
Course Author(s): Ann L. McCann, RDH, PhD; Emet D. Schneiderman, PhD
CE Credits: 2 hours
Intended Audience: Dentists, Dental Hygienists, Dental Assistants, Dental Students, Dental Hygiene Students, Dental Assistant Students

Date Course Online: 06/10/2002  Last Revision Date: 1/24/2018  Course Expiration Date: 1/23/2021
Cost: Free  Method: Self-instructional  AGD Subject Code(s): 550
Online Course: www.dentalcare.com/en-us/professional-education/ce-courses/ce46

Disclaimer: Participants must always be aware of the hazards of using limited knowledge in integrating new techniques or procedures into their practice. Only sound evidence-based dentistry should be used in patient therapy.

Introduction
This continuing education course describes the criteria for judging the quality of a research report and provides guidelines for interpreting the research information. These skills will help oral healthcare professionals decide whether or not to incorporate the research findings into their patient therapy and practice procedures.

Conflict of Interest Disclosure Statement
• Dr. Ann McCann has done consulting work for P&G.
• Dr. Emet Schneiderman reports no conflicts of interest associated with this course.

ADA CERP
The Procter & Gamble Company is an ADA CERP Recognized Provider.

ADA CERP is a service of the American Dental Association to assist dental professionals in identifying quality providers of continuing dental education. ADA CERP does not approve or endorse individual courses or instructors, nor does it imply acceptance of credit hours by boards of dentistry.

Concerns or complaints about a CE provider may be directed to the provider or to ADA CERP at: http://www.ada.org/cerp

Approved PACE Program Provider
The Procter & Gamble Company is designated as an Approved PACE Program Provider by the Academy of General Dentistry. The formal continuing education programs of this program provider are accepted by AGD for Fellowship, Mastership, and Membership Maintenance Credit. Approval does not imply acceptance by a state or provincial board of dentistry or AGD endorsement. The current term of approval extends from 8/1/2017 to 7/31/2021. Provider ID# 211886
Course Contents
• Overview
• Learning Objectives
• Criteria for Judging a Research Report
  ◦ When Was the Work Published
  ◦ Where Was it Published
  ◦ Are the Qualifications of the Authors Appropriate?
  ◦ Is the Purpose Clearly Stated?
  ◦ Is the Experimental Design Clearly Described and Appropriate?
  ◦ Have All Possible Influences on the Findings Been Identified and Controls Instituted?
  ◦ Has the Sample Been Appropriately Selected?
  ◦ Has the Reliability of the Scoring Been Assessed?
  ◦ Are the Comparisons Between Groups Appropriate?
  ◦ Is the Investigation of Sufficient Duration?
  ◦ Is the Statistical Analysis Appropriate to Answer the Research Questions or Hypotheses?
  ◦ Have the Research Questions or Hypotheses Been Answered?
  ◦ Do the Interpretations and Conclusions Logically Follow the Experimental Findings?
  ◦ Is there a Scientific Basis for the Recommendation of New Therapy?
• Conclusion
• Course Test
• References
• About the Authors

Overview
Oral healthcare professionals need to ascertain the quality of journals and to read articles critically, distinguishing good from poor research. Reading critically involves having a working knowledge of research while maintaining a healthy skepticism. Brian Burt describes such skepticism as "somewhere between blind acceptance and blanket distrust." Just because a study is in print or on the Internet does not mean the conclusions about a therapy or product should be accepted at face value. The art of literature criticism is a skill that can be learned, and the practice of this art will contribute to knowledge of the literature.

The oral health practitioner must continually review dental research articles to learn about new options for therapy and oral health products. How does the practitioner know when he/she should incorporate these new procedures into their practice? This continuing education course presents guidelines for critically evaluating the research and for making decisions about how to apply the research findings to clinical practice. This course builds on the previous course, “Using Research for Clinical Decision-Making: Elements of a Research Report,” that describes the various components of a research report.

Learning Objectives
Upon completion of this course, the dental professional should be able to:
• Explain why the oral health practitioner needs to know how to judge the quality of a research report.
• State why the oral health practitioner should know when the research report was published.
• Explain why the oral health practitioner should know where the research report was published.
• State why the oral health practitioner should review the qualifications of the authors.
• Explain why it is important to evaluate whether the purpose is clearly stated.
• Explain why it is important to evaluate whether the experimental design is clearly described and appropriate.
• Explain why the oral health practitioner should determine if the sample is appropriate with regard to the number and characteristics of the subjects.
• Discuss why the oral health practitioner should determine if the reliability of the scoring has been assessed.
• Explain the importance of evaluating whether the interpretations and conclusions logically follow the experimental findings and to what population the findings may be generalized.
• Determine the basis for adopting new therapy or products in clinical practice.
Criteria for Judging a Research Report

