EVIDENCE-BASED DENTISTRY SERIES

How to evaluate a dental article about harm

Rhonda F. Jacob, DDS, MS,[a] and Patrick M. Lloyd, DDS, MS[b]

CLINICAL SCENARIO

An edentulous 58-year-old woman presents to your office stating that she has been contemplating dental implants. She is considering implants in the maxillary and mandibular arches. You explain that you will perform a diagnostic workup that will include a consultation with a surgeon. Afterward you will discuss the various implant alternatives. On the health questionnaire, the patient reports that she smokes 2 packs of cigarettes daily, and a recent annual physical examination revealed no health problems. She is taking no medications. You consider that her smoking habit may affect the successful outcome of her implants and decide to examine the literature on harmful effects of smoking and dental implants.

SEARCH OF THE LITERATURE

A literature search required examination of specific headings with one or more cross-references. A MEDLINE search using the Ovid search engine (Ovid Technologies, Inc, New York, N.Y.) is initiated with "dental implants" as the keyword and limited to "human" topics. MEDLINE offers an opportunity to "explode" the search and then select various subheadings. The selected subheadings were "adverse events," "contraindications," "standards," "utilization," and "statistics and numerical data," which yielded 502 articles. A combination of these articles with "smoking" as a keyword revealed 10 possible articles. When clinical trials were used to further limit the 10 articles, no articles were selected. It was then necessary to display the 10 abstracts. One article by Bain and Moy entitled "The association between the failure of dental implants and cigarette smoking"[1] appeared that it might answer the question for this patient. It was a cohort study of a single private practitioner who treated patients with dental implants over a 6-year period: totaling 2194 titanium-screw implants placed in 540 subjects. The overall failure rate in the smokers was 11.3%, whereas the failure rate in the nonsmokers was 4.8%. Failure rates for smokers versus nonsmokers were also derived for the anterior and posterior positions of both arches.

INTRODUCTION

It is important to understand that "harm" from the patient's perspective can be defined as any encounter that results in a harmful or unwanted outcome. The encounter must always proceed the outcome. Dentists also interview patients who may have experienced harmful exposures, leading to harmful events. Two classic clinical findings, "nursing-bottle caries" and "tetracycline staining" define the exposure and the harmful event in its clinical name. Life-style exposures or a patient characteristic / risk factor can result in a harmful outcome that must be countered by dental therapy.

It is human nature that clinicians are less likely to define iatrogenic events or "unwanted side effects" that occur as a result of dental treatment as "harming" the patient.

Are subgingival margins associated with increased risk for gingivitis? Are patients who use removable partial dentures more prone to dental caries? Are some patients at risk for jaw fracture after placement of dental implants. When pondering questions about harm, prosthodontists must consider the study design used to evaluate the association of the putative cause and the harmful outcome, the validity of the data reported in the studies, the strength of the causal association, and the relevance to patients in their practice. A search of the literature and application of study guidelines specific to evaluating an article on harm will allow the clinician to select the best available evidence to make clinical decisions.

EVALUATING AN ARTICLE ON HARM

Selection of the appropriate research design for studying a harmful outcome or adverse event is dependent on the research question and the feasibility of conducting the trial. Even though the appropriate design may have been used in the trial, it is important to determine whether the study was of a high quality related to gathering and assessing the comparison groups. Levine et al[2] and Sackett et al[4] have described users' guides for the medical literature that assist in evaluating the quality of a health care study concerning harmful outcomes or adverse events. These guides are enumerated in Table I in the form of 3 main questions that consider the validity of the study, the strength and precision of the results, and whether the results will help the practitioner in caring for his or her patients.

This article is adapted from Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V, for the Evidence-Based Medicine Working Group. Users' guide to the medical literature IV. How to use an article about harm. JAMA 1994;271:1615-9.

[a]Professor, Dental Oncology, University of Texas, M.D. Anderson Cancer Center.

[b]Associate Professor, Chair of Family Dentistry, University of Iowa College of Dentistry, Iowa City, Iowa.

