Methods for Comparing the Results of Different Studies

Steven E. Eckert, DDS, MS[1]/Yong-Geun Choi, DDS, MPH, MPH[2]/Sreenivas Koka, DDS, MS, PhD[3]

The reader of oral and maxillofacial implant literature is challenged to be cognizant of the quality of clinical research data. The large variety of possible study designs utilized by clinical researchers requires the reader to have a fundamental awareness of the advantages, disadvantages, and limitations of commonly utilized study designs. This article aims to provide the reader with information to make an informed decision regarding the quality of a clinical research paper, so that he or she can judge whether the information presented in any given manuscript was obtained in a manner that truly minimizes bias, in the form of systematic or random errors, and whether it warrants serious consideration for clinical decision making. Special consideration is given to scientific literature pertaining to the use of oral and maxillofacial implants. In addition, the reader is presented with a method for placing any single manuscript within a "hierarchy of evidence" enabling an "estimate of confidence" in a particular therapy. By utilizing such methods to appraise the quality of research data, clinicians and patients will be better informed when making treatment-planning decisions. INT J ORAL MAXILLOFAC IMPLANTS 2003;18:697–705

Key words: *data quality, dental implants, evidence-based dentistry, evidence-based medicine, literature hierarchy, research design, review literature*

The practice of dentistry involves decision making on the part of the clinician and of the patient. From an ethical and a legal standpoint, a patient is required to provide informed consent before any treatment is rendered, but it is difficult to define the volume and quality of information that should be presented to appropriately inform the patient. Since the practitioner must make decisions regarding materials, techniques, procedures, and treatment options to present information to the patient in the form of treatment options and consequences, clearly the dentist must be knowledgeable about the credibility of this information. Therefore, the practitioner is challenged to be aware of and to understand a myriad of alternatives before treatment planning can proceed.

How does the clinician gain necessary knowledge? How does the clinician determine what knowledge is credible? In the past, this may have been a process of passing on information from one generation to the next, with knowledge of techniques predominating over scientific principles. Today, dental science demands the need for evidence to assist practitioners in decision making; in an era of evidence-based dentistry, the clinician is constantly faced with the prospect of determining which path of evidence should be followed.

Unfortunately, not all evidence is of equal value to the clinician. Some forms of evidence, although apparently compelling, are based on opinion rather than on sound scientific principles. Conversely, the best forms of evidence are those that are applicable to the general dental population and have been validated through comparisons with other techniques, materials, or procedures. Although it is clear that this statement is true, it is also obvious, through scrutiny of the dental literature, that controlled

[1]Associate Professor, Department of Dental Specialties, Mayo Clinic, Rochester, Minnesota.
[2]Director, Department of Restorative Dentistry, Bundang Jesaeng Hospital, Bundang City, Kyunggi do, South Korea.
[3]Associate Professor, Department of Oral Biology, University of Nebraska Medical Center, College of Dentistry, Lincoln, Nebraska.

Reprint requests: Dr Steven Eckert, Mayo Clinic, 200 First Street SW, Rochester, MN 55905. Fax: +507-284-8082. E-mail: seeckert@mayo.edu

clinical trials, which compare different treatment approaches, are either largely unavailable or are of dubious quality. Consequently, the clinician is forced to compare information on products, materials, or devices through scrutiny of published material that by itself does not compare these products, materials, or devices. Said another way, the clinician must make qualitative decisions without comparative studies to establish the qualitative differences.

The manner in which the clinician evaluates different studies may set the path toward appropriate or inappropriate clinical decision making. This article will describe some important areas of concern related to comparing studies and will suggest methods that can be used to determine an "estimate of confidence" in the final assessment.

