EVIDENCE-BASED DENTISTRY SERIES

Users' guides to the dental literature: How to get started

Alan B. Carr, DDS, MS,[a] and Glen P. McGivney, DDS[b]

CLINICAL SCENARIO

You are a general practitioner who has just returned from a Continuing Education (CE) course reviewing the latest information about restorative dentistry. While you recognize that most treatments you provide for patients appear to be successful, there are times when you are uncertain about what should have been done. How you resolve such uncertainty is becoming increasingly bothersome, as the information you are exposed to regarding patient care, dental materials, and techniques seems to be overwhelming. In addition, as you get farther from your professional training you worry about how "refined" and up to date your clinical skills are. You have recently heard about a movement in health care, now recognized in dentistry, that attempts to help busy clinicians deal with issues of staying current[1] and devote some time trying it out.

INTRODUCTION

Clinicians who want to stay abreast of significant changes in their respective areas of health care need help in dealing with the large volume of published literature, which professes to be of importance to the reader but is often confusing. This series of articles is intended to help clinicians work through the confusion associated with recognizing and understanding important clinical research and to give the individual clinician a means to achieve an increased level of confidence in clinical decision making. The series of articles to follow (Table I) are patterned after the series published in *The Journal of the American Medical Association* (JAMA),[2] which served as the curriculum for learning the skill of critical appraisal, one of the hallmarks of the 5-step evidence-based process (Table II).

Recognizing that not all published material is useful to us or applicable to patient needs, this series is designed to help clinicians identify whether studies are valid (close to the truth) and applicable to their patients' needs. The intent is not to make the reader a researcher, but to help them become critical "consumers" or users of research. The process of creating an understanding of what constitutes worthwhile research is accomplished by providing a framework of questioning that reveals the level of uncertainty associated with the published evidence. The goal is to provide a process for continual learning that fosters maintenance of professional knowledge and skills in this age of expanding information.

Table I. Dental users' guides

Introductory units
Design
Clinical disagreement and measurement
Core EB units
Diagnosis
Prognosis
Therapy
Harm
Overview

EB = Evidence based.

Table II. Evidence-based 5-step process[3]

1. Translation of clinical problems into answerable questions
2. Conducting an efficient search for the best evidence
3. Critical appraisal of the evidence for its validity and clinical applicability
4. Application of the results of the critical appraisal in clinical practice
5. Evaluation of one's performance

In this initial article, the general approach to critical appraisal will be reviewed as part of the previously mentioned 5-step process (Table II). In addition, this article will review some basic skills required to effectively practice evidence-based (EB) care, defined as the conscientious, explicit, and judicious use of current best evidence from published health care research combined with clinical expertise in the management of individual patients.[3] The skills for practicing evidence-based care include: how to phrase clinical questions to allow effective and efficient searching, how to find useful articles, and what specific guidelines you can use to decide what to read. The remaining articles in the series will deal with specific clinical topics important to the various aspects of patient care (Table I).

To illustrate how critical appraisal is applied in everyday practice, most of the units will follow a specific format: a clinical scenario will serve to identify a knowledge gap (some uncertainty), formulation of an answerable question will focus a search for the best available evidence, and a description of the process of

Adapted from the Oxman, JAMA 1993;270:2093-5.

[a]Director, Advanced Prosthodontics and Maxillofacial Prosthetics, College of Dentistry, The Ohio State University, Columbus, Ohio.
[b]Former Director of Postgraduate Prosthodontics, Department of Restorative Dentistry, School of Dental Medicine, SUNY at Buffalo, Buffalo, N.Y.

Table III. Where do clinical questions come from?[3]

1. Clinical evidence: how to gather clinical findings properly and interpret them soundly
2. Diagnosis: how to select and interpret diagnostic tests
3. Prognosis: how to anticipate the patient's likely course
4. Therapy: how to select treatments that do more good than harm
5. Prevention: how to screen and reduce the risk of disease
6. Education: how to teach yourself, the patient, and the family what is needed

critical appraisal for the respective unit will be the major focus of each unit.

