Evaluation of the Literature: Evidence Assessment Tools for Clinicians

Giovanpaolo Pini Prato a, Umberto Pagliaro b, Jacopo Buti c, *, Roberto Rotundo d, and Michael G. Newman e

aDepartment of Surgery and Translational Medicine, University of Florence, Italy
bPrivate Practice, Florence, Italy
cSchool of Dentistry, The University of Manchester, UK
dPeriodontal Unit, Eastman Dental Institute, University College of London, UK
eSection of Periodontics, UCLA School of Dentistry, Los Angeles, CA, USA

ABSTRACT

The progressive improvement in the quality of scientific articles has led to an increase in difficulty in reading and interpreting them so that now clinical knowledge and experience must be complemented by methodological, statistical and computer skills. The aim of this article is to offer practitioners the tools, the simplest keys, that will allow them to understand and critically judge the results of scientific studies.

The “peer-review” process of a clinical article submitted to a journal is described and the Science Citation Index and the Impact Factor are presented to the reader as essential instruments to evaluate a specific article’s impact and the impact of a given journal on the scientific world, respectively. An article should be evaluated on the basis of some key issues which include, at least, an assessment of methodological aspects, a critical analysis of the statistical component and a proper understanding of the clinical impact of the study outcomes.

The standard approach for evaluating the quality of individual studies is based on a hierarchical grading system of research design which represents an essential tool to identify the strength of the evidence of an article. Many different biases may affect the reliability of study results. Randomized Control Trials (RCTs) and Systematic Reviews (SRs) are able to minimize the number of biases and thus are at the highest level of the scale of evidence representing the final steps of a treatment’s “career.”

Finally, moving from research to clinical practice, attention on the clinical impact of study’s outcomes is of paramount importance as the literature contains studies (including RCTs) that present statistically significant results but which, from the clinical standpoint, are only relatively or not at all significant. Clinical Practice Guidelines represent a useful tool for practitioners assisting...
the decision-making process when choosing the most appropriate treatment for their patients.

Keywords: Evidence-based dentistry, Grading System, Clinical Significance, EB Tools.

INTRODUCTION

The last years of the twentieth century witnessed extraordinary advances in scientific medical knowledge and an increasing number of publications on research and studies.

The rapidly increasing production of scientific data, the almost real-time availability of information online, the possibility for direct exchange during congresses, meetings and courses have fostered the creation and proliferation of almost countless tools for disseminating knowledge, including scientific journals and reviews.

The progressive improvement in the quality of scientific articles has led to an increased difficulty in reading and interpreting them, so that now clinical knowledge and experience must be complemented by methodological, statistical and computer skills. These issues have generated a gap between the increasingly high level of scientific research and the practitioners who, in many cases, cannot keep pace with the availability and assessment of the new scientific proposals. But, now more than ever, in this era of the philosophy of evidence-based medicine in which the patient’s requests, the operator’s skill and the appropriateness of the treatment must be supported by accredited scientific proof (evidence). The practitioner who has to decide on how to treat a patient is duty-bound to keep up-to-date with current advances in medical research has to use the essential tools from the deontological, ethical and forensic standpoints.

The aim of this article is to offer practitioners the tools, the simplest keys, that will allow them to understand and critically judge the results of scientific studies, to evaluate the effectiveness of one treatment approach with respect to another and, above all, to verify the results of sound and controlled studies against information from uncontrolled studies or data published in journals without any peer review process.

“Scientific knowledge is a body of statements of varying degrees of certainty – some most unsure, some nearly sure, none absolutely certain.”

Richard P. Feynman, (Nobel Prize for Physics).1

THE CLINICAL ARTICLE

Types of Articles

Evidence about clinical topics consists of the results of studies published in descriptive and scientific articles. Scientific articles may be reviews of the literature and original research. The reviews summarize the information obtained from previously published articles on a given topic. Narrative reviews are based on an arbitrary selection of articles on a given topic, which are evaluated and summarized without any predefined strategy. Systematic Reviews, on the other hand, call for articles selected on the basis of inclusion and exclusion criteria. The selected studies are then critically evaluated and the results summarized according to a predefined protocol. When it is possible to scientifically compare clinical data obtained from the review of the literature, a statistical analysis called meta-analysis is used.

Original articles are based on observations and experiments; they have a standard format (Introduction, Materials and Methods, Results, Discussion, Conclusion); they may be case report-series or controlled or randomized studies.