FACTORS RELATED TO STUDY DESIGN

Efficacy Versus Effectiveness

The terms *efficacy* and *effectiveness* are often used interchangeably, but this should not be the case. Studies that demonstrate *efficacy* are those that show that a treatment "works"; however, these studies often have a number of inclusion criteria that limit the reader's ability to generalize results. Efficacy is generally thought of as an idealized result following a trial in which many variables are controlled. Efficacy studies contain specific inclusion and exclusion criteria that are applied to the patient population before the patients enter into the treatment. Likewise, defined evaluation periods and defined interim treatments may be required. This is in contrast to an *effectiveness* study, in which the treatment is offered to the general population by an uncontrolled group of clinicians under broad guidelines with less specific follow-up care. In some practices, the terms may be indistinguishable, but in most settings there are significant differences. In general, the efficacy study produces the best results possible for a given therapy, while clinical effectiveness, since it applies to a broader group of patients, some of whom may possess unfavorable prognostic traits or behaviors, rarely reaches a similar level of clinical success.

The results of a randomized clinical trial can be analyzed and presented in 2 ways: according to the treatment to which the patients were randomized or according to the treatment they actually received.[1] The planned treatment of the former demonstrates the efficacy of a treatment, while the pragmatic latter establishes its effectiveness. The explanatory analysis is based on the principle that avoidance of bias is a prime goal in the randomized clinical trial, that randomization is the prime mechanism for avoiding bias, and that bias can arise from any post-randomization made by patients or by investigators.[2] More simply, the effectiveness study may be more prone to bias. The disadvantage is that anything that happens to a patient after a treatment has been randomly assigned is thereafter included in the events ascribed to that treatment, regardless of why and how the events occurred.[3] Most randomized clinical trials adopt the explanatory analysis. Therefore, clinicians should check the method of data analysis used in the literature when they read the results of a randomized clinical trial.

As a clinician reviews articles describing results of various treatments, it is imperative that the nature of the report be made clear. Two different reports using the same products and techniques can give distinctly different results if one takes place in a controlled environment, such as a dental school or research facility, and the other is performed in the private practice sector. The former is more likely to describe efficacy, while the latter is more likely to describe effectiveness. Differing results would be anticipated, but the clinician/reader, in an attempt to compare the data, may find the differences in results to be contradictory unless the study setting is fully appreciated. More problematic yet is the comparison of 2 slightly different products in the 2 different settings. In that situation it may be difficult or impossible for the reader to compare the results of the 2 reports. Therefore, a review of literature that included studies that differ in these ways would find comparison difficult or impossible.

In reality, the clinical effectiveness study may be more pertinent to the private clinician than the efficacy study. The exception occurs if the private clinician practices in the same way as described in the efficacy study. However, even then there could be a difference relative to the study population.

Differences in Study Populations

Study populations can differ greatly. Obvious factors such as age, gender, and ethnicity may be coupled with less obvious factors related to cultural differences, dietary differences, and environmental factors that might lead to remarkable differences in patient response to treatment. The clinician must evaluate the stated demographic information to determine its relevance to that of his or her patient population. Even if there is apparent similarity, subtle differences could account for significant differences in results.

Furthermore, the way in which one patient group is maintained could differ from that of another group of patients. For example, if a study includes frequent dental hygiene visits, the patients

in that study may gain a significant therapeutic effect from this treatment. If the article describing the study fails to disclose the frequency of these hygiene visits, a clinician reading the report may incorrectly interpret results as being applicable to clinical settings in which such frequent recall appointments are not provided. Conversely, a different study without any ongoing maintenance program may produce unsatisfactory results. Often, articles appear with statements describing "routine follow-up" that do not define this follow-up procedure. Consequently, the reader cannot determine whether this is similar or dissimilar to his/her own practice.

Prospective Versus Retrospective Design

Studies may be conducted *prospectively* or *retrospectively*. Although a well-designed prospective study generally possesses more internal validity than a well-designed retrospective study, blanket statements regarding validity, without consideration to study design, can be misleading. Neither study design is consistently superior to the other; each has distinct advantages and disadvantages.

The most distinctive difference between prospective and retrospective studies is that retrospective studies are more prone to bias—typically, recall bias. Since treatments and outcomes have already occurred in retrospective studies, susceptibility to bias, both in assessment of treatments and outcomes, is higher than for prospective studies.[4] Therefore, clinicians should ascertain whether the potential biases are well controlled when reading about studies that were retrospective.