HOW TO PHRASE CLINICAL QUESTIONS

The information used to solve clinical problems comes from the clinical encounter with the patient (diagnostic and related clinical findings) and external sources deemed important to the decision-making process. However, the quality of the information resulting from the clinical encounter is limited by clinical expertise. To provide the most effective and efficient treatment for a patient, the clinician often looks to external sources such as textbooks, advice from experts, or the published literature. Reliance on textbooks may not meet the needs of the patient if the information is out of date or not specific to the patient's needs. Expert advice may be appropriate for specific problems, but for everyday clinical problems, expert opinion has been shown to be less objective,[4] difficult to extrapolate, and confusing, particularly if the experts disagree.[5] The last alternative, to search the published literature, is not without its own problems. Many clinicians believe this resource is too difficult and time-consuming to find a valid and applicable answer to a clinical question. To use this resource effectively and efficiently, one needs to ask questions that are answerable. Clinical questions arise in all aspects of patient care (Table III) so they need to be clearly phrased to allow precise answers. Components of a well-phrased question include: the patient or problem being addressed, the treatment or exposure being considered, any comparison interventions or exposure (when relevant), and the clinical outcome of interest.[3] By way of illustration, questions could be phrased as follows:

- Would a specific appliance (intervention) reduce chronic snoring (outcome) in a middle-aged man (patient)? —(a question of therapy).
- Would a selective group of probing depths (intervention) performed in all patients (patient) improve / equal the diagnosis of periodontal disease (outcome) as accurately as complete mouth probing (comparison)? —(a question of diagnosis).
- Does a single episode of unilateral facial pain (exposure) in a 22-year-old coed (patient) increase the likelihood of developing future temporomandibular dysfunction (outcome)? —(a question of prognosis).
- Do over-the-counter whitening agents (intervention) increase the likelihood of tooth sensitivity (outcome) in patients who use them chronically (patient)? —(a question of harm).
- Would implant-supported dentures (intervention) improve patient outcome (outcome) in an edentulous 65-year-old patient (patient), compared with conventional dentures (comparison)? —(question of therapy).

The greatest return on time invested will come with questions that deal with frequently encountered clinical problems, those that have important clinical implications (are they more costly or are associated with treatment failure?) and those that cause confusion. It is possible to have a patient that brings to mind several questions. How does one decide which to pursue? In these situations, to help you identify which question may be most critical consider one of the following: What is the most important question from the patient's perspective? Which question do you think should be addressed first? And, which question, when answered, will help the most? Once it has been determined which to address first focusing the question as described helps find useful articles and, at the same time, it allows one to omit those that will not be helpful.

It may not always be easy to formulate good clinical questions. This is especially true when dealing with situations that are not routinely familiar. In such situations, we can consider that our questions take 1 of 2 forms; those that are "background" and those that are "foreground" in nature.[6] Background clinical knowledge would include basic knowledge such as, "What is this disorder?" "What causes it?" "How does it present?" Considering such background clinical knowledge, we might develop a foreground question such as, "in patients with severe xerostomia, would a course of pilocarpine improve oral comfort and the quality of life (QOL) (over doing nothing) to be worth the potential side effects and cost?" Although foreground questions usually have 3 or 4 parts, background questions do not. These usually start with what, where, when, why, how, and who, and end with a clinical entity, such as a health state or health intervention.

The reason for the different types of questions is likely due to the level of experience the clinician has with the condition in question. In Figure 1 the rectangle represents the universe of potentially relevant clinical knowledge, and the diagonal denotes the rough division between background and foreground knowledge. On the scale beneath are 3 levels of experience. "A" is a learner with little clinical knowledge or experience, whose needs are largely of the background type

Table IV. Guides for selecting articles that are most likely to provide valid results[2]

Primary studies
Therapy
* Was the assignment of patients to treatments randomized?
* Were all of the patients who entered the trial properly accounted for and attributed at its conclusion?
Diagnosis
* Was there an independent, blind comparison with a reference standard?
* Did the patient sample include an appropriate spectrum of the sort of patients to whom the diagnostic test will be applied in clinical practice?
Harm
* Were there clearly identified comparison groups that were similar with respect to important determinants of outcome (other than the one of interest)?
* Were outcomes and exposures measured in the same way in the groups being compared?
Prognosis
* Was there a representative patient sample at a well-defined point in the course of disease?
* Was follow-up sufficiently long and complete?
Integrative studies
Overview
* Did the review address a clearly focused question?
* Were the criteria used to select articles for inclusion appropriate?

