EVIDENCE BASED DENTISTRY

Design Architecture

Catherine Hayes, DMD, DMSc

Dentists often need to make clinical decisions based on limited scientific evidence. To base a clinical dental practice on scientific evidence more effectively, clinicians must have the skills to evaluate the dental literature critically. In dentistry and dental education, clinical decision making is traditionally based on expert opinion. These opinions usually coincide with standard practice. Recently, however, there has been a shift to support expert opinion or standard practice with evidence.

The shift toward evidence based dentistry provides an opportunity for the transfer of scientific information into clinical decision making (Fig. 1). Simply defined, evidence based dentistry focuses on scientific evidence in guiding clinical decisions. The practice of evidence based dentistry requires reviewing the results of all research relating to a particular clinical issue and assessing the validity of the findings. An additional step is to determine if the study's results will help in caring for a particular patient or group of patients or assessing the external validity (generalizability) of the study. For example, if a particular study evaluates the effect of a specific treatment on a limited patient population, the findings may not be applicable to the practice of a particular clinician.

TYPES OF RESEARCH STUDIES

To evaluate research studies critically, clinicians must have a working knowledge of the principles of scientific research and an understand-

From the Department of Oral Health Policy and Epidemiology, Harvard School of Dental Medicine, Boston, Massachusetts

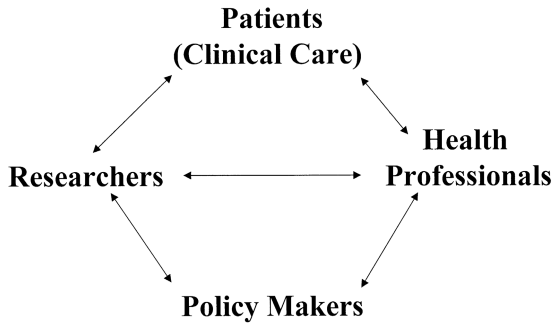

Figure 1. Clinical decision making.

ing of the various types of research studies. Briefly, there are two broad categories of research: basic science and clinical research. The principles that govern the validity of scientific research are common to both branches of scientific research. It is more challenging to ensure that a study is free of bias with clinical research than with basic science or laboratory research, because in the laboratory the researcher has more control over the environment and other variables that may influence the results of the study. This article focuses on assessing the validity of clinical research studies.

It is important to understand the hierarchy of evidence in clinical research. All clinical research studies are encompassed under the broad heading of epidemiologic studies. *Epidemiology* is defined as the study of the distribution and determinants of disease frequency in human populations.[2] The *distribution of disease* refers to who is at risk for a particular disease. For example, older men have the highest risk for oral cancer. The *determinants of disease* are the factors that affect the individual's risk of developing a disease. For example, tobacco use increases an individual's risk for developing oral cancer and is thus considered a risk factor. A risk factor may increase an individual's likelihood of developing a disease (as smoking increases the risk of lung cancer), or it may decrease an individual's likelihood of developing disease (as fluoride decreases the risk of dental caries). In clinical research the aim is to quantify risk relationships as well as benefits of specific treatments to improve the health of the public.

Epidemiologic studies include studies that follow the natural course of disease or treatment effects as well as studies in which the investigators intervene in assigning a treatment for a particular condition or in using a preventive agent to decrease likelihood of disease. These studies can be categorized into two broad categories: descriptive and analytical studies.

Descriptive Studies

Descriptive studies describe the general characteristics of the distribution of a disease, particularly in relation to person, place, and time.

Descriptive studies commonly seen in the dental literature are case reports and case series studies, which are detailed reports of an individual patient (case report) or a group of patients (case series) with a particular disease or who have received a particular treatment. Case series studies abound in the dental literature. An example of a case series study is one in which investigators report on patients treated in their practice with a particular implant system. This report may be a long-term study in which the investigator reports on a variety of treatment outcomes. Although this study may provide interesting information to clinicians, it cannot demonstrate the superiority of one treatment over another without the use of an appropriate comparison group. It is impossible to know what effect a particular treatment has on these outcomes without making a comparison with another treatment. This comparison is possible only with an analytic study design, described later.

