

EVIDENCE-BASED DENTISTRY SERIES

How to evaluate an article about therapy

Gary R. Goldstein, DDS,^a and Jack D. Preston, DDS^b

CLINICAL SCENARIO

You have been treating a 54-year-old woman for many years and despite the excellence of your fixed partial denture restorations, the intense routine maintenance by her periodontist, and good homecare, she has been experiencing a continued deterioration of her periodontal tissues. Her attempts to quit smoking have been unsuccessful; otherwise she is in good health and taking no medications. Because you are her primary care dentist, she has questioned you about her current dilemma. The periodontist has suggested a 3-week course of doxycycline therapy to control her latest exacerbation of periodontal disease, but she is concerned about Food and Drug Administration (FDA) reports asking for prudent use of antibiotics. How do you advise this patient?

THE SEARCH

Using your acquired EBD skills, you do a literature search on your computer for appropriate evidence to support your advice. A search of doxycycline results in 2759 "hits." The addition of the key word "periodontitis" limits the citations to 32. Your desire to pursue clinical evidence leads you to the article "Randomized controlled trial of doxycycline in prevention of recurrent periodontitis in high-risk patients: antimicrobial activity and collagenase inhibition" by McCulloch et al, 1990.¹ After reading the abstract, you order the full-text article. The study evaluated 82 subjects who received surgical treatment within the last 3 years and, despite regular subgingival scaling and prophylaxis, still experienced either frequent periodontal abscesses, continued loss of periodontal attachment, or tooth loss resulting from their disease. The study design was randomized, placebo-controlled, and double-blinded. Three subjects who reported ingestion of tetracycline or doxycycline within the last year were excluded at the screening stage. Subjects were monitored for up to 12 months from baseline measurements and 17 of those, who had no recurrence of disease, were dismissed from the study. Doxycycline (200 mg initial dose followed by 100 mg/day for 3 weeks) was administered to 30 subjects and a lactose-containing

placebo to 25 subjects. Subjects were monitored for gingival attachment level, pocket probing depth, bleeding or suppuration on probing, gingival crevicular flow, plaque index, gingival index, antimicrobial activity, and collagenase inhibition. Seven subjects dropped out, 6 subjects from the placebo group and 1 from the test group. After 7 months, 15 of the remaining 19 subjects (79%) in the placebo group and 13 of the remaining 29 subjects (45%) in the test group exhibited recurrence of disease.

Today's practitioner is deluged by an insurmountable volume of literature and information, some of it "cutting edge," some contradictory, and some useless. Clinical decision making mandates that evidence rather than empiricism dictate treatment. Evidence-based dentistry (EBD), adopted from our medical colleagues,²⁻⁴ presents guidelines to determine the validity of the results and whether they can be applied to clinical practice.

This article provides an overview of EBD rules to assist the clinician in determining the appropriate dental therapy.

HOW TO EVALUATE AN ARTICLE ABOUT THERAPY

Most practitioners fall into a patient treatment routine, and base their therapy on procedures they know best and with which they are comfortable. When dramatically new procedures become established, such as has happened with dental implant therapy, patients have a right—moral, ethical, and legal—to know the risks and benefits of any therapy that is recommended. When considering the merits of an information source, the reader or listener must clearly understand what the purpose of the study was and how the investigators sought to establish their premise, and the results of the study must directly relate to this purpose statement. Even though the result of a study may seem to offer evidence that indicates that the conclusions of the study were justified, the reader must ascertain whether the investigators used credible methods to arrive at their conclusions. Guyatt et al^{3,4} divided the assessment of validity into primary and secondary considerations.

VALIDITY: WERE THE RESULTS VALID?

Primary: Was the assignment of patients to treatment randomized?

There are many methods by which subjects are selected and assigned to a study. The strongest evidence is gained by assigning participants by randomization—assuring that all who enter into the investigation have an

Presented at the annual meeting of The American Academy of Fixed Prosthodontics, Chicago, Ill., February 1999.

^aProfessor and Director, Advanced Education Program in Prosthodontics, New York University College of Dentistry, New York, N.Y.