Prospective studies generally have inclusion and exclusion criteria. In some instances, prospective studies may be designed to eliminate known risk factors that could contribute to higher failure rates. Well-designed prospective studies consider the anticipated differences in treatment outcomes when the number of subjects for enrollment is considered; this is generally described as a *power analysis*. Most prospective studies are analogous to efficacy studies described above. When a variable is studied prospectively, there is a comprehensive list of factors that must be evaluated. Failure to recognize factors that should have been evaluated means that these factors will not be recorded, assessed, or accounted for during statistical analysis and/or data interpretation. Likewise, follow-up procedures must be described and rigorously completed. Patients who fail to comply may be termed protocol failures and could then be eliminated from the study.

Retrospective studies are less likely to consistently follow every factor for every patient. These studies are quite likely to record major adverse events. Complications are normally well documented in retrospective studies, and as risk factors toward failure become evident, these too are recorded. Retrospective studies are dependent upon the quality of information recorded by the clinicians involved in the study. Since different clinicians may practice differently, the retrospective study may demonstrate a larger variability of techniques or materials. In some instances, retrospective studies will be initiated once clinical problems are encountered. These studies may be performed to investigate causality of a specific outcome. Many times, it is through retrospective studies that adverse outcomes are documented; this can result in dramatic changes in treatment modalities.

Data Analysis

With any study there is a need for data analysis. Whether data are followed forward (prospective study) or backward (retrospective study), they must be collated and evaluated in such a way as to make them meaningful. Raw data can be easily distorted to demonstrate results that do not adequately describe clinical performance. Statistical analysis of data, particularly time-dependent analysis, is more valuable if there is a risk of time-dependent deterioration, complication, or failure. Commonly used forms of time-dependent analysis in clinical research include survival curves. A practitioner may want to have an estimate of the percentage of subjects who will survive at a given time point. The estimate can be obtained from a survival curve.

Although there are a variety of statistical methods for generating survival curves, the Kaplan-Meier (KM) method is the most prevalent in the dental literature. For example, for a new heart transplant protocol, the method permits estimation of subject survival over time, even when subjects are lost to follow-up (eg, move away, no longer want to participate) or followed for different lengths of time. At each time point, eg, every year after the heart transplant surgery, the KM method typically involves computing the number of subjects that had died at each time point (numerator) divided by the total number of subjects that were still in the study at that time point (denominator). Subjects who are lost to follow-up are no longer included in the denominator from the first time point at which they were lost. It is important to realize that the KM method allows the practitioner to estimate the probability of a subject surviving at a given time point, eg, 5 years, from the cumulative probability of a subject surviving each preceding time interval (years 1, 2, 3, and 4). Therefore, although a probability calculated in a

given interval may not be accurate because of a relatively small number of subject deaths in that interval, the overall probability of surviving to each time point is more accurate.

In addition, with statistical analysis, the generation of *confidence intervals* (CIs) is useful because it can tell the reader how precise the estimate is, or in the case of heart transplant therapy, how precise the percentage survival estimate actually is. A commonly used CI is the 95% CI, which means the range over which there is a 95% chance that the estimate is true. Therefore, if a heart transplant protocol has a 91% survival rate with a 95% CI of 4%, there is a 95% chance that the real survival rate falls between 87% and 95%. A CI is dependent upon 2 key pieces of data: *(1)* how many observations (eg, subjects) were included in the study; and *(2)* the spread in the data (routinely measured as the standard deviation). A study with a large number of observations and a narrow data spread will likely have a small 95% CI, resulting in a survival estimate with a 95% chance of falling in a narrow range, which is consequently "more believable." Armed with an estimate and a 95% CI, the clinician is much better able to determine how powerful the data presented in a clinical research article actually are.