Cross-sectional surveys are another type of descriptive study that report the status of an individual with respect to the presence or absence of both exposure and disease assessed at one point in time. These studies are also limited in their ability to demonstrate definitively the benefits of a particular treatment or the significance of a particular exposure. For example, a study that examined 500 individuals, including a complete oral examination, a medical examination, and an interview regarding a variety of health, dietary, and sociodemographic factors, reports on the association between oral health and diet. The investigators report that individuals with good oral health also had a healthy diet, indicating that a healthy diet contributes to adequate oral health. With a cross-sectional study it is impossible to conclude anything about causality. Adequate oral health might enable a person to consume more fruits and vegetables that constitute a healthy diet, a conclusion that is quite different from the conclusion that adequate diet results in good oral health. Essentially, in a cross-sectional study it is impossible to determine if A causes B or vice versa; this situation is analogous to the "chicken-and-egg" phenomenon. In summary, descriptive studies are often referred to as *hypothesis-generating* studies. They are often the first step in investigating a particular scientific question.

Analytic Studies

Analytic studies differ from descriptive studies in that they include an appropriate comparison group that permits the testing of epidemiologic hypotheses. Causality can be investigated with analytic studies. The two broad subcategories of analytic studies are intervention and observational studies.

Intervention studies or clinical trials are considered to be the gold standard for clinical research studies. Because the examiner assigns the exposure or treatment, it is often possible to blind both the subject and the examiner to the treatment assignment, creating a double-blinded

study that minimizes bias of the study findings. Also, the ability to assign subjects randomly into treatment groups ensures that the only difference between study groups is the intervention being evaluated. In a randomized study, each subject has an equal likelihood of being assigned to any of the study groups, thus reducing the influence of bias. This process creates groups that are relatively similar with respect to all variables except for treatment, thus balancing the study groups in terms of known and unknown confounders. Randomization to create similar study groups is possible only with clinical trials and therefore significantly increases the validity of these studies in comparison with other clinical research study designs. Whenever possible, a clinical research question should be addressed with a double-blind, randomized, controlled clinical trial. Such a trial is not always feasible for ethical or logistic reasons, leading investigators to choose one of the other study designs.

In observational studies, investigators observe the natural course of events, noting which subjects are exposed or not exposed, which have had a particular treatment and which have not, and which have or have not developed the outcome. There are two subcategories of observational studies: cohort studies and case-control studies. In a cohort study, subjects are selected on the basis of presence or absence of a particular exposure (treatment) and then followed to determine the association between the exposure (treatment) and outcome. All subjects must be free of the disease of interest at the time the exposure is defined. Cohort studies are efficient for the study of rare exposures, such as occupational exposures (e.g., to asbestos), provide the ability to examine multiple effects of a single exposure, and provide the ability to determine the temporal relationship between exposure and disease. Cohort studies also have disadvantages: they are inefficient for the study of rare diseases, they may be expensive and time consuming, and they have the potential for loss-to-follow-up bias that may affect the validity of the study.[2] An example of a cohort study in dental research is following individual smokers and non-smokers to determine their risk for developing periodontal disease. The study subjects must be free of periodontal disease when the study begins.

The second class of observational studies are case-control studies in which subjects are selected on the basis of whether or not they have the disease of interest. Case-control studies are efficient for studying rare diseases and diseases with long latency periods and have the ability to examine multiple causes of a single disease. The disadvantages of case-control studies include their inefficiency for the study of rare exposures, the difficulty in establishing a temporal relationship between exposure and disease, and their susceptibility to selection and recall bias.[2] An example of a case-control study is a study examining the association between oral cancer and smoking. Oral cancer cases are compared with a similar group of individuals who do not have oral cancer to determine the difference in smoking rates between the groups. This approach was used when it was first discovered that smoking is a significant risk factor

for lung cancer. It is important that cases and controls be selected from the same source population to ensure that study subjects are similar except in respect to the diagnosis of the study disease.

In summary, the study design chosen to address a specific research question must take into account the nature of the exposure or treatment and the nature of the outcome as well as ethical and logistic considerations. For example, if one were studying the effect of two treatments on a particular disease, to randomize subjects ethically to one treatment or the other, there must be sufficient belief that either treatment may offer benefits to the study participant and that neither treatment poses any risk. This assurance is often not possible, and researchers therefore choose one of the other analytic approaches. It is important to decide if the disease or outcome is considered rare and thus decide which observational design is most efficient in addressing the specific questions, keeping in mind that bias and confounding are of greater concern in observational than in intervention studies.

STUDY SAMPLES

Clinical research is conducted using samples of subjects selected from the population of individuals who have the disease of interest. For example, if investigators are interested in evaluating a specific treatment for the replacement of missing teeth, a sample of subjects who meet the study criteria are selected. Each investigator determines a priori the inclusion and exclusion criteria and the size of the sample to be used in a particular study. For example, a study may include adults over age 40 years who have at least six missing teeth. These characteristics are referred to as *inclusion criteria*. Anyone who smokes, who has received antibiotic therapy within the past 6 months, or who has a history of diabetes is not eligible to participate. These characteristics are referred to as *exclusion criteria*. The inclusion and exclusion criteria are based on characteristics that the investigator believes, from previous research or clinical experience, may affect the results of the study. Samples are used to estimate population values, because it is not practical to measure all individuals in a population. Most of the application of statistics in medicine and epidemiology involves making inferences from samples to populations.

MEASURES OF ASSOCIATION

In epidemiologic studies, it is important to quantify the relationship between exposure and outcome. This quantification is accomplished by calculating a relative risk or odds ratio, values that are referred to as *measures of association*. Table 1 demonstrates the method of calculation. First, the relative risk is defined as the ratio of the incidence of disease in the exposed group divided by the incidence of disease in the nonexposed

Table 1. DISEASE STATUS

Exposure Status	Positive	Negative	Total
Positive	a	b	a + b
Negative	c	d	c + d
Total	a + c	b + d	

$$\text{Relative risk} = \frac{\text{incidence of disease in exposed subjects}}{\text{incidence of disease in nonexposed subjects}} = \frac{a/(a+b)}{(c+d)}$$

group. If there is no association between exposure and disease, the relative risk is equal to 1. If the exposure increases the incidence of disease, the relative risk is greater than 1. If the exposure is protective, the relative risk is less than 1.

Example. A randomized, controlled clinical trial was conducted to evaluate the effect of two treatments (scaling and root planing versus systemic antibiotic therapy) on periodontal disease outcomes. Successful treatment was considered to be that which resulted in a probing pocket depth of less than 4 mm at the end of 12 months of follow-up. Of the 500 subjects in the scaling and root planing group, 350 were classified as successful cases, compared with 250 in the antibiotic treatment group. The results are shown in Table 2.

The relative risk of 0.71 indicates that the standard therapy, scaling and root planing, is more beneficial in treating periodontal disease. The classification of treatment success may be considered arbitrary, and the investigator may wish to evaluate several outcomes, such as actual attachment loss in millimeters.

In case-control studies, a relative risk cannot be used, because by definition the cases in a case-control study already have disease. Instead an odds ratio is calculated using the same 2×2 table format. Essentially, the odds ratio determines the odds of being exposed among cases and controls.

Example. A case-control study was conducted to determine the association between cigarette smoking and periodontal disease. Subjects with periodontal disease were compared with a similar group of subjects free of any periodontal

Table 2. EFFECT OF ANTIBIOTIC AND SCALING AND ROOT PLANING THERAPY ON PERIODONTAL DISEASE

Treatment	No. of Successful Outcomes	No. of Treatment Failures
Antibiotic	250	250
Scaling and root planing	350	150
Total	600	400

$$\text{The relative risk} = \frac{250/500}{350/500} = 0.71$$

disease. The participants' smoking status was then ascertained by self report and validated by coltinine levels. Of the 1000 subjects with periodontal disease, 400 were smokers, compared with 200 of the controls subjects. The results are shown in Table 3.

The interpretation of the odds ratio is the same as the relative risk. Therefore, in this example, the conclusion is that smokers are 2.7 times more likely to have periodontal disease than nonsmokers.

CONFIDENCE INTERVALS

The measures of association are calculated with data from the sample of individuals being studied; however, it is the population estimate of risk that is of interest. To estimate the population value of the measure of association, a confidence interval is calculated. A confidence interval is one method of statistical inference that allows statements to be made about the population using data from the sample. The most commonly used method is that of calculating a 95% confidence interval. The methods of calculation are beyond the scope of this discussion; interested readers are referred to a statistical text.[1, 5] Briefly, the data can be used to calculate an interval that includes lower and upper limits. For example, in a study conducted to examine the association between diabetes and tooth loss, the relative risk was calculated to be 1.9, and the 95% confidence interval was calculated to be 1.2 to 2.7. That is, the data indicate that there is approximately a twofold increase in the risk of tooth loss among diabetics as compared with nondiabetics. It can be concluded with 95% confidence that the true risk is between a 20% increase and a 2.7-fold increase. Because the null value of 1.0 is not included in this interval, this result is statistically significant.

