

Evidence-based Dentistry: Part V. Critical Appraisal of the Dental Literature: Papers About Therapy

• Susan E. Sutherland, DDS •

A b s t r a c t

Evidence-based dentistry involves defining a question focused on a patient-related problem and searching for reliable evidence to provide an answer. Once potential evidence has been found, it is necessary to determine whether the information is credible and whether it is useful in your practice by using the techniques of critical appraisal. In this paper, the fifth in a 6-part series on evidence-based dentistry, a framework is described which provides a series of questions to help the reader assess both the validity and applicability of an article related to questions of therapy or prevention.

MeSH Key Words: dental research/methods; evidence-based medicine; human research design

© J Can Dent Assoc 2001; 67(8):442-5
This article has been peer reviewed.

The need for valid and current information for answering everyday clinical questions is growing. Ironically, the time available to seek the answers seems to be shrinking. In addition, a surprising amount of published research “belongs in the bin.”¹ Critical appraisal can be used to rapidly assess and discard reports of research studies that are irrelevant or of poor quality. The purpose of the next 2 papers in this series is to introduce the tools used to critically appraise papers according to the type of clinical question addressed by the study. These concepts and tools were developed by the evidence-based medicine group at McMaster University^{2,3} and are used worldwide in the practice of evidence-based care in many of the health sciences professions. In this paper, techniques to evaluate research studies related to questions of therapy will be discussed. In the final paper in the series, critical appraisal techniques will be presented for the evaluation of papers about diagnostic tests, causation and predicting prognosis.

Questions Relating to Therapy

When considering a new therapeutic or preventive intervention, common sense dictates that the highest levels of evidence — randomized controlled trials (RCTs) and systematic reviews — should be sought before subjecting patients to possibly useless, and perhaps even harmful,

treatments. The RCT is the strongest design for a clinical study because randomization of patients to the comparison groups minimizes bias by ensuring that the patients in each group are as similar as possible in all respects, except for the treatment under investigation. As more RCTs studying a particular question become available, it is more difficult for the reader to process and synthesize all of the information to find the answer to a clinical question. Systematic reviews (sometimes called “secondary” publications or integrative research) summarize, analyze and report the combined results of a number of RCTs. They are done with the same rigour that is expected of primary studies, but the “unit of analysis” is the individual study rather than the individual patient.

Randomized Controlled Trials

Asking the following questions will help you to assess the validity and the importance of a study about a treatment or a preventive intervention.^{4,5}

Was the allocation of patients to study groups random?

The first thing to consider is whether or not the treatment allocation was truly randomized. Was the assignment of each patient to either the treatment group or the control

group decided completely by chance, by the flip of a coin or by some other similar method? This assignment helps to ensure that people in the treatment and the control groups are similar at the outset and that differences at the end of the trial are due to the intervention and not to some “selection” factor. Look for words like random allocation, randomly assigned or randomized trial in the title or abstract. If absent, go on to the next title. In the methods section, look for a description of the way randomization was done. If this was done by the flip of a coin, coded and sealed envelopes, random number tables or a computer-generated sequence, randomization was appropriate. Any method of allocation where the sequence could be guessed by anyone is inappropriate. Unfortunately, randomization methods are not often described and you are left to wonder about the details. When reading these papers, you might want to remember that research has shown that inadequate randomization can exaggerate the estimate of treatment effect by 41% and that even if the paper states that the study is randomized, but the description of the randomization methods is unclear, the estimate of the effect is exaggerated, on average, by 30%.⁶

Were all the patients who entered the trial accounted for and analyzed at the end of the study?

