The Use of the Evidence-Based Approach in a Periodontal Therapy Contemporary Science Workshop

Michael G. Newman,* Jack G. Caton,† and John C. Gunsolley‡

* University of California Los Angeles, School of Dentistry, Section of Periodontics, Los Angeles, California.
† University of Rochester, Eastman Dental Center, Department of Periodontics, Rochester, New York.
‡ Department of Periodontics, College of Dental Surgery, University of Maryland, Baltimore, Maryland.

Background: When appropriately evaluated and carefully managed, the integration of emerging technology into practice can improve health and enhance the quality of life. Since the last American Academy of Periodontology Workshop in 1996, great technological advances in the areas of data access, retrieval, and management have been made. The World Wide Web has “exploded” with great possibilities for gathering data from many sources. Evaluation methods such as meta-analysis and modeling have likewise improved, permitting a more objective and useful assessment of the retrieved information. The purpose of this paper is to demonstrate how the evidence-based (EB) approach was used to plan and implement a consensus conference on periodontal therapy, the Workshop on Contemporary Science in Clinical Periodontics.

Rationale: The methodologies and philosophies associated with the EB approach provided the ideal framework for assessing the applicability of the newest clinical research to patient therapy.

Methods: Evidence-based systematic reviews on 15 topics associated with contemporary clinical periodontal practice were conducted prior to the Workshop. High standards of scientific rigor and scholarly ideals were stressed throughout the process. At the highly structured conference the reviews served as the basis for development of consensus reports that include implications for practice and research.

Material Covered:
1. The rationale, design, and implementation of a conference on contemporary clinical periodontics using an evidence-based approach.
2. Data management, clinical versus statistical significance, and the challenges of technology transfer and dissemination.
3. The benefits and limitations of using the EB approach in a consensus conference.

Conclusions:
• The consensus statements resulting from the conference should serve to augment clinical decision-making, research priorities, education, and reimbursement.
• The evidence-based approach removed much of the subjectivity traditionally associated with classical reviews of the literature and allowed participants to focus on substantive issues.


KEY WORDS
Periodontal diseases/therapy; data monitoring; evidence-based dentistry; review literature; meta-analysis; decision making.

Periodontology has a rich history and a strong passion for science. The substantial and extensive periodontal information base, developed over the years, has provided a rational basis for choosing the best treatment for patients. Appraisal of this information has been an ongoing and continuous effort by the American Academy of Periodontology (AAP) to ensure that the most accurate and efficacious concepts and technologies are used to provide care and stimulate innovation. The following information provides the background on how an evidence-based approach was used to plan and implement a consensus conference on contemporary clinical science associated with periodontal therapy.

RAPID CHANGE AND NEW KNOWLEDGE
Changes in clinical periodontics have emerged at a rapid rate. Many diagnostic methods and treatment options can be incorporated into the management of the patient’s oral condition.¹ The AAP’s 2020 Vision² suggests that the rate of new products and treatments will continue to increase, placing greater demands on practitioners to become more familiar with ways that facilitate their...
Evidence-Based Approach

Volume 8 • Number 1 • December 2003

objective evaluation and proper implementation. An EB systematic review and the consensus of experts can be used to assess the appropriate use of emerging technology.

New technology and ideas are not considered by all practitioners in the same way. One might think that a proven “new” idea that decreases morbidity, improves the quality of life, and provides a broader scope of options would be welcomed by all care givers. This is not the case and has not been the case for almost 125 years as practitioners are generally slow to adopt new ideas and technologies.

Eisenbeg’s well quoted Ten Lessons for Evidence-Based Technology Assessment describes how the stethoscope was received in the medical community: Soon after Rene Laennec invented the stethoscope in 1816, it was met with “suspicion and distrust...by those who were practicing medicine when it was introduced.” As late as the 1850s, skeptics described the new diagnostic aid as “a dangerous instrument.”

While most agree that health care technology has advanced the clinician’s ability to improve patients’ health and quality of life, there has been considerable disagreement about which technologies to use, how much is too much (or not enough), and whether the technologies practitioners use are providing value for the money spent. In 1827, a commentator on the stethoscope wrote, “The extent of its usefulness is, as yet, far from being ascertained.” Most new technologies since then have been greeted by similar initial doubts.

The stethoscope is a good example of a “new technology” that proved to be clinically useful. However, clinicians must also be suspicious about new technologies unless the evidence indicates that adoption of the technology in clinical practice makes sense. Barriers to implementing new concepts and ideas into clinical practice, like those that will come out of the Workshop, are a complicated issue and are discussed in more detail below (see From Workshop to Practice).

