EVIDENCE-BASED DENTISTRY SERIES

The overview: An article that interrogates the literature

David A. Felton, DDS, MS,[a] and Brien R. Lang, DDS, MS[b]

CLINICAL SCENARIO

Your practice has attracted more and more patients who are concerned about caries and the effect of oral hygiene materials being presented as preventative. While attending a dental meeting, you pass a commercial display advertising what appears to offer compelling evidence for the effectiveness of chlorhexidine treatment as a preventative medicament. The sales representative offers you a list of studies to support the applications for this material; however, the list is much too long for you to read every article. Instead, you decide to search the literature yourself for an overview article on chlorhexidine.

A comprehensive structured review of the published literature, that has an explicit and focused question, rules for inclusion of primary studies to present as evidence, an explanation of the strength of the evidence, and a summary of the collective findings from the primary studies is called an overview. When the data from these primary studies allow combining in rigorous statistical analyses it is called a meta-analysis.

Overviews of the scientific literature are very useful to clinicians and researchers who are responding to questions on the efficacy of patient therapy, or formulating a research project to address a basic science or transitional research question related to dentistry.[1-3] An overview that addresses a well-defined question is more helpful to a reader than a haphazard or biased selection of the available citations.

Overviews must be read as critically as a primary research article. An overview uses specific methodologic criteria and should be viewed "as a study in itself." It must pose a question, gather criteria-based data on the question from published studies of the highest quality, analyze the data, and then draw conclusions from the analysis. The fundamental difference between an overview and primary articles reporting the results of a study is the unit of analysis, not the scientific principles that apply.

STRATEGIES FOR ASSESSING AN OVERVIEW

It is extremely important for the reader to follow a specific strategy when assessing an overview article or an article that incorporates a meta-analysis in the review. The basis elements of such a strategy should determine whether: (1) a clearly defined question was addressed, (2) specific search strategies were used, (3) preparation of the results was detailed, (4) the overall results of the overview were appropriate, and (5) the results can be applied to patient care.

Unless an overview clearly states a question, the reader has no way of knowing if the overview is relevant to a particular treatment option the clinician may be evaluating. The question in the overview must have 3 basic parts: (1) what is being reviewed (patient histories, elements of the physical examination, diagnostic tests, treatment modalities, treatment outcomes), (2) in whom (the experimental population of interest), and (3) for what outcomes (consequences).

The overview should provide an explicit statement of the search strategies used by the author. Such strategies should include, but not necessarily be limited to, one or more bibliographic databases such as PUBMED, MEDLINE, or GREATFUL-MED, and include a statement of which key words were used and in what order they were used. The criteria used to select studies for an overview should be as rigorous as the criteria used to include or exclude subjects in any clinical study. If information is lacking on how the search was conducted, the quality of the search is in question and the credibility of the overview is suspect. Methods that enabled the author to search for classical articles, such as those found in the Index Medicus, should also be used. A system for searching the references of relevant articles for other references that may have an impact on the experimental question should be included. Having 2 or more people participate in selecting articles for inclusion in an overview guards against bias and errors. If there is good agreement between the 2 participants, then the reader will have more confidence in the results of the overview.

A listing of inclusion/exclusion criteria used to select articles must be included. Precise criteria such as experimental or clinical outcomes, treatment maneuver, physical or experimental tests, histories, or prognostic factors are important. In addition, key elements in the study design, such as the use of randomized clinical trials, case series, or cohort studies are examples of "quality filters" authors should use in identifying the inclusion or exclusion criteria.

The majority of the information contained in this article has been adapted from materials developed by the Evidence-Based Medicine Working Group, Department of Clinical Epidemology, McMaster University, Hamilton, Ontario, Canada.

[a]Professor and Chair, Department of Prosthodontics, School of Dentistry, University of North Carolina, Chapel Hill, N.C.

[b]Professor Emeritus, Department of BMS/Prosthodontics, School of Dentistry, University of Michigan, Ann Arbor, Mich.

Searches involving languages other than English are critical. Personal contact with experts in the field of research addressed by the overview also should not be overlooked. There are 2 important reasons for using personal contacts. The first is to identify published studies that might have been missed. The second is to identify unpublished studies. Although the inclusion of unpublished studies is controversial, their omission can lead to "publication bias" because there is a higher likelihood that only studies with "positive" results are published. If unpublished studies are included in an overview, they should be handled in the same way as published studies. Full written reports should be obtained, and the validity of both published and unpublished studies should be appraised.