• When Was the Work Published
• Where Was it Published
• Are the Qualifications of the Authors Appropriate?
• Is the Purpose Clearly Stated?
• Is the Experimental Design Clearly Described and Appropriate?
• Have All Possible Influences on the Findings Been Identified and Controls Instituted?
• Has the Sample Been Appropriately Selected?
• Has the Reliability of the Scoring Been Assessed?
• Are the Comparisons Between Groups Appropriate?
• Is the Investigation of Sufficient Duration?
• Is the Statistical Analysis Appropriate to Answer the Research Questions or Hypotheses?
• Have the Research Questions or Hypotheses Been Answered?
• Do the Interpretations and Conclusions Logically Follow the Experimental Findings?
• Is there a Scientific Basis for the Recommendation of New Therapy?

When was the Work Published?
If the research article is older than three to five years, it is quite possible the therapy or product discussed has been superseded by a newer investigation. However, some older papers have definitively answered important questions and should be regarded as classic studies.

Where was it Published?
Although textbooks are often the first sources to be consulted on a given subject, they are often dated. When the material is delivered to the publisher it may already be several years old, and the publication process often takes several additional years. Textbooks are particularly problematic in that the sources of information are often absent and incorrect information is sometimes propagated from one textbook to another. Increasingly popular electronic textbooks have the potential to be more current than in the past. For these reasons, journals represent the most current source of information in any discipline. The quality of research in the scientific journals, such as the Journal of Dental Research and the Journal of Periodontal Research is consistently high.

Schemes have been developed by Journal Citation Reports (JCR, Thomson Reuters, Philadelphia) and others to rate the relative impact of journals, i.e., a ratio of citations to articles published. In 2016, the impact factor (IF) for the New England Journal of Medicine was 72.4, but it ranged from 2.8 and 4.2 for the top ten dental journals (out of 82 journals). Higher ratings are to be expected for journals that serve much large audiences than oral health care professionals. IFs are most appropriately compared to those within the field of dentistry. Ratings below 1.0 are considered low impact. These rating are helpful but not the end-all in characterizing the quality of journals and are readily available through some university libraries. More about IF calculation by JCR is available on YouTube.²

The most critical indicator of quality is whether or not a journal is peer reviewed. Peer review means that several experts in an appropriate field critically evaluate the manuscript and provide recommendations to the editor. Ideally, the editor will maintain complete anonymity between authors and reviewers. Most journal papers that are eventually published are returned to the authors at least once for revisions and many are rejected. According to Brian Burt¹, the top dental journals publish fewer than half of the manuscripts they receive. You can determine if a journal has a peer review process by looking for the listing of an editorial review board at the front of the issue or by reading the journal's guide to authors. This is usually published once a year, and the explanation of the submission process will probably indicate if peer review is involved. Examples of peer reviewed journals include the research journals listed in the preceding paragraph as well as the Journal of Dental Hygiene, the Journal of the American Dental Association, and the Journal of Dental Education.

Are the Qualifications of the Authors Appropriate?
Look for evidence of (1) at least one well-known researcher, (2) a history of research training, and (3) affiliations with a reputable institution. While little-known investigators can do good work and poor work can issue from prestigious institutions, these guidelines should
be helpful in guiding you to credible research. Funding from the U.S. National Institutes of Health (NIH), Health Resources and Services Administration (HRSA) or comparable agencies is another important criterion; obtaining highly competitive grants from these agencies speaks to the credibility of the authors.

Is the Purpose Clearly Stated?
If research questions or hypotheses are not provided or stated clearly, this suggests the investigation was not well planned. This raises suspicions about the scientific merit of the experiment. The purpose should also be stated in an objective manner. Be suspicious if it appears the author had already decided what the results would be at the beginning of the investigation. This assumption by the investigator could so influence or bias the conduct of the research that the true effects of the experiment may not be identified.

Is the Experimental Design Clearly Described and Appropriate?
An experimental or prospective investigation is one where the investigator manipulates a factor or factors (independent variable) and observes the effect on another factor or factors (dependent variable) over time. This design is considered to be the ideal research design. For example, the investigator tests the effect of an antiplaque/antigingivitis mouthwash (independent variable) on plaque deposits and gingivitis (dependent variables). Sometimes, due to ethical or other reasons, the investigator is unable to manipulate the variables. Under these circumstances, a retrospective investigation is conducted in which data are collected on a phenomenon that has already occurred. For example, a researcher identifies groups who differ on a condition such as oral cancer, one group with it and one without, and works backwards to identify the causative factors for the condition, such as a history of tobacco and alcohol use.