Table I. User's guides[2]

I. Are the results of the study valid?
Primary guides:
Are there clearly identified comparison groups similar with respect to important determinants of outcome other than the one of interest?
Are the outcomes and exposures for both groups measured in the same way?
Is follow-up sufficiently long and complete?
Secondary guides:
Is the temporal relationship of cause and effect correct, consistent, and reasonable?
Is there a dose-response gradient?
II. What are the results?
How strong is the association between exposure and outcome?
How precise is the estimate of the risk?
III.Will the results help me in caring for my patients?
Are the results applicable to my practice?
Is the magnitude of the risk clinically relevant?
Should I attempt to stop the exposure or discontinue the therapy?

Table II. Directions of inquiry and key methodologic strengths and weaknesses for different study designs[2,3]

Design	Starting point	Assessment	Strengths	Weaknesses
Randomized controlled trial	Exposure (usually a treatment)	Follow forward in time for adverse event	Internal validity: easier to control bias, equal study groups, blind assessments	Feasibility, generalizability, may be small number of adverse events and therefore not statistically significant
Cohort	Exposure	Follow forward in time for adverse event	Feasible when randomization of exposure not possible	Possible threats to internal validity: ? equal study groups, difficult to blind
Case control	Adverse event	Follow backward in time to obtain exposure history	Overcomes temporal delays, may only require small sample size	Possible threats to internal validity: ? equal study groups, difficult to blind

ARE THE RESULTS OF THE STUDY VALID?

Primary guides

Are there clearly identified comparison groups similar with respect to important determinants of outcome other than the one of interest? For a study to identify harmful exposures or treatments that may cause harm, comparison groups are necessary. The usual human comparison groups are (1) patients who have experienced the exposure but may or may not have the harmful outcome, compared with (2) patients who have not experienced the harmful exposure but may or may not have the harmful outcome. The ratios of these comparison groups yield the "relative risk" that the exposure may have caused or is associated with the harmful outcome. The credibility of the study results is influenced by the choice of comparison groups, how those comparison-group subjects are gathered for the study, and how the subjects are observed during the study.

The clinical question that is being asked determines the study design that would give the strongest evidence. For instance, does the question involve a harmful side effect of a therapy, or does it involve the harmful outcome of a patient's lifestyle or risk factor that is not under the control of the clinician? In turn, the design of the study determines how the comparison groups are gathered. The basic study designs that clinicians encounter when assessing whether their patients have been or might be exposed to a potentially harmful factor are reviewed in this article (Table II).[2,3] The confidence that a practitioner can place in the conclusions of a study potentially depends on the classification of the study design. The reader is urged to expand on the hierarchy of research study designs, in the third article in this series entitled "Hierarchy of research design used to categorize the 'strength of evidence' in answering clinical dental questions."[3]

Randomized controlled trials (RCTs). Important clinical information related to patient harm may be

gleaned from data collected in a study that assesses therapeutic intervention and reports the side effects or harmful effects of the therapy. The advantages of randomization are 2-fold: (1) subjects have an equal opportunity to receive either the causal agent or some alternative experience, and (2) study groups are similar with respect to the known and unknown determinants of outcome. The results section should describe how successful the randomization process was in "equally distributing" the population variables. The treatment and assessments rendered can be designed and controlled to be as similar as possible in both groups as the intervention is under the control of the investigator. The subjects can be followed forward in time for therapy outcomes and adverse event outcomes. A prospective study protocol allows monitoring and some control to be exercised concerning competing interventions that patients may undertake. Whenever possible, data collectors and subjects can be blind to the interventions and measurements.

Unfortunately, it is sometimes necessary to render treatment with possible harmful side effects (ie, lip paresthesia from various mandibular surgeries, bone marrow suppression from cancer therapies). Because RCTs are at the pinnacle of research design hierarchy, a clinician can be confident in the truth of the relationship between a treatment and an adverse outcome if it is demonstrated in an RCT. However, the RCT is primarily designed to study a treatment effect. Because the treatment effect is more likely to occur than the adverse event, the number of patients needed to demonstrate the adverse event may be much larger than the number of subjects enrolled to demonstrate a treatment effect. Therefore, the causal relationship between the therapy and the adverse event may not be a conclusive finding, due to the small number of patients and events.

Cohort studies. This study design is useful when it is not ethical to randomly assign subjects to a harmful exposure. In cohort studies, the investigator identifies exposed and unexposed groups and follows them forward in time, observing the subjects for the development of the outcome. It is important to note that subjects in cohort trials are usually self-selected to the exposure of interest or are selected by circumstance (ie, patients who are currently using over-the-counter tooth-whitening solutions who might be evaluated for tooth sensitivity, patients on antiseizure medication who might be evaluated for gingival changes). It is imperative that the investigators thoroughly document the characteristics of the exposed and unexposed groups. This information is used to judge the comparability of the groups with particular interest to those subject characteristics that may affect the occurrence of the outcome (eg, smoking status, age, gender, exposure time). If any of these other characteristics also affect the outcome, the true relationship between the exposure in question and the outcome may be masked.

Statistical techniques can be used to adjust imbalances in the distribution of characteristics so that the exposed and unexposed groups are "comparable" in these other factors. However, there may be factors that are unknown or have not been documented that could potentially influence the outcome. In cohort studies, there is always the possibility that the comparison groups differ in some characteristic or additional exposure that could influence the outcome. Because these subjects are followed forward in time, depending on the exposure and outcome, it might be possible to blind those making outcome assessments to the intent of the study. It is difficult to blind subjects to the exposure, however, as they usually must give a history of the exposure. One must be ingenious in the methods of assessment to minimize bias in these trials.

Case-control studies. This type of design is useful for assessing causation of an outcome event that occurs very rarely, when the outcome of interest does not occur until many years after the time of exposure, or when the outcome is catastrophic. In this design, the subjects already have the adverse outcome of interest (eg, fractured root in an endodontically treated tooth, periodontitis, pulpal necrosis). The investigators are working backward in time, often using existing medical and dental records and patient recall to determine the causal agent or exposure. The investigators select a comparison group, whose subjects do not have the outcome of interest, but are similar in the known variables that may influence the outcome (ie, age, gender, concurrent medical conditions, oral hygiene habits, opposing occlusion). For instance a clinician might wonder what is the cause for an apical fistula occurring years after endodontic therapy. She searches the records of all the patients in her practice who had endodontically treated teeth in the past 10 years and returned for retreatment because of apical fistulas. She could then search endodontic records of patients treated in the past 10 years who did not require retreatment. This selected control group should be comparable in as many variables as possible, except for the outcome of interest (apical fistula). The 2 groups are then evaluated retrospectively for exposure to a causal agent or a specific patient/treatment factor that may have necessitated the endodontic retreatment.

This type of study may be based on patient recall of an exposure or examination of past medical or dental records. As in the cohort studies, the investigator cannot make the 2 comparison groups comparable with respect to unknown variables that affect the outcome of interest. Case-control studies may suffer from additional biases, including the chance that a subject with a particular problem may be more likely to recall an

exposure. Subjects are often not blind to the outcome, and therefore often have their own explanation for causation "worked-out" and this biases the information an interviewer can obtain from the subjects. Often the medical or dental records are incomplete, as the data in the records were recorded for routine treatment purposes, not for research purposes. Patient-treatment data may not have been recorded because of time constraints or patient unavailability. This lack of data may not have hindered routine treatment, but does hinder the validity of the subsequent retrospective research. Bias can also occur when an interviewer is aware of the study hypothesis. It is not uncommon to find that the interviewer probes more vigorously for an exposure history in subjects who have the outcome compared with those not having the outcome, thus creating a systematic bias. For these reasons, inferences drawn from case-control studies are less strong and may have limited value for making treatment decisions.

Case series and case reports. These studies do not have comparison groups and therefore cannot be considered to provide reliable data about causal relationships. Often the data are presented in percentages for a particular group exhibiting a certain finding. The causal relationship of any particular variable with an outcome of interest is weak. An example of such a weak relationship is the reduction in multiple sclerosis symptoms after the removal of dental amalgams. Changing one's clinical practice based on case reports that support such weak relationships are not prudent decisions and may deny patients the benefits of proven efficacious treatments (ie, dental amalgam has a long history as a reliable, cost-efficient restorative material).

Case series and case reports do play a vital role in the process of scientific thinking and speculation. Within the scientific community, they generate questions related to causal relationships and therapeutic outcomes. They expose possible interactions, thereby encouraging investigators to design better studies (eg, RCTs) to answer the clinical questions that have been raised.

Are the outcomes and exposures for both groups measured in the same way?

It is important that both groups are treated similarly in terms of method of discerning exposure history and method of measuring or grading the outcome. As mentioned previously, subjects or interviewers with knowledge of the study hypothesis may bias study results and therefore should be blind to the study hypothesis or group assignments whenever possible. The interviewers may prompt subjects in the exposure group to recall exposure history producing an investigator bias in the data collection. Likewise, subjects may recall exposure to a number of agents they believe may be causative of their disease process, compared with a control group who is not concerned about a health risk. For instance, this has lead to patients commonly describing a trauma to the jaw, when searching for a "reason" for the diagnosis of osteogenic sarcoma; however, clinicians describe this as a recall bias with potentially little relevance to causation. Tests and examinations to determine the outcome should be similar in both groups. Clinicians may perform more frequent or complex tests in the group they believe is at an increased risk for a particular outcome and therefore may detect disease that would go unnoticed, or not be detected until a later time under routine recall conditions. This is described as "surveillance bias."

Is follow-up sufficiently long and complete?

The investigators must have a sound hypothesis for the causal relationship of an intervention / exposure and an adverse event, to be assured that the patients have been followed for a long enough time. Studies that end before a sufficient number of subjects can achieve the outcome event will likely fail to demonstrate the appropriate causal relationship. The longer the follow-up period, however, the more difficult it is to control other interventions, and calibrate and train interviewers and individuals making assessments. The longer the follow-up required for the outcome event to occur, the greater the chance for subjects to be lost to follow-up. When subjects are lost to follow-up before the outcome occurrence can be assessed, one must make a determination as to whether the subjects lost have influenced the final conclusions. One method of final analysis is to consider the lost subjects in the exposed group as though none of them had the harmful outcome. The lost subjects in the unexposed group are tallied as though all of them did have the harmful outcome. These subjects are then analyzed along with the subjects who completed the trial. If this consideration does not change the final conclusions, then the number of subjects lost to follow-up was probably not excessive. If this approach does alter the study conclusions, then a careful analysis of the subjects lost to follow-up must be accomplished: First, to determine that those subjects who were lost to follow-up were similar in baseline characteristics to the remainder of study subjects; and second, to determine that the lost subjects in both groups had comparable prognostic factors in terms of important determinants of outcome. Statistical techniques may be used to adjust for the loss of subjects, as long as the number of the lost subjects is few.

Secondary guides

Is the temporal relationship of cause and effect correct, consistent, and reasonable? The casual agent or exposure should precede the outcome. When determining cause and effect, it is imperative to understand not only the temporal relationship, but also understand

Table III. Determining odds ratio

Exposure	Adverse event (case)†	No adverse event (control)†
Exposed	a	b
Unexposed	c	d

RR = (a/[a+b])/(c/[c+d]).
†Odds ratio = (a/c)/(b/d).

Table IV. Adverse events

Exposure	Left study	Remained in study
Pilocarpine	11	64
Placebo	9	78

that there may be exposures that frequently occur simultaneously or consecutively, followed by the outcome. The investigator may be making observations on one exposure, when it is truly the other exposure that is causing the outcome. Phenytoin therapy has been evaluated as a cause of gingival hyperplasia. Poor oral hygiene has also been implicated along with phenytoin as the cause of hyperplasia. One must determine that the hyperplasia is not a condition of subjects with seizure disorders that occurs despite the type of antiseizure medication. Some case control studies have also failed to determine whether hyperplasia occurred because of poor oral hygiene or poor oral hygiene occurred because of gingival hyperplasia.[5] It is important that analysis of several studies produces similar conclusions and that conclusions are consistent with what we understand of cell biology and organ systems.

Is there a dose/response gradient?

The causal relationship is more convincing, if the risk of the outcome increases with an increase in dose of the harmful agent or increased time of exposure to the harmful agent. Likewise, decreased dose of the offending agent or decreased time of exposure should decrease the risk of the outcome.

WHAT ARE THE RESULTS OF THE STUDY?

How strong is the association between exposure and outcome?