Accurate assessment of the quality of the methods used in the primary studies cited is essential for correctness of the conclusions drawn in the overview. For example, a number of case series and cohort studies suggested that occlusal contacts were maintained and wear was not a problem after placing Class II composite restorations in maxillary premolars. This served as justification to conduct a larger randomized clinical trial that demonstrated wear resistance was not effective in proximal/occlusal restorations. The results of an overview that included only the case series and cohort studies lead to an incorrect conclusion that restoring teeth with proximal/occlusal defects using composite materials was acceptable. This emphasizes the need to use the strength of the randomized clinical trial research design for questions related to therapy.

The reader can only make decisions about the quality of a primary study when the study methods are reported in sufficient detail. The reader must be able to determine whether the study met minimal scientific criteria that would allow strong inferences to be drawn from the results. If for example, the conclusions drawn differ significantly from other cited studies, can these differences be explained on the basis of the methodologic quality? The truth most likely lies in the conclusions drawn from the studies in which the methodologic quality was not questionable and the patient population/subjects and outcome measurements in the included studies were similar. However, the reader must recognize that differences could also be due to differences in subjects in the primary study or differences in the exact outcome measurement used in that particular study.

Before scrutinizing how the overview author combined the results of the individual studies, the reader must be satisfied that it makes sense to combine this set of studies. Are the patients, treatments, outcomes, and research methods sufficiently similar that it makes clinical and biologic sense to combine them? If not, hopefully the overview author refrained from trying to combine the results statistically. The author must state explicitly the basis for any conclusions reached, and provide a unifying explanation for conflicting results. If the overview of dissimilar studies combines their results, the reader is cautioned about accepting the conclusion.

Some overviews are initiated to settle controversy. Therefore, it should come as no great surprise that the primary articles included in the overview would differ and have opposing results. In a number of instances, individual articles may agree on the effect of a factor being examined in an experimental question, but disagree on the magnitude of the effect. Good overviews confront these differences and try to explain them. Differences in study results generally arise from 5 sources: (1) different sorts of patients (with different oral environmental conditions or responsiveness to treatment), (2) different histories, ways of performing the diagnostic tests, or performing treatments (including the extent of the therapy, lengths of the test periods, combinations of treatment, and compliance), (3) different outcomes (defined and measured in different ways and with different measurement systems or degrees of sensitivity), (4) different study methods (with different rigor and power), and (5) the play of chance.

All 5 of these possible explanations should have been considered by the author in interpreting individual study results to ensure that it makes clinical sense to combine the studies. In addition, it is possible to test the extent to which differences among results of individual studies are "significant" (greater than could be expected due to chance alone). The statistical analyses that are used to do this are called tests of homogeneity. The more significant (closer to zero) the test, the less likely it is that the observed differences are due to chance alone. When there is "statistically significant" heterogeneity (a low probability of the differences being due to chance alone), both the "average" effect and the confidence interval around the average effect need to be interpreted cautiously.

Unfortunately, tests of homogeneity have limited power to detect differences. So, a nonsignificant test does not necessarily rule out important differences, and large differences between study results still dictate some degree of caution in interpreting the overall findings. At the same time, conclusions that are drawn on the basis of between study comparisons should be viewed skeptically. Even when there are large differences between the results of different studies, a summary measure from all the best available studies may provide a better estimate for clinical use than the results of any 1 study. However, the strength of inference associated with the overall effect is weaker than when study results are consistent.

The results of individual primary studies should be reported in sufficient detail that the reader is able to critically assess the basis for the reviewers' conclusions.

Tables summarizing crucial aspects of methods and results can be helpful in quickly deciding whether conclusions really are consistent with data. Clinical and statistical significance should receive appropriate attention in the overview.

In clinical research, investigators collect data from individual patients. In overviews, investigators collect data from individual studies. These data must also be summarized, and investigators are using quantitative methods to do so. Simply comparing the number of "positive" studies to the number of "negative" studies is not an adequate way to summarize the results, for it is possible for a study to be counted as "positive" in one overview and "negative" in another. There is also a tendency to overlook small but clinically important effects when studies are statistically "nonsignificant" (but potentially clinically important) results are counted as "negative."

Typically, meta-analyses weigh studies according to the precision of their results, that is, the overall results represent a weighted average of the results of the individual studies. Occasionally, studies are also given more or less weight depending on their quality. If there are large differences in the quality of the studies, poorer quality studies may be given a weight of zero (excluded) either in the primary analysis or in a "sensitivity analysis." In such a sensitivity analysis, the investigators would leave out the poor quality studies from the analysis, and find out whether that made an important difference in the overall estimate.