The Peer Review Process

An author who wants to publish a study must send the article to the journal’s editor who then submits it to reviewers often including a statistical reviewer. This process is known as “peer review” and its purpose is to assess the scope of the study, the correctness of the materials and methods used (internal validity), the congruity of the results as well as whether or not the article is pertinent to the journal and its readership (external validity). The article may be accepted by the reviewers as is, or returned to the author(s) with requests for modifications prior to publication, or it may be considered unsuitable for publication and rejected (Fig 1).

Some journals do not have a peer review process and therefore they have less scientific value. The existence of peer review is an important requisite for guaranteeing that a journal is included in the most important scientific literature databases. Indeed, one measure of scientific value is a judgement by the Institute for Scientific Information (ISI), which comprises committees of experts who enter and maintain in the databases only those journals that meet well-defined international standards.

Science Citation Index

Once a scientific article has been published in an international journal, it will be evaluated and perhaps cited by other authors. A specific article’s impact on the scientific world can be assessed by the Science Citation Index.
(SCI) that keeps track of the citations made over a given period of time.

**Impact Factor**

The huge number of journals has made it necessary to evaluate the impact of a given scientific journal on a specific field of interest. One system, The Impact Factor (IF), was created for this purpose. The IF is a number that indicates the frequency with which colleagues cite an article published in a given journal over a given period of time.

The IF of a journal (R) for any given year is calculated using the formula:

$$\text{IF} = \frac{n}{C_{14}}$$

where $n$ is the number of citations of journal R (international journal databank) over the 2 previous years and $C_{14}$ is the number of articles published in journal R over the 2 years preceding year $N$.

The success of the Impact Factor in the international scientific community has strengthened the belief that journals with a high IF publish the best articles or are the most selective in their choice of papers to publish. Therefore, the Impact Factor should not be used as a criterion for evaluating a scientific paper, but rather as an indicator of the standards of the journal in which the article is published.

**METHODOLOGIC CONSIDERATIONS USED IN THE DESIGN OF THE STUDY**

**Bias**

Any scientific study or research may contain errors that reduce its validity. There are three types of errors:

1. deliberate error (fraud);
2. accidental error;
3. and, the error known as bias.

Bias is “... any systematic error that results in inaccurate estimation of the effect of an exposure on an outcome.”

Bias can affect any phase of a study (Table 1).

**TABLE 1. The most frequent types of bias in scientific studies.**

<table>
<thead>
<tr>
<th>Study phase</th>
<th>Bias basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature review</td>
<td>Foreign language</td>
</tr>
<tr>
<td>Study design</td>
<td>Patient selection</td>
</tr>
<tr>
<td>Conduction of the</td>
<td>Learning curve</td>
</tr>
<tr>
<td>study</td>
<td></td>
</tr>
<tr>
<td>Data collection</td>
<td>Measurement error</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>Inappropriate statistics</td>
</tr>
<tr>
<td>Interpreting the</td>
<td>Significance</td>
</tr>
<tr>
<td>results</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecologic</td>
</tr>
<tr>
<td></td>
<td>Surrogate variables</td>
</tr>
<tr>
<td>Publication</td>
<td>Data “fishing expedition”</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Publication/journal</td>
</tr>
<tr>
<td></td>
<td>Conflict of interest</td>
</tr>
</tbody>
</table>

The following is a brief summary of some of the most frequent types of bias.

**Foreign Language Bias.** The authors of the study may make the error of not considering previous publications in less commonly spoken languages.

**Patient Selection Bias.** If the group of patients selected for the study is not sufficiently representative of the population with the disease to study. Even an error in randomization (for the correct distribution of the patients in the test and control groups) can lead to such a lack of homogeneity among the characteristics of the different groups as to invalidate the results of the study.

**Learning Curve Bias.** If it is a new surgical technique that is being tested, there are usually improvements in results as the clinician’s confidence with the procedure increases. If the groups are not randomized, or are incorrectly randomized, the results may be invalid, e.g. because all the control cases are treated at the beginning of the study.

**Measurement Bias and Blinding.** To avoid this error, it is essential that the person taking the measurements be a clinical expert, whose accuracy and reliability have been evaluated beforehand (calibrated) and that he/she is not apprised of the treatment administered to each person, i.e. that the study is blinded.