When there are serious concerns about bias in a study, the magnitude of the risk or strength of association between an exposure and an outcome does not help the clinician determine the true cause-and-effect relationship, because biases may magnify apparent associations. However, where there is no obvious bias in the study design or the study methods, the strength of association may increase one's confidence in the results. For cohort studies, one can determine the relative risk (RR) of the outcome of interest occurring in the exposed population, compared with the unexposed population. For the calculations, use the incidence of the event in the exposed group divided by the incidence of the event in the unexposed group.

$$RR = (a/[a+b])/(c/[c+d])$$

For case-control studies one calculates the exposure rate for those with the outcome of interest (cases) compared with subjects who do not have the outcome (controls) to produce an odds ratio. For very rare events the odds ratio closely approximates the relative risk. The odds ratio is determined by using the odds of a "case" having the event divided by the odds of the "control" having the event (Table III).

$$\text{Odds ratio} = (a/c)/(b/d)$$

An RR of greater than 1 represents an increase in risk associated with exposure and a value less than 1 represents a reduction in risk. As an illustration of calculating the RR we can consider adverse events in an RCT by LeVeque et al.[6] They used pilocarpine hydrochloride to stimulate saliva production in patients with irradiated head and neck cancer. Individuals withdrew from the placebo group and from the pilocarpine group because of various side effects that the subjects contributed to "the pill" they were taking. Usually an adverse event is described as a single entity, ie, bleeding, seizure, or infection, but for this study, the authors considered "leaving the study" because of medical complications or physical side effects as the adverse event (Table IV).

In the placebo group, 9 of 87 and in the pilocarpine group 11 of 75 had adverse events. The RR of having an adverse event when taking pilocarpine was 1.41 (RR = [11/75]/[9/87]). When calculating RR, it is necessary that the proportion of patients with the outcome can be determined in both groups, ie, that both the numerator (those who experienced the outcome) and the denominator (all subjects in the test group who did and did not experience the outcome) are known. In case control studies, only the numerators are known (because only patients who have the outcome are selected from the general population), the appropriate expression of association is the odds ratio (the relative odds of exposure in the case subjects compared with control subjects).

When there is a small increase in RR or odds ratio and the study design may have been weak, ie, case-control study, clinicians should wait for stronger evidence before changing their clinical practice. However, if there is a large relative risk and the strong association of cause and effect has been controlled as much as the particular outcome event allows, unknown variables are less likely to be confounding factors or cause study bias.

In the case of the pilocarpine study, medical complications and patient complaints of side effects were

significant enough for subjects to leave the study. The risk of these adverse complications in the test group compared with the placebo group was 1.41. A value of 1 would indicate identical risk in the 2 groups, and indicate that the side effects were unlikely a cause of the study drug. This RR indicates a 41% increase of complications in the test group. This study included dose escalation, and the escalation of dose increased the number of complications. The types of physical complications could also be explained by the pathophysiology of the drug. Given this information, the complications, although not medically serious, seemed to be attributable to the drug, and did cause patients to leave the study.

How precise is the estimate of the risk?

In addition to the relative risk, the estimate of the precision of the relative risk can be determined. This estimate is called the "confidence interval." Many clinicians have a better understanding of *P* values than they have of confidence intervals. The *P* value describes the risk of the false-positive conclusion that a treatment is efficacious when it is not. In other words, the *P* value tells how often these results would have occurred by chance if the experimental treatment were really no different from the control. For instance, if a report states that a therapy improved subjects' ability to masticate by 30% with a P=.06, this *P* value implies that, if the experiment was conducted 100 times, but there was no real difference between the 2 therapies, 6 of 100 times the difference between the 2 therapies would reveal a 30% (or exceed 30%) improvement by chance alone. If, however, the P=.04, 4 of 100 times a 30% improvement in mastication would occur by chance alone.

Although the difference between 4 of 100 and 6 of 100 is minimal, science consistently describes statistical significance at .05. This would declare that a trial generating a 30% improvement and a P=.04 is positive, and establishes that the experimental and control treatments are different. In converse, the trial that generates this same 30% improvement with a P=.06 is interpreted as negative and does not establish that the experimental and control therapies are different from one another. In the former case, the result might influence practice one way, and in the latter case influence practice another way. Focusing on *P* values can have limitations in clinical decision making. Authorities have debated whether P<.05 represents an appropriate way of branding studies as simply "positive" or "negative."

Because of this concern, there is increasing interest in expressing study results with their associated confidence intervals (CIs).[7,8] Use of a CI around the RR helps clinicians decide the range within which they can be confident of the RR estimate. Also somewhat arbitrarily, science invokes a 95% CI in clinical research.

Table V. Hypothetical cohort study of smokers and nonsmokers who undergo periodontal surgery

	Adverse event		
Exposed smokers	Yes	No	Total
Yes	9 a	11 b	20
No	4 c	16 d	20
Total	13	27	40

The range of values within that interval includes the true RR 95% of the time. Confidence intervals also help the clinician decide just how large an adverse effect might be present, despite the failure to show a statistically significant association in the study. A statistically significant association may not be revealed because of the small numbers of subjects in the study. It is seldom that the true RR lies at the extremes of the interval and the true RR will lie outside these values only 5% of the time.

Referring again to the pilocarpine study, adverse reactions occurred with a relative risk of 1.41 (confidence intervals of 0.62 to 3.23). This tells the clinician that in only 5% of cases will the relative risk of these adverse events lie outside 0.62 or 3.23. Because the relative risk confidence interval includes the value of 1, there may actually not be an increased risk of complications for our test subjects. Therefore, this study is not definitive enough to answer the question of relative risk of an adverse event. On the basis of estimation of the CI in this investigation, we could conclude that there is insufficient evidence to support a change in a potentially beneficial treatment because of the risk of an adverse event.

The 2 × 2 table in Table V demonstrates a hypothetical cohort study of smokers and nonsmokers who undergo periodontal surgery.

Nine of 20 smokers experienced an adverse event of flap dehiscence at the suture line, compared with 4 of 20 nonsmokers for an RR of 2.25. Because the RR is greater than l, there is an increased risk of an exposed individual (smoker) having an adverse event. With only 40 subjects, the confidence interval reveals the precision of the estimate of the RR (ie, the RR could be as low as 0.83 or as high as 6.13). Because the confidence interval includes l, the clinician cannot exclude "no statistical difference" in risk for the 2 populations. However, despite the possibility of no cause-and-effect relationship, the confidence interval suggests a large risk for an adverse event. Because the lower boundary of the confidence interval of the relative risk is very near l, and a large upper limit is noted, the clinician may suspect that a risk of flap dehiscence for smokers is probable, but this study does not statistically confirm this impression. Note that only 5% of the time is value

of the RR at the boundaries of the CI. The best estimate is not the boundaries, but the stated RR. The RR in this small cohort of 40 patients suggests a 125% difference between healing of smokers and nonsmokers. This great a risk of a nonhealing wound in a smoker should give cause for concern and the clinician should begin the search for more well-conducted clinical trials that have an increased number of subjects.

In the statistical analysis of a study, an increased number of subjects serves to narrow the upper and lower boundaries of the CI. Even if the relative risk remained the same, increasing the number of subjects in a second study would narrow the CI, giving the clinician a more precise estimate of the strength of the association between exposure and outcome. For instance, if the periodontal surgery study was increased to 80 patients in each cohort and even if the proportions of flap dehiscence and smokers / nonsmokers remained the same (RR = 2.25), the new confidence interval range would narrow from 1.36 to 3.71. Now the confidence interval does not include l, and an assessment of an increased risk of an adverse event for smokers is evident. In this case, the clinician is quite certain that RR = 2.25 and knows that enough subjects have been enrolled to demonstrate a difference in the risk of adverse events between the exposed and nonexposed groups. With this information, a clinician may confidently counsel surgical patients concerning healing complications related to smoking.

WILL THE RESULTS HELP ME IN CARING FOR MY PATIENTS?

Are the results applicable to my practice?

After the clinician determines that the study is appropriate and that the associations of a cause and effect are strong, one should examine the population of the study and decide how it compares with his or her patient population. Is there a similar age range? Are race and gender distributions comparable? Do your patients have differences in confounding dental or medical conditions similar to those of the described population? For example, the risk of caries is lower in a community with a high level of fluoride in the public water source, therefore an investigation examining recurrent caries may not be applicable if the fluoride exposure was dissimilar to patients in your practice. Do the clinicians rendering the care have the same skill level as you, and are you able to duplicate the therapy techniques? If you are satisfied that the populations and settings are similar, you may wish to change your clinical practice after asking yourself the next 2 questions.

Is the magnitude of the risk clinically relevant?

