

Evidence-based Dentistry: Part VI. Critical Appraisal of the Dental Literature: Papers About Diagnosis, Etiology and Prognosis

• Susan E. Sutherland, DDS •

A b s t r a c t

Critical appraisal methods assist the reader in assessing the validity (closeness to the truth) and the relevance (usefulness in everyday practice) of research findings. The specific techniques of critical appraisal can vary somewhat, depending on the nature of the research question. In this paper, the final in a 6-part series on evidence-based dentistry, frameworks are presented to enable the judicious reader of the dental literature to apply sensible questions to the evaluation of papers related to diagnosis, etiology and prognosis.

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

© J Can Dent Assoc 2001; 67(10):582-5
This article has been peer reviewed.

In the preceding article, the concept of critical appraisal — discovering whether a research study is both believable and useful for your patient or your practice — was introduced. Guidelines were also provided to assist the reader in critically appraising articles related to therapeutic or preventive interventions. These guidelines were based on a series of questions developed by the McMaster Evidence-based Medicine Group.^{1,2} In this paper, tools are provided to help determine the validity and usefulness of research papers about diagnostic tests, causation or prognosis.

Assessing Articles About Diagnostic Tests

When considering a new diagnostic test, it is important to remember that tests are rarely 100% accurate; there will be false positives and false negatives with any test. The best tests are the ones that are good at detecting most of the people with the condition (high sensitivity) and at excluding people who don't have the condition (high specificity). The most useful tests help to establish an accurate diagnosis, which supports the most appropriate treatment leading to the best outcome for the patient. The questions below will help you to decide if a paper that claims to validate a diagnostic test is believable and useful.³⁻⁵

Was the test compared blindly and independently with a “gold” standard?

The reference or “gold” standard is considered to be the

“truth.” This standard may be a biopsy or autopsy finding, a well-established blood test or some other “proof” that the condition does or does not exist. No reference standard is perfect and for many conditions, there is no gold standard. If that is the case, it is up to the authors to explain the cluster of criteria or the theoretical construct against which they are comparing the new test.

It is important that the reference standard and the new test be interpreted by 2 different investigators, neither of whom should know the results of the other test, the conclusions of the other researcher or the details of the case history. All dentists have experienced the value of a bitewing radiograph in the diagnosis of an interproximal lesion that was suspected in the clinical exam (or vice versa). While knowledge of the results of a cluster or a sequence of tests (such as exam findings *plus* radiographic findings) is appropriate and indeed useful in the clinical setting, such information introduces bias in the research setting, when the diagnostic test is being developed and evaluated. The researchers should be blinded to the findings of other tests or pertinent patient information at this point.

Was the test evaluated in a range of patients representative of a clinical practice setting?

An appropriate patient sample is one in which mild, moderate and severe forms of the condition exist, as well as

conditions that are different from each other but commonly confused. A test is not needed to differentiate incipient caries from gross caries; however, a test that helps the practitioner decide at what point to intervene in the progression of caries is valuable. Similarly, tests that help to differentiate odontogenic pain from facial neuralgia are helpful. A test's predictive value changes with the prevalence of the target disorder. If a diagnostic test is validated in a highly specialized practice setting, such as a university or a hospital, where the condition may be much more common than in a community practice, the test may perform "better" in that setting than in yours. The authors should tell you about the study setting and patient selection.

Did everyone who received the new test get the gold standard?

Some studies will only give patients the gold standard test if the new test is positive. If the outcome of the new test influences whether or not confirmation of the results with the gold standard is carried out, validation of the properties of the new test will be distorted and biased.

Can the test be replicated in my practice?

The paper should tell you exactly how to perform and interpret the test, and should address all issues related to preparation of the patient, precautions to be undertaken and possible side effects and complications.

Do the results of the test apply to my patient?

a. Will the test have the same accuracy for my patient as for the study patients?

If the practice setting in the study was similar and the patient selection criteria broad, the answer is "probably."

b. Will the results change my treatment approach?

You need to decide if the test actually supplies new diagnostic information you didn't already have, whether this information will change how you manage the particular problem and, finally, whether this change provides any benefit to the patient. If the answer to any of these fundamental questions is "no," then the accuracy of the test is irrelevant.

Assessing Articles About Causation

Understanding cause and effect relationships, particularly how they relate to harmful exposures, is important in the daily practice of dentistry. What is the risk, for example, of using local anesthetic with epinephrine in a patient with moderate, stable angina? What is the risk of *not* using it? What is the risk to the unborn fetus of a pregnant dental assistant if nitrous oxide sedation is used on a regular basis in the office? The following guides are provided to help you critically appraise an article on causation or harm.⁶

Were the comparison groups similar?

Besides exposure to the suspected causal agent, a number of other "confounding" factors can influence the outcome of a study. It is important that these other factors be similar in the comparison groups. It would be unethical to design a randomized trial to study a harmful exposure, so most often we have to rely on the next most powerful design — the cohort study — in which exposed and non-exposed patients are assembled, followed forward in time, and monitored for the outcome of interest. If the outcome is rare or takes a long time to develop, case-control studies are done in which cases and similar, but non-affected, controls are identified. Exposure to the agent is assessed in a retrospective manner and the results are compared between the 2 groups. Both of these non-experimental designs suffer from the absence of randomization, so there is no guarantee that the 2 groups are similar. Furthermore, the retrospective nature of the case-control study makes this design susceptible to significant bias. Case reports and case series, although thought-provoking and often the stimulus for further research, lack a comparison group and cannot provide evidence for cause and effect relationships.

Were the exposures and the outcomes measured in the same way in both groups?

Bias can be introduced in the measurement of either the outcome or the exposure. For instance, when clinicians are aware that patients have been exposed to a risk factor, they have a tendency to be more diligent in their assessments ("surveillance bias"). In cohort studies, this bias can be minimized by blinding the clinicians doing the assessments as to the exposure status of the patient. In case-control studies, clinicians might ask more detailed questions about the exposure if they know the patient tests positive. Similarly, patients who test positive may be more motivated to recall events leading to exposure, or may wish to downplay an exposure (e.g. smoking exposure, or drug and alcohol use), especially if they perceive they may be judged. Both patients and clinicians in case-control studies can be blinded as to the hypothesis of the study to control these kinds of biases.

Did the exposure precede the outcome?

This criteria is more readily applied in cohort studies than in case-control designs, but is not always clear-cut. For example, are depressed patients more likely to develop atypical facial neuralgia, or is depression a consequence of constant, severe pain?

Is there a dose-response relationship?

Increased quantity or duration of exposure should lead to an increased risk for or severity of outcome.

Does the association make sense?

Have other explanations been ruled out? Does the association make biological sense and is it in keeping with our current understanding of the basic sciences? Does it fit with what we already know?

Can I apply the results to my practice?

If the characteristics of your patients are similar to those in the study, if the treatments or exposures described in the paper are similar to those of your patients, and if the study design was strong, the findings described in the study may be quite relevant to your practice. Whether or not you change your current practice depends on the magnitude of the risk, the strength of the evidence and the availability of a safe, effective and realistic alternative.

Assessing Articles About Prognosis

The possible outcomes of a disease or condition and the anticipated frequency of those outcomes define “prognosis.” Patients frequently ask dentists questions related to prognosis and, more often, how a planned intervention might alter the prognosis. For example, a parent may ask if his or her child’s teeth will remain straight forever after orthodontic treatment. A patient may enquire how long an implant and crown will last; if oral leukoplakia will progress to oral cancer; or if periodontal disease will cause tooth loss.

Prognostic factors are characteristics about the patient (for example, demographics or biological makeup), the condition, and other coexisting or comorbid conditions (for instance, diabetes in a patient with periodontal disease) which help to predict — not necessarily cause — the outcome. Rather, their presence is associated with increased or decreased risk for the development of the outcome.

The best research design for studying prognosis is the cohort study. In the event of rare outcomes or a lengthy duration from the first evidence of a prognostic factor to the development of the condition, a case-control design can be used, but the inferences that can be made from its findings are much weaker. The following questions can help you to decide if the results of a study of prognosis are valid and suitable.⁷

Were the patients well described, representative and at a similar point in the development of their disease?

The condition of interest must be adequately described and the criteria for deciding whether or not a patient has the condition should be clearly stated. For example, if the investigators have assembled a group of patients with localized juvenile periodontitis, we would want to know exactly what diagnostic criteria were used to include a patient in the cohort. Since we will probably be interested in the

outcome of this condition in all patients who have it, the results from a population-based study (perhaps with patients enrolled by community periodontists) where all degrees of the condition are represented will be more informative. If the study includes only severe, unusual or refractory cases referred from practice to a university setting, the outcomes for these patients will not be as good and the prognosis of the disease will appear to be much worse than it really is. In addition, patients should be identified and entered into the cohort at a uniform, early stage of the disease (at the “inception”). If patients are entered at various stages of their clinical course, the prognosis of the disease becomes distorted. If patients are entered later, teeth may already have been lost. Since data are only being collected prospectively (to avoid all the bias associated with retrospective designs) these adverse outcomes will not be counted and the prognosis of the disease will seem better than it really is. Likewise, those patients who are entered late, but have retained their teeth throughout a prolonged course of the disease will not have this favorable period of survival counted and the overall prognosis will appear to be worse.