**SINCE THE 1966 WORLD WORKSHOP**

In most health professions, high quality clinical care progresses despite gaps in knowledge, but having periodic assessments keeps the information base up-to-date and more accessible. The AAP regularly monitors treatments, products, and concepts to insure that even though they have been evaluated once, they are still the best available or even as useful as originally envisioned. Changing patterns of disease and improvements in treatment may render a once invaluable approach now inappropriate. On the other hand, technologies also evolve. A test, device, drug, procedure, or intervention for which there is new or mounting evidence for broad use may prove to be important only after thorough evaluation, continued development and use in the field. A major goal of the evidence-based Workshop described here is to thoroughly evaluate contemporary diagnostic and treatment modalities so that the information can be used to augment clinical decision-making, research priorities, education, and reimbursement.

Over the years AAP workshops have stressed the importance of examining the literature. Often these have been exhaustive reviews accompanied by expert narration and evaluation of the material gathered by the reviewer. The results were scholarly works that guided a whole range of activities including, but not limited to, patient management, establishment of insurance benefits, education, product development, identification of standards of care, formation of clinical research protocols, and investment in research and policy development. This approach served multiple stakeholders and helped to guide progress including the improvement of clinical periodontal treatment.

The American Academy of Periodontology consensus conferences have been conducted since 1966 when the first World Workshop in Periodontics was held in Ann Arbor, Michigan. The proceedings of every conference have served to fill gaps in knowledge and guide interested parties to advance the state of knowledge about periodontal health and disease. Practitioners, researchers, companies, policy makers, students, third party payers and the Academy itself regularly use these periodic evaluations for decision making.

Since the last Academy World Workshop in 1996, great technological advances in the areas of data access and retrieval have been made. The World Wide Web has “exploded” with great abilities to gather data. This has resulted in more thorough and complete analysis of data collections and information. Methods of evaluation of information have likewise improved and complex statistical tools such as meta-analysis and modeling of complex clinical studies can be readily used to give greater confidence to decision-makers.

Even with all of these advances, the translation of clinical and laboratory research into patient treatment remains at a relatively slow pace. The recent literature discussing the transfer of technology documents a myriad of reasons why this slow pace is still the norm in clinical practice of dentistry and medicine (see From Workshop to Practice).

**THE EVIDENCE-BASED APPROACH**

During the 1990s, a process for reviewing large volumes of clinical and scientific data, the EB approach, emerged in medicine and other health fields. EB is “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based [healthcare] means integrating individual clinical expertise with the
best available external clinical evidence from systematic research."\(^7\)

Although EB was originally developed in medicine, its principles are the same for dentistry. The American Dental Association defines the term “evidence-based dentistry” (EBD) as an approach to oral health care that requires the judicious melding of systematic assessments of clinically relevant scientific evidence, relating to the patient’s oral and medical condition and history, with the dentist’s clinical expertise and the patient’s treatment needs and preferences.\(^8\) A relative weighting of information is used in the evidence-based approach, making information derived from scientific inquiry more important than information based on intuition, authority or custom.\(^9\)

The EB approach was used to guide the development of systematic reviews and consensus discussions in this international Workshop on Contemporary Science associated with periodontal therapy. Evidence helps to close the void between what is known and what is practiced, and for “newer” treatments, this gap may be substantial. The availability of evidence-based technology assessments is not enough to improve practice, reduce variation, improve physical function, relieve anxiety and stress, and achieve better outcomes.\(^3\) In other words, evidence may be necessary, but it is certainly not sufficient for complete decision-making.\(^10\) The findings of research need to be translated into information that is considered better and necessary because the number of new scientific insights that emerge each year is overwhelming (Table 1).

The “features” of the EB approach are aimed at helping policymakers, patients, and practitioners make decisions with regard to health care by basing their decisions on the best evidence available. Compared to other methods of collecting and assessing information (Table 2), a consensus conference based on EB methods offers a high degree of certainty about its conclusions and recommendations.

### EVIDENCE-BASED SYSTEMATIC REVIEWS AND GUIDELINES

Practical, evidence-based recommendations on how to manage health problems are seen by practitioners, payers, and policymakers as potentially power-ful tools to achieve effective and efficient care, provided that they are well developed and implemented.\(^10\) “Provided they are well developed” is critical to the validity of the conclusions. This is a major stumbling block for many well-meaning groups that review and report on large databases because they typically do not have the leadership and/or the financing to conduct the in-depth analysis necessary. Systematic reviews require considerable resources and time.