ASSESSING VALIDITY

In interpreting the results of any research study, one must consider three possible alternative explanations for research findings.[2, 4] These alternative explanations are chance, bias, and confounding. *Chance* refers

Table 3. INCIDENCE OF PERIODONTAL DISEASE IN SMOKERS AND NONSMOKERS

	Disease Status	
Tobacco Use	***Positive***	***Negative***
Smokers	400	200
Nonsmokers	600	800
Total	1000	1000

$$\text{Odds Ratio} = ad/bc = \frac{(400)(800)}{(200)(600)} = 2.67$$

to the probability that the results observed may be a chance occurrence and not necessarily the result of the treatment under study. Chance is assessed by statistical analysis of the research data and by calculating a *P*-value. The *P-value* is defined as the probability that what was observed, or something more extreme, occurred by chance alone. In scientific research, the cutoff for statistical significance has traditionally been set at 0.05. That is, if a *P*-value is 0.05 or less, the observation is considered to be statistically significant. Numerous statistical tests are used to calculate the *P*-value. The type of test used depends on the type of data being analyzed. Many statistical texts are available that provide details of statistical tests.[1, 5] *Bias* refers to the divergence from the truth. In epidemiologic studies, investigators aim to determine the true relationship between a specific exposure and a specific outcome. Anything that obscures this true association may result in bias. For example, if investigators know the treatment status of a subject, they may pay closer attention to their evaluation of the outcome, thus introducing observation bias into the study and interfering with the results. Standardization and calibration of examiners as well as blinding of the examiner or investigator and the subjects are important steps that can be taken to decrease bias in clinical research studies. A more detailed discussion of bias is provided in the paper by Jacob and Carr.[5] *Confounding* refers to the influence of a second variable or factor on the relationship between an exposure and outcome. This factor must be associated with both the exposure and the outcome. For example, in a study examining the relationship between smoking and oral cancer, alcohol intake could be considered a potential confounder, because it is an independent risk factor for disease and is also associated with smoking. Adjustment should be made for both known and suspected confounders in multivariate analysis of the data.

CRITERIA FOR CAUSALITY

If the findings of a study do not seem to be the result of chance, bias, or confounding, one must attempt to determine if a causal relationship exists. Several criteria are used in epidemiologic research to determine if an association is causal.[3] These criteria include (1) consistency, (2) biologic plausibility, (3) strength of association, (4) temporal relationship, and (5) dose-response relationship. *Consistency* refers to the body of evidence from multiple studies. The results of the present study must be compared with previous similar studies to determine if the results are consistent. If, in fact, several studies demonstrate similar results, this consistency lends credence to the association's being a causal one. The relationship between the exposure and outcome must make sense biologically. Also, the association should be strong. For example, a relative risk of 5.0 is more indicative of a true relationship than a relative risk of 1.2. If it can be demonstrated that an exposure during a specific window of time is related to the outcome, this temporal relationship provides evi-

dence for causal relationship. For example, in the prevention of neural tube defects, it has been demonstrated that women who consume folic acid during the time before neural tube closure have a lower likelihood of giving birth to child with a neural tube defect than women who take folic acid outside this critical period, thus providing evidence of a temporal relationship between the exposure and the outcome. Similarly, demonstrating that the relationship becomes stronger with increasing amounts of the exposure also lends credence to a causal relationship. It is often not possible to satisfy all the criteria for causality. It is the overall body of evidence regarding the association between a particular exposure and outcome that allows the inference of causality to be made. A single study cannot demonstrate causality, because in a single study only one sample is taken from the entire population of individuals with a particular condition. It is possible that the sample is not representative of the entire population. Thus, several studies must consistently demonstrate similar findings before a conclusion of causality can be made.

SUMMARY

It is important for clinicians to understand the type of clinical studies that appear in the literature and the inherent strengths and limitations of each study. The three possible alternative explanations, chance, bias, and confounding, must be considered for any research study. Thus, it is important to evaluate research studies critically in light of this discussion and not simply to summarize the findings. Finally, conclusions about causality can only be made on the body of evidence, not on any single study.

References

1. Campbell M, Machin D: Medical Statistics: A Common Sense Approach, ed 2. New York, John Wiley & Sons, 1993
2. Hennekens CH, Buring JE: Epidemiology in Medicine. Boston, Little Brown, 1983
3. Hill AB: Principles of Medical Statistics. New York Oxford University Press, Chapter XXIV, 1966
4. Hulley SB, Hulley SR: Designing Clinical Research. Baltimore, Williams & Wilkins, 1988
5. Jacob RF, Carr AB: Hierarchy of research design used to categorize the "strength of evidence" in answering clinical dental questions. J Prosthet Dent 83:137–152, 2000
6. Weintraub J, Douglass C, Gillings D: Biostats: Data Analysis for Dental Health Care Professionals. Chicago, Joshi International, 1985

Address reprint requests to

Catherine Hayes, DMD, DMSc
Department of Oral Health Policy and Epidemiology
Harvard School of Dental Medicine
188 Longwood Avenue
Boston, MA 02115

e-mail: catherine_hayes@hms.harvard.edu