The gold standard for testing new therapeutic agents is the randomized clinical trial (RCT). This is an experimental design used to test the hypothesis that a particular agent or procedure favorably alters the natural history of a disease. Two designs are commonly used in clinical trials, the parallel design and the crossover design. With the parallel design (the most common of the two), an experimental group with the new treatment and a control group with the standard or placebo treatment (inactive substance) are used, with pre-treatment and post-treatment measurements of health (Figure 1).

![Paralle Design](image_url)

Figure 1.
With the crossover design, the initial experimental group switches to the control group and the control group to the experimental group halfway through the experiment. This occurs after a washout period, where each subject's physiological condition is allowed to return to baseline (Figure 2). An advantage of this latter design is that no patient is denied the experimental treatment. Another advantage is that each subject serves as his/her own control; fewer subjects are needed for such designs. The crossover design...
can only be used with respect to diseases or conditions that recur when treatment or medication are withheld, such as gingivitis.

There are two potential designs for studying phenomena that occur over extended periods of time, for example, growth of the mandible or the course of periodontitis. In a cross-sectional design (Figure 3a), a sample(s) of a population (cross section) is assessed at one time; with a longitudinal design (Figure 3b), the same sample of individuals is assessed at several different time points. While a cross-sectional design is often more expedient, the longitudinal design usually provides better information, since the actual amount of growth or change in each individual can be assessed. Lacking a true time element, cross-sectional designs are considered unsuitable for making conclusions about cause and effect (causation).

**Have All Possible Influences on the Findings Been Identified and Controls Instituted?**

A good research design controls or minimizes bias, the influence of factors other than the experimental ones on the results of the study. For example, a researcher plans to study the effect of toothpaste claiming a new chemical agent on plaque and gingivitis with the hypothesis that it will reduce plaque formation and gingival bleeding upon probing. In this case, the researcher wants assurance that, if the hypothesized results occur, they are due to the effect of the toothpaste and not some other factor, such as another mouthwash or drug. If such controls could not be imposed, then the researcher could not conclude with any certainty that the reduction in plaque formation and gingival bleeding was due to the chemical agent under study. The strategy is to identify as many factors as possible up front that could influence the results and systematically impose controls for them, such as preventing the subjects from using other drugs. For all the unidentified factors (confounders) that could possibly influence the results, of which there could be many, additional design strategies are employed.

One of these strategies is random assignment of subjects to treatment groups. There is usually an experimental group that receives the experimental therapy and a control group that receives no treatment, a placebo, or the currently recognized standard of therapy. Both groups are needed so comparisons can be made. With random assignment, every subject has an equal chance of being assigned to either group. The assumption can then be made that the unidentified influences are equal for both groups, and they differ from each other only in terms of the therapy being studied. The strategy is to identify as many factors as possible up front that could influence the results, of which there could be many, additional design strategies are employed.

An additional control strategy is the use of a double-blind design (Figure 4). In this design, the subjects should not be able to determine whether they are in the control or the experimental
group; they are “blind” to their assigned group. Additionally, the members of the research team who work directly with the subjects will not know (be “blind”) which of the subjects is receiving the experimental treatment and which is receiving the standard or placebo treatment. In the case of the subjects, knowledge of their group assignment may cause them to behave differently than normal, which would introduce new and uncontrolled influences on the results. For instance, individuals may perform oral hygiene better if they know they are using the newly designed toothbrush. This phenomenon of subjects performing better when they know they are in an experiment is known as the Hawthorne effect. Researchers may also perform differently with the knowledge of subject assignment, such as under assessing the level of gingival disease in an experimental group using a new antiplaque toothpaste.

Investigators are concerned about identifying cause and effect relationships between variables so prevention and treatment strategies for disease can be developed. The extent to which investigators can conclude causation between

Figure 3b.
Has the Sample Been Appropriately Selected?

The selection of the subjects for an investigation is a vital research process, since the degree to which research findings can be generalized is largely dependent on the appropriateness of the sample. The target population is the entire group to which the results of the experiment can be generalized, such as all dental hygienists or all individuals with cerebral palsy. Since it is usually not feasible or even necessary to collect data on the entire population, a representative portion or sample is chosen. The best method for choosing a sample is random sampling, where every member of the population has an equal chance of being selected for the sample. Often access to the entire population is not available to the researcher, and a convenience sample of those members who are available is used. In this case, the researcher will try to identify subject characteristics that may have influence on the study and either verify statistically the groups are equal in regard to these characteristics or adjust for these differences during the analysis of the data with statistical tests such as analysis of covariance (ANCOVA)\(^4\) and logistic regression.