The overview guides make an implicit or explicit reference to investigators' need to evaluate the validity of the studies that they are reviewing to produce their integrative article. The validity criteria one would use in making this evaluation would depend on the area being addressed (diagnosis, prognosis, therapy, harm), and are those that are presented in the part of the Table dealing with primary studies.

portrayed by the vertical dimension of the rectangle. "B" has increased knowledge and experience, and the needs are more evenly divided. "C" has extensive knowledge and experience, and the majority of the knowledge needs would be foreground. In this diagram, please notice the diagonal is placed to show that clinicians are never too inexperienced to ask foreground questions (we cannot let our learners off so easily!), or too experienced to ask background questions. It is the condition of the patient that determines the knowledge needs. Clinicians may be at "C" for frequently encountered problems, at "B" for occasional problems, and at "A" for new disorders or those outside their special area of interest.

TRACKING DOWN USEFUL ARTICLES

Once the clinical question has been phrased, the next step is to find the most current, best available evidence. Several options are available, which could include asking a colleague (or expert), checking textbooks and their references, looking through articles in journals, or searching through a bibliographic database.

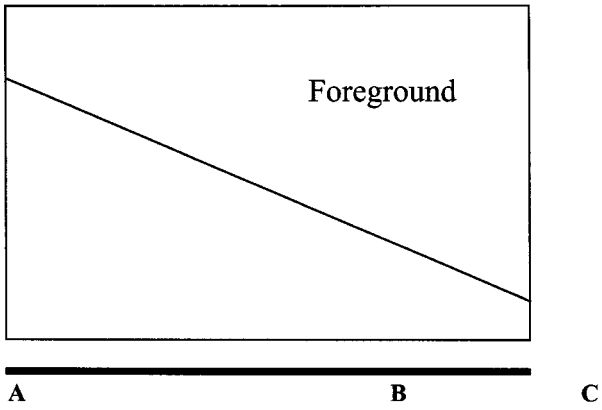

Fig. 1. Background and foreground questions.

As stated previously, if you are seeking answers to a clinical problem, asking a colleague is a risky and inefficient way to get an answer because whether it is a valid answer is unknown. Textbooks will likely not have the most current information and may suffer from the personal biases of the chapter or text authors. Unless you spend an enormous amount of time updating a broad personal file of useful articles, this method also may be less than adequate for providing current best evidence.

The final method, searching a bibliographic database, such as Medline, is a common means of obtaining clinical evidence in health care and is a critical skill required to practice evidence-based care. A variety of databases and searching software is currently available to help clinicians develop these skills. Using the terminology in your carefully phrased question, along with knowledge of key search strategies, is a suggested way to start.[3] The goal of a search is to provide studies that are specific to your question, while excluding those not relevant. From such a search, one can review the titles and/or abstracts for those that appear to be the most valid and do a more careful appraisal of the individual articles you select.

Articles can be searched by any word listed on the database (including those in the title, abstract, author's name, or institution) and by a restricted thesaurus of medical titles known as medical subject headings (MESH).[7] Searching by text word can supplement a MESH search and increase the sensitivity of a search. Maximum sensitivity, specificity, and precision of searches by category of interest (diagnosis, prognosis, therapy, etc) are possible using critical appraisal quality filters.[3] *Sensitivity* is defined as the proportion of studies in Medline meeting criteria for scientific soundness and clinical relevance that are detected by the search. *Specificity* is defined as the proportion of less sound/relevant studies that are excluded by the search strategy. *Precision* refers to the proportion of all citations retrieved that are both sound and relevant.