Sometimes the outcome measures that are used in different studies are similar but not exactly the same. For example, different trials might measure wear resistance of several different composites. If the subjects in the different studies and the interventions are reasonably similar, it might still be worthwhile to estimate the average wear resistance of the composite materials across studies. One way of doing this is to summarize the results of each study as an "effect size." The effect size is the difference in wear between the different composites and the control groups (usually amalgam) divided by the standard deviation. Hence, the results of each study are summarized for the number of standard deviations of difference between the composites and control groups. In such a situation, it is possible to calculate a weighted average of effect sizes from studies that measured an outcome of interest (wear resistance) in different ways. However, it is difficult to interpret the clinical importance of an effect size. Meta-analyses that use an effect size to summarize the results should also report the results in a way that clearly conveys practical importance (for example, by translating the summary effect size back into natural units).

As of April 2000, a total of 767 review articles in dentistry have been reported, but only 64 meta-analyses have been published in dental journals and cited in PUBMED. "Meta-analysis is the retrospective systematic identification, appraisal, and comparison of all clinical studies relevant to a selected topic of interest, using statistical methods to combine and summarize results from the related studies."[4] It is perhaps the best method for minimizing bias from related investigations. Clinical trials often have small sample sizes that preclude the use of strong statistical analysis of the results. Combining the results of multiple similar trials using meta-analysis may increase the statistical power sufficiently to allow detection of small but significant effects of therapy that the individual trials could not determine.

One of the advantages of an overview is that a diverse range of patients is usually included across all studies reviewed, and if the results are consistent across studies, they apply to this wide variety of patients. However, the clinician may still be left with doubts about the applicability of the results. For example, the subject in our study may present with only maxillary premolar teeth, a slightly different situation than any of the subjects included in the trials on the wear of different composite materials against natural posterior teeth. One might question whether the composites have a larger effect on molar teeth than premolar teeth.

Although it is a good idea to look for focused overviews because they are more likely to provide valid results, this does not mean that you should ignore outcomes that are not included in an overview, which would be important to the subjects, such as esthetics. When making a clinical decision, the expected benefits must be weighed against the potential harms and costs. Although this is most obvious for deciding whether to use a therapeutic or preventive intervention, it is important to remember that providing patients with information about causes of dental disease or prognosis can also have both benefits and harms.

SEARCHING FOR AN "OVERVIEW"

The National Library of Medicine's PUBMED citation listing can be accessed through its Internet address: www.ncbi.nlm.nih.gov/PUBMED. The use of the PUBMED system requires some basic knowledge of the use of key words to allow access to the articles of interest, along with the skills to properly limit the scope of the search without eliminating useful references. For example, a query of the PUBMED system for articles related to the use of chlorhexidine as a preventative treatment for caries is summarized in Table I.

The busy clinician does not have the time to review 200+ citations to determine whether he/she should consider the use of chlorhexidine as a treatment modality for caries reduction in his/her dental practice. Unfortunately, the use of Reviews in the search query may also fail to provide the reader with a sufficient analysis of the articles cited in the Review manu-

Table I. Query results in PUBMED system

Query no.	Key words	Number of citations in PUBMED
1	Chlorhexidine and caries	236
2	Chlorhexidine and dental caries	208
3	Chlorhexidine and caries and reviews	3
4	Chlorhexidine and dental caries and reviews	2
5	Chlorhexidine and caries and meta-analyses	2
6	Chlorhexidine and dental caries and meta-analyses	1

script to allow careful interpretation of the data presented. For example, in Table I, the 3 articles noted under Query 3[5-7] did not provide the references by van Rikjom et al[8] discovered in Query 6, which was a meta-analysis on chlorhexidine therapy.

In the overview by van Rijkom et al,[8] the authors stated that they would assess the caries-inhibiting effect of chlorhexidine treatment. Specifically, the authors would examine factors that potentially could modify the effect of chlorhexidine in caries prevention, that is, the application method, application frequency, target population, fluoride regime, tooth surface involved, and caries diagnosis. Although the authors identified "what was being reviewed" and for "what outcomes" they did not specify "in whom" in their study objectives.

The authors used the MEDLINE database to begin their search of the literature for articles on the effect of chlorhexidine treatment in caries prevention. The search was limited to articles that appeared in print between the years 1975 and 1994 in English, French, or German. Twenty-four articles were found in the search. Two examiners using comparability for dental and methodologic reasons independently reviewed the articles. The 2 examiners subjected the 24 articles to the criteria of: (1) chlorhexidine applied to permanent teeth from 11- to 15-year-old children; (2) studies performed in clinical trials with randomly assigned treatment groups, including generally treated experimental groups as well as treatment focused on subjects in the experimental group with *Streptococcus mutans* $> 25 \times 10^5$/mL, saliva; (3) the availability of caries incidence data; (4) a treatment duration of at least 1 year; and (5) evaluation at the end of the treatment period. These "quality filters" were used to identify the "in whom" part of the experiment question, and to the narrow overview, and identification of potential factors that could modify the effect of chlorhexidine. Using the criteria, a total of 10 articles were eliminated from further review.