The selection of the subjects for an investigation is a vital research process, since the degree to which research findings can be generalized is largely dependent on the appropriateness of the sample. The target population is the entire group to which the results of the experiment can be generalized, such as all dental hygienists or all individuals with cerebral palsy. Since it is usually not feasible or even necessary to collect data on the entire population, a representative portion or sample is chosen. The best method for choosing a sample is random sampling, where every member of the population has an equal chance of being selected for the sample. Often access to the entire population is not available to the researcher, and a convenience sample of those members who are available is used. In this case, the researcher will try to identify subject characteristics that may have influence on the study and either verify statistically the groups are equal in regard to these characteristics or adjust for these differences during the analysis of the data with statistical tests such as analysis of covariance (ANCOVA)\(^4\) and logistic regression.

the variables will depend on how well extraneous variables are controlled. For instance, in an experimental study of an antiplaque/antigingivitis mouthwash, where brushing habits, diet, and other key factors are carefully controlled across study groups, it may be valid for the authors to conclude the use of this mouthwash has the effect of reducing plaque deposits and gingivitis. Drawing conclusions of causation are more difficult in retrospective studies and especially in cross-sectional studies. For example, if a study showed a positive correlation (a statistical measure of association) between the amount of sugary beverages consumed and caries incidence in a sample of children, the investigators would need to be more cautious in making statements such as “sugary beverages cause caries.” Other unknown factors that may systematically associate with the beverage consumption may actually be the cause of the caries. Well-controlled prospective studies, in which the same subjects are evaluated at two or more time points, have the greatest potential for determining causation. In general, being absolutely certain of cause and effect remains one of the most vexing problems for scientists.
The size of the sample must be large enough to (1) accommodate the expected loss of subjects during the investigation and (2) demonstrate differences between groups by statistical logic. If the groups are too small, a real difference may not be detected with statistics. Possible misinterpretations are that the experimental treatment or product is not effective (even though it is) or the treatments or products used for both groups are equally effective (when one is actually superior). A rule of thumb is the minimal size of each group should be 26-30 subjects. Conversely, if the groups are too large, a trivial difference that does not have any therapeutic importance may reach statistical significance.4

A formal method for determining sample size is called power analysis, which provides more assurance that real differences or effects, if present, will be identified. Computer programs such as G*Power that perform power analysis are now readily available. They allow both an investigator to choose an appropriate sample size and a reader to decide whether a conclusion of “no difference” or “no effect” is warranted, given the actual results of the study. Investigators who perform proper power analyses when designing their studies are much more likely to produce (and publish) clear and unambiguous results, from which conclusions of “effect present” or “effect absent” can be drawn. Studies in which “power” was not addressed directly are more likely to yield inconclusive results.5-7

A research report should clearly describe the characteristics of the sample so the reader can evaluate whether or not the sample is representative of the population and appropriate for the experiment. This is particularly true if the sample was not randomly selected. Included should be such characteristics as age, sex, race or ethnic group, relevant medical conditions, oral hygiene habits, etc. If the reader is not provided with this information, the extent to which the conclusions apply to any other group of individuals or situations remains ambiguous. For instance, a researcher might generalize about the incidence of congenital heart disease in the United States population based on an experimental sample. However, if a description of the sample indicated it was composed of institutionalized patients (who have a much higher incidence of the disease), the reader would know such a generalization was inappropriate. The report should also describe the size of the sample, both at the beginning and the end of the investigation.

Has the Reliability of the Scoring Been Assessed?
Examiners may use an instrument to assess disease levels (such as a periodontal disease index) or perform some type of rating during investigations. This practice brings up the issue of reliability or consistency of the rating process. Consistency is essential, since experiments and their results should be reproducible. Intra-examiner reliability, the ability of an examiner to rate the same conditions in the same way over time, is usually acceptable with training and experience. Inter-examiner reliability, consistency among examiners, is more difficult to achieve.

When multiple examiners are involved, they should undergo a program where they are trained to use the rating criteria. This training process is often referred to as calibration. In the research report, the author should describe the efforts toward achieving a high level of reliability and provide the numerical value obtained for reliability whenever possible. For example, Cohen’s Kappa Coefficient of Reliability (κ) yields a score of 0 to 1, where 1 represents perfect reliability or agreement among raters. A κ value greater than 0.7 is typically considered a good level of agreement.