DECIDING WHICH ARTICLES TO READ

When you conduct a search, how do you quickly know which article(s) to read? Are there key features to look for which can guide you to the strongest evidence? By using 3 key questions from the format of the Users' Guides, one can screen the titles and abstracts from a search to decide which are worthy of more careful study:

1. What are the results?
2. Are the results valid?
3. Will the results help me in caring for my patients?

Once these articles are identified, and if careful evaluation reveals that the results are of interest and possibly applicable to the question, then the research methods can be evaluated to determine whether they are valid or close to the truth. Table IV provides 2 key questions to use as guides for determining validity for both primary studies (those providing original data on a topic) and integrative studies (summaries of primary studies).

As described below, this common thread of questioning provides the reader with a framework for identifying what to read and then for determining how useful the information is to the question at hand. Clinicians will soon recognize that simple "yes" or "no" answers to the above questions may not always be found in an article. This may be confusing at first because our goal was to learn evidence-based skills to help reduce the uncertainty associated with clinical care. How then can one resolve the fact that our primary questions cannot be answered with a yes or no? The questions are intended to focus our attention on the features of research methods that are most critical to providing valid (truthful) and useful (applicable) results. Because methods used in research vary in how strong or weak inferences can be made, understanding this "strength of evidence" is at the heart of critical appraisal. In an ideal world, one would find very strong evidence for all-important questions and have less uncertainty to deal with. However, when there is no such rigorous research evidence for all the important questions, it is still important to understand what level of uncertainty remains (the shades of gray in the evidence). Because decisions must be made, understanding the limitations associated with the best available evidence equips us with critical information to ensure our patients are receiving the best available care. Table II shows critical appraisal as part of the 5-step process of evidence-based practice. The evidence-based process includes a 5-part sequence that includes the translation of clinical problems into answerable questions, conducting an efficient search for the best evidence, critical appraisal of the evidence for its validity and clinical applicability, application of the results of the critical appraisal in clinical practice, and evaluation of one's performance.

For the subsequent core units, the 3 primary questions (or guides) will each have secondary questions, which focus on the clinical problems at hand. The language used for these secondary questions will reflect the important issues of validity and applicability for each respective topic.

SUMMARY

The goal of this series is to help busy clinicians, like the one described in the clinical scenario, learn the basic skills necessary to remove the uncertainty associated with patient care. The next two articles in this series provide information about design and measurement. The design article may help unravel some of the mystery of determining valid evidence based on an understanding of strength of evidence. The design article is followed by information describing why clinical disagreement is so common and how we can be more confident in clinical measures we make and those we read about in research.

The core critical appraisal units of diagnosis, prognosis, therapy, harm, and overview then follow these introductory units and present guides for appropriately selecting and using articles of each respective unit type. The guides or rules of evidence in effect provide the justification and application of the critical questions used to determine whether the results of an article are valid and applicable to the questions you have phrased. As stated previously, the challenge comes when answers are not as clear as yes or no. The reward, however, does not come from being able to find yes and no answers but from removing the uncertainty associated with our clinical questions and replacing the uncertainty with the current best evidence.

REFERENCES

1. Robbins JW. Evidence-based dentistry: what is it, and what does it have to do with practice? Quintessence Int 1998;29:796-9.
2. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature I. How to get started. JAMA 1993;270;2093-5.
3. Sackett DL, Richardson S, Rosenberg W, Haynes RB. Evidence-based medicine: how to teach and practice EBM. Edinburgh, U.K.: Churchill Livingston; 1999.
4. Mulrow C, Langhorne P, Grimshaw J. Integrating heterogeneous pieces of evidence in systematic reviews. Ann Intern Med 1997;127:989-95.
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell P, editors. Clinical epidemiology. A basic science for clinical medicine. 2nd ed. Boston: Little, Brown & Co; 1991. p. 174, 340.
6. Richardson WS, Wilson MC. On questions. Background and foreground. Clin Epidemiol Biostat Newsletter 1997; Nov:6.
7. Greenhalgh T; How to read a paper. The Medline database. BMJ 1997; 315:540-3.

GLOSSARY

Absolute risk - observed or calculated probability of an event occurring in a population that is under study. Absolute risk difference: the difference in the risk for disease or death between an exposed population and an unexposed population.