^bProfessor Emeritus, The University of Southern California School of Dentistry; and Executive Vice President, Dental/Medical Diagnostic Systems, Los Angeles, Calif.

equal opportunity to be assigned to one of the study groups. If subjects are not randomized to treatment, there is serious potential for bias, which would invalidate the results. For example, it would be easy for the principal investigator to select those with a poorer prognosis to receive the secondary therapy or placebo, while those with an apparently better prognosis would receive the primary therapy. Proper assignment means that regardless of the time at which they come into the study, at what site they enter (when multiple sites are involved), regardless of their personal status, every subject has the potential of being placed in a study group, according to random assignment methods. Randomization ensures that the variables over which the study has control, and any operative unknown variables that could potentially influence the study outcome are evenly spread throughout all the groups. Of course, randomization necessitates a prospective study. The validity of the results also demands that the investigators assemble a group large enough to follow during the course of the investigation and that a significant amount of those subjects will be able to complete the course of therapy.

It is imperative to realize that randomization cannot make up for a poorly planned and implemented study. Just because the design was a randomized controlled trial (RCT) does not mean that the reader is released from the necessity of examining the methodology. "Evidence-based medicine is not restricted to randomized trials and meta-analyses. It involves tracking down the best external evidence (from systematic reviews when they exist; otherwise from primary studies) with which to answer our clinical questions."⁵

Although randomization eliminates allocation bias, it necessitates a compelling study population ("N"). Brunette⁶ stated that "one problem with the random-groups design is that it is possible, because of random sampling fluctuations, that the 2 groups are not identical or even similar." Feinstein⁷ questioned the blind faith often put in randomized trials, and suggested that prognostic stratification is critical to the use of the data. He maintains that if data are to be evaluated in prognostic subgroups, those subgroups should be identified, where possible, before the study starts and that subjects be allocated to those subgroups before they are randomized.⁸ For example, if a study were being performed in which smoking was a major confounding variable, it is prudent to identify the smokers and non-smokers ahead of time to insure that "chance" does not put a significantly large number of smokers in one group. Although it would be excessively cumbersome and nearly impossible to stratify for every potential variable, key factors that have the potential to alter the study should be identified at the start of the trial. In the doxycycline study, subjects were randomized by a computer-generated random allocation method. No stratification for confounding variables was performed.

Were all patients who entered the trial properly accounted for and attributed at its conclusion?

All subjects who enter the study must be accounted for, regardless of the reason for which they did not complete the study. One sometimes reads studies in which a substantial number of participants are lost, and no effort is made to explain their loss. It is sometimes implied that such losses are inconsequential and that the only matter of importance is the assessment of those who completed the study. If a participant dies, an assessment of the cause of death must be made to ensure the cause of death was not related to the test treatment. If the number of subjects in the study is large enough, random loss to death should occur equally for all groups of the study. However, subjects who choose to drop out of a study may do so because of germane reasons and, in therapy trials, unhappiness with the treatment is a significant factor. The number of those beginning and completing the study must be clearly stated and any differences recorded.

A rule of thumb is a dropout rate of greater than 20% compromises the validity of the study. There are many methods of dealing with dropouts. One method is to use the mean of the control group for the dropouts in the test group and the mean of the test group for the dropouts in the control group. Another is to assume that all dropouts in the test group did poorly and all dropouts in the control group did well. Although there is controversy among epidemiologists over how to deal with dropouts, what is agreed on is that the method should be decided at the inception of the study, not after the data are collected. It is also the researchers' obligation to describe the method of dropout assessment for the reader.

In our clinical scenario, all patients who entered the trial were accounted for at the end of the trial. Seven patients dropped out and statistical analyses were performed and reported with and without the dropouts.

Secondary: Were patients, their clinicians, and study personnel "blinded" to treatment?

Bias is difficult to overcome, and even well-meaning investigators might skew the results of a trial toward the outcome they inherently favor. Patients might be given clues that influence their response or outcome assessment may be skewed toward the desired goal. Bias must be recognized and methods of intervention initiated in the design, measurement, and analysis of any research. Once the assignment is made, all evaluations of the groups' progress should be made without any observer knowledge as to which group is receiving which treatment. Only then can an unbiased assessment be made.