It is not uncommon to read a study which began with a certain number of patients and ended with a lesser number, with a mere statement that a particular number of patients were “not available for follow-up.” The reasons for loss to follow-up may be extremely important. In fact, patients who do not complete trials may provide more information about the intervention than those who do. Patients may have dropped out because of side effects (even to the placebo) or perhaps because they benefited from the intervention and with the resolution of their problem or condition, chose not to return for follow-up. Even when loss to follow-up is accounted for and explained in the paper, follow-up of less than 80% of the patients enrolled at the beginning is generally considered unacceptable.³

It is also important that patients be analyzed in the group to which they were originally randomly allocated, even if they switched groups or were noncompliant with either the experimental or the control treatment. This is the intention to treat principle and it serves to preserve the powerful function of randomization; factors we cannot know about will remain reasonably equally distributed between the 2 groups. This consistency prevents the intervention from appearing to be effective when it is not and makes the results of the study more conservative and more believable.

Were patients, clinicians and study personnel “blinded”?

Patients should be blinded as to whether they are in the active or the control group to minimize the placebo effect. To reduce “measurement bias,” the clinician assessing the outcome should also be blinded. The greater the extent of blinding of all study personnel, the more rigorous the trial.

Were the groups similar at the outset and treated equally throughout the study?

Randomization does not always create groups that are balanced for known prognostic factors, especially in small studies. The investigators should present baseline data on all patients in each group and if there are significant differences, assure the reader that these differences were adjusted for in the statistical analysis.

Co-interventions are additional treatments other than those being investigated that are used by or given to patients. Co-interventions are problematic if they are given differentially to either the treatment and the control group and are much less of a problem in double-blind studies. It is helpful to the reader if allowed co-interventions are described in the methods section and if the extent of use of non-permissible co-interventions is documented in the results. The success of blinding can be assessed by the investigators by asking both clinicians and patients after completion of the study what group they thought the patient was in and comparing the answers with the actual allocation. If the results of this analysis show that more patients or clinicians guessed correctly than one would expect by chance (say, more than one person in 20 guessed correctly, or $p > 0.05$), then the methods used for blinding didn't really work.

Were clinically important outcomes assessed?

A clinically important outcome is one that is important to the patient. A carious tooth that requires treatment is important to a patient; a cariogenic bacterial count generally is not. Mobility and loss of teeth are important to patients; radiographic measurements of bone loss are not. Microbiological and radiographic end points are “surrogate” or secondary end points, not primary clinical ones. Although these substitute outcome measures are important to study early on in the research of a disease to help understand the disease process, they are often chosen inappropriately in more definitive trials because a difference can be shown between the treatment and the control group using smaller sample sizes and shorter follow-up times. The difference shown, however, may not be relevant to the patient. There are many examples in the biomedical literature where subsequent large trials fail to show the effectiveness of an intervention when clinically relevant outcomes, as opposed to surrogate ones, are measured.⁷

Can the results of the study be applied to my patient(s)?

By looking at the study's inclusion and exclusion criteria, you can make a reasonable judgment as to whether or not the results of the study are useful in the management of the patient problem at hand. If the results can be generalized to your patients, it is important to consider if the benefit is greater than any potential harm, added cost or inconvenience.

Systematic Reviews

Systematic reviews (also known as overviews or as meta-analyses if results of the primary studies can be combined mathematically) differ from traditional journal or textbook reviews.⁸ Systematic reviews have most often been done for questions relating to therapy, although they can and have been done for all types of questions. While widely accepted standards have been developed⁹ for the conduct of systematic reviews for issues related to therapeutic questions, agreed-upon standards and critical appraisal techniques for reviews which synthesize the results of observational studies remain undeveloped at this time. The following guidelines will enable you to judge the validity and usefulness of a systematic review^{10,11} of RCTs addressing issues of therapy.

Was a clearly stated question asked?

The question being addressed by the review should be focused in terms of the population being studied, the intervention given and the outcomes being considered. If these key elements are not present in the title or the abstract, you should go on to the next title.

Were the inclusion criteria appropriate?

Specific inclusion and exclusion criteria related to the population, intervention, outcome and acceptable study design must be well defined and clearly stated. This allows the reader to decide if the appropriate studies were included. In addition, this permits the review to be replicated and avoids preferential citation of studies that support a particular viewpoint.