A clinician understands that adverse events usually described in reports can occur in both the nontreated and treated groups. The relative risk and odds ratio do not describe the frequency with which an adverse event occurs. They merely describes how often the adverse event occurs in an exposed group compared with an unexposed group. Clinicians can use harm data from an RCT or cohort study to assist in making a clinical judgment about the adverse event.

This formula determines the absolute risk increase of an adverse event in the 2 groups by taking the difference (1) of the proportion of adverse events in the exposed group (Y = exposed group or therapy test group) and the adverse events in the unexposed group (X = an unexposed group receiving either a placebo or the standard of care treatment). The reciprocal of this absolute risk increase tells us how many patients one would have to treat with the therapy or must be exposed to a harmful agent before one adverse event occurs, compared with the unexposed group. Considering the pilocarpine article again, the frequency of adverse events in the pilocarpine group Y: 11 / 75 or 0.146 minus the frequency of adverse events in the placebo group X: 9 / 87 or 0.103 = 0.043. The inverse of this difference, $1/(Y - X)$ or $1/0.043 = 23$. The clinician understands that 23 patients would have to be treated with pilocarpine to yield one additional adverse event that was related to the pilocarpine. This is not a very high rate of adverse events related to the pilocarpine, and the events were relatively minor physical symptoms that were reversible, without residual effects, when the "drug wore off." A patient might desire to take the risk of these symptoms, if there was a proven efficacy of the test therapy over the standard of care therapy. However, would the clinician and patient be willing to take the risk, if the harmful exposure was smoking and the risk was loss of dental implants?

If my therapy causes harm, should I attempt to discontinue the therapy?

The clinician not only determines the magnitude of the adverse events of the exposure or therapy, but should also consider what benefits of therapy will be lost if the therapy is discontinued. To determine whether the therapy will do more harm than good, one can compare the adverse event outcomes of the therapy with the reduction of risk of the measured outcome in the therapy trial. As an example of discussing clinical relevance, we can again consider the pilocarpine study that examined its use to counter the effects of xerostomia in irradiated head and neck cancer patients. One must examine the efficacy of the therapy to consider the absolute risk reduction of xerostomia that occurred by using pilocarpine compared with placebo. The "numbers needed to treat" (NNT) calculation tells the clinician how many patients would need to receive the therapy to

Table VI. All implant locations in both arches

Exposure	Implant failure	No implant failure	Total
Smoking	44 a	346 b	390
No smoking	86 c	1718 d	1804

RR = 2.37 (1.67-3.35).

Table VII. All maxillary implants

Exposure	Implant failure	No implant failure	Total
Smoking	35 a	161 b	196
No smoking	64 c	819 d	883

RR = 2.46 (1.68-3.61).

Table VIII. All mandibular implants

Exposure	Implant failure	No implant failure	Total
Smoking	9 a	185 b	194
No smoking	22 c	899 d	921

RR = 1.94 (0.01-4.15).

Table IX. All anterior mandibular implants

Exposure	Implant failure	No implant failure	Total
Smoking	4 a	81 b	85
No smoking	1 c	366 d	367

RR = 17.27 (1.10-152.56).

actually recognize the improvement of xerostomic symptoms in 1 additional patient. This calculation requires the risk of having xerostomia without treatment (placebo group) X: 57/77 (0.74) and then the risk of having xerostomia after being treated with pilocarpine Y: 37/69 (0.54). The absolute risk reduction is X – Y (0.74 – 0.54 = 0.20). The number needed to treat to improve 1 person's xerostomia symptoms is 1/(X – Y) or 1/0.20 = 5.

From the previous calculations, it was determined that for every 23 patients receiving pilocarpine, 1 patient would suffer an adverse event directly related to the medication that was serious enough to withdraw from the study. It was also determined that for every 5 patients receiving pilocarpine, 1 would experience a clinical improvement in xerostomia. With this information, a clinical judgment can be made. If the drug is not too expensive or too difficult to administer, the trade off between the side effects and the therapeutic improvement might be worth it for both patient and clinician. If, however, the adverse event for every 23rd patient was more serious (eg, requiring hospitalization), clinicians might feel substantially different about dispensing this medication.

Table X. All posterior mandibular implants

Exposure	Implant failure	No implant failure	Total
Smoking	5 a	104 b	109
No smoking	21 c	533 d	443

RR = 1.2 (0.466-3.14).

