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

- Susan E. Sutherland, DDS -

Abstract

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

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.” 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 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.

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
Believable. The results of the study are more conservative and more
from appearing to be effective when it is not and
ment is performed. This consistency prevents the inter-
will remain reasonably equally distributed
treatment. This is the powerful function of randomization; factors we cannot
between the 2 groups. This consistency prevents the inter-
undergo a mere statement that a particular number of patients
didn't really work. Correctly, or
were “not available for follow-up.” The reasons for loss to
referred to as "selection" factor. Look for words like random allocation,
sealed envelopes, random number tables or a computer-
sequence, randomization was appropriate. Any
before any two patients were assigned to the same group. If this was done by the flip of a coin, coded and
method of allocation where the sequence could be guessed
randomization can exaggerate the estimate of treatment
effect by 41% and that even if the paper states that the
methods is unclear, the description of the randomiza-
are often chosen inappropri-
important to the patient. A carious tooth that requires treatment is
are measured.7
were the groups similar at the outset and
treated equally throughout the study?
Randomization does not always create groups that are
balance of known prognostic factors, especially in small
patients in each group and if there are significant differ-
ience, assure the reader that these differences were adjusted
for in the statistical analysis.
Co-interventions are additional treatments other than
patients, clinicians and patients after
for in the statistical analysis.

Were the 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

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 inter-
vention and with the resolution of their problem or condi-
tion, 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 inter-
vention from appearing to be effective when it is not and
makes the results of the study more conservative and more
believable.

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 gener-
ally is not. Mobility and loss of teeth are important to
patients; radiographic measurements of bone loss are not.
Microbiological and radiographic end points are “surro-
gate” 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 under-
stand the disease process, they are often chosen inappropri-
ately 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 litera-
ture where subsequent large trials fail to show the effective-
ness 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. 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 developed 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 of RCTs addressing issues of therapy.

Was a comprehensive literature search done?

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 the same, the 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.

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 the same, the 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

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.