Was a comprehensive literature search done?

It is important that all pertinent studies are included and that important ones have not been missed. There is evidence that a number of high-quality, methodologically sound studies remain unpublished ("publication bias") because their results are negative.¹² The authors of the overview should clearly state their search strategy, including key words and databases used. Ideally, the search should include other sources, such as multiple databases, reference lists from relevant papers, conference proceedings and personal contacts with experts.

Was the validity (quality) of the primary studies assessed?

It is important to know the quality of the included studies. If many of the studies were weak, their combined results will not be believable. It is helpful to the reader if a study-by-study critique is presented in a table or if there is a thorough discussion of the methodological quality of the included studies.

Was the assessment of the studies reproducible and free of bias?

Decisions regarding which studies met the inclusion criteria, the validity of each primary study and the meaning of the data within each study involve judgment on the part of the reviewer. All such judgments are susceptible to error and unintentional bias. To overcome this, 2 or more authors of the review should perform each of these steps independently, blind to each other's decisions, and then come to agreement by consensus. Ideally, the names of the authors of the primary studies and their affiliated institutions should be deleted during the review process.

Were the results similar from study to study?

Even with fairly strict inclusion criteria, there is bound to be some variation in the results of the eligible studies. The authors should present the salient features of each study in terms of the included patients and the stage or severity of their disease, the intervention (for example the dose, route or timing) and the way in which the outcome was measured, and they should try to explain the variability of the results.

Were the findings of the studies combined appropriately?

As a reader, you will want to know if it was reasonable for these studies to be combined in a systematic review, keeping in mind that no 2 studies would (or should) ever be exactly the same. If the studies seem too dissimilar, they should not be combined mathematically. A statistical test can be done to see if the results are different merely by chance. If this test indicates that the study results are similar enough to combine mathematically, a meta-analysis is done. A "vote count," that is, a vote counting the number of positive studies versus the number of negative studies is not appropriate. The reason for this is that small studies may be "underpowered," i.e., there may not have been a large enough sample size for the study to have sufficient power to detect a difference in treatment effect between the experimental and the control groups. One of the major advantages of meta-analysis is that the results of a number of small but similar studies can be combined to achieve a large enough sample to detect an effect.

Were the authors' conclusions supported by the data?

The results of each study must be reported in enough detail to allow the reader to judge the grounds for the reviewers' conclusions. Are the conclusions justified, given the methodological quality of the studies? Do the results and conclusion answer the original question asked?

Will the results help in caring for patients?

As with all research, you need to decide if your patients and your practice setting are similar to the patients of studies included in the review. Are you able to implement the intervention in your practice and are the potential benefits worth any potential harm or cost?

Conclusion

A well-designed randomized controlled trial is the strongest research design for clinical trials. The systematic review is a powerful way to assemble multiple studies and synthesize their findings. In both cases, however, the credibility of the research needs to be determined through the use of critical appraisal techniques.

In the final paper in this series, critical appraisal methods and their application to studies related to other types of clinical questions commonly encountered in dental practice — questions related to diagnostic tests, to etiology, causation or harm, and to prognosis — will be discussed. ♦

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.

Correspondence to: Dr. Susan E. Sutherland, Department of Dentistry, 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 opinion or official policies of the Canadian Dental Association.

References

1. Greenhalgh T. How to read a paper: getting your bearings (deciding what the paper is about). *BMJ* 1997; 315(7102):243-6.
2. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston: Little, Brown and Company; 1991.
3. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. London: Churchill Livingstone; 1997.
4. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1993; 270(21):2598-601.
5. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA* 1994; 271(1):59-63.
6. 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.

7. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996; 125(7):605-13.
8. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997; 126(5):376-80.
9. 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.
10. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *CMAJ* 1988; 138(8):697-703.
11. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *JAMA* 1994; 272(17):1367-71.
12. Felson DT. Bias in meta-analytic research. *J Clin Epidemiol* 1992; 45(8):885-92.