**Inappropriate Statistics Bias.** This can be found, for example when analyzing incorrect statistical units. One relatively frequent example occurs in implantology studies when a patient has more than one implant and the statistical unit is the implant rather than the patient.

**Significance Bias.** This can be found when the results have a statistical significance which has a negligible impact from the clinical standpoint (clinical significance).
Ecologic Bias. Ecological bias is when conclusions of studies proven true for populations are applied to individuals. In this regard, Nieri et al. have shown how conducting meta-analyses on aggregate data (means) rather on individual patient data can lead to error.

Surrogate Variable Bias. This occurs, for example, if only response-predictive variables (surrogate variables) are used in order to reach conclusions for long-term events. Studies on the effectiveness of periodontal treatment that use reduction of bleeding on probing as the variable instead of tooth loss over time due to periodontitis are an excellent example of this type of bias.

“Fishing Expedition” Bias. Sometimes researchers collect a large number of data for too many variables which they then analyze looking for statistical significance: this is known as “fishing expeditions.” To avoid this type of bias it is essential to establish and record the study protocol beforehand.

Publication Bias. The unpublished, or grey, literature contains more frequent negative results of the tested treatments than the published literature. This is due to the tendency among researchers and journals not to publish results that are negative with respect to the current trends in the literature.

Conflict of Interest Bias. Some corporate-sponsored studies may not be published if the results for the tested product are negative. In conclusion, no scientific study or research can be totally error-free. For this reason, the strength of a study depends on the authors’ ability to reduce biases to the minimum.

The Grading System for Study Quality

The standard approach for evaluating the quality of individual studies is based on a hierarchical Grading System of research design; it can be represented as a scale on which the different studies may be positioned according to an increasing quality of evidence. This has been discussed extensively in the literature but bears repeating for completeness in the context of this article. The higher the position of a trial on the scale, the greater the quality of evidence of the study, while the number and the relevance of biases are progressively reduced (Fig 2). The concepts of “control” and “randomization” make it possible to identify three levels of increasing quality in the Grading System:

Studies with low-value of scientific evidence are grouped at the lowest level (Table 2):

Expert Opinion. An eminent clinician, particularly expert in a specialty, expresses his/her opinions on a therapy based on personal experience and clinical knowledge. For instance, in 1986, Miller dealing with root coverage with the free gingival graft, stated “while some factors are based on accepted surgical principles, others are derived from clinical experience, and their significance… must be considered conjecture.”

Case Report. Case Report is defined as a “report on the treatment of a single patient.” It is a presentation of a new technique or a previously unpublished clinical case with relevant clinical impact. A case report may also describe a new biological principle or a rare condition. For example, Pasquinelli, in a paper entitled “The Histology of New Attachment Utilizing a Thick Autogenous Soft Tissue Graft in an Area of deep Recession: A Case Report,” concluded that “…placement of a thick autogenous gingival graft can result in regeneration of connective tissue attachment, cementum and bone on a root surface with long-standing recession.”

Case Series. Case Series, defined as “Collections of reports on the treatment of individual patients,” report a higher number of homogeneous and standardized cases which show a certain reproducibility of the results. For instance, Harris, in the paper entitled “The Connective Tissue with Partial Thickness Double Pedicle Graft: The Results of 100-consecutively Treated Defects,” stated that “…esthetic root coverage can be accomplished in a predictable manner.”

However, the above mentioned studies do not include any control group and no comparisons can be drawn. Moreover, in these studies, statistical analysis to test differences between test and control groups cannot be performed, conclusions cannot be generalized and several biases (publication, selection bias) may affect the study’s quality. Therefore the reader should carefully consider the low level of evidence of case reports and case series when transferring their results to clinical practice.

The intermediate level of the Grading System consists of studies where a test group is compared to a control group: Retrospective Controlled Trial and Prospective Controlled Trial (cohort) (Table 3).
**TABLE 2.** Studies with low-value of scientific evidence: positive and negative aspects.

<table>
<thead>
<tr>
<th>Type of study</th>
<th>Favorable aspects</th>
<th>Unfavorable aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert opinion</td>
<td>• Illustration of a new therapeutic technique;</td>
<td>• Lack of data useful for statistical analysis.</td>
</tr>
<tr>
<td></td>
<td>• Observation of a new phenomenon.</td>
<td></td>
</tr>
<tr>
<td>Case report</td>
<td>• Illustration of a new therapeutic technique;</td>
<td>• Lack of data useful for statistical analysis;</td>
</tr>
<tr>
<td></td>
<td>• Observation of a new phenomenon;</td>
<td>• Conclusions cannot be generalized;</td>
</tr>
<tr>
<td></td>
<td>• Verification of a biological principle.</td>
<td>• Patient selection bias;</td>
</tr>
<tr>
<td>Case series</td>
<td>• Greater number of homogeneous and standardized clinical cases;</td>
<td>• Publication bias.</td>
</tr>
<tr>
<td></td>
<td>• Reproducibility of results.</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 3.** Studies with intermediate-value of scientific evidence: positive and negative aspects.

<table>
<thead>
<tr>
<th>Type of study</th>
<th>Favorable aspects</th>
<th>Unfavorable aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrospective controlled trial</td>
<td>• Presence of comparison of different treated groups of population.</td>
<td>• Lack of randomized allocation of the patients to different groups;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Patient selection bias;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The researcher knows a priori the results of the therapies;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Spurious variables (i.e. smoke habits) can affect the treatment.</td>
</tr>
<tr>
<td>Prospective controlled trial</td>
<td>• Presence of comparison of different treated groups of population;</td>
<td>• Lack of randomized allocation of the patients to different groups;</td>
</tr>
<tr>
<td></td>
<td>• The researcher does not know a priori the results of the therapies.</td>
<td>• Spurious variables can affect the treatment.</td>
</tr>
</tbody>
</table>

**Retrospective Controlled Trial.** *Retrospective Controlled Trial* is a study in which “…patients affected by a specific pathology (or receiving a treatment or both) are compared to unaffected individuals.”11 For instance, Cordioli et al., in a paper entitled “Comparison of 2 Techniques of Subepithelial Connective Tissue Graft in the Treatment of Gingival Recessions,”14 concluded that “…the treatment outcome relative to the percentage of root coverage was comparable for the 2 techniques.”

**Prospective Controlled Trial.** *Prospective Controlled Trial (cohort)* is a “…Trial in which patients affected by a specific pathology (or receiving a treatment or both) are followed during time and compared to unaffected individuals.”11 For example, Pini Prato et al. in a paper entitled “Guided Tissue Regeneration versus Mucogingival Surgery in the Treatment of Human Buccal Gingival Recession”15 concluded “…The membrane procedure compared favorably with the mucogingival surgery in the treatment of deep recession.”

Even if the above mentioned studies have a control group and comparisons among groups can be performed, several biases can affect the strength of these studies. Selection bias (inclusion of patients used for other studies), the “a priori” knowledge of the results in the retrospective trials and lack of randomization of the groups may affect the quality of the study.

The highest level of evidence consists of controlled studies where patients are randomly allocated to groups: *Randomized Controlled Trials (Individual/Multicenter). Systematic Reviews of RCTs and Clinical Practice Guidelines (CPG)* also belong to this level of evidence (Table 4).

**Randomized Controlled Trial (RCT).** *Randomized Controlled Trial (RCT)* is defined as “…An experiment in which two or more interventions, possibly including a control intervention or no intervention, are compared by being randomly allocated to participants.”16 Usually, participants are assigned to a test group that is treated by the investigated therapy; and a control group that is treated by an
alternative therapy. The groups are then followed up for a certain time to explore possible differences among groups. This type of study must follow specific international rules based on the Consolidated Standards of Reporting Trials (CONSORT). For example, Slot et al., in the article entitled “Maxillary overdenture supported by four and six implants in the anterior region; 1-year results from a Randomized Controlled Trial,” stated that “Bar-supported overdenture on four implants in the anterior maxillary region are not inferior to overdentures supported by six bar-connected implants.”

Systematic Review (SR) of RCTs. Systematic Review (SR) of RCTs is a “Systematic and quantitative review of the results of a series of individual trials with the aim of data integration.” For example, Lang et al., in a study entitled “A Systematic Review of the effects of full-mouth debridement with and without antiseptics in patients with chronic periodontitis,” stated that full mouth debridement with and without the use of antiseptics associated with root planing and conventional staged debridement “…may be recommended for debridement in the initial treatment of patient with chronic periodontitis.”

RCTs and SR of RCTs represent the highest level of evidence. In fact, in RCTs, randomization assures an allocation of patients to test or control group and minimizes the effect related to the presence of possible unknown variables. Implementation of randomization aims at obtaining homogeneous and comparable groups, the only difference being the tested treatment. SR of RCTs, summarizing data from different studies in different experimental conditions makes it possible to achieve greater statistical power. Cochrane reviews and ADA Evidence-Based reviews are internationally recognized as high standard Systematic Reviews in evidence-based health care.