Similarly, patients should not know to which group they have been assigned. This might be difficult if there are ascertainable differences in the physical

appearance, taste, odor, and so forth, of the test treatment or when comparing different prosthesis designs. Not all investigations can be blinded. If one group receives a given physical therapy routine, it will obviously be different from that undertaken by a second group. Every effort should be made to ensure that patients are blinded to their group assignment to preclude *participant bias* from occurring. Patients may have a preestablished preference or dislike for a given therapy, regimen, or material and may, purposely or inadvertently influence the results.

While blinding is an ideal, there are some studies in which it would be difficult to ensure. For example, in a study comparing implant-retained overdentures with conventional complete dentures, it would be impossible to blind either the patient or the observer. In the doxycycline study, both the patients and investigators were blinded to treatment.

Were the groups similar at the start of the trial?

Cohort uniformity is essential in clinical studies. If there are age or gender variations, ethnic differences, or substantial variations in general health, these factors may influence the results. The larger the number of subjects, the greater is the chance that randomization will account for these uncontrolled variables. However, a cautious researcher will assess the groups to which subjects were assigned to ascertain whether there might be factors that might have a bearing on the outcome. Sample size then becomes important and, together with the factors that might cause the loss of subjects from any group, must be considered when assessing the merit of a study result.

When evaluating the doxycycline study, we cannot be sure if the groups were similar. Although the authors reported that the distribution of the tooth types was similar (a critical fact because molars exhibited active disease more than anterior teeth) we know nothing about, for example, smoking and parafunctional habits, number of remaining teeth, style of occlusion, and type of restorations. Theoretically, randomization eliminates allocation bias and should distribute these factors evenly among both groups given that the "N" was sufficiently large. The authors of the study should show in the methods section how the "N" was calculated and the power of the study. The reader can then determine whether they feel the "N" of this study assuages this concern.

Aside from the experimental intervention, were the groups treated equally?

Even though a study may be well designed, the subjects randomly assigned to the treatment groups, and the investigators blinded to the evaluation, it is possible to inadvertently invoke bias during the study. When a

therapy is thought to have some side effect, investigators may follow that treatment group more closely, or ask questions not given to other groups. Simple questions regarding taste, feel, or appearance may be asked—giving subtle clues to the subject that they are/are not in a specific group. This is more probable when the groups are not blinded, and the manner in which the groups are treated may alter the outcome.

When something is studied, the act of studying may alter that which is being studied. If an investigator recalls a "test group" more often than a control group, or if the "test group" receives additional adjunctive therapy (such as prophylaxis, antimicrobial therapy or a second drug given to counteract a side effect), the impact of the cointervention on the primary observation is unknown and affects the validity of the study. Every effort should be made to ensure that all groups receive equal treatment whenever possible. In the article we are evaluating, it appears that all groups were treated equally.

VALIDITY: WHAT WERE THE RESULTS?

This is the essential fact that must be established. Sometimes a report may be written to cause the reader to focus on what the investigator hopes will be seen and ignore other factors that may have equal primacy. The actual results can be well established if some simple but direct questions are asked and valid answers sought.

How large was the treatment effect?

Not all studies have results that are as easy to determine as life or death, cure or continued disease. Most investigations arrive at conclusions that require some cautious thought to determine the actual magnitude of the effect. If the groups within the study were equal (and that must be first established) and the outcome measure validated, then the results may be statistically analyzed, graphed, or otherwise assessed. However, statisticians may not agree on how results are best analyzed, and some investigators will use a method that best presents the result they would prefer.

Readers must ascertain the degree of negative outcomes that invariably are the consequence of any therapy or treatment regime by establishing the risk involved (percentage experiencing negative effects in the study) and from that the *absolute risk reduction* (ARR) between the treatment group and the others in the study. Using X = negative outcomes in the control group, and Y = negative outcomes in the treatment group, the $ARR = X - Y$. Associated with absolute risk reduction is *relative risk*, which is calculated by dividing the risk with therapy by the risk without therapy ($RR = Y/X$). A common expression of efficacy is the *relative risk reduction* (RRR). This evaluation is

expressed as a percentage, and the larger the RRR, the more effective is the therapy ($RRR = 1 - (Y/X) \times 100$).

These calculations are important but not easily arrived at in prosthodontic therapy. Where death is concerned, the negative, or adverse, outcome is readily apparent. In dentistry, positive and negative outcome assessments are not as obvious and often controversial.