Accuracy - in measurement, a measure is accurate if it reflects the "true" state of the attribute being measured, without bias. In diagnostic tests, the proportion of test results that agree with the gold standard (a+d/a+b+c+d).

Allocation - assignment of a patient who meets the inclusion criteria of a study to the groups being followed, most often an experimental treatment group and a control (or usual treatment) group.

Alpha error - *see type 1 error* (the error that results if a true null hypothesis is rejected or if a difference is concluded when there is no difference.)

Alternative hypothesis - the opposite of the null hypothesis. It is the conclusion when the null hypothesis is rejected.

Beta error - *see type II error* (the error that results if a false null hypothesis is not rejected or if a difference is not detected when there is a difference).

Bias -any systematic error that results in an incorrect estimate of the association between treatment or exposure and the result. The error related to the way the targeted and sampled populations differ; which threatens the validity of a study.

Blind study - an experimental study in which subjects do not know the treatment they are receiving; investigators may also be blind to the treatment subjects are receiving; see also *double-blind trial*.

Case-control study - an observational study that begins with patient cases who have the outcome or disease being investigated and control subjects who do not have the outcome or disease and then looks backward to identify possible precursors or risk factors.

Case-series study - a simple descriptive account of interesting or intriguing characteristics observed in a group of subjects.

Categorical observation - a variable whose values are categories (an example is type of anemia). See also *nominal scale.*

Chance agreement - a measure of the proportion of time 2 or more rates would agree in their measure or assessment of phenomena.

Chi-square (χ^2) test - the statistical test used to test the null hypothesis that proportions are equal or, equivalently, that factors or characteristics are independent or not associated.

Clinical significance - in contrast to statistical significance, clinical significance is a difference between observations that is a tangible benefit to the patient. A comparison of groups could demonstrate a statistically significant difference without there being any tangible clinical difference perceived by the patient(s).

Clinical trial - an experimental study of a drug or procedure in which the subjects are humans.

Coefficient of variation (CV) - the standard deviation divided by the mean (generally multiplied by 100). It is used to obtain a measure of relative variation.

Cohort - a group of subjects who remain together in the same study over a period.

Cohort study - an observational study that begins with a set of subjects who have a risk factor (or have been exposed to an agent) and a second set of subjects who do not have the risk factor or exposure. Both sets are followed prospectively through time to learn how many in each set develop the outcome or consequences of interest.

Co-intervention - interventions, other than the treatment under study, that are applied differently to the treatment and control groups. Co-intervention is a serious problem when double blinding is absent or when the use of very effective nonstudy treatments is permitted.

Concurrent controls - control subjects assigned to a placebo or control condition during the same period of time that an experimental treatment or procedure is being evaluated.

Confidence interval (CI) - the interval computed from sample data that has given probability that the unknown parameter, such as the mean or proportion, is contained within the interval. Common confidence intervals are 90%, 95%, and 99%.

Confidence limits - the limit(s) of a confidence interval. These limits are computed from sample data and have a given probability that the unknown parameter is located between them.

Confounding variable - a variable more likely to be present in one group of subjects than in another that is related to the outcome of interests and thus potentially confuses, or "confounds," the results.

Continuous scale - a scale used to measure a numerical characteristic such as values that occur on a continuum (an example is age).

Control group - in a clinical trial, subjects assigned to the placebo or control condition; in a case-control study, subjects without the disease or outcome.

Controlled trial - a trial in which subjects are assigned to a control condition as well as to an experimental condition.

Correlation coefficient - r – (Pearson product moment) a measure of the linear relationship between 2 numerical measurements made on the same set of subjects. It ranges from –1 to +1, with zero indicating no relationship.

Crossover study - a clinical trial in which each group of subjects receives 2 or more treatments in different sequences.

Cross-sectional study - an observational study that examines a characteristic (or set of characteristics) in a set of subjects at one point in time; a "snap-shot" of a characteristic or condition of interest; also called survey or poll.

Dependent variable - a known variable, the value of which will affect the outcome of a study; also called a response or criterion variable.

Descriptive statistics - statistics such as the mean, the standard deviation, the proportion, and the rate used to describe attributes of a set of data.

