EVIDENCE-BASED DENTISTRY SERIES

Hierarchy of research design used to categorize the "strength of evidence" in answering clinical dental questions

Rhonda F. Jacob, DDS, MS,[a] and Alan B. Carr, DDS MS[b]

The purpose of this article is to highlight important features of research design that clinicians can use to determine which articles are useful when attempting to answer clinical questions and determine the best therapy for a particular patient. This article offers a systematic means of categorizing the quality of research reports for clinicians and clinical investigators. A recurring clinical theme of hygiene education is used to exemplify how phrasing the clinical question determines the type of study design that could be used. The article describes the continuum of research reports, and categorizes them by their inherent strengths and weaknesses. The report describes why the research designs in the supreme position of the research hierarchy, are the most valuable to clinicians seeking evidence that defines the best therapy for their patients.

Our calling as health care practitioners is to offer the best possible care for our patients. In dentistry, one might wonder how one could define "best possible care." Because few clinical situations are life-threatening, the impetus to perform rigorous clinical research to compare efficacy of dental therapies may not seem as important as in medical therapies. In many instances, the final treatment decision is a function of what therapy options are offered to the patient (provider preference), and which treatment option the patient finds most appealing. Therapies that are offered are often those that can best be described as the practitioners' routine, or the treatment the practitioner has determined is the therapy that the patient would most likely choose.[1] The "appeal" of a particular therapy is related to the cost the patient can afford, the esthetics the patient desires, and the expediency with which the care can be delivered. From a patient's perspective, the dental materials and prostheses may appear to be equal in function, with the primary differences being esthetics, cost, and sophistication. A patient frequently assumes that the dentist is only offering the best treatment, or the only treatment available for his or her condition. One can see that the definition of the best therapy is a function of the primary considerations of the patient and the clinician.

As stated in the introductory article to the evidence-based dentistry series,[2] when a practitioner searches the literature for the "best possible therapy" he or she should formulate a question to guide the literature search. This question should not only focus on efficacy and complications, but also focus on patient desires and constraints. Searching fixed hybrid implant prostheses literature for a patient, who indicates that he/she does not have financial means for this prosthesis design, would not be a relevant search. Formulating the appropriate question allows the practitioner to find the relevant literature.

The purpose of this article is to discuss the various clinical research designs that might be included in that relevant literature. The quality of those various research designs can be placed in a "research design hierarchy." The position of each research design within the hierarchy is a function of the strengths and weaknesses of features within each design. Many research designs are available in health care research, and there are ideal research designs to answer particular clinical questions. When the ideal design is used, the strength of the conclusions is great and this report exemplifies the "best available evidence" for making a treatment decision. However, for many clinical questions, the ideal research design may have never been used; as it cannot always be achieved because of constraints in population management, money, ethics, and time. Various design compromises have been advanced to combat some of these constraints, but clinicians and researchers must understand that the confidence in the final research conclusions can be weakened by these design compromises. This leads to weaknesses in the ultimate clinical inferences and decisions that can be derived from that research. Understanding "strength of evidence" is at the heart of evidence-based health care. By using this hierarchial analysis of the research that comprises the relevant clinical literature, clinicians can ultimately determine a treatment plan based on the best-available evidence. This hierarchial analysis will also be relevant for study designs concerning the cause of disease, where exposure to a causal agent or event precedes the disease.

[a]Professor of Dental Oncology, Maxillofacial Prosthodontics, M.D. Anderson Cancer Center, Houston, Texas.
[b]Professor, School of Dentistry, Ohio State University, Columbus, Ohio.

PRIMARY CONSIDERATIONS WHEN EXAMINING THE "BEST EVIDENCE"

Consideration of the quality of any research requires an assessment of internal validity and lack of bias within the study. Internal validity is defined as the correctness of the study results for the study population. It is influenced by how well the methods, outcome measurement, and data analyses are carried out in the study. A study's internal validity is threatened by bias and random variation.

The ability of randomized controlled trials to control bias and random variation is what places them in the premier position of research hierarchy when examining effectiveness of therapies. When selecting articles to offer best evidence, the question "Are the results valid" [2,3] is answered by examining the control of bias and its effect on internal validity.