Are the Comparisons between Groups Appropriate?
Groups of subjects being compared to one another should be as similar as possible. For example, one would not want to apply an experimental treatment to a group of first graders and use a group of second graders as the control. In order to make valid comparisons between treatments, it is important the therapies of the groups are fairly compared. If not, the experimental therapy could appear superior to the control, even though it is not. This is a particular concern with investigations that have a commercial sponsor. Hypothetically, a study could be conducted to determine if plaque removal with a new mouthwash and
tooth brushing (experimental group) is superior to removal with a toothbrush alone (control group). If the control group is given less time for brushing than the experimental group, or if only the experimental group is allowed to use toothpaste, the control therapy is being set up to show less effectiveness. The procedures used by the groups should be comparable and described clearly in the methods section so the reader can make these judgments.

Is the Investigation of Sufficient Duration?
The duration of the investigation must be long enough to permit detection of effects, which vary according to the purpose of the study. In order to gain its Seal of Acceptance, the Council on Scientific Affairs of the American Dental Association requires that clinical trials be at least six months long for chemotherapeutic products to control gingivitis. However, the association requires only thirty to ninety days for testing manual or powered toothbrushes. For dental caries trials, where the development of disease is a more lengthy process, a duration of several years is needed. For studying the durability of restorative material or implants, investigators may be interested in longer term results, such as eight years or more. Once again, the duration of the experiment should be clearly stated in the methods section.

Is the Statistical Analysis Appropriate to Answer the Research Questions or Hypotheses?

Descriptive Statistics
Sufficient data in the way of tables and figures need to be provided to the readers. These findings are often summarized by using descriptive statistics. Measures of central tendency describe what is typical for the sample. The most commonly used measure of central tendency is the mean, which is simply the arithmetic average of a set of numbers. The mean is ideal for describing a variable that has been measured on a continuous or interval scale (e.g., height, weight, or age). Two other measures of central tendency are the median and the mode. The median is the middlemost value in a set of ranked data (50th percentile), while the mode is the most frequently occurring value in a data set. The median is often preferable for describing the typical response for an ordinal variable (e.g., an ordered scale with five levels ranging from greatest pain to least pain upon probing). The mode is often best for describing what is “typical” for a categorical variable (e.g., sex, race, eye color).

In conjunction with measures of central tendency, one must know how observations within a sample vary around that “typical” value. The standard deviation expresses the variability around the mean. It tells us how much a member of the sample, on average, deviates from the mean. A large standard deviation indicates the scores of a data set are widely dispersed around the mean, while a small standard deviation indicates tight clustering around the mean. By definition, the collection of individuals that falls between one standard deviation below and one above the mean comprises approximately 68% of a normally distributed population; similarly, plus or minus two standard deviations comprises about 95% of the population and is commonly used to define the limits for what is “normal” for a particular population. The variance is simply the standard deviation squared. The standard deviation provides a much better summary of the variability within the data than does the observed range (the span between the lowest and highest values), as the latter tells us nothing as to where the majority of observations fall. For the median, quartiles function analogously to the standard deviation; each quartile comprises 25% of the population. The inter-quartile range (IQR) from the 25th to the 75th percentile is a useful way to characterize the range of values that are most typical for a population. If a research report provides only means (or medians) without measures of variability, this indicates a serious omission and should raise a red flag as to the quality of the research.

Inferential Statistics
The techniques of inferential statistics enable the researcher to infer or generalize from a sample to the larger population from which it was drawn. With a few exceptions, all scientific studies that seek to answer questions such as “is there a difference between therapies” or “is one superior to another,” require inferential statistics. The selection of statistical tests by the
researcher prior to initiating the investigation often requires consultation with a statistician. Each method of analysis, such as chi square, t-test, and ANOVA (analysis of variance), requires certain assumptions about the data, such as whether or not the scores are normally distributed around the mean. Violations of these assumptions may yield distorted results.

**Estimation**

In this fundamental process of statistical reasoning, sample statistics are computed to estimate their counterparts for the population, known as parameters. The parameters, or the true characteristics of the population, are represented by the Greek letters \( \mu \) the mean and \( \sigma \) for the standard deviation. The corresponding sample statistics are often represented with capital Roman letters: \( N \) for the sample size, \( X \) for the mean, and \( SD \) for the standard deviation. Standard errors (SE) can be computed for sample statistics; these reflect the precision with which the parameters have been estimated. As the sample size becomes very large, and therefore more representative of the population, the standard errors will become very small; this indicates great precision. Confidence intervals can also be computed to reflect the precision of estimates. For example, a study reported a mean DMFT score (decayed, missing, and filled teeth) of 2.23 with a 95% confidence interval from 2.03 to 2.43 for a sample of children. This indicates the true population mean, \( \mu \) falls within that range 95% of the time.