A further methodological “quality leap” is achieved with the Clinical Practice Guidelines (CPG) developed in medicine to transfer the results of the best scientific research (RCTs and SRs) to clinical practice and to provide useful information for institutions, universities, insurances companies, clinical practitioners, and mostly for patients (see paragraph “From research to clinic”).

The “Pathway” of a Treatment

The pathway of a treatment can be described as the “career” of a certain therapy/intervention. This usually starts from the initial presentation to the scientific community and may proceed throughout several steps of increasing scientific validity till eventually arrive to a conclusion on its effectiveness based on high-quality studies.

Case reports are frequently published in the health care literature. Case reports are the first publication of a therapeutic concept, the first presentation of a new technique or statement of a biological principle; they are also reliable, understandable and generally appreciated by practitioners.

The case report study should be validated by RCT through the different steps of the Grading System. In fact, thanks to the use of the control group randomization and a more rigorous scientific method, RCTs are less subject to bias and offer more reliable conclusions. This process (career of treatment) may require a long period of time (even up to 20 years!). A treatment’s pathway or “career” can lead to 3 different results: 1. evidence of the treatment’s validity; 2. evidence of a lack of treatment validity; 3. insufficient evidence.

<table>
<thead>
<tr>
<th>TABLE 4.</th>
<th>Studies with high-value of scientific evidence: positive and negative aspects.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of study</strong></td>
<td><strong>Favorable aspects</strong></td>
</tr>
<tr>
<td>Randomized Controlled Trial (RCT)</td>
<td>• The randomization process minimizes the presence of spurious variables.</td>
</tr>
<tr>
<td>Systematic Review (SR) of RCT</td>
<td>• SR uses data from several studies;</td>
</tr>
<tr>
<td></td>
<td>• SR makes it possible to aggregate different experimental conditions;</td>
</tr>
<tr>
<td>Clinical Practice Guidelines (CPG)</td>
<td>• CPG uses data from several studies;</td>
</tr>
<tr>
<td></td>
<td>• CPG makes it possible to aggregate different experimental conditions;</td>
</tr>
<tr>
<td></td>
<td>• CPG transfer the results of the best scientific research (RCIs and SRs) to clinical practice.</td>
</tr>
</tbody>
</table>
One typical example is the case report published by Nyman in 1982\textsuperscript{23} who was the first to discuss the clinical concept of periodontal regeneration and described the use of non-resorbable membrane. Subsequently, case series, controlled clinical studies, RCTs,\textsuperscript{24} and SRs\textsuperscript{25} demonstrated that the therapy first proposed in 1982 was indeed effective. This confirmatory process lasted more than 20 years (Fig 3).

On the other hand, the use of citric acid for conditioning the root surface in the treatment of gingival recession proposed by Liu & Solt in 1980,\textsuperscript{26} despite some initial positive results from case series,\textsuperscript{27} was subsequently found ineffective and useless for achieving greater root coverage by Roccuzzo et al. in an SR in 2002\textsuperscript{28} (Fig 4).

Sometimes, evidence reporting the validity of a certain treatment may prove inconclusive. For example, regarding the management and treatment of peri-implant infections, a Cochrane Systematic Review recently published in the European Journal of Oral Implantology\textsuperscript{29} shows that the number of high quality trials investigating the treatment of peri-implantitis is very limited. The conclusions are that “...there is no reliable evidence suggesting which could be the most effective interventions for treating peri-implantitis,” and “...larger well-designed RCTs with follow-up longer than 1 year are needed.”

### STATISTICS FOR THE CLINICIAN

The concept of probability was born in 1660, but statistical thinking along with its modern implications only developed during the nineteenth century.\textsuperscript{30} Then, in the first half of the twentieth century, the increase in scientific knowledge led scientists to modify their approach toward the interpretation of natural phenomena.

The application of statistics to medical research developed dramatically during the course of the twentieth century, so that, to date, it is hard to find an article which does not include some statistical procedure. The presence of statistics in the biomedical literature has continuously increased. However, this has not been accompanied by a proper understanding of its language by the main users, the clinicians. In fact, “…Many people think that all you need to “do” statistics is a computer and appropriate software. This view is wrong even for analysis, but it certainly ignores the essential consideration of study design, the foundations on which research is built. Caregivers need not be experts in statistics, but they should understand the principles of sound methods of research.”\textsuperscript{31}

Therefore, at least some basics are needed for an understanding of any scientific paper.