McCulloch et al¹ used acceptable negative outcomes (abscess and/or >2 mm gingival attachment loss) and reported a relative risk reduction of 43% excluding the dropouts and 28% with the dropouts included. Confidence Intervals were not reported and would have assisted the reader in applying the results to the general population.

WILL THE RESULTS HELP ME IN TREATING MY PATIENTS?

Once the result is known and the treatment effect quantitatively evaluated, readers must determine whether the suggested therapy really applies to the patients in their practices. Was the population studied representative of the patient(s) under consideration to receive the suggested therapy or procedure? For example, was the study performed using a military (male) population having a mean age of 22 years, or in a tertiary care facility using cancer patients with a questionable prognosis and mean age of 62 years?

It is important to read the results section, not just the conclusions. The reader may find that he/she interprets the data differently from the investigators. Most dental therapy articles are descriptive and do not allow for comparative statistics. Results are usually expressed in positive numbers that can be misleading to the clinician. For example, a report of an 85% success rate might sound impressive, but when viewed as a negative (15% failure rate) may have more impact on the decision-making process.

Do the inclusion/exclusion criteria for the study allow extrapolation to your own patient population? If it appears that these parameters for extrapolation are met, then the primary consideration is simply will your patients benefit from this therapy? The answer to this question gives meaning to the entire process. If the answer to all the other questions is positive, and yet the study is of no benefit to the patients for whom you are considering the procedure, then the research may merely be "interesting" or "promising" but there is nothing to move the reader to immediately apply the results in the expectation of patient benefit.

It may well be that the benefit is too minimal to undertake the risk, or the statistically significant change was of too little clinical value to initiate the procedure. There is a substantial difference between statistical significance and clinical significance. For example, an investigator may be using an extremely accurate measurement device that can report bone loss

around implants in tenths of a millimeter. After 5 years, their study may show a statistically significant bone loss of 0.01 mm associated with Brand X implants. But are the results clinically significant, especially if there is a cost, time, or procedural difference between Brand X and Y? Such reports may point the way to future studies and portend enhanced success with improved methods, but are set aside without their recommendations being implemented.

Another clinically relevant concern is the difference between a biologic response and a clinical response. Two examples are tetracycline cords and glass ionomer cements. Tetracycline impregnated materials may demonstrate the ability to kill microbes (a biologic response), which may have no clinical relevance, but if they demonstrate a decrease in inflammation or probing depth (a clinical response), then the data may be meaningful to the reader. The same thought process holds true for glass ionomer cements. To say that they release fluoride (a biologic response) is meaningless to the clinician, but if they cause a decrease in caries (a clinical response), then the results have implications for the restorative dentist.

Relevance is the most difficult question for the clinician to establish. In our test scenario, the article in question is very well designed. The double-blind randomized control study is at the highest level on the hierarchy of evidence. Limiting the study to high-risk patients allows the reader to make direct correlation to patients who might need this care the most. The authors' attention to detail and the meticulous manner in which they procured their outcome assessments gives validity to their research and findings. When it appears that the recommended therapy (regimen, material, procedure) is of great enough value to merit trial, then further consideration should be given 2 additional factors.

Were all clinically important outcomes considered?

Whereas a drug (therapy, material, regimen) may be used for a benefit that is anticipated, other ramifications may be encountered. The most dramatic unexpected outcome would be death. Most correlated events, especially in restorative dentistry, are more subtle. The use of base metal alloys as gold substitutes offered improved physical properties, but the question arose whether allergies and tissue toxicity were correlated with their use. A medication may relieve the patient's pain at the risk of gastric distress. A study on a new type of all-ceramic restorations may concentrate on color and longevity, but may not evaluate wear of the opposing dentition. Secondary outcomes may only be observed in larger population samples and may not be immediately apparent in smaller trials.

Chronology bias deals with the length of time in which the maneuver is evaluated. If one was conduct-

ing a study of maxillary osseointegrated implants and decided, based on mandibular studies, to initiate stage II surgery after 3 months, one might conclude that the success rate in the studied cohort does not justify its use in the general population. In reality, extending the time between stage I and II to 6 months produces more favorable results. Clinical reports on the 1- or 2-year survival rates of new restorative materials may give us useful information, but the trials have not been conducted long enough to ensure that other clinically important outcomes have surfaced.