International groups, such as the Cochrane Collaboration, prepare, maintain, and promote the accessibility of systematic reviews of the effects of health care interventions with the expectation that clinicians will base their decisions on the best evidence, as opposed to any evidence, and that they will also consult databases containing such reviews. In medicine, EB is presented as a method for continuous learning and for

<table>
<thead>
<tr>
<th>Table 1.</th>
<th>Why The Evidence-Based Approach Is Better Than Other Assessment Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>The evidence-based approach:</td>
<td></td>
</tr>
<tr>
<td>• Is objective</td>
<td></td>
</tr>
<tr>
<td>• Is scientifically sound</td>
<td></td>
</tr>
<tr>
<td>• Is patient focused</td>
<td></td>
</tr>
<tr>
<td>• Incorporates clinical experience</td>
<td></td>
</tr>
<tr>
<td>• Stresses good judgment</td>
<td></td>
</tr>
<tr>
<td>• Is thorough and comprehensive</td>
<td></td>
</tr>
<tr>
<td>• Uses transparent methodology</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2.</th>
<th>Other Methods of Collecting and Assessing Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Purpose</td>
</tr>
<tr>
<td>Surveys</td>
<td>Determines practice patterns, attitudes.</td>
</tr>
<tr>
<td>Expert opinions</td>
<td>Provides guidance in areas in which data may be inadequate.</td>
</tr>
<tr>
<td>Narrative literature review</td>
<td>Overview of reviewers’ interpretation of the subject. Collection of evidence is determined by reviewers’ personal experience.</td>
</tr>
<tr>
<td>Systematic evidence review</td>
<td>Comprehensive, objective search and analysis of all evidence including unpublished data. More reliable and accurate conclusions and inferences.</td>
</tr>
<tr>
<td>Consensus based on EB systematic review</td>
<td>Combines expertise of multiple stakeholders with systematic review. Highest level of evaluation and most useful.</td>
</tr>
</tbody>
</table>
improving care. However, there is a growing consensus that although dentists and physicians like to be informed about scientific results, they have problems acquiring the skills needed to search and review the relevant literature or to consult databases within the context of their daily work.11-14

The evidence-based consensus reports that were produced in this Workshop were essentially scientific and clinical assessments that differ only slightly from a clinical guideline development. As defined by the Institute of Medicine, guidelines are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.”15 The inclusion of scientific evidence within clinical practice guidelines has now become more or less standard in the Western world. The reports from this Workshop do not give specific algorithms or explicit guidelines. They provide information that can be used to guide decision-making in clinical practice.

THE WORKSHOP DEVELOPMENT PLAN
There are 5 interconnected stages that describe the planning and execution of the Workshop (Fig. 1).

Stage 1. Determine Goals, Develop Conference Structure, Topics, and Select Participants
In 2000, the Academy Board of Trustees approved a plan to conduct an EB workshop that would gather available information about clinical treatment modalities associated with emerging, newer, and less established treatments.

Goal 1. To establish a sound and evaluated database of contemporary technology in the form of systematic reviews using EB methodologies. In order to accomplish this, evidence-based systematic reviews would be conducted that could serve as the basis for a consensus development on the implications for practice and research. The systematic review incorporates “a process of systematically locating, appraising and synthesizing evidence from scientific studies in order to obtain a reliable overview.”9

Goal 2. A second, but equally important goal, was to empower clinicians, research workers, funding agencies, payers, product developers, and policy makers with insights that would facilitate using, improving, and developing new products. Newer products inevitably undergo continued refinement and innovation as practitioners become familiar with what to expect from incorporating them into their treatment armamentarium. Having a well evaluated scientific basis, as will be available from the proceedings of this Conference, enhances appropriate use of products and treatments and facilitates creativity. Having a good foundation is also important for the continuous evaluation and improvement.

Conference structure. The Conference was structured to ensure a scholarly approach and the incorporation of the highest standards of peer review. A single source for biostatistical evaluation was used. One hundred participants were invited to the conference. A balance between clinical and scientific expertise was sought for each subject group.

Stage 2. EB Methodology Training and Calibration
The Academy conducted orientation, and systematic review training and calibration meetings with the reviewers, secretaries, chair, Organizing Committee, and staff. Facilitators Ian Needleman and Richard Niederman trained the groups to encourage consistency. Reviewers and group members developed focused questions that would serve to direct the systematic reviews. The report from the NHS Centre for Reviews and Dissemination,9 served as the guide for conducting the systematic reviews and as a general reference for the Organizing Committee and section chairs.

Stage 3. Perform Systematic Review and Analysis
The strength and generalizability of the evidence is based on evaluation according to the EB system that incorporates a hierarchy of studies or observations, starting with the lowest level of laboratory findings to the highest level consisting of systematic reviews of randomized controlled clinical trials. The systematic review of all available evidence and its objective analysis using predetermined criteria and rules is a major
difference from a narrative review of the literature (Table 3).

The systematic review, a 4-step method. Step 1. The topics to be evaluated by the reviewers and the consensus conference were developed by the Conference Organizing Committee and approved by the Academy Board of Trustees. A clinically relevant and focused question was developed for each topic. The format used is called the PICO format, an acronym composed of the 4 essential elements that must be part of the question:

- **P** = Population or patient
- **I** = Intervention
- **C** = Comparison
- **O** = Outcome

Once the PICO question(s) was finalized a formal search protocol was developed before the search was conducted. Similar to a protocol for a clinical study or laboratory experiment, the protocol clearly identifies the details of the search so that others can evaluate the strategy and replicate it. The protocol also included a plan for managing the biostatistical aspects of the findings (see below). The protocol was submitted by the reviewers, evaluated by the Workshop co-chairs and then approved by the Workshop Organizing Committee. Development and approval of a protocol are key differences from traditional narrative reviews of the literature that are based on the efficiency and preferences of the author-reviewer.

Step 2. This step focused on systematically conducting searches for the best evidence from all studies, published or unpublished, in the English language that may help to answer the clinically relevant PICO question.

“Best evidence” is a term that refers to information obtained from the following sources in order of most valid: human randomized controlled clinical trials; non-randomized controlled clinical trials; cohort studies; case-control studies; crossover studies; cross-sectional studies; and case studies. In the absence of scientific evidence, the consensus opinion of experts in appropriate fields of research or clinical practice is used.

After selecting, summarizing, and synthesizing all relevant studies that directly answered the focused clinical PICO question, the strength of the available scientific evidence was graded using predefined criteria.

Step 3. The management of the analyses was centralized for consistency and the details are described below (Data Management and Analysis). Qualitative and/or quantitative analyses were conducted to determine the validity and usefulness of the findings. In the EB approach, 2 types of validity are considered:

- Internal validity refers to the intrinsic soundness of the scientific and clinical rationale, the structure of the study, determination of bias, and statistical approach. This evaluation often “makes or breaks” the study in terms of its inclusion or exclusion within the context of the systematic review.

Inevitably, large numbers of studies are not included in the

### Table 3.

Differences Between Traditional Literature and EB Systematic Reviews

<table>
<thead>
<tr>
<th></th>
<th>Narrative Review of the Literature</th>
<th>Traditional Systematic Review</th>
<th>AAP EB Consensus Conference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td>Subjective</td>
<td>Objective</td>
<td>Objective</td>
</tr>
<tr>
<td>• Comprehensive search</td>
<td>Not always</td>
<td>Always</td>
<td>Always</td>
</tr>
<tr>
<td>• Statistical evaluation of</td>
<td>Uncommon</td>
<td>When appropriate</td>
<td>When appropriate</td>
</tr>
<tr>
<td>findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Assess entire body of evidence</td>
<td>Varies</td>
<td>Always</td>
<td>Always</td>
</tr>
<tr>
<td>Strength of recommendations</td>
<td>Weak</td>
<td>Strong</td>
<td>Strongest</td>
</tr>
<tr>
<td>• Emphasis on future development</td>
<td>Varies</td>
<td>Not usually</td>
<td>A major goal</td>
</tr>
<tr>
<td>• Generalizability</td>
<td>Weak</td>
<td>Good</td>
<td>Varies</td>
</tr>
<tr>
<td>Reliability of clinical</td>
<td>Varies</td>
<td>Varies</td>
<td>Excellent</td>
</tr>
<tr>
<td>conclusions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research suggestions</td>
<td>Not usually focused</td>
<td>Varies</td>
<td>A major goal</td>
</tr>
<tr>
<td>Expedites transfer of new</td>
<td>Usually reviews</td>
<td>Often</td>
<td>A major goal</td>
</tr>
<tr>
<td>information</td>
<td>established area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limitations</td>
<td>Subjectivity</td>
<td>Strictness of methodology</td>
<td>Strictness of methodology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>may eliminate useful</td>
<td>may eliminate useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td>information</td>
<td>information</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expensive and time-consuming</td>
<td>Expensive and time-consuming</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Requires efficient</td>
<td>Requires efficient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dissemination since</td>
<td>dissemination since</td>
</tr>
<tr>
<td></td>
<td></td>
<td>adoption of new</td>
<td>adoption of new</td>
</tr>
<tr>
<td></td>
<td></td>
<td>technology is</td>
<td>technology is</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inherently slow</td>
<td>inherently slow</td>
</tr>
</tbody>
</table>
EB analysis because they did not meet the predetermined criteria for internal validity set up in the protocol of the systematic review. The reasons for including or eliminating a study are made clear to the reader. However, just having explicit criteria does not ensure that reviewer was correct. In fact many of the disagreements and discussion between participants at a consensus conference, as well as readers of the proceedings, revolve around study design and inclusion and exclusion criteria.

External validity is the extent to which the evidence is relevant and generalizable to the population and conditions of typical dental practice.

Step 4. The findings of the systematic reviews were submitted for peer review. The authors’ opinion on the quality and strength of evidence used in the review and implications for practice were included. Gaps and opportunities in the knowledge base that require further research were identified by most authors.

Stage 4. Conference
The Conference was focused on translating the findings from systematic reviews into general consensus statements that described scientific and clinical assessments and implications for research and practice.

The primary goal was to develop non-ambiguous assistance to dentists, hygienists, physicians, and others until later research or reviews become available. In this regard, the same 5 questions were given to each consensus group to guide their evaluation of the evidence and to maintain consistency between groups:

1. Does the section agree that the evidence-based systematic review is complete and accurate?
2. Has any new information been generated or discovered since the evidence-based search cut-off date?
3. Does the section agree with the interpretations and conclusions of the reviewers?
4. What further research needs to be done relative to the focused questions of the evidence-based review?
5. How can the information from the evidence-based review be applied for patient management?

Consensus statements for question 5, developed by the workshop participants, were ranked according to the strength of the evidence (Tables 4 and 5). Upon initial review of the Consensus Reports, it may appear to the reader that recommendations from one section may contradict the recommendations on a similar topic in another section. In fact, many of the contemporary therapies evaluated by the consensus groups have not had sufficient time to generate the studies that would be the basis for a Strong grade. However, many of these therapies are safe and effective and can be used after evaluation of the patient’s circumstances and preferences. Having a strong recommendation is not a prerequisite for a particular therapy’s use. In some situations several consensus groups may have evaluated a particular therapy and graded the evidence differently. This is perfectly acceptable since the evidence used to support the application of the therapy in one situation may not have been intended to be applied to a totally different therapy or context.

Stage 5. Publish and Disseminate Findings
In addition to publishing Workshop proceedings in this issue of the *Annals of Periodontology*, wide dissemi-

---

**Table 4.**

<table>
<thead>
<tr>
<th>Level</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Evidence obtained from at least one properly designed randomized, controlled trial (RCT).†</td>
</tr>
<tr>
<td>II-1</td>
<td>Evidence obtained from well-designed controlled trials without randomization.</td>
</tr>
<tr>
<td>II-2</td>
<td>Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than 1 center or research group.</td>
</tr>
<tr>
<td>II-3</td>
<td>Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments.</td>
</tr>
<tr>
<td>III</td>
<td>Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.</td>
</tr>
</tbody>
</table>

* Adapted from references 16, 17, and 18.
† Generally, this refers to an RCT that is also sized large enough so that the results are significant and compelling.

---

**Table 5.**

<table>
<thead>
<tr>
<th>Strength</th>
<th>Strength of Evidence</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>Strong evidence to support the recommendation</td>
<td>Consistent Level I studies†</td>
</tr>
<tr>
<td>Moderate</td>
<td>Moderate evidence to support the recommendation</td>
<td>Consistent Level II-1 or II-2 studies or extrapolations‡ from Level I studies</td>
</tr>
<tr>
<td>Limited</td>
<td>Limited evidence to support the recommendation</td>
<td>Level II-3 studies or extrapolations from Level II-1 or II-2 studies</td>
</tr>
<tr>
<td>Incomplete or Insufficient</td>
<td>Incomplete or insufficient evidence to support the recommendation</td>
<td>Inconsistent or inconclusive studies of any level; Anecdotal evidence only, Level III</td>
</tr>
</tbody>
</table>

* Adapted from references 16, 17, and 18.
† See Table 4.
‡ Extrapolations are when data are used in a situation which has potential clinically important differences from the original study situation.
nation of the information developed at the Workshop will be coordinated by the Academy.

**DATA AND INFORMATION MANAGEMENT**

In this Conference, systematic reviews followed specific criteria when selecting studies for inclusion in possible meta-analyses of the PICO questions. Very strict criteria were set up around a research question to screen studies. Criteria were also used to evaluate the rigor of the studies. Randomization, masking, entrance criteria, therapy provided, patient population, outcome variables, observation time, controls, and a myriad of other issues that may influence the results of the studies were evaluated.

The choice of studies, outcome variables, and criteria for inclusion in the reviews was the task of the individual authors.

- Each author used the specific PICO question as a guide to determine the criteria for study inclusion and a search strategy.
- Authors provided the data to be analyzed in an electronic form.
- A determination of whether there were sufficient data to do a formal meta-analysis was made for each research question by the authors and the data analyst.
- The decision was based on having a sufficient number of studies with the same outcome variable to provide insight into the PICO question.
- A predetermined number of studies for review or analysis was not set. Instead, the studies were evaluated for their similarity, and based on the nature of the question, it was determined if there were a sufficient number of studies. In general, the broadness of the questions required a larger number of studies in order to provide reasonable answers.

The use of broader research questions produced potential and real analytical problems for multiple areas of the review and analysis process. An example will help point out the general approach used for this Conference. The example will use the question regarding the use of systemic antibiotics as an adjunct for periodontal therapy. The very broad question derived from the PICO question was: Do systemically administered antibiotics improve a primary clinical outcome measure such as gains in clinical attachment level?

The broad nature of the question allowed inclusion of studies that evaluated the use of antibiotics adjacent to any therapy (or even no therapy), for any population and for any antibiotic.

- There were 29 studies that could be used for this purpose.
- Some of those studies had multiple active arms (e.g., with different antibiotics).
- Studies that had multiple active arms allowed multiple comparisons, thus 36 actual comparisons were available.
- Due to other excluding factors (e.g., insufficient measurements, follow-up period), 27 comparisons were selected for meta-analysis.
- These studies had:
  - Multiple patient populations (i.e., aggressive and chronic periodontitis) and multiple mechanical therapies (i.e., none, root planing and scaling, and surgical therapy)
  - Multiple antibiotics (e.g., doxycycline, tetracycline, clindamycin, spiramycin, penicillin, amoxicillin plus clavulanate, metronidazole, and metronidazole plus amoxicillin)
  - Almost no study employed the same regimenation and/or dose of antibiotics.
- If strict criteria were used, very little information could be gleaned due to the variation in studies.

However, the approach taken was to enable an evaluation of the overall question, “Do systemically administered antibiotics improve a primary clinical outcome measure such as gains in periodontal clinical attachment level?” The answer to that question was positive with 23 of the comparisons favoring the antibiotic groups. While this is certainly an interesting answer, the broad question does not provide clinicians specific guidance for the clinical situations in which antibiotics are useful.

For that reason, subgroup analysis of the data was done. In the case of systemic antibiotics, the narrower questions asked were:

- Did the patient population matter?
- Did the mechanical therapy matter?
- Did the type of antibiotic matter?

There were insufficient data to answer questions about the optimal antibiotic regimen. However, the systematic review by Haffajee and Socransky and the related Consensus Report provided clinically useful information on the subject.

**Analysis Methods**

The meta-analysis used for this conference utilized a software program. The majority of the reports had outcomes that were continuous variables; i.e., clinical attachment level, probing depth, and bone level.

The data were analyzed using a standardized difference as described by Fleiss. The results were checked with both a fixed effects model and a random effects model and the results were consistent. Both Cohen's d (unadjusted) and Hedges's g (adjusted) were used to test for heterogeneity. Both tests had to be non-significant to support the lack of heterogeneity.

There were a number of the reports that utilized a dichotomous response (such as Yes, No). For these reports, the data were analyzed using a method that was first described by Mantel and Haenszel. This

§ Comprehensive Meta-Analysis, version 1.025, Biosoft, Englewood, NJ.
method was later used for a wider class of problems and has been adapted for meta-analysis. The results were confirmed by Peto’s method for combining odds ratio.

**The Forest Plot (Figure 2)**

For each meta-analysis a forest plot is provided that graphically demonstrates the results of the meta-analysis. The majority of the studies had effects that were continuous variables and the forest plot described below is for continuous variables. The forest plot represents the difference of effect between a control arm of a study and the active arm of the study. A horizontal line with a short vertical line crossing it mid-point represents this difference for each study. The length of the line represents a 95% confidence interval (CI) for the study results and the short line represents the normalized mean effect.

It is important to understand that the normalized mean effect is the mean difference divided by the standard deviation of the difference. Thus, the “real” mean is adjusted by the variability of the study. The plots generated for the meta-analyses show a summary for each subgroup of studies and for the composite of all of the studies.

There is a long vertical line on the forest plot. That line represents a zero effect. Thus, if the mean effect is to the right of the line, then the study has an effect that favors the active arm. If the 95% CI does not cross the zero vertical line, then the effect is significant.

When evaluating a forest plot, the observer should look at the entire graphical picture, looking for a significant effect, but, more importantly, looking to determine the consistency of the effects. The power of the forest plot is in its graphic representation and overview of the data.

**Analysis Assumptions**

In order to use the above analysis and to include as many studies as possible, a number of assumptions had to be employed. Since the majority of the outcomes in systematic literature reviews were continuous variables, the summary of those studies were means. The following discussion of assumptions is based on continuous data. Our general approach was to use conservative assumptions, to avoid inflating the statistical power of the meta-analysis when assumptions had to be made.

1. Some of the studies in some of the reports had more than 1 active arm (for a total of 3 or more arms). In these cases, the control data were used for each active arm and both arms were included. Each active arm with the control group was considered as a separate study for analysis purposes.

2. The analytical methods used for this conference required a normalized difference between study arms (active–control). This is a difference between treatments divided by the standard deviation of the difference. In determining the standard deviation of the difference, a number of compromises had to be made.

Some studies do not report a standard deviation (SD), but a standard error (SE). The SEs in these cases were converted to SDs by multiplying by the square root of N. In a number of studies the authors report very small standard errors. The small SEs were not due to better measurement techniques, but were based on measuring each site of interest multiple times. In these cases we converted the SE back to a SD by using an N that reflected what we believed the author did. If we had any doubt, we used the number of readings which would result in a conservative assumption for the standard deviation.

Many studies report an initial and final measurement, but not a change over time in an outcome measurement. Other studies report a change over time, but not initial and final measurements. We developed a change score for each group if it did not exist from the initial and final readings. We then had to determine a standard deviation for the change. The equation for the SD of a difference is the square root of the sum of the variances of the measures, minus the covariance of the measurements. In calculating the standard deviation of change, we used a conservative estimate by ignoring the covariance term, since we did not have a measurement of the covariance between the measurements.

In the regeneration studies and bone grafting studies, there was frequent use of multiple defects in each patient’s mouth and split-mouth designs in which each patient received both therapies (control and active). For sample size purposes, we always used the patient, not the number of defects. Additionally, in determining the standard deviation of difference between groups we again used the conservative assumption that the covariance between therapies was zero.

In a number of systematic reviews there was interest in comparing results from cases series, since there were not a sufficient number of randomized clinical trials to evaluate the PICO questions. Additionally, even...
randomized studies had effects that influenced both treatment groups, such as different recall schedules, patient populations, and surgical technique.

In order to evaluate these issues, we did some exploratory analyses of the data using the number of studies as the sample size, which made the study the unit of analysis. We then used analysis of variance to evaluate effect between studies that may have had influences on the outcomes of those studies. It should be noted that any findings with this type of analysis are associations. Similar to epidemiologic studies in which, at best, associations can be determined, more rigorous studies (randomized clinical trials) are needed to verify these findings. While some would argue that we should have not included these types of analyses in our reports, we believe that argument is outweighed by the information provided, permitting a stronger evaluation of some of these issues than has been done by traditional review articles.

**CLINICAL VERSUS STATISTICAL SIGNIFICANCE**

No proceedings of a consensus conference on clinical topics would be complete without some comment on the issue of clinical significance. Defining and measuring it have been elusive and troublesome for many. However, recent literature on the subject offers new insights about how to think about clinical significance.

The answer sought in evaluating data from clinical studies is to know if the effect produced by the new intervention is relevant and big enough to be important scientifically, clinically, and practically.

For clinicians this is a value judgment and it is made in conjunction with the patient’s preferences and the context of the clinical treatment. In other words, deciding if something is clinically significant is an individual decision that is based on individual circumstances. Clinical significance can be defined as the smallest worthwhile effect for the variable in question in the context of the individual patient’s circumstances.

The key obstacle to accepting this working definition is that, unlike a yes/no decision based on some arbitrary $P$ value, it requires the decision-maker to give genuine thought and intellectual rigor to determine the smallest worthwhile effect for the variable in question.\(^{28,29}\)

Statistical significance is defined by a probability or $P$ value. The smaller the $P$ value, the less likely the effect was due to chance. One problem of testing the classical null hypothesis, as is done in most clinical trials, is that even for the most minute true difference between 2 treatments, the $P$ value will indicate statistical significance if the sample size is big enough.\(^{28}\)

The determination of accepting or rejecting treatment is context-specific for each patient. Context is decisive when selecting treatments and making decisions.\(^{30}\) Hopkins has provided an example that illustrates this point:

You have analyzed the data in a new manner that takes into account clinical or practical significance. Here is the outcome of the analysis for the average person in the population you studied: an 80% chance the effect is clinically beneficial, a 15% chance that it has only a clinically trivial effect, and a 5% chance that it is clinically harmful. The effect has a good chance of helping people. Indeed, it has 16 times more chance of helping than of harming. If you think that the 80% chance of helping is too low or that the 5% risk of harming is too high (it will depend on the nature of the help and harm), you could get more data before you publish. So what is the $P$ value for the above data? Incredibly, it’s 0.20.\(^{28}\)

Statistical significance data derived from a clinical trial as the sole criterion for deciding if a product or treatment method is useful may eliminate some clinically valuable interventions from being considered. It is often not possible to mimic specific clinical problems in a clinical trial, thus a clinician may be presented with a combination of patient characteristics that cannot be studied. Periodontists often have to determine that a particular patient will likely benefit from: occlusal adjustment, gingival augmentation, chemical root modification, distal wedge surgery, use of surgical dressing, certain suture techniques, ultrasonic versus hand instrumentation, and other procedures even though they cannot cite strong evidence that justifies high probabilities of success.

The $P$ value is one piece of information. The precision of estimation using confidence intervals for the true value of the effect is another important consideration for determining clinical significance. However, the classical 95% confidence limits may give an impression of too much uncertainty for some clinically useful effects. Even 90% might be too conservative in this respect, but there is something appealing about limits that define the true value correctly 9 times out of 10 times.\(^{28}\) This does not mean that 9 of 10 patients will achieve the desired outcome; it means that the confidence interval would contain the true value in 9 of 10 trials.

Hopkins\(^{28}\) has devised an excellent table (Table 6) for assigning plain language descriptions to probability levels typically encountered in reading articles about therapy. “It is important to emphasize that the probabilities refer to the value for the mean effect in the population. They do NOT in general refer to the probability that the effect will benefit or harm an individual. They will do that only if there are no substantial individual responses to the treatment in question. If there ARE individual responses, chances of benefit or harm for the individual get closer to 50%. In the extreme case of massive normally distributed individual responses—that is, a huge standard deviation representing individual responses, much bigger than the mean and much
Evidence-Based Approach
Volume 8 • Number 1 • December 2003

Between 15% and 20% adoption. Once the product
the ascending portion of the adoption curve, often
appear to acquire their own momentum somewhere on
ficult to stop a change from spreading further. Changes
vation has a “tipping point,” after which it becomes dif-

Understanding the dynamics and critical issues around
sounds good, but the reality is that a new concept or new
treatment decisions automatic. In health care, inven-
available for application to practice does not make
management process. Having the best quality evidence
helps to support the patient

FROM WORKSHOP TO PRACTICE
Good quality information helps to support the patient
management process. Having the best quality evidence
available for application to practice does not make
treatment decisions automatic. In health care, inven-
straitforward to say that “full integration
of the Conference findings enhances the likelihood for
optimal clinical outcomes and better quality of life.” That
sounds good, but the reality is that a new concept or new
treatment is slow to reach its destination, the patient.
Understanding the dynamics and critical issues around
the barriers to diffusion and acceptance of new concepts
is a first step towards minimizing the discrepancy
between current information and clinical practice.

The curve that describes the dissemination of innovation
has a “tipping point,” after which it becomes dif-
cult to stop a change from spreading further. Changes
appear to acquire their own momentum somewhere on
the ascending portion of the adoption curve, often
between 15% and 20% adoption. Once the product
or treatment reaches that level, it usually becomes inte-
grated as usual and customary in clinical practice.

In his widely quoted article on the diffusion of innov-
ation, Berwick asks: “Why is the gap between knowl-
edge and practice so large? Why do clinical care
systems not incorporate the findings of clinical science
or copy “best known” practices reliably, quickly, and
even gratefully into their daily work simply as a matter
of course?28

To answer those questions it is necessary to under-
stand the major areas that influence the rate of spread
of a change. If the reader wishes to get maximum value
from the proceedings of this Conference, understanding
what comprises the barriers and influencers of adopting
new technology will likely be very helpful.

Perceptions of the Innovation
Perceptions predict between 49% and 87% of the vari-
ance in the rate of spread of a new idea or product.32
Five characteristics of perception seem to be most
influential.

1. The perceived benefit of the change is the most
powerful characteristic. Individuals are more likely to
adopt an innovation if they think it can help them. This
is a more complicated idea than it appears, because for
most people who accept or reject an innovation, bene-
fit is a relative matter—a matter of the balance between
risks and gains, and this is defined in many ways.28 A
common aversion to new technology is not knowing the
predictability of the intervention in individual patients.

2. Compatibility. In order to diffuse rapidly, an inno-
vation must be compatible with the values, beliefs, his-
tory, and current needs of individuals.

3. Complexity. The rate of diffusion is correlated to
the complexity of the proposed innovation. Generally,
simple innovations spread faster than complicated
ones. In a successful diffusion process, the original
innovation itself mutates and matures into many dif-
ferent but related innovations.28 A good example is in
the area of osseointegrated implants, where the first
implants were indicated primarily for mandibular ante-
rior areas in fully edentulous patients under hospital
operating room conditions.

4. Trialability (whether or not a proposed adopter
believes he or she can find a way to test the change on
a small scale without implementing it everywhere at
first). Big learning curves, such as learning to do com-
plex surgery, are impediments to the majority of practi-
tioners unless the benefits are perceived to be high. When
periodontists were beginning to routinely use implants,
the “price of entry” was very high. Surgical kits and armamentarium cost thousands of dollars and there was little opportunity for trialability. Implants could not be placed unless the kits were purchased.

5. Observability, the ease with which potential adopters can watch others try the change first.

The results of this Conference should positively influence perceptions and facilitate a more rapid integration of new practices into patient management protocols.

CONCLUSIONS

The EB approach used in the Conference provides readers with information that was derived from an explicit process, where objectivity and reduction of bias were stressed.

Different from traditional systematic reviews, the emphasis in this Workshop was on emerging technology and procedures. This placed greater emphasis on the reviewers to seek evidence from alternative sources such as researchers and companies.

Similarly, the consensus development placed significant importance on validating the products and procedures and determining what additional information should be developed as they evolve and mature.

Data and information management used in the preparation of systematic reviews and consensus development are important and complex. Understanding clinical significance and the factors associated with enhancing diffusion of the results are critical for maximizing the outcomes of the Conference.

REFERENCES


Correspondence: Dr. Michael G. Newman, Suite 303, 12218 Montana Ave., Los Angeles, CA 90049-7505. E-mail: mgnewman@aol.com.