

Evidence-based Dentistry: Part IV. Research Design and Levels of Evidence

• Susan E. Sutherland, DDS •

A b s t r a c t

Previous papers in this series on evidence-based dentistry have discussed the first 2 steps in seeking answers to clinical problems — formulating a clear question and strategically searching for evidence. The next step, critical appraisal of the evidence, is made easier if one understands the basic concepts of clinical research design. The strongest design, especially for questions related to therapeutic or preventive interventions, is the randomized, controlled trial. Questions relating to diagnosis, prognosis and causation are often studied with observational, rather than experimental, research designs. The strongest study design should be used whenever possible. Rules have been established to grade research evidence. This paper, the fourth in the series, presents an overview of research methodology most commonly used in the dental literature.

MeSH Key Words: dentistry; evidence-based medicine; research design

© J Can Dent Assoc 2001; 67:375-8
This article has been peer reviewed.

The principles and methods of evidence-based dentistry give dentists the opportunity to apply relevant research findings to the care of their patients. The key to finding evidence is to start with a focused, well-built clinical question.^{1,2} A clear question will help you to identify key words for use in your strategic search. Once evidence has been found, you need to decide if the results are believable and whether the findings can be applied to your patient. Assessing the *validity* (closeness to the truth) and the *relevance* (importance and usefulness) of the evidence is called critical appraisal. The purpose of this paper is to discuss the concept and rationale for “levels of evidence” and the types of research designs that are appropriate for answering the clinical questions most commonly encountered in dental practice.

The Evidence Hierarchy

Evidence-based practice involves tracking down the available evidence, assessing its validity and then using the “best” evidence to inform decisions regarding care. Rules of evidence have been established to grade evidence according to its strength.³⁻⁵ Systematic reviews and randomized controlled trials represent the highest levels of evidence, whereas case reports and expert opinion are the lowest (Table 1). This “ladder of evidence” was developed to a

large extent for questions related to interventions or therapy. For questions related to diagnosis, prognosis or causation, other study designs such as cohort studies or case-control studies will often be more appropriate. For these types of studies, it is useful to think of the various study designs not as a hierarchy, but as categories of evidence, where the strongest design which is possible, practical and ethical should be used.

It should be noted that using the “rules” or categories of evidence only helps classify studies based on the type of research design. The quality of each individual study still needs to be assessed for strengths and weaknesses using the techniques of critical appraisal.

Basic Concepts of Research Design

Clinical research can be experimental or observational. In experimental studies, the intervention is under the control of the researcher, whereas in observational studies, the researcher observes patients at a point in time (cross-sectional studies) or over time (longitudinal studies). If the observations are made by looking forward and gathering new data, the study is prospective; if the data already exist (for instance, in dental records or as census data), the studies are retrospective.

Table 1 The evidence ladder^a

High-quality systematic reviews
Large randomized trials with clear-cut results
Small randomized trials with uncertain results (i.e., positive trends without statistical significance)
Nonrandomized trials with contemporary controls
Nonrandomized trials with historical controls
Cohort studies
Case-control studies
Dramatic results from uncontrolled studies (e.g., the treatment of infections with penicillin in the 1940s)
Case series and other descriptive studies
Reports of expert committees and opinions of respected authorities, based on clinical experience

^aBased on references 3, 4 and 5

Experimental Studies

Experimental studies can be either controlled (there is a comparison group) or uncontrolled. Uncontrolled studies provide very weak evidence and should not be used to guide practice. These studies may be carried out early in an area of research to explore the safety of a new intervention, to identify unanticipated effects and to gather baseline data for the planning of more definitive trials. For similar purposes, a study may use a historical control group, where data would be gleaned from a chart review or a previous study. These designs are generally weak because many factors may have changed since the data were gathered and there are no assurances that bias was not introduced in the collection, recording or retrospective interpretation of the data.