External validity, also called generalizability, relates to the ability to generalize the findings in the sample population to the general population. It is also a prime consideration for the practitioner asking the question, "Will the results of this study help me in caring for my patients?"[3] External validity is what allows the observations from the study to hold true in other clinical settings. To be certain that a research report is relevant to the clinician, it requires that the report clearly define the patient population.

Finally, one must consider feasibility of the ideal research design when searching for literature relevant to a clinical question. Although a research design of "higher hierarchy or quality" would be ideal to answer the question, such a design may have never been reported in the literature. In this situation, the clinician may need to make treatment decisions based on a report using lesser quality research.

INTRODUCTION TO THE CLINICAL RESEARCH DESIGN HIERARCHY

The research hierarchy classifications[4] are based on 3 key features of study design that are fundamental to "bias control" in a study: (1) the manner in which the subjects were assembled for the test groups, (2) whether exposure to the intervention or putative causal factor was under the control of the investigator, and (3) whether the outcome of interest was present at the time of enrollment (prospective or retrospective evaluation). These features are all influenced by bias and random variation. The higher the study design ranks in the research hierarchy, the better the study design minimizes bias and distributes random variation equally between the study groups. Group A designs have attempted to maximize internal validity compared with groups B, C, and D designs. A recurring clinical theme of hygiene education is used to exemplify how phrasing the clinical question determines the type of study design that could be used. In this clinical scenario, group A designs have attempted to maximize internal validity in comparison to the designs of groups B through D.

Group A

Key features

- Exposure to intervention or putative causal factor under control of the investigator
- Concurrent control group
- Outcomes not present at the time of study enrollment

Study designs

- Randomized concurrent controlled trial
- Quasi-randomized concurrent trial
- Randomized before and after trial

Question: *Can subjects with gingivitis learn how to perform adequate oral hygiene and achieve better outcomes of gingival health (defined as no bleeding on probing of the gingival collar) (a) after viewing an oral hygiene educational video in addition to personal oral hygiene instruction, or (b) after receiving personal hygiene instruction only?*

Study design.

All subjects presenting to a dental school clinic with at least 10 teeth per arch and with bleeding on probing the gingival collar of at least 1 tooth per quadrant were stratified according to number of tooth surfaces that exhibit bleeding on probing the gingival collar. Subjects who agreed to enter a trial of oral hygiene education were then randomized to (a) view a video on oral hygiene and receive personal oral hygiene education, or (b) receive personal oral hygiene education only. The education was repeated at each return visit. The subjects were followed prospectively at 2-month intervals, for up to 6 months, to compare the presence or absence of bleeding on probing.

Randomized concurrent controlled trial

In a randomized concurrent controlled trial (RCT), assignment of patients into study groups is a random allocation. Cohorts of individuals are either exposed or not exposed to the experimental maneuver and then both groups are followed for a specified time, assessing the outcome of interest. A concurrent control group allows both groups to be treated and measured in a standard and blind fashion, thereby decreasing bias and increasing internal validity. Depending on the question under investigation, there are often known or suspected variables that could have a large influence on outcome and might mask or interact with the experimental maneuver (smoking or diabetes in a periodontal therapy study). Before randomization, the subjects with these known variables are stratified, and a complicated randomization process (often requiring computer assistance) will attempt to place subjects with equal allocation of these known variables into each test

Table I. Ideal design features of a randomized control trial

	Randomized control trial	
Ideal design features	**Advantages**	**Disadvantages**
Feasibility		
- Manageable costs		- Expensive in terms of time, money, people
- Manageable time		- Long time for outcome and possibly long time to enroll patients
- Available subjects		
Sampling direction		
- Outcome absent at sampling	- Absence of outcome can be assessed at beginning	
- Standardized eligibility criteria	- Can be assessed easily and aids in determining generalizability of study	
Equivalence of groups		
- Concurrent control groups	- Concurrent control groups allows standardization	
- Random allocation		
- Prognostic stratification	- Method avoids risk of conscious and unconscious bias that occurs with other allocation methods	
	- Known confounding variables can be equally distributed with stratified randomization and unknown variable are likely balanced using randomization	
Maneuver		
- Standardized protocol for administration	- Prospective design allows same treatment for all subjects	
- Control of co-intervention or other exposures	- Contamination with other exposures that might influence outcomes can be minimized	
Outcomes		
- Standardized criteria for measurement	Blind assessments with standardized measurement criteria decreases bias as all subjects assessed similarly	
- Blind assessments		
Data analysis	Statistical tests are often based on assumption of random allocation	
Ethics		Professionals are sometimes concerned about withholding treatment from control group
Other		Concern that patients who participate in RCT differ from the general public to whom results would be applied