### Variables

In statistics, a variable is a characteristic of a unit being observed in a study. It may assume more than one of a set of values which represent the single measurements. Generally, in controlled trials, one single variable is chosen to represent the mean of comparison among therapies. This variable, which has to be of primary importance for the aims of the study, is also called “outcome variable” or “dependent variable” or “endpoint.” The choice of the outcome variable determines the scale used for measurement.

### Types of Data and Scales of Measurement

Data can be of three main types:

- **Categorical (nominal) data:** values or observations that can be sorted into groups or categories. The difference among categories is qualitative: for example sex (male/female), bleeding on probing (positive/negative), blood group (A, B, AB, O) or ethnicity. The categories have no ranking or numerical relationship to each other.
- **Ordinal data:** values or observations that can be sorted into groups or categories and these categories have an intrinsic ranking or ordering. An example can be the classification of periodontitis (mild/moderate/severe).
• **Numerical data**: values or observations that can be measured. The measurements can be either discrete or continuous variable. Discrete numerical data can only take certain values, usually integers (e.g. number of smoked cigarettes, tooth loss per year); continuous variable data can theoretically assume any numerical value, but usually fall within a certain range (e.g. probing pocket depth expressed in mm).

**Descriptive Statistics**

Descriptive statistics provide simple summaries about the numerical information obtained by data collection. They can only be used to describe the sample that is being studied and the observations that have been made. That is, the results cannot be generalized to any larger group (external validity). Demographic or clinical characteristics of the subjects such as the average age, the proportion of subjects of each sex, and the baseline measurements of clinical variables of interest are usually summarized in tables. Categorical data are easily described by histograms or pie charts.

The information obtained from a series of observations is usually described by the following values representative of the so-called “central tendency”:

- **Mean.** Arithmetic average, i.e., the sum of all the values divided by the number of observations.
- **Median.** Value such that half of the observations’ values are less than and half are greater than that value. The median is also called the 50th percentile or the 0.5 quantile.
- **Mode.** The most commonly occurring observation.

**Variability** is a measure of the spread of the observations of a data set. It can be described by:

- **Standard Deviation and Variance.** Measures of the variability (spread) of measurements across subjects. Standard deviation is the square root of the variance.
- **Range.** The range gives the maximum and minimum values of the observations.

**Inferential Statistics**

**Sample and Population.** When conducting an experiment, i.e., comparing two therapies/procedures, researchers and clinicians are interested in knowing which of the two tested therapies is better on the average than the other in the entire patient population. In other words, the main aim of any investigation is to understand what would happen in all the population of individuals with characteristics similar to those included in the study sample.

Inferential statistics are concerned with making predictions or inferences about a population from observations and analyses of a sample. The results of an analysis of a sample are generalized to the larger population that the sample represents. In order to do this, however, it is imperative that the sample be representative of the group to which it is being generalized. For example, a periodontist may want to know if a specific surgical flap is effective in a sample of, say, 30 patients included in his trial, but he and the scientific community will be surely more interested in knowing if any other sample of 30 patients with the same characteristics of the study participants will obtain, in average, the same results following the same intervention. This is the aim of inferential statistics.

**Standard Error of the Mean (SEM).** In theory, many different samples may be extracted to test the difference between the therapies of interest. Thus, each sample will be described by its own mean and standard deviation. The standard deviation of the sample means is called the standard error of the mean (SEM).

**Confidence Interval (CI).** Once the standard error is calculated, the confidence interval is determined by multiplying the standard error by a constant that reflects the desired level of significance, based on the normal distribution. The constant for 95 percent confidence intervals is around 2.

Thus the CI for a sample mean \( x \) will be: \( x \pm 2 \times SEM \).

That is, if we had 100 different samples, we could obtain 100 different estimates of this range; in about 95 of these, the true population mean would be within this range, and in about five of these estimates the true population mean would be beyond it.

**Understanding the meaning of a confidence interval is of primary relevance in the interpretation of study results.** In SRs, the effect size expressed in terms of mean difference between treatments and related CI is widely used and represented in graphs called forest plots. For example, when considering the difference between CAF and CAF + CTG, the point estimate expresses the mean difference and the width of the CI should be considered by the reader as the “confidence” that can be given to the estimated value; the larger a confidence interval for a particular estimate, the more caution is required when using the estimate (Fig 5).

**Hypothesis Testing.** Statistical hypotheses can be tested in order to generalize the results of a sample to the population. In statistics, a hypothesis is a claim about something.

For example, a clinician dealing with root coverage procedures might want to answer the question “is coronally advanced flap (CAF) combined with connective tissue graft (CTG) more effective than coronally advanced flap alone in terms of Recession Reduction (RecRed)?” (Fig 6a, b, c).

Hypothesis testing is used to check whether or not there is likely to be a difference between one or more compared data sets. In the example, a sample providing
information on CAF + CTG and another on CAF alone are selected and compared.

The method by which samples are selected to learn more about characteristics in a given population is called hypothesis testing. Hypothesis testing involves four key steps:

Step 1: State the hypotheses.

Step 2: Set the criteria for a decision.

Step 3: Compute the test statistic.

Step 4: Make a decision.

Step 1: state the hypothesis. The Null Hypothesis ($H_0$) is stated first, and it assumes that there is no difference in the treatment effects of the tested therapies. In the above example, the null hypothesis is that “CAF + CTG is no more effective than CAF alone.” The basis of the decision is to determine whether this assumption is likely to be true.

The Alternative Hypothesis ($H_1$) is a statement that directly contradicts the null hypothesis. In this example, $H_1$ states that “there is a difference in the treatment effects between CAF + CTG and CAF alone.”

Step 2: set the criteria for a decision. To set the criteria for a decision, the level of significance for a test has to be stated: it refers to a criterion of judgment upon which a decision is made regarding whether to accept or reject the null hypothesis. The level of significance is the probability of rejecting the null hypothesis when it is true. The criterion or level of significance is typically set at 5%. This means that there is a 5% chance that you will reject the null hypothesis when it is true.

In the example, it means that there is a 5% chance that the null hypothesis of no difference between CAF + CTG and CAF is rejected.

Step 3: Compute the test statistic. A statistical test provides a mechanism for making decisions about a research question using a mathematical formula. The choice of the correct statistical test for an experiment mainly depends on the nature of the outcome variable of interest.

In the example, the outcome variable (RecRed) is continuous (expressed in mm) and a t-test is chosen for the analysis.
Step 4: make a decision. The value of the test statistic is used to make a decision about the null hypothesis.

*p*-value (*p*). Assuming that there is no difference (*H₀*) between two therapies (CAF + CTG and CAF in the example), the *p*-value indicates the probability of obtaining a difference which is at least equal to the one obtained in the study. The *p*-value for obtaining a sample outcome is compared to the level of significance and there may be two scenarios:

*H₀ is rejected:* If the *p*-value is less than 0.05 (level of significance), the difference between CAF + CTG and CAF is statistically significant for the outcome variable RecRed.

*H₀ is accepted:* If the *p*-value is greater than 0.05, then there is a reasonable probability that CAF + CTG and CAF do not differ in terms of RecRed, thus it can be concluded that the difference between the two therapies is not statistically significant.

**FROM RESEARCH TO CLINICAL PRACTICE**

**Clinical Versus Statistical Significance**

When a practitioner reads a scientific article, in addition to the methodological and statistical aspects that permit him/her to evaluate the quality of the study, s/he must focus attention on the clinical impact of study’s results. The literature contains studies (including RCTs) that present statistically significant results but which, from the clinical standpoint, are only relatively or not at all significant.

For example, a double-blind multicenter RCT\(^{33}\) evaluated the efficacy of a controlled-release biodegradable chlorhexidine chip used as an adjunct to scaling and root planing (SRP) versus placebo or scaling and root planing alone on reducing probing depth and improving clinical attachment level in adult periodontitis. The results showed a significant reduction of probing depth at 9 months in the chlorhexidine chip plus scaling and root planing treatment group compared with both control groups, scaling and root planing alone and placebo chip plus scaling and root planing, respectively. The difference versus SRP alone was 0.30 ± 0.07 mm (*p* = 0.00001) and versus placebo chip plus SRP was 0.26 ± 0.08 mm (*p* = 0.00056). Based on these results, the authors stated that “the adjunctive use of the chlorhexidine results in a reduction of probing depth and a significant improvement in attachment level compared with scaling and root planing alone and suggest that the chlorhexidine chip is a safe and effective adjunctive chemotherapeutic agent for the treatment of adult periodontitis.”

Reading this article, even in presence of a great statistical result (0.00001), the clinician may be confused evaluating its clinical impact, with only 0.3 mm of difference between the 3 different approaches: will the clinician follow this clinical approach on his/her patients?

This example shows how important it is to explain the terms “significance” or “clinical relevance” which is also called “clinical impact.”

Probably, the best definition of clinical significance is the one proposed by Greenstein\(^{34}\):

“clinical significance denotes a change that may alter how a clinician will treat a patient, and this value judgment varies depending on the situation.”

Among the many factors that can influence the clinical relevance of a treatment, we must mention the therapeutic effect of the treatment, the cost/benefit ratio (economic and biological), the operator’s skill, the environment in which he/she works and the characteristics of the patient (social and psychological). Unfortunately, these factors are still difficult to estimate and even today, in periodontology, clinical significance is judged only on the basis of statistical values (e.g. the *p*-value).

**Conflict of Interest**

Another important aspect is the assessment of the presence of a possible conflict of interest. Today, more than ever, this matter is very crucial. The global economic crisis has led to a reduction in government spending on
The Role of Clinical Practice Guidelines

Practitioners today can obtain additional benefits by applying Clinical Practice Guidelines (CPG) to their decision-making process when choosing the most appropriate treatment for their patients.

In 2011, the US Institute of Medicine defined the CPG as “statements that include recommendations intended to optimize patient care that are informed by Systematic Reviews of evidence and an assessment of the benefits and harms of alternative care options.” The CPG are produced under the auspices of a medical specialty association, relevant professional society, public or private organization, government agency at the federal, state, or local level, or health care organization. Clinical Guidelines must meet several criteria according to the US National Guidelines Clearinghouse. As a consequence, CPG are located at the highest level of the Grading System for therapies.

The above-described CPG differ from the generally called Clinical Guidelines in therapy. In fact, the latter often derive from consensus reports of expert opinions; they are not based on the highest level of evidence, and they do not take social impact into account.

In periodontology, due to the recent increase in the quality of clinical studies, only 2 CPG have been published. In 2008, the first CPG was developed to determine the most appropriate surgical techniques for periodontal patients with infrabony defects ≥3 mm, in terms of efficacy, complications, and patient opinions. With regard to the efficacy of treatments, the final recommendations indicated that Guided Tissue Regeneration (GTR) and Emdogain (EMD) can yield better results than open flap debridement (OFD) in terms of clinical attachment level (CAL) gain, reduction of pocket depth (PD), and bone gain after 1 year of follow-up. The available data are insufficient for an evaluation of bone or bone substitute grafts after 1 year. Considering the complications and patient opinions related to the investigated treatments, the available data are insufficient.

More recently, a second CPG has been developed with the aim of determining the most appropriate surgical techniques, in terms of efficacy, complications and patient opinions, for the treatment of buccal single gingival recessions without loss of interproximal soft and hard tissues. The final recommendations indicated that CAF + CTG and CAF + EMD are more effective for achieving complete root coverage than CAF alone. In the presence of cervical abrasion and root sensitivity CAF + CTG + Restoration causes less sensitivity than CAF + CTG. In terms of complications, CAF produces less post-operative discomfort, but there is little information about patient aesthetic satisfaction.

SUMMARY AND CONCLUSIONS

Clinicians and researchers should be able to understand and critically judge the results of scientific studies. At this purpose, a paper should be evaluated on the basis of some key issues which include, at least, an assessment of methodological aspects, a critical analysis of the statistical component and a proper understanding of the clinical impact of the study outcomes.

Grading Systems represent an essential tool to identify the strength of the evidence of an article and RCTs, SRs and CPG are at the highest level of this scale. The concepts of sample and population, the difference between descriptive and inferential statistics and the proper interpretation of the p-value are key elements to correctly evaluate the results from trials. However, attention on the clinical impact of study’s outcomes is of paramount importance as the literature contains studies (including RCTs) that present statistically significant results but which, from the clinical standpoint, are only relatively or not at all significant. Clinical Practice Guidelines represent a useful tool for practitioners to their decision-making process when choosing the most appropriate treatment for their patients.

REFERENCES

22. ADA Center for Evidence-based Dentistry. Available at: http://ebd.ada.org/SystematicReviews.aspx; Accessed 03.08.13.