Randomized Controlled Trials

Randomized controlled trials (RCTs) are the gold standard by which all clinical research is judged. The fact that randomization keeps study groups as similar as possible from the outset, together with other features of the design, such as blinding, sample size justification, appropriate outcome measures and statistical analysis, means that RCTs have the greatest potential to minimize bias. Bias is any factor or process that acts to deviate the results or conclusions of the study away from the truth, causing either an exaggeration or an underestimation of the effects of an intervention.⁶ In fact, methodology research has shown that, most often, bias and weak designs cause trials to conclude that a treatment is effective when it really may not be, and to overestimate the effect, even when it is true.⁷⁻⁹

Randomization of treatment allocation is what makes the RCT one of the simplest and most powerful tools of scientific research.¹⁰ In any study involving people there are potentially many unknown factors — genetic or lifestyle factors, for example — which can have a bearing on the outcome. Randomization, if done properly, reduces the risk that these unknown factors will be *seriously* unbalanced in the various study groups. The allocation sequence *must* be

randomly allocated. This can be by the flip of a coin, or more usually, by using random number tables or computer-generated sequences. Dates of birth (even or odd), chart numbers or any other alternating type of sequence is inappropriate, because there is the potential for people associated with the study, either directly or indirectly, to guess the sequence. Although sometimes called “pseudo-” or “quasi-randomized,” these trials are nonrandomized.

Blinding is another key feature of RCTs. The “double-blind” trial is one in which both the researcher and the patient do not know whether the patient is in the experimental group or the control group. This design is most useful when the control group is receiving an identical placebo drug or “sham” intervention, but falls down in many types of important studies. Few patients would agree to participate in a study where the control group received “sham” orthognathic or TMJ surgery. Surgical trials, by necessity, are “open” trials, since both the investigator performing the surgery and the patient know the intervention. However, there are 3 other groups or individuals who can be blinded. The investigator evaluating the outcome must not be the surgeon who performed the operation and should be kept unaware of the intervention (the patient must be thoroughly informed about the importance of not “dropping hints”). Although a surgical scar is usually a giveaway, the outcome measure should be planned with this in mind. The other 2 groups who can be kept blinded are the statistician(s) doing the data analysis and the investigators who write the results of the trial. To do this, the allocation code is not broken until these components are completed. Although blinding of the statistician is being done with increasing frequency, blinding of the investigator writing the report is rarely done.

Two special types of RCTs — cross-over studies and split-mouth designs, have been used in dental research, particularly in the periodontal literature. Although these designs require smaller sample sizes to detect a treatment effect, their use is fraught with peril and may be inappropriate, unless certain criteria are met. A discussion of issues related to these designs is beyond the scope of this article, but interested readers are referred to several excellent papers.¹¹⁻¹⁵

Observational Studies

RCTs cannot answer all clinical questions. There are situations where they may not be necessary, appropriate, ethical or feasible, or they simply may not have been done yet. In general, questions of therapy are best answered by RCTs, or even better, meta-analyses if available, whereas questions of diagnosis, prognosis and causation may be best addressed by observational (sometimes called “epidemiological”) studies. Observational studies, which are frequently undertaken in dentistry, can be even more challenging to design and execute, in terms of controlling bias.

Therefore, it is very important to use critical appraisal methods (presented in part 6 of this series) to assess the validity of these studies.

To a large extent, the type of observational study done depends on the rarity of the disease or condition and on issues related to human resources and economics. Usually several methods of answering the question are possible and the strongest design should be used. The following are some of the most common types of observational studies.

The Cohort Study

In a cohort study, it is known at the outset whether people have been exposed or not to a treatment or possible causal agent (i.e., a vaccine, a drug or an environmental toxin) and are divided into groups or cohorts (treated or exposed versus nontreated or nonexposed) on this basis. They are then followed forward in time (prospectively) for years or even decades to see how many in each group develop a particular disease or other outcome. These studies are usually less expensive and easier to administer than RCTs. They may also be ethically more acceptable, because a potentially beneficial treatment is not withheld, and conversely, a possibly harmful treatment is not given. The major disadvantage is that we can never be sure that the cohorts are well matched and that there are not other factors, such as social class or occupational exposure, that may influence the results. In addition, for rare disorders, the sample size or length of follow-up needed to show an effect may be prohibitively large.

One of the most famous cohort studies¹⁶ followed 40,000 British doctors in 4 cohorts (non-smokers, light, moderate and heavy smokers) for 40 years, from 1951 to 1991. This study, which achieved 94% follow-up, was instrumental in establishing the causal link between smoking and lung cancer and other diseases, as well as the dose-response relationship between smoking and lung cancer. This study showed the tremendous strength of a well-designed cohort study.¹⁷

A variation of a cohort study is a longitudinal study in which there is only one group. Included in the group (called the *inception cohort*) are people who have a positive screening test (for example, for a new genetic marker) or who have all been diagnosed with an early stage of a disease (for example, multiple sclerosis). They are then followed and evaluated on a repeated basis to assess the development of the disease (i.e., in the example of the genetic marker), or the time frame for particular outcome measures, in the case of a chronic disease.

The Case-control Study

In this type of study, people with a particular condition (the “cases”) are matched with a group of people who do not have the disorder (the “controls”) and the researchers look back in time to determine the proportion of people in

each group who were exposed to the suspected causal factor. This is a relatively quick and inexpensive study and is often the best design for rare disorders or when there is a long time lag between the exposure and the outcome. An example of an important case-control study is the one that examined the relationship between the development of vaginal cancer in young women and the use of diethylstilbestrol by their mothers during pregnancy.¹⁸ The major disadvantage of this type of study is that it relies on memory (“recall bias”) or on medical records, which may be inaccurate or incomplete.

Cross-sectional Studies

This design attempts to establish an association between a possible causal factor and a condition, by determining an exposure to the factor and “caseness” at the same time. For instance, a large cross-section of women might be interviewed to determine if they had given birth to a baby with a cleft palate *and* if they had taken a particular drug during pregnancy. Although this type of study is relatively easy and inexpensive to carry out and ethically acceptable, it can only establish an association, not a cause and effect relationship. In addition, both “exposure” and “caseness” may depend on accurate recall of past events.

Case Reports and Case Series

Case reports and case series are often used to describe a condition (usually a rare disorder or a novel aspect of a less rare condition), a new treatment or innovation, or adverse effects of an intervention. They often provide a richness of information which cannot be conveyed in a trial. The description of cases may alert the world to important new problems and then allow hypotheses to be developed, leading to focused studies of stronger design. Case studies and case series are relegated to the lowest rungs of the evidence ladder, however, because isolated observations are collected in an uncontrolled, unsystematic manner and the information gained cannot be generalized to a larger population of patients.

Integrative Studies

Basing important clinical decisions on single trials, especially when the result is a change in treatment policy, is risky. Because of the numbers of patients needed to detect small to moderate differences for clinically important outcome measures, definitive answers may not be found in single studies, unless they are well-designed “large simple trials.” These “mega” trials, which usually involve many thousands of patients, have rarely been carried out in dentistry.

When the information from all relevant trials addressing the same question is combined using well-established, rigorous methodology,¹⁹ the result is a systematic review or overview. If the results of each trial were reported in such a

way that they can be combined statistically by the researcher, the result is a quantitative systematic review or meta-analysis. Although systematic reviews are observational, retrospective research studies, they employ scientific methods to control bias and, in doing so, provide potent methods for synthesizing and summarizing data. In fact, systematic reviews are considered the highest level in the evidence hierarchy.

Conclusion

Once research has been published, especially in a respected, peer-reviewed journal, it achieves a certain level of respectability and credibility. Unfortunately, methodological research has shown that acceptance of the findings of many published studies is not always deserved.^{9,20,21} For the unsuspecting reader of the dental literature, this can be a frightening revelation. Fortunately, most clinical studies can be assessed easily by using the techniques of critical appraisal. Critical appraisal of studies that address the various types of questions encountered in dental practice is the subject of the final 2 papers in this series. ♦

Dr. Sutherland is a full-time active staff member of the department of dentistry at the Sunnybrook and Women's College Health Sciences Centre, University of Toronto in Toronto.