group. Unknown variables that may influence outcome will also be equally placed in each test group by the randomization process. Because no pattern of allocation can be "perceived" by caregivers or the patients, bias in delivering the maneuver and in data collection is reduced and internal validity is thus increased. Extraneous exposures and therapies, called co-interventions, that patients might become involved in during the test period, can also be controlled or monitored (Table I).

Quasi-randomized concurrent control trial

The quasi-randomized concurrent control trial (QRCT) is similar to the RCT, but differs in that the subjects are not stratified as to known, confounding variables or are they randomly assigned to the test groups. Subjects are assigned using various quasi-random methods, such as an-every-other subject assignment, every-other-day assignment, by odd-even birth dates, hospital numbers, and so forth. Problems arise with this method because recruitment can be influenced. If a researcher believes that a subject would benefit from one of the therapies, the subject can be entered on that particular day. If a researcher believes that a subject would not benefit from a particular therapy, and his or her birth date or hospital number requires assignment to that therapy, the researcher or the practitioner might withdraw the subject from inclusion in the study. This "guiding" of allocation may be altruistic, but can offer considerable bias to the trial. This type of allocation is also difficult to keep blind from the caregivers and the subjects. Standardiza-

Table II. Ideal design features of a quasi-randomized control trial

	Quasi-randomized control trial	
Ideal design features	**Advantages**	**Disadvantages**
Feasibility - Manageable costs - Manageable time - Available subjects		- Expensive in terms of time, money, people - Possibly long time for outcome to occur and possibly long time to enroll patients
Sampling direction - Outcome absent at sampling - Standardized eligibility criteria	- Absence of outcome can be easily assessed at beginning of study - Can be assessed easily and aids in determining generalizability of study	
Equivalence of groups - Concurrent control groups - Random allocation - Prognostic stratification	- Concurrent control group aids standardization of methods - If quasi-randomization is allowed to work without influence from researchers then allocation bias is reduced and known (prognostic variables) and unknown variables have a chance for equal distribution	- Allocation is not truly random, hence chance for conscious and unconscious bias of variables and increased likelihood of imbalance between known and unknown variables
Maneuver - Standardized protocol for administration - Control of co-intervention or other exposures	- Provided allocation code is unknown, standardized treatment possible for all subjects - Contamination with other exposures that might influence outcomes can be minimized	- Likelihood that caregiver may know of allocation method and bias could negatively impact standardized treatment
Outcomes - Standardized criteria for measurement - Blind assessments	- Standardized measurement criteria assures all subjects assessed similarly - Blind assessment possible if allocation code not known or “broken”	- Standardized measurements may be biased by inability to maintain blindness to allocation code - Increased likelihood that blindness cannot be maintained due to known allocation method or “broken code”
Data analysis		Potential to violate statistical assumption of randomized allocation
Ethics		Concern about withholding treatment from control group
Other		Concern that patients who participate in QRCT differ from the general public to whom results would be applied

tion of care and data collection can be compromised (Table II).

Randomized before and after study

A before and after study allows investigators to offer 2 treatments to the same population of patients. Distribution of known and unknown patient variables are “inherently carried along with the patient” as he or she is given one therapy, and then given the other therapy. In this way, patients serve as their own control group. Critical design features include that the first therapy outcomes must be completely reversible (back to baseline) and “washed out of the patient’s system” before delivery of the second therapy, and which therapy a subject receives first is through random assignment. Both therapies are concurrently delivered in the trial. Random treatment assignment creates a 50% chance of receiving therapy A followed by B, or therapy B followed by A. Having random assignment of subjects to the first therapy (1) creates a concurrent control group, allowing for standardization of care and outcomes assessment, (2) allows for blinding of the therapy to patients and outcomes assessor, and (3) controls for a learning curve to patients going through the same outcomes assessment twice, it will be equally distributed to the 2 groups.