With statistical analysis, the clinician may determine whether a moderate difference in clinical outcome with a large number of enrolled subjects and a large number of patients lost to follow-up is meaningful. Without statistical analysis, the reader is apt to guess about the significance of results. When a procedure is repeated in 1 patient, the outcome of 1 treatment may be dependent on the outcome of another in that same patient. This is known as *clustering of results.* Whenever a study is performed, it is important for the reader to understand whether the statistical methodology controls for clustering of events or if this potential outcome has not been considered. Thus, is it the *patient* or the *procedure* that represents the number of subjects described in a study? If it is the latter, then the risk of clustering must be considered. Wei and coworkers[5] have suggested the establishment of a "robust *P* value" as a method to account for clustering of events; however, this approach has not been received with universal support. The reader must determine in the course of an article whether clustering was considered or ignored, a robust *P* value was used, or the patient or a specific device was considered the study subject, since each analytic method carries with it a different level of significance.

One risk with retrospective studies relates to study length. These studies tend to be conducted over longer periods of time. During those time periods there is a greater likelihood that minor changes in materials, techniques or designs can alter the therapy enough to create distinctly different treatment groups. With these changes, survival data may also change. If this is recognized, it is possible that the authors may describe specific "before and after" data, but if it is not recognized, significant improvements in design may not be apparent to the reader. Comparison of pooled data in such a study could create erroneous impressions when the reader compares the data to the results of other studies. For these reasons, the use of time-dependent statistical analysis may be even more critical to the establishment of meaningful conclusions with retrospective studies.

Prospective studies are generally more expensive to complete since the study must enroll specific patients and must exclude others. Data must be maintained. The dependence on personnel to complete these tasks is greater with prospective studies. These studies are generally performed over shorter time periods and often have closer patient follow-up. The ability to discover late complications or compliance related to complications with prospective studies is thereby compromised.

Study Duration and Intervals

When a reader compares studies, it is critical to compare similar time periods and time intervals. For example, a study describing time-dependent results over a 10-year interval could be compared to a similar study with 5-year results, if both studies are evaluated for the 5-year interval only. In this situation, the 10-year study data would probably be assessed at the end of the initial 5 years. This can be accomplished if all data are available in table form, or it may be possible if data are available in graphic form relative to time.

Also, if a study begins many years prior to a second study, it is distinctly possible that results of the 2 studies may differ simply because of changes in the standards of practice that occurred during the time period covered by the 2 studies. When 2 studies are conducted in 2 distinctly different time periods, even if the duration of the studies is similar, distortion of results is possible. Consider an article describing 5-year results written on information gathered in the mid-1980s, and contrast this with an article describing 5-year results with data gathered from 1997 to 2002. It is distinctly possible that the earlier study may represent a different generation or version of the device being tested and therefore may have little relevance to the more current study. Materials, devices, and techniques used in the 1980s study may not even be available for use a decade

and a half later. Therefore, comparisons that lead to conclusions about the 2 different approaches should not be made.

STATURE OF PUBLICATION/ STRINGENCY OF PEER REVIEW

Information is disseminated in many different ways. Lectures and continuing education courses provide valuable information, but the audience may not always hear what the speaker had to say. Speakers may present preliminary data without statistical analysis or may even present opinions that could have been formed without any data analysis. Written information is generally more reliable than oral presentations, since it can be re-assessed as needed. Journal editorial policies differ, with some journals being quite stringent and others being relatively lax. Peer review alone does not ensure that material coming to press is technically accurate, but it does demonstrate a level of scrutiny that is external to that provided by the authors alone. Journals that are highly dependent upon industrial advertisements may be more liable to bias, since publication of an article that is unfavorable to a specific advertiser could result in loss of that specific sponsor. This is not to say that advertisements are undesirable, since they do defray the cost of publication, but a journal that is highly dependent on such advertisements may also be subject to an editorial bias. Most, but not all, scientific journals accept advertisements that make substantiated claims and publish those advertisements in sections of the journal that are distinct from the scientific literature. Some journals exert editorial control over their advertisers, while others include unedited advertisements in such close proximity to specific papers that they are virtually indistinguishable from the article itself.

IMPLANT-SPECIFIC LITERATURE

There are a number of factors that are specific to the literature covering dental implant usage. Prior to 1980, for the vast majority of early articles, implant survival was the primary outcome measure. Most of these articles described implant survival or failure but did not distinguish between the number of implants that failed to achieve osseointegration and the number of patients who experienced failure of an implant or implants. These articles did not provide information relative to clustering of failures. Implants were treated as independent variables, even though 1 patient may have received more than 1 implant. In such a situation, the implants do not perform as totally independent variables. Instead, they are dependent upon a number of patient factors such as bone quality and quantity, patient health and habits, and other factors that could impact implant survival. It is important that the reader be able to distinguish between articles that describe implant failure alone, those that describe methodology to account for event clustering, and those that describe the percentage of patients who have experienced implant loss. Results will vary depending on the method of data analysis and presentation; it is difficult to directly compare articles using different methods.

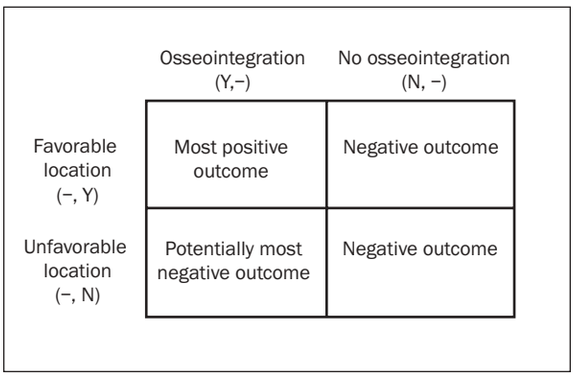

Fig 1 Effect of implant integration status and location on treatment outcome.

In essence, an implant has few immediate outcomes of interest. The implant will either achieve osseointegration or it will not, and the implant will either be placed in a favorable location or it will be placed in an unfavorable location. These events may be assessed in a table (Fig 1). Clearly, the most favorable initial result occurs when an implant achieves osseointegration and is placed in a location that is favorable toward the planned definitive restoration. Failure to achieve osseointegration, regardless of whether the implant was favorably or unfavorably located, carries a similar consequence: the treatment plan must be altered to another prosthetic intervention, or a replacement implant must be considered. Perhaps the most perplexing situation occurs when an implant achieves osseointegration but is placed in a position that is unfavorable for the definitive prosthesis. In some instances this will require compromise of optimal prosthesis design. The most severe consequence occurs when an implant achieves integration but is placed in a position that prevents restoration. In this situation, the only courses of action involve implant disuse or surgical removal of the implant, a condition that removes more bone and potentially creates a larger defect than had been initially encountered (Fig 1).

Table 1 Classification of Implant Complications

Complication severity/type	Correction
Minor	
Prosthetic screw loosening or abutment screw loosening	If prosthesis is screw retained, correction requires screw tightening.
Prosthetic screw fracture or abutment screw fracture	If prosthesis is screw retained, correction requires screw retrieval and replacement.
Moderate	
Abutment screw loosening	If the prosthesis is cement retained, there is no access to the abutment screw. Access must be provided, the screw tightened, and the access opening in the prosthesis repaired.
Material fracture	Prosthesis must be retrieved and repaired. Retrieval of screw-retained prosthesis is accomplished more predictably than retreival of a cement-retained prosthesis.
Implant failure	If implant is nonessential to the integrity of the prosthesis, then simple removal of the failed implant is needed.
Severe	
Abutment screw loosening	If a prosthesis is retained by connection to an abutment that has no re-orientation capacity, the restoration must be replaced in the event of abutment screw loosening, since the abutment cannot be accurately re-oriented to the original position.
Implant failure	If the implant is essential to integrity of prosthesis then it must be removed and replaced prior to refabrication of new prosthesis.
Implant fracture	Implant remnant must be surgically removed.

Since implant-supported restorations are thought to be long-term prosthetic solutions for complete and partial edentulism, it is appropriate to assess the clinical performance of these prostheses over long time periods. The most favorable long-term outcome occurs when an implant maintains osseointegration and provides support, retention, and stability for an optimally designed dental prosthesis. Therefore, clinical parameters for prosthesis success should be applied to studies that report implant-supported rehabilitation. Complications need to be classified as those that can be resolved with minor, moderate, or severe intervention (Table 1).

Patient compliance is an important factor that determines the success of any medical or dental therapy. Therefore, with all other factors being equal, implant therapy that demands less compliance is more favorable to therapy that demands a higher level of patient compliance. Compliance may take the form of daily oral hygiene, frequent dental prophylaxis, periodic retentive maintenance, or other procedures that demand ongoing patient cooperation. Studies should provide sufficient detail regarding ongoing recall procedures to allow the reader to evaluate similarity to their normal clinical practices.

Implant failure is always an adverse event for the patient, surgeon, and restoring clinician, but some failures may be more challenging than others. Failure to achieve osseointegration, if discovered before prosthesis fabrication, is the least unfavorable of the failure scenarios, because it results in minimum biologic toll in terms of lost bone. Beyond these factors it is also the most likely failure pattern to be treated by the possible placement of a larger implant immediately upon its discovery. Failure after prosthesis fabrication may require implant removal, implant replacement, and fabrication of a new prosthesis. These treatments increase the economic impact of the failure. Failure related to implant fracture is the most unfavorable failure pattern in that it demands a surgical procedure for retrieval of the fractured implant followed by a protracted healing period prior to implant replacement and prosthesis re-fabrication. When a clinician is considering implant survival studies, it is therefore critical to understand the time of implant failure. For example, 2 studies that utilize pooled data demonstrate similar survival rates for different implant designs. However, one design achieves this result while manifesting predominantly an early failure pattern, and the other demonstrates a late failure pattern. Therefore, a better way to demonstrate these data is to show implant survival plotted against time on a survival curve. This method of data presentation will allow the clinician to better predict the long-term costs of implant care, a factor that may be critical to the decision-making process.

Large numbers of enrolled subjects in long-term studies provide comforting data when the performance is adequate. However, rigid adherence to the need for large, long-term studies may diminish the clinician's willingness to read critical information from short-term studies that report adverse events. By means of statistical analyses, it is possible in these short-term studies to establish CIs that allow a reader to determine whether these adverse events were caused by chance or if they were the result of other factors.

Assessment of the implant dentistry literature demonstrates that few randomized controlled clinical trials (RCT) have been performed. Esposito and coworkers, in a review of these studies, concluded that the quality of RCTs in implant dentistry is poor and needs to be improved.[6] The authors described errors in blinding, randomization and concealment allocation, and reporting of patient withdrawals.

Another issue that is often ignored is sample size. Because implant survival rates are generally high, sample sizes need to be large to demonstrate statistically significant differences for meaningful clinical differences in implant survival performance. However, prior consultation with an experienced statistician would permit a clinician researcher to predict the number of implants that need to be included in a study for it to provide data that can be analyzed in a statistically meaningful manner (a power analysis). For example, suppose a clinician is designing a study in which the ability to detect a difference in implant survival from 85% to 95% at the end of the study, with a power, $1 - \beta$ and $\alpha = .05$ is desired. With appropriate statistical calculations, it can be determined that the clinician researcher needs to include 135 implants in each arm of the study, for a total sample size of 270 implants.[7]

Implant literature has traditionally presented case series and case reports. Few direct comparisons of different implant designs are available, and these studies rarely utilize randomization for patient assignment. Within the last 5 to 10 years, an increasing number of studies have performed time-dependent statistical analysis of data. Some of these may be considered to be nonrandomized cohort studies, since implant performance is compared within the patient population on the basis of anatomic location, prosthetic design, or other factors critical to clinical performance. As evidence hierarchies are developed, it is important to consider how implant literature is placed within the hierarchy. This is crucial so as to understand the quality of the evidence presented in any single article, and to establish an estimate of confidence in evidence-based decisions.

METHOD TO ESTABLISH AN EVIDENCE HIERARCHY

It is clear that a review of the dental literature, even that describing data obtained as a result of excellent experimental design, is fraught with potential problems. Somehow, the clinician must establish a method for review that helps him or her gain confidence when comparing the results of different studies. Because of the volume of literature that is published, it is not always possible to wait for a systematic literature review aimed at demonstration of scientific "truth." The prudent clinician must be capable of judging literature and placing it into a useful hierarchy.

In the medical literature there are numerous articles that present a hierarchy of evidence. Virtually all of these examples describe a spectrum of evidence, with the most reliable information derived from RCTs and the least reliable information derived from expert opinion. The hierarchy itself does not indicate that expert opinion is bad and that controlled clinical trials are good; it simply places an estimate of confidence around the different sources of information.

One approach to stratifying clinical trials is presented on the Internet by the Centre for Evidence-Based Medicine of the University Department of Psychiatry at Warneford Hospital in Oxford, United Kingdom (Table 2).[8] This system assigns different levels to different types of studies, with the lowest numbered levels associated with the studies that invoke the greatest level of reader confidence. In most situations, single studies will not provide enough definitive information to allow a clinician to clearly choose one definitive therapeutic path. This may occur because of study differences, as described previously in this article. Therefore, the reader will need to compare numerous papers that may not agree. Ultimately, the reader must compile the information and form an opinion—in essence an "estimate of confidence" in this opinion—eg, "very confident," "reasonably confident," or "unsure." Again, the Centre for Evidence-Based Medicine provides a basis for grading confidence in reviews of papers (Fig 2).[8] Through the use of these 2 methods, evaluation of the level of the article, and determination of a grade of recommendation regarding a specific therapy, a clinician can estimate confidence in a treatment plan and be better prepared to provide patients with the necessary amount and appropriate quality of information.

Clinicians should develop 2 capabilities when defining literature according to the suggested hierarchy. First, they should be able to define the type of

Table 2 Classification of Evidence Level For Different Types of Clincal Research Studies

Level of evidence	Type of study	Main features	Potential problems
1a	Systematic review of randomized controlled trials that demonstrate consistent results	Assessment of the validity of published randomized controlled trials regarding the thoroughness of controlling random errors and systematic errors in research methods; sampling, data collection, and data analysis	Dependent on the capability of reviewers and publication bias
1b	Individual randomized controlled trials (with narrow confidence interval)	An experiment in which subjects are randomly allocated into either study or control groups to receive or not to receive an experimental preventive or therapeutic procedure	Transfer bias if substantial (> 20%) number of subjects is lost to follow-up
2a	Systematic review of cohort studies that demonstrate consistent results	Assessment of the validity of publised cohort studies regarding the thoroughness of controlling random errors and systematic errors in research methods; sampling, data collection, and data analysis	Dependent on the capability of reviewers and publication bias
2b	Individual cohort study	Observations of subjects with different exposure levels by no random allocation over a long period for the comparison of incidence rate of outcome	Transfer bias related to many patients (> 20%) lost to follow-up; confounding related to incomplete control of the effect of prognostic baseline characteristics
3a	Systematic review of case-control studies that demonstrate consistent results	Assessment of the validity of published case-control studies regarding the thoroughness of controlling random errors and systematic errors in research methods; sampling, data collection, and data analysis	Dependent on the capability of reviewers and publication bias
3b	Individual case-control study	Comparison of case (subjects having outcome) and control (subjects not having outcome) with regard to the level of the exposure	Recall bias related to differences in accuracy of recall to memory of past exposures; transfer bias related to many (> 20%) lost to follow-up; confounding related to incomplete control of the effect of prognostic baseline characteristics
3c	Cross-sectional studies	Measurement of 2 variables conducted at one particular time, with no follow-up	The temporal sequence of the 2 variables cannot necessarily be determined; transfer bias related to impossibility to discern the number of lost subjects
3d	Ecologic studies	Study units are groups of people rather than individuals	Ecologic fallacy because an association observed between variables on an aggregated level does not necessarily represent the association that exists at an individual level
4	Case report, case series	Description of the experience of a single patient or a group of patients; neither internal nor external comparison groups	Random sampling error, confounding related to lack of comparison group
5	Expert opinion without explicit critial appraisal, or base don physiology, bench research or "first	No supporting data from organized research provided, but opinion from personal experiences or belief asserted	No control over systematic and random error in samping, data collection, and data analysis

Levels of evidence adapted from Centre for Evidence-Based Medicine, University Department of Psychiatry, Warneford Hospital, Oxford, United Kingdom (http://www.cebm.net).[8]

study design by evaluation of the materials and methods of the article, rather than depending entirely upon the author's description. For example, a case series actually can be called a cohort study if there are cohorts within the case series that are compared. Likewise, without defined follow-up periods and study duration, a "prospective" study is inaccurately described as such. Second, assignment to a higher level of study design within the hierarchy does not guarantee superior study quality. RCTs, for example, can be worse than a case-control study, depending on the crucial control of the potential bias. A higher-hierarchy study type indicates only that there are more and easier opportunities to control bias than in a lower-hierarchy study type. If some researchers using the higher-hierarchy study type did not make the most of more opportunities, while other researchers using the lower-hierarchy study type did make the most of a smaller number of opportunities, the quality of lower-hierarchy study could be much better than that of higher-hierarchy study. The second capability—the ability to recognize bias in dental literature—has often been ignored in education regarding evidence-based dentistry. Therefore, clinicians should identify educational resources to help develop this capability.

CONCLUSIONS

Dental clinicians depend upon the availability of reliable and unbiased literature to assist them in making clinical treatment decisions. Comparison of scientific articles is remarkably challenging for a number of reasons outlined herein. This situation would be improved if all clinical reports were published using statistically analyzed data presented relative to time and, where appropriate, reported with CIs. In this way, after reading a group of articles relating to a particular clinical scenario, the practitioner can assign each article a level of confidence. Based on each article's assigned level and grade of recommendation for different therapies, the clinician then can determine an overall "estimate of confidence" for a given treatment plan. Consequently, both practitioner and patient are better prepared to make decisions that will increase the likelihood of successful clinical treatment.

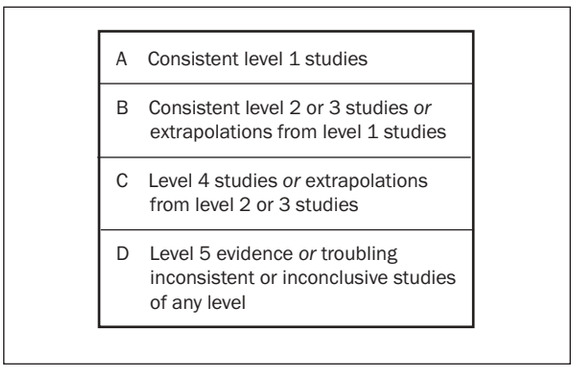

Fig 2 Grades of recommendation for a particular therapy (Adapted from Centre for Evidence-Based Medicine, University Department of Psychiatry, Warneford Hospital, Oxford, United Kingdom).[8]

REFERENCES

1. Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology: The Essentials, ed 3. Baltimore: Williams & Wilkins, 1996:151.
2. Feinstein AR. Clinical Epidemiology. Randomized clinical trial. The Architecture of Clinical Research, ed 1. Philadelphia: Saunders, 1985:683–718.
3. Feinstein AR. Clinical Epidemiology. Implementation of the outline: Outcome events. The Architecture of Clinical Research, ed 1. Philadelphia: Saunders, 1985:311–351.
4. Gordis L. Epidemiology, ed 1. Philadelphia: Saunders, 1996:166.
5. Wei LJ, Lin DY, Weisfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc 1989;84:1065–1073.
6. Esposito M, Coulthard P, Worthington HV, Jokstad A. Quality assessment of randomized controlled trials of oral implants. Int J Oral Maxillofac Implants 2001;16(6):783–792.
7. Glantz SA. Primer of Biostatistics, ed 5. New York: McGraw-Hill, 2002:412.
8. Centre for Evidence-Based Medicine, University Department of Psychiatry, Warneford Hospital, Oxford, United Kingdom (http://www.cebm.net).