Correspondence to: Dr. Susan E. Sutherland, Department of Dentistry, Suite H126, Sunnybrook and Women's College Health Sciences Centre, 2075 Bayview Ave., Toronto, ON M4N 3M5. E-mail: susan.sutherland@swchsc.on.ca

The views expressed are those of the author and do not necessarily reflect the opinions or official policies of the Canadian Dental Association.

References

- Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995; 123(3):A12-3.
- Sutherland SE. The building blocks of evidence-based dentistry. *J Can Dent Assoc* 2000; 66(5):241-4.
- Woolf SH, Battista RN, Anderson GM, Logan AG, Wang E. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. A report by the Canadian Task Force on the Periodic Health Examination. *J Clin Epidemiol* 1990; 43(9):891-905.
- Sackett D. Rules of evidence and clinical recommendations. *Can J Cardiol* 1993; 9(6):487-9.
- Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1992; 102(4 Suppl):305S-311S.
- Jadad A. Bias in RCT's: beyond the sequence generation. In: *Randomized Controlled Trials: A User's Guide*. London: BMJ Publishing; 1998. p. 28-45.
- Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983; 309(22):1358-61.
- Antczak AA, Tang J, Chalmers TC. Quality assessment of randomized control trials in dental research. I. Methods. *J Periodontol Res* 1986; 21(4):305-14.
- Antczak AA, Tang J, Chalmers TC. Quality assessment of randomized control trials in dental research. II. Results: periodontal research. *J Periodontol Res* 1986; 21(4):315-21.
- Jadad AR, Rennie D. The randomized controlled trial gets a middle-aged checkup. *JAMA* 1998; 279(4):319-20.
- Chilton NW, Fleiss JL. Design and analysis of plaque and gingivitis trials. *J Clin Periodontol* 1986; 13(5):400-10.
- Goldberg JD, Weiss AI, Koury KJ. Design of clinical trials for chronic diseases: implications for periodontal disease. *J Clin Periodontol* 1986; 13(5):411-7.
- Hujoel PP, Moulton LH. Evaluation of test statistics in split-mouth clinical trials. *J Periodontol Res* 1988; 23(6):378-80.
- Antczak-Bouckoms AA, Tulloch JF, Berkey CS. Split mouth and cross-over designs in dental research. *J Clin Periodontol* 1990; 17(7 Pt 1):446-53.
- Newcombe RG, Addy M, McKeown S. Residual effect of chlorhexidine gluconate in 4-day plaque regrowth crossover trials, and its implications for study design. *J Periodontol Res* 1995; 30(5):319-24.
- Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ* 1994; 309(6959):901-11.
- Greenhalgh T. How to read a paper. Getting your bearings (deciding what the paper is about). *BMJ* 1997; 315(7102):243-6.
- Herbst AL, Anderson S, Hubby MM, Haenszel WM, Kaufman RH, Noller KL. Risk factors for the development of diethylstilbestrol-associated clear cell adenocarcinoma: a case-control study. *Am J Obstet Gynecol* 1986; 154(4):814-22.
- Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized controlled trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol* 1995; 48(1):167-71.
- Altman DG. The scandal of poor medical research. *BMJ* 1994; 308(6924):283-4.
- Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodologic quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273(5):408-12.

C D A R E S O U R C E C E N T R E

The following texts on evidence-based medicine are available on loan to CDA members:

Evidence-based medicine: how to practice and teach EBM, by David L. Sackett; *Evidence-based practice: a primer for health care professionals*, by Martin Dawes and others; *The evidence-based medicine workbook: critical appraisal for clinical problem solving*, by Robert A. Dixon and others. For more information, contact the Resource Centre at tel.: **1-800-267-6354** or **(613) 523-1770**, ext. 2223; fax: **(613) 523-6574**; e-mail: info@cda-adc.ca. (Shipping charges and taxes apply on all loans.)