Dichotomous observation - a nominal measure that has only 2 outcomes (examples are gender: male or female; survival: yes or no); also called binary.

Discrete scale - a scale used to measure a numerical characteristic that has integer values (an example is number of pregnancies).

Distribution (population) - the frequency of occurrence for values of a characteristic or variable. Distributions may be based on empirical observations or may be theoretical probability distributions (eg, normal, binomial, chi-square).

Double-blind trial - a clinical trial in which neither subjects nor the investigator(s) know which treatment subjects have received.

Effect or effect size - the magnitude of a difference or relationship. It is used for determining sample sizes and combining results across studies in meta-analysis.

Effectiveness - a measure of the benefit resulting from an intervention for a given health problem under usual conditions of clinical care for a particular group; this form of evaluation considers both the efficacy of an intervention and its acceptance by those to whom it is offered, answering the question, "Does the practice do more good than harm to people to whom it is offered?"

Efficacy - a measure of the benefit resulting from an intervention for a given health problem under the ideal conditions of an investigation; it answers the question, "Does the practice do more good than harm to people who fully comply with the recommendations?"

Estimation - act or process of using a sample of information from a population to draw conclusions about the parameters of that population.

Event - a single outcome (or set of outcomes) from an experiment.

Experimental study - a comparative study involving an intervention or manipulation. It is called a trial when human subjects are involved.

Factor analysis - an advanced statistical method for analyzing the relationships among a set of items or indicators to determine the factors or dimensions that underlie them.

Factorial design - in ANOVA, a design in which each subject (or object) receives one level of each factor.

False negative - a test result that is negative in a person who has the disease.

False positive - a test result that is positive in a person who does not have the disease.

Frequency distribution - list of (d) values that occurs, along with the frequency of (d) occurrence, in a set of numerical observations. It may be set up as a frequency table or as a graph.

Generalizability - the extent to which the findings of a study from a sample of a population can be representative of, or inferred to, the total population (also called external validity).

Gold standard - in diagnostic testing, a procedure that always identifies the true condition—diseased or disease-free—of a patient.

Historical cohort study - a cohort study that uses existing records or historical data to determine the effect of a risk factor or exposure on a group of patients.

Historical controls - in clinical trials, previously collected observations on patients used as the control values against which the treatment is compared.

Homogeneity - the situation in which the standard deviation of the dependent (Y) variable is the same regardless of the value of the independent (X) variable; an assumption in ANOVA and regression.

Hypothesis test - an approach to statistical inference resulting in a decision to reject or not to reject the null hypothesis.

Incidence - a rate giving the proportion of people who develop a given disease or condition within a specified period.

Independent events - events whose occurrence or outcome has no effect on the probability of each other.

Independent variable - the explanatory or predictor variable in a study. It is sometimes called a factor in ANOVA.

Inference (statistical) - the process of drawing conclusions about a population using a sample of observations of the population.

Interaction - relationship between 2 independent variables that have a different effect on the dependent variable; ie, the effect of one level of a factor *A* on the level of factor *B*.

Interrater reliability - the reliability between measurements made by two different persons (or raters).

Interval scale - a measurement scale that sorts and orders, like an ordinal scale, but there is a fixed unit of measurement associated with the scale.

Intervention - the maneuver used in an experimental study. It may be a drug or a procedure.

Intrarater reliability - the reliability between measurements made by the same person (or rater) at 2 different points in time.

Kaplan-Meier product limit method - a method for analyzing survival for censored observations. It uses exact survival time in the calculations.

Kappa (κ) - a statistic used to measure interrater or intrarater agreement for nominal measures.

Level of significance - the probability of incorrectly rejecting the null hypothesis after testing the hypothesis. Also see *alpha value* and *P value*.

Life table analysis - a method for analyzing survival times for censored observations that have been grouped into intervals.

Likelihood ratio - the ratio of true-positives to false positives, in diagnostic testing.

Longitudinal study - a study that takes place over an extended period.

Matching (or matched groups) - the process of making 2 groups homogenous on possible confounding factors. It is sometimes done before randomization in clinical trials.

Measurement error - the amount by which a measurement is incorrect because of problems inherent in the measuring process; also called systematic error bias.

Meta-analysis - a method for combining the results from several independent studies of the same outcome so that an overall *P* value may be determined.

Morbidity rate - the number of patients in a defined population who develop a morbid condition over a specified period.

Mortality rate - the number of deaths in a defined population over a specified period. It is the number of people who die during a given period divided by the number of people at risk during the period.

Nominal scale - the simplest scale of measurement. It is used for characteristics that have no numerical values (examples are race and gender). It is also called a categorical or qualitative scale.

Nonrandomized trial - a clinical trial in which subjects are assigned to treatments on other than a randomized basis. It is subject to several biases.

Null hypothesis - the hypothesis being tested about a population. *Null* generally means "no difference" and thus refers to situations in which there is no difference (eg, between the means in a treatment group and a control group).

Number needed to treat (NNT) - the number of patients who must be exposed to an intervention before the clinical outcome of interest can be expected to occur; for example, the number of patients needed to treat to prevent one adverse outcome.

Numerical scale - the highest level of measurement. It is used for characteristics that can be given numerical values; the differences between numbers have meaning (examples are height, weight, blood pressure level). It is also called an interval or ratio scale.

Observational study - a study that does not involve an intervention or manipulation. It is called case-control, cross-sectional, or cohort, depending on the design of the study.

Odds - the probability that an event will occur divided by the probability that the event will not occur; ie, odds = P/(I-P), where P is the probability.

Odds ratio (OR) - an estimate of the relative risk calculated in case-control studies. It is the odds that a patient was exposed to a given risk factor divided by the odds that a control was exposed to the risk factor.

Ordinal scale - used for characteristics that have an underlying order to their values; the numbers used are arbitrary (an example is plaque scores).

Outcome (in an experiment) - the result of an experiment or trial.

Overview - a structured review of the published literature, which has an explicit and focused question, rules for inclusion of primary studies to present as evidence, an explanation of the strength of the evidence, and a summary of the collective findings from the primary studies. When the data from the primary studies allow combining in rigorous statistical analysis, it is called a meta-analysis.

Paired *t* test - the statistical method for comparing the difference (or change) in a numerical variable observed for 2 paired (or matched) groups. It also applies to before and after measurements made on the same group of subjects.

Placebo - a sham treatment or procedure. It is used to reduce bias in clinical studies.

Population - the entire collection of observations or subjects that have something in common and to which conclusions are inferred.

Post hoc comparisons - methods for comparing means after analysis of variance.

Post-test odds - (in diagnostic testing) - odds that a patient has a given disease or condition based on a diagnostic procedure. They are similar to the predictive value of a diagnostic cast.

Power - the ability of a test statistic to detect a specified alternative hypothesis or difference of a specified size when the alternative hypothesis is true (ie, $1 - \beta$ and β is the probability of a type II error). More loosely, it is the ability of a study to detect an actual effect or difference.

Precision - the range in which the best estimates of a true value approximate the true value. See *Confidence interval*.

Predictive value of a negative test - the proportion of time that a patient with a negative diagnostic test result does not have the disease being investigated.

Predictive value of a positive test - the proportion of time that a patient with a positive diagnostic test result had the disease being investigated.

Pretest odds - in diagnostic testing the odds a patient has a given disease or condition before a diagnostic procedure is performed and interpreted. They are similar to prior probabilities.

Prevalence - the proportion of people who have a given disease or condition at a specified point in time. It is not truly a rate, although it is often incorrectly called a prevalence rate.

Probability - the number of times an outcome occurs in the total number of trials. If *A* is the outcome, the probability of *A* is denoted *P(A).*

Prognostic factors - demographic, disease-specific, or comorbid characteristics associated strongly enough with a condition's outcomes to predict accurately the eventual development of those outcomes. Compare with risk factors. Neither prognostic nor risk factors necessarily imply a cause-and-effect relationship.

Prognosis - the possible outcomes of a disease or condition and the likelihood that each one will occur.

Proportion - the number of observations with the characteristics of interest divided by the total number of observations. It is used to summarize counts.

Prospective study - a study designed before data are collected.

***P* value** - the probability of observing a result as extreme as or more extreme than the one actually observed from chance alone (ie, if the null hypothesis is true).

Random assignment - the use of random methods to assign different treatments to patients or vice versa.

Random error or variation - the variation in a sample that can be expected to occur by chance.

Randomization - the process of assigning subjects to different treatments (or vice versa) by using random numbers.

Randomized clinical trial - an experimental study in which subjects are randomly assigned to treatment groups.

Random sample - a sample of *n* subjects (or objects) selected from a population so that each has known chance of being in the sample.

Random variable - a variable in a study in which subjects are randomly selected or randomly assigned to treatments.

Relative risk (RR) - the ratio of the incidence of a given disease in exposed or at risk persons to the incidence of the disease in unexposed persons. It is calculated on cohort or prospective studies.

Reliability - a measure of the reproducibility of a measurement. It is measured by kappa for nominal measures and by correlation for numerical measures.

Repeated-measure design - a study design in which subjects are measured at more than one point in time. It is also called a split-plot design in ANOVA.

Representative population (sample) - a sample of the population that is similar to the population to which the findings of a study are being applied.

Retrospective cohort study - see *historical cohort study.*

Risk factor - a term used to designate a characteristic that is more prevalent among subjects who develop a given disease or outcome than among subjects who do not. It is generally considered to be causal.

ROC (receiver operating characteristic) curve - in diagnostic testing, a plot of the true-positives on the *Y* axis versus the false-positives on the *X* axis; used to evaluate the properties of a diagnostic test.

Sample - a subset of the population.

Sampled population - the population from which the sample is actually selected.

Sampling distribution (of a statistic) - frequency distribution of the statistic for many samples. It is used to make inferences about the statistic from a single sample.

Scale of measurement - the degree of precision with which a characteristic is measured. It is generally categorized into nominal (or categorical), ordinal, and numerical (or interval and ratio) scales.

Sensitivity - the proportion of time a diagnostic test is positive in patients who have the disease or condition. A sensitive test has a low false-negative rate.

Specificity - the proportion of time that a diagnostic test is negative in patients who do not have the disease or condition. A specific test has a low false-positive rate.

Statistic - a summary number for a sample (eg, the mean), often used as an estimate of a parameter in the population.

Statistical significance - generally interpreted as a result that would occur by chance; eg, 1 time in 20, with a *P* value less than or equal to .05. It occurs when the null hypothesis is rejected.

Statistical test - the procedure used to test a null hypothesis (eg, *t* test, chi square test).

Stratified random sample - a sample consisting of random samples from each subpopulation (or stratum) in a population. It is used so that the investigator can be sure that each subpopulation is appropriately represented in the sample.

Survey - an observational study that generally has a cross-sectional design; a commonly used design to collect points.

Survival analysis - the statistical method for analyzing survival data when there are censored observations.

Systematic error - a measurement error that is the same (or constant) over all observations. See also *bias.*

Test-retest (reliability) - a measure of the degree to which an instrument or test provides a consistent measure of a characteristic on different occasions.

Trial - an experiment involving humans, commonly called a clinical trial. It is also a replication (repetition) of an experiment.

True-negative - a test result that is negative in a person who does not have the disease.

True-positive - a test result that is positive in a person who has the disease.

***t* test** - the statistical test for comparing a mean with a norm or for comparing 2 means with small sample sizes ($n < 30$). It is also

used for testing whether a correlation coefficient or a regression coefficient is zero.

Type I error - the error that results if a true null hypothesis is rejected or if a difference is concluded when there is no difference.

Type II error - the error that results if a false null hypothesis is not rejected or if a difference is not detected when there is a difference.

Uncontrolled study - an experimental study that has no control subjects.

Validity - the property of a measurement that measures the characteristic it purports to measure. Internal validity refers to the quality of rigor of a study, which has limited the effects of bias on the association between the treatment or exposure and result.

Variable - a characteristic of interest in a study that has different values for different subjects or objects.

Variance - the square of the standard deviation.

Variance (within subject) - the variability in measurements of the same object or subject. It may occur naturally or may represent an error.