**Measures of Risk from Epidemiology**

Certain measures are increasingly used to characterize oral disease, its causes and effectiveness of treatments. These terms and measures are derived from population-oriented research – epidemiology – and are somewhat different from those employed in traditional laboratory research. **Prevalence** is the proportion of individuals with a particular disease or condition at a point in time. In contrast, **incidence** rate is the number of new cases, typically of a disease, in a population over a specified period of time; incidence reflects the rate of increase (or decrease) of a disease in a population. Risk can be characterized in a number of ways that express the association between an exposure and an outcome. The exposure can be a disease, a treatment, behavior or environmental factor. **Relative risk** (RR) is the ratio of the risk of an outcome of those exposed to those who are unexposed. For instance, the RR computed in a hypothetical prospective study that looked at fluorosis in those exposed to well-water with an over-abundance of fluoride versus those with typical municipal tap water was 2.5. This estimate of RR tells us that those with the over-florinated water run 2.5 times the risk of developing fluorosis versus those drinking municipal water. Along with RR, investigators should present its 95% confidence interval, e.g., CI: 1.8-5.0. This tells us that we are 95% confident that the true value of RR falls between 1.8 and 5.0. Importantly, in this example, the interval does not include 1. Had it included 1 that would tell us that the risk was equivalent in the two groups, i.e. no association between water supply and fluorosis.

Another and more widely used measure of risk is the **Odds Ratio** (OR); it too should be presented with its 95% CI and has the same interpretation if it equals 1 or the CI includes it. OR is the ratio of the odds of an outcome occurring in the exposed group to the odds of that outcome in the unexposed group; this is the preferred measure for retrospective studies. Both RR and OR use categorical (nominal) variables. OR is often calculated in the context of logistic regression; the advantage of this approach is that an adjusted odds ratio (AOR) can be computed that expresses the association between two categorical variables while adjusting for other, potential confounders.

**Number Needed to Treat** (NNT) has gained some traction in studies on new therapies. NNT expresses the number of patients that need to be treated with the therapy over a specified period of time in order to achieve one additional good outcome. Practitioners want this to be as low as possible, ideally one. On the other hand in studies concerned with adverse side effects of new/experimental therapies, the measure of interest is **Number Needed to Harm** (NNH). Specifically, NNH is the number of patients that, if they received the treatment, would lead to 1 additional patient being harmed during a specified period. We typically want this number to be very large, particularly if the harm is serious, relative to the benefit of the treatment.
**Hypothesis Testing**

The purpose of this area of inferential statistics is to provide an objective means for determining whether the findings of a study (e.g., differences between treatments) are real or due to chance. It is not sufficient merely to state the mean of one group is larger (or smaller) than another, thereby concluding that a treatment effect was present. With hypothesis testing, the probability or risk can be determined that conclusions, based on the data, are wrong.

It is important the text state clearly which statistical tests were used to answer what questions or hypotheses and whether or not these tests yielded statistically significant results (less than the predetermined probability value). The **probability level** (p), often set at 0.05, must also be stated. In the case of a study concluding an experimental mouthwash prevented gingivitis, a result of p < 0.05 means that there is no more than a 5% chance the finding could have been due to chance rather than real differences. Alternately stated, these statistical results indicate there is an excellent probability (>95%) that the effect of the mouthwash is real. In this example, the results are statistically significant, since the p value was less than 0.05. Sometimes other p values (e.g., smaller and more stringent) are required, and the author should indicate why these are being used. Contemporary statistical software packages for personal computers will compute exact p values for each test (e.g., p = 0.0378). Ideally, the authors will have reported these for all statistically significant tests, rather than just state that p < 0.05, as the magnitude of p values provides additional useful information. For example, if p=0.001, this means that there is a 1 in 1000 chance of falsely concluding the presence of an effect, whereas if p=0.05, there is a 1 in 20 chance of being wrong. Other items typically reported are the numerical value of the computed test statistic (e.g., χ² for chi square, F for ANOVA) and the corresponding **degrees of freedom** (df), which reflects the relationship between the sample size and the number of estimations involved in the test.

For a good presentation of the fundamentals of descriptive and inferential statistics, see Dawson and Trapp, Elston and Johnson, and Glantz. For a thorough discussion of evidence-based dentistry, see Hackshaw et al. Epidemiological measures and related concepts are outlined comprehensively in Guyatt et al.

**Have the Research Questions or Hypotheses Been Answered?**

Since the article started with a stated question or hypothesis, an answer needs to be provided. The Results section reports the answer in terms of statistical logic, while the **Discussion** and the **Summary and Conclusions** sections should explain what that answer means with regard to the research sample, to a larger population, and to the practice of healthcare.

**Do the Interpretations and Conclusions Logically Follow the Experimental Findings?**

The interpretations (Discussion section) and the conclusions should be based on the data. Trust your judgment here. If the interpretations and conclusions are not logical - if they appear to be unrelated to the findings or generalize beyond the realm of the investigation - then their value is questionable.

The extent to which the researchers can generalize to a broader population and real life circumstances will depend on the exact conditions of the study itself. For instance, in a study of the effect of antiplaque/antigingivitis mouthwash on plaque deposits and gingivitis, was the study sample limited to a group of 20-30-year old dental hygiene students who tend to brush their teeth at least twice a day and are extremely health conscious and compliant? Perhaps in the broader population of adults, that includes the full array of ages and brushing habits, the “effect” of the positive mouthwash would be overwhelmed by the generally lower compliance and attentiveness to oral hygiene found in the general population.

**Is there a Scientific Basis for the Recommendation of New Therapy?**

An issue with which the investigator must wrestle is the clinical application of the experiment and whether to recommend the experimental therapy or product. A new therapy or product should not be recommended by a health professional until investigations demonstrate a statistically significant superiority to the current therapy or product. A belief by an investigator or
clinician that a new therapy is better, based on his or her observations, is not enough evidence to support a change in practice.

Conversely, statistical significance does not mean necessarily the findings of an investigation are clinically important, or the new therapy or product has enough value that the profession should adopt it. Hypothetically, an investigation comparing the gain of attachment levels following periodontal surgical therapy versus periodontal debridement showed a superiority of surgical therapy that was statistically significant; however, the actual clinical difference was so small it did not make sense for the investigator to recommend one form of therapy over the other. Furthermore, the cost/benefit ratio must be considered when weighing the value of research findings. Since, in this example, surgical therapy was much more expensive and uncomfortable for patients, it was not logical or scientifically sound to recommend surgery over periodontal debridement.

When it comes to making decisions about a new therapy, take an attitude of “prove it to me.” Do not be the first to use a new therapy in your practice on the premise that newer is better. We all know oral health professionals who follow this course of action. In regard to new dental products, it is important to know the history and practices of the companies that manufacture these products. If they have a history of poorly conducted research, of making unsupported claims, or of withholding information, the practitioner should probably pursue the products of a more reputable company. The most rational approach for treatment decision-making is to wait until there is a volume of scientific reports that supports a particular therapy or product before deciding to incorporate it into patient care.

Conclusion
The key to providing the best care for your patients is to adopt treatment regimens that have been proven effective. Evidence of effectiveness is generated in clinical trials and published in the dental and scientific literature. The oral health practitioner needs to stay abreast of the most recent literature and practice the skills of critically evaluating the research evidence. This course and the Elements of a Research Report have presented the basic skills necessary for a critical review of research evidence, and the reward for practicing them will be the knowledge that you are providing the most effective care for your patients.
**Course Test Preview**

To receive Continuing Education credit for this course, you must complete the online test. Please go to: [www.dentalcare.com/en-us/professional-education/ce-courses/ce46/start-test](http://www.dentalcare.com/en-us/professional-education/ce-courses/ce46/start-test)

1. **Which is NOT true of an experimental or prospective research design?**
   a. Dependent variables affect independent variables.
   b. Cause and effect relationships can be inferred.
   c. The experiment can be planned before the effect has occurred.
   d. It is the ideal research design.

2. **Which of the following correctly characterizes the difference between a parallel clinical trial and a crossover trial?**
   a. Only the parallel design has a washout period.
   b. All participants have the advantage of the experimental treatment in the crossover design.
   c. The crossover design only has one experimental phase.
   d. The parallel design does not have a control group.

3. **Which of the following correctly characterizes the difference between cross-sectional and longitudinal designs?**
   a. It is easier to show a cause and effect relationship with the longitudinal design.
   b. The cross-sectional design assesses the same individuals at different time points.
   c. It is quicker to do the longitudinal design.
   d. The longitudinal design assesses a sample of individuals at one time.

4. **Which of the following is a method for minimizing bias?**
   a. Nonrandom assignment of subjects to treatment groups.
   b. Have the subjects know to which group they have been assigned.
   c. Have the investigators know which individuals will be in each group.
   d. Place controls on as many influencing factors as possible up front.

5. **Identification of cause and effect relationships is best facilitated by ___________.**
   a. a prospective research design
   b. a retrospective research design
   c. allowing extraneous variables to influence results
   d. allowing the Hawthorne effect to influence results

6. **What is power analysis?**
   a. A method for assigning subjects to treatment groups.
   b. A method for determining the reliability of scoring.
   c. A method for determining sample size.
   d. A statistical test for determining differences between treatments.