Was follow-up sufficiently long and complete?

Clinically important outcomes such as a carious lesion requiring restoration or tooth loss can take a long time to occur after the identification of a prognostic factor. Therefore, the follow-up period needs to be long enough to detect the endpoint of interest. In addition, if follow-up is incomplete, the validity of the study may be severely threatened. Loss to follow-up in clinical studies is often poorly reported, but is an extremely important validity issue. How do you know if the validity is threatened because of loss to follow-up? One rule of thumb is to have serious reservations about the results of the study if more than 20% of the patients did not complete the study. You may also consider the proportion of patients lost to follow-up in relation to the proportion of patients who have suffered the adverse outcome and assume a “worst-case scenario” — that is, assume that all unavailable patients have suffered the bad outcome. Remember, 2 of the major reasons patients drop out of studies is that they get better or they get worse. Assuming the worst-case scenario is more conservative than assuming that only some (how many?) of the unavailable patients did poorly. If the proportion of patients lost to follow-up is large and the proportion of remaining patients developing the adverse event is small, then the validity of the study is questionable. For example, consider a study where 8% of the patients are lost to follow-up. If the proportion of remaining patients who develop the adverse endpoint is 30% and we assume a worst-case scenario for the 8% of unavailable patients, the true rate of patients with the bad outcome may be as high as 38%. On the other

hand, if the event rate in the remaining patients is only 2%, the impact of the 8% of lost patients, all of whom could possibly have suffered the adverse event, is much greater. The "worst-case scenario" in this instance suggests that the true rate of bad outcomes could be as high as 10%, rather than 2%. You would be justified in questioning the validity of this study. The investigators should compare the clinical and demographic characteristics of all patients lost from the study with patients who completed the study, to see if there are major differences. In addition, it is important to know the reasons for loss to follow-up. For instance, patients who simply don't keep appointments may be generally "non-compliant" and this, in itself, may be an important prognostic factor for many disorders.⁸

Were the outcome criteria explicit and applied objectively?

The outcome must be clearly defined. For instance, "implant failure" can have many meanings unless specific criteria are defined. If any clinical judgment is involved in assessing the outcome, the clinician should be blinded to any other features of the patient which might influence interpretation of the outcome.

Were extraneous prognostic factors adjusted for?

Factors such as age and socioeconomic status can interfere with the assessment of prognosis. Although these factors are not the cause of the outcome, they may be associated with or be markers for the true prognostic factors. For instance, age does not cause rampant caries, but it may be associated with caries because of other age-related factors such as dietary alterations, medication-induced xerostomia and increased functional dependency. The authors should state that these other variables have been adjusted for in the analysis.

Were the study patients similar to my own? Will the results help to select or avoid therapy or provide advice for patients?

Patient characteristics should be described in detail to permit you to judge how similar they are to your own. Knowing the expected clinical course of a condition can help you to decide if and when to intervene and what to tell your patient.

Conclusion

In this series, we have highlighted the principles and discussed the tools needed to practise evidence-based dentistry. By formulating a focused clinical question, executing an efficient literature search, evaluating the evidence and, if relevant, applying it to the patients in your practice, you can meet the challenge of continuing to

provide quality oral health care in a rapidly changing environment head on. ♦

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

1. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston: Little, Brown and Company; 1991.
2. Sackett D, Richardson W, Rosenberg W, Haynes R. Evidence-based medicine: how to practice and teach EBM. London: Churchill Livingstone; 1997.
3. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994; 271(5):389-91.
4. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; 271(9):703-7.
5. Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. *BMJ* 1997; 315(7107):540-3.
6. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. *JAMA* 1994; 271(20):1615-9.
7. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. *JAMA* 1994; 272(3):234-7.
8. Haynes RB, Dantes R. Patient compliance and the conduct and interpretation of therapeutic trials. *Control Clin Trials* 1987; 8(1):12-9.

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

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

Evidence-based practice: a primer for health care professionals, by Martin Dawes and others; *Evidence-based medicine: how to practice and teach EBM*, by David L. Sackett and others; *The evidence-based medicine workbook: critical appraisal for clinical problem solving*, by Robert A. Dixon and others. (Shipping charges and taxed apply on all loans.) The Resource Centre has prepared an information package on *Evidence-based dentistry*. The package is available to CDA members for \$10.00 (plus tax). 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.