If a study fails to mention secondary outcomes, it may only mean that the investigators did not have a large enough population to demonstrate such events or that the study ended too soon, not that such events do not arise. When properly reported, authors will call these omissions to the attention of the reader rather than blithely reporting apparent positive results. On reviewing the doxycycline study, it becomes apparent that all clinically important outcomes were evaluated.

Are the likely treatment benefits worth the potential harm and costs?

Although the primary effect of a drug or procedure may offer measurable benefit, it might also have side effects—obvious or covert—that negate or temper the advantages. These side effects may not be experienced by all the treatment population and they may be specific to 1 group or subset. Three patients in the doxycycline group reported gastrointestinal upset that was not severe enough to prevent them from taking the medication.

This is another instance in which the initial structure of the investigation is important, because the side effect may not be definitely attributable to the treatment unless there is a control group against which the trial can be compared. The magnitude of the benefit and the frequency with which that benefit might be expected must be balanced against the risk of harm, and the frequency with which such negative effects might occur. Harm can come in many forms. Death and adverse medical side effects are obvious, but monetary implications, additional surgical procedures, and increased time of therapy are of major importance to our patients. Risk-benefit analysis is essential to ascertain whether a therapy should be initiated, and the patient must be made aware of the negative aspects of a course of therapy and its potential benefits. A circumspect and intelligent analysis of all the aspects of any new therapy is an essential precursor to recommending or undertaking any course of therapy.

RETURNING TO OUR SCENARIO

From reading the article in the search, it appears that in a high-risk patient having an exacerbation of recurrent periodontitis, the study validates a single regimen of doxycycline therapy. The unanswered

question, which is pertinent to the clinician, is “what happens after repeated use of the low-dose antibiotic?” Will there be an alteration in the normal oral flora? Will drug resistant bacteria emerge?

SUMMARY

In restorative dentistry, we change therapy because of improvements in products and/or techniques or because the “product” we use is no longer on the market. As dentists, we are usually doing the “field testing” for commercial companies and the “anecdotal” data supplied by newsletters, nonpeer-reviewed publications, and Internet chat groups are often, and unfortunately, the first documentation of poor clinical performance and/or adverse secondary effects. The problems with root fracture when cementing dowels and the fracture of complete coverage porcelain restorations with resin/ionomer luting agents are examples.

Randomized controlled studies are the “gold standard” of research design. Unfortunately, many can be of long duration and great expense. However, the therapy that is not efficacious is also expensive. Because most advances are made with small case studies, which are already at a lower level of evidence, it is imperative that what we read has the greatest possible validity. Sackett⁵ has concluded “...some questions about therapy do not require randomized trials (successful interventions for otherwise fatal interventions) or cannot wait for the trials to be conducted. And if no randomized trial has been carried out for our patient’s predicament, we follow the trail to the next best external evidence and work from there.” EBD gives the reader the tools to evaluate that evidence.

The goal of Evidence Based Dentistry is better quality, clinically oriented and relevant research that provides better information for the clinician and better treatment for the patient.

REFERENCES

- McCulloch CA, Birek P, Overall C, Aitken S, Lee W, Kulkarni G. Randomized controlled trial of doxycycline in prevention of recurrent periodontitis in high-risk patients: Antimicrobial activity and collagenase inhibition. *J Clin Periodontol* 1990;17:616-22.
- Oxman AD, Sackett DL, Guyatt GH. Users’ guide to the medical literature I. How to get started. *JAMA* 1993;270:2092-5.
- Guyatt GH, Sackett DL, Cook DJ. Users’ guide to the medical literature II. How to use an article about therapy or prevention A. Are the results of the study valid? *JAMA* 1993;270:2598-601.
- Guyatt GH, Sackett DL, Cook DJ. Users’ guide 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? *JAMA* 1994; 271:59-63.
- Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: How to practice and teach EBM. New York: Churchill Livingstone; 1997. p. 4,5.
- Brunette DM. Critical thinking understanding and evaluating dental research. Chicago: Quintessence Publishing Co; 1996. p. 152.
- Feinstein AR. An additional basic science for clinical medicine: II. The limitations of randomized trials. *Ann Intern Med* 1993;99:544-50.
- Feinstein AR. An additional basic science for clinical medicine: III. The