7. **Which of the following methods is NOT appropriate for generating random numbers?**
   a. Drawing names from a hat.
   b. Picking every 7th name from a list.
   c. Random number software program.
   d. Random number table.
8. Which of the following is NOT appropriate in determining the size of the sample for an investigation?
   a. Large enough to accommodate the expected loss of subjects during the investigation.
   b. Large enough to demonstrate the differences between groups by statistical logic.
   c. Make the sample very large so trivial differences reach statistical significance.
   d. Minimal size of the group should be 26 to 30 subjects.

9. The measures of central tendency include which of the following?
   a. mean
   b. median
   c. mode
   d. All of the above.

10. The collection of individuals that falls between one standard deviation below and one above the mean comprises approximately ____ of a normally distributed population.
    a. 95%
    b. 100%
    c. 68%
    d. 15%

11. When the p-value is set at 0.05 in an experiment comparing two different therapies, this typically means that there is ____________.
    a. no more than a 5% risk that a statistical finding of a “difference” is due to chance
    b. no more than a 0.05% risk that a statistical finding of a “difference” is due to chance
    c. a greater than 50% risk that a statistical finding of a “difference” is due to chance
    d. a greater than 95% risk that a statistical finding of a “difference” is due to chance

12. In an experiment comparing the efficacy of two different therapies where a t-test demonstrates a p value of 0.001, this means ____________.
    a. no significant difference was found
    b. a significant difference was found
    c. the two therapies were equivalent
    d. there is a 0.1% risk that the difference is real

13. Which of the following information should be included for each inferential statistical test?
    a. p value
    b. If p value is statistically significant.
    c. Value of the test statistic.
    d. All of the above.

14. Which of the following is a scientific basis for the recommendation of new therapy?
    a. The belief by the investigator that a new therapy is better.
    b. A volume of scientific reports supports the new therapy or product.
    c. The cost is greater than the benefit.
    d. Results are statistically significant but not clinically important.

15. Why do examiners, scoring oral disease levels in an experiment, need to be calibrated?
    a. So that all the examiners use the same criteria for scoring.
    b. So that the measurement of disease has a reliability level as close to “1” as possible.
    c. So that the results are an accurate measure of the disease level of the sample.
    d. All of the above.
References
About the Authors

Ann L. McCann, RDH, PhD

Ann L. McCann is Professor and Director of Planning and Assessment at the Texas A&M University College of Dentistry in Dallas. She is nationally known for her contributions to assessment, having directed two federal grants from the Health Resources and Services Administration (HRSA) to develop assessment methods and tools for health profession programs. She also served as the assessment expert on an NIDCR research grant for developing evidence-based practitioners and scientists and on many HRSA grants for developing dental pipeline programs for under-represented minority students. She has written a book on assessment, a chapter on women’s oral health and over 80 other publications. Dr. McCann has won the Teacher of the Year and the Institutional Service Excellence Award at the College of Dentistry. She also was a fellow in the Leadership Institute of the American Dental Education Association and former Chair of Dental Hygiene at University of Detroit-Mercy. Dr. McCann currently directs three online courses in the Master’s degree program Education for Health Professionals at Texas A&M University and mentors the dental students enrolled in that program. She also monitors the College strategic plan and is in charge of all assessment activities at the College, including course evaluations and many online surveys.

Email: amccann@tamhsc.edu

Emet D. Schneiderman, PhD

Dr. Emet D. Schneiderman is Professor of Biomedical Sciences at Texas A&M University College of Dentistry in Dallas. His research has focused on the use of digital imaging and biostatistics to study craniofacial growth and development, particularly in individuals with cleft-lip and palate. Dr. Schneiderman's work in educational research has focused on program evaluation, use of technology in the curriculum and academic integrity. Dr. Schneiderman has been an investigator on NIH, NLM, HRSA and Texas TIF grants and has published 70 articles, a book, and two-dozen computer programs. He currently teaches evidence-based dentistry, applied biostatistics, research design and human gross anatomy. From 1996 until 2006 he served as Executive Director of Information Technology at the College of Dentistry. In this role he pioneered the high speed networks and digital imaging technologies that are now in use at all 300 operatories throughout clinics. Dr. Schneiderman did his Bachelors and Masters work at Northwestern University and earned his PhD at the University of Michigan, all in biological anthropology. He has served as an evaluator for the Southern Association of Colleges and Schools, a member of Soredex's Digital Imaging Advisory Group and reviewer for NIH-funded North and Central Texas Clinical and Translational Science Initiative. Dr. Schneiderman currently chairs Texas A&M University's Institutional Review Board in Dallas.

Email: eschneiderman@tamhsc.edu