Further, the use of 2 investigators to independently review the studies identified in the MEDLINE search assisted in the elimination of any bias in the selection of articles for inclusion. Each investigator determined whether an article met both dental and methodologic criteria. Studies that were out of the age range, were not clinical trials, contained no caries incidence data on surfaces involved, or used treatment that lasted less than 1 year were eliminated from consideration. Studies that specified application methods, application frequencies, caries risk, fluoride regime, caries diagnostic levels, and included tooth surface were considered in the overview. Studies that contained both experimental and control populations needed to be evident.

In the van Rijkom overview, investigators described the caries-inhibiting effect of chlorhexidine by a prevented fraction calculated for each reported study. The prevented fraction was the difference in number of new decayed and filled surfaces between the control group and the chlorhexidine group, divided by the number of new decayed and filled surfaces in the control group. It indicated the percentage reduction of caries incidence in the chlorhexidine group and was considered less sensitive to experimental circumstances than the absolute reduction. Studies that included follow-up years indicating the total length of the study as well as the treatment duration were excluded because the prevented fraction was assumed to be independent of the duration of the study.

To compare data across studies by meta-analysis, the investigators calculated the 95% confidence intervals for the prevented fraction from the data reported for each study. This was possible given the reported mean data and the standard deviations in each study. The authors were able to pool data at the 95% confidence interval from the selected studies to demonstrate the homogeneity of the published results. Although only 8 articles were included in the analysis, the number of subjects in the individual reports varied from 9 to 72. Combining the articles into a meta-analysis allowed the assessment of 306 total subjects, giving strength to the statistical analysis of the results. The total sample size, appropriate inclusion / exclusion criteria for the articles reviewed, 95% confidence interval selected, and homogeneity of the individual study results allow the reader to clearly assess that the use of chlorhexidine in this particular patient population will have a dramatic effect on the caries incidence, and thus should be applicable to the reader's practice.

An important finding in the meta-analysis that should be of invaluable use to the reader related to the "application method" for use of chlorhexidine in this patient population. In the 8 articles included in the analysis, application of topical gels (professionally applied), rinsing with 0.5% chlorhexidine rinse, use of chlorhexidine containing tooth pastes, or combinations of toothpaste and rinsing, were studied. The meta-analysis demonstrated no differences in the application

method. This enables the clinician to tailor treatment based on patient compliance rather than a single prescribed therapy. Thus, this meta-analysis provides useful information for the clinician in the treatment of his juvenile patients.

SUMMARY

Although the question on the efficacy of chlorhexidine therapy for the clinician's adult patient may not have been totally answered, a greater perspective on this treatment modality was gained by the reader because of the quality of the overview cited, and the evidence-based approach used by the authors in performing the overview.

REFERENCES

1. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. How to read reviews and economic analyses. In: Clinical epidemiology—a basic science for clinical medicine. Boston: Little, Brown and Company; 1991. p. 379-94.
2. Oxman AD, Guyatt GH, the Evidenced-Based Medicine Working Group. How to use an overview. In: Evidence-based medicine: users' guides to the medical literature. Chapter VI. Hamilton, Canada: McMaster University Health Sciences Centre, 1993.
3. Oxman AD, Cook DJ, Guyatt GH, for the Evidenced-Based Medicine Working Group. Users' guide to the medical literature. VI: How to use an overview. JAMA 1994;272:1367-71.
4. Palmer AJ, Sendi PP. Meta-analysis in oral health care. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1999;87:135-41.
5. Ismail AI. Prevention of early childhood caries. Community Dent Oral Epidemiol 1998;26(Suppl 1):49-61.
6. Kidd EAM. Role of chlorhexidine in the management of dental caries. Int Dent J 1991;41:279-86.
7. Hogg SD. Chemical control of plaque. Dent Update 1990;17:332-4.
8. Rijkon HM, Truin GJ, van't Hof MA. A meta-analysis of clinical studies on the caries-inhibiting effect of chlorhexidine treatment. J Dent Res 1996;75:790-5.

0022-3913/2000/$12.00 + 0. **10/8/108651**

doi:10.1067/mpr.2000.108651