnostic variables. Oral Surg Oral Med Oral Pathol Oral Radiol Endod, in press
3. Clive J, Woodbury MA: Continuous and discrete global models of disease. Mathematical Modeling 7:1137–1154, 1986
4. Diggle PJ, Liang K-Y, Zeger SL: The Analysis of Longitudinal Data. New York, Oxford University Press, 1994
5. Elashoff J: nQuery Advisor® Version 4.0 User's Guide. Los Angeles, CA, 2000
6. Feinstein A: Clinical Biostatistics. Boca Raton, FL, CRC Press, 2002
7. Friedman G: Primer of Epidemiology, ed 4. New York, McGraw Hill, 1994
8. Glantz A: Primer of Biostatistics, ed 4. New York, McGraw Hill, 1997
9. Imrey PB, Chilton NW: Design and analytic concepts for periodontal clinical trials. J Periodontol (suppl) 63:1124–1140, 1992
10. Kilburn KH, Alsmundsson T: Anteroposterior chest diameter in emphysema. Arch Intern Med 123:379–382, 1969
11. Littell RC, Milliken GA, Stroup WW, et al: SAS® System for Mixed Models. Cary, NC, SAS Institute, 1996
12. Löe H, Anerud A, Boysen H, et al: Natural history of periodontal disease in man. Rapid, moderate and no loss of attachment in Sri Lankan laborers 14 to 46 years of age. J Clin Periodontol 13:431–440, 1986
13. Moses LE, Louis TA: Statistical consulting in clinical research: The two-way street. Stat Med 3:1–5, 1984
14. Neely A: The natural history of periodontal disease in man. Risk factors for progression of attachment loss in subjects receiving no oral health care. J Periodontol 72(8):1006–1015, 2001
15. Piantadosi S: Clinical Trails: A Methodological Perspective. New York; John Wiley & Sons, 1997
16. Riegelman RK: Studying a Study and Testing a Test, ed 4. Philadelphia, Lippincott Williams & Wilkins, 2000
17. Sackett DL, Haynes RB, Tugwell P, et al: Clinical Epidemiology ed 2. Philadelphia, Lippincott Williams & Wilkins, 1991
18. Sheehan TJ: The medical literature: Let the reader beware. Arch Intern Med 140:472–474, 1980
19. Spilker B: Guide to Clinical Trials. New York, Raven Press, 1991

20. Ten Have TR, Landis JR, Weaver SL: Association models for periodontal disease progression: A comparison of methods for clustered binary data. Stat Med 14:413–429, 1995
21. Voelkl KE, Gerber S: Using SPSS for Windows. New York, Springer-Verlag New York, 1999

Address reprint requests to

Jonathan Clive, PhD
Department of Biostatistical Consultation
University of Connecticut Health Center MC3805
Farmington, CT 06030

e-mail: clive@nso1.uch.edu