Table XI. All posterior maxillary implants

Exposure	Implant failure	No implant failure	Total
Smoking	17 a	72 b	89
No smoking	48 c	319 d	439

RR = 1.78 (1.06-2.89).

Table XII. All anterior maxillary implants

Exposure	Implant failure	No implant failure	Total
Smoking	18 a	89 b	107
No smoking	16 c	428 d	444

RR = 4.67 (2.46-8.85).

RESOLUTION OF THE SCENARIO

The article by Bain and Moy[1] is a cohort trial that evaluated clinical records of patients who were treated by 1 practitioner during a 6-year period from 1984-1990, with prosthetic restoration of at least 1 year. The demographics of the subjects were available as to gender, mean age, and positive or negative smoking history at time of initial examination. The average age was 54 to 55. No information as to health problems, such as diabetes is reported. Approximately 50% of the implants were placed in the maxilla. It was not reported how many implants were placed per subject, the absolute numbers or types of prostheses placed, or the average length of the implants, but the authors reported no difference among any of these variables between the smoking and nonsmoking groups. One might assume that single tooth implants, splinted implants with fixed prostheses, and overdentures were inserted. The adverse event was the loss of an implant or bone loss in excess of 50%.

In reviewing the generalizability of these data to another clinician's practice, there are questions about prostheses types, health of the population, and number of implants per patient that make generalizability difficult to discern. How much the patients smoked, and if they refrained from smoking during the initial wound healing period is unknown. Numbers of subjects lost to follow-up and length of follow-up per subject are in question, so it is difficult to discern if failure rate in both groups may actually be greater than reported. Also, it is unknown if more implants were lost in subjects treated

earlier in the study; which might be indicative of a treatment learning curve.

The 2×2 tables (Tables VI through XII) depict the analysis for relative risk with the associated confidence intervals of smokers and nonsmokers in the Bain and Moy article.[1] The individual implant is the unit of measure, which allows one to determine whether particular locations within the arch are more susceptible to implant failure. The various tables consider the implant locations.

The magnitude and precision of the estimate are quite varied for the different locations in the mouth. There is very little precision in many of the RR estimates. It is possible that subjects lost implants for reasons other than smoking, and this cause for failure could confound the findings in this report. The number of implants lost per subject was not reported and the number of failed implants in some locations is small in both groups of subjects. This is particularly true for the anterior mandibular implants that have only 4 implant losses in the smoking group (Table IX). If these came from 1 subject, and that subject had not been included in the study, the RR would have dropped from a possible 17.27 to 4.4. Despite some of the population and methodology questions generated by this report, it appears that there may be cause for concern. The clinician should revisit this topic in the future, always diligent for updated literature. All reports should be assessed for quality as well as magnitude of risk.

REFERENCES

1. Bain CA, Moy PK. The association between the failure of dental implants and cigarette smoking. Int J Oral Maxillofac Implants 1993;8:609-15.
2. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guide to the medical literature: IV, how to use an article about harm. J Am Med Assoc 1994;271:1615-9.
3. Jacob RF, Carr AB. Hierarchy of research design used to categorize the "strength of evidence" in answering clinical dental questions. J Prosthet Dent 2000;83:137-52.
4. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Deciding whether your treatment has done harm. In: Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston: Little, Brown and Company; 1991. p. 283-302.
5. Klar LA. Gingival hyperplasia during dilantin therapy; a survey of 312 patients. J Public Health Dent 1973;33:180-5.
6. LeVeque FG, Montgomery M, Potter D, Zimmer MB, Rieke JW, Steiger BW, et al. A multicenter, randomized, double-blind, placebo-controlled, dose-titration study of oral pilocarpine for the treatment of radiation-induced xerostomia in head and neck cancer patients. J Clin Oncol 1993;11:1124-31.
7. Sackett DI, Haynes RB, Guyatt GH, Tugwell P. Deciding on the best therapy. In: Sackett DI, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston: Little, Brown and Company; 1991. p. 217.
8. Potter RH. Significance level and confidence interval. J Dent Res 1994;73:494-6.

0022-3913/2000/$12.00 + 0. **10/8/108496**

doi:10.1067/mpr.2000.108496