Table III. Ideal design features of a randomized before and after study

	Randomized before and after study	
Ideal design features	**Advantages**	**Disadvantages**
Feasibility - Manageable costs - Manageable time - Available subjects	- Ideal for therapies that have immediate and reversible outcomes, yielding decreased time and costs - Subjects volunteer because they can try both therapies	
Sampling direction - Outcome absent at sampling - Standardized eligibility criteria	- Absence of outcome (baseline status) can be assessed before each therapy - Can be assessed easily and aids in determining generalizability of study	- Cannot be used if outcomes are not reversible or subject does not return to baseline (often a problem with dental therapies)
Equivalence of groups - Concurrent control groups - Random allocation - Prognostic stratification	- Both interventions occurring concurrently - Method avoids risk of possible treatment order bias (learning curve) and patients act as own control - Confounding variables are equally distributed because same patients act as own control	
Maneuver - Standardized protocol for administration - Control of co-intervention or other exposures	- Prospective design allows same treatment for all subjects - Contamination with other exposures that might influence outcomes can be minimized, especially when study time is short	
Outcomes - Standardized criteria for measurement - Blind assessments	Blind assessments with standardized measurement criteria decreases bias	
Data analysis	- Statistical tests are often based on assumption of random allocation - Equal sized-groups	
Ethics	No withholding of treatment as subjects receive both maneuvers.	
Other		Concern that volunteer subjects different than general population to whom study results applied

The difficulty of using this research design to answer a clinical question is that many dental maneuvers are not reversible. Drug trials are possible using a before and after study, allowing the drug effects and blood levels to return to baseline between therapies. The hygiene example is difficult to perform as a before and after study, as the returning of the subjects to a baseline level of oral hygiene is difficult. Getting the first hygiene results to "wash-out of the patient's system" is problematic. Sometimes subjects are asked to discontinue all hygiene methods for several weeks before initiating the study and between therapies to establish a similar baseline. Learning the first hygiene method might sensitize the patient to improved performance or learning the second hygiene method is a likely bias. The internal validity of the study related to treatment order bias is increased by the fact that this possible influence of treatment order is balanced in both groups because all subjects received the maneuvers in a random sequence. This type of trial is ideal for outcomes that occur in a relatively short period. In this way, co-interventions and differences that can occur in a patient as a result of elapsed time, namely, new comorbid conditions, lack of compliance, and so forth, can be minimized (Table III).

Group B

Key features

- Exposure to intervention or putative causal factor not under control of the investigator
- Control group (may or may not be concurrent)
- Outcomes not present at the time of study enrollment

Study designs
- Cohort (nonequivalent group) concurrent study
  - Comparison group derived from same patient pool
  - Comparison group derived from different patient pool
- Before–after study
  - Same patients are used in both before and after phase
  - Different patients are used in the before and after phase

Cohort (nonequivalent group) concurrent study

Question: *Do subjects who choose to receive personal hygiene instructions from dental personnel, have better outcomes of gingival health (defined as no bleeding on probing), than subjects who choose to receive oral hygiene instructions of viewing an oral hygiene instruction video and receiving additional personal hygiene instructions?*

Study design. Subjects who had previously presented to a dental school clinic (same patient pool), with at least 10 teeth per arch and with bleeding on probing of the gingival collars of at least 3 tooth surfaces per arch, were asked if they would like to participate in an oral hygiene education study of either (a) viewing a video on oral hygiene instruction with additional personal oral hygiene instruction, or (b) receive personal oral hygiene instruction only. Subjects were allowed to choose their education method. Subjects who entered each group were matched with subjects in the other group by 2 education levels, and by 2 levels of oral health (number of tooth surfaces per arch that bled on probing). Subjects were educated with the method of their choice, and were educated on each follow-up visit. They were followed prospectively at 2-month intervals, for up to 6 months, to compare the level of bleeding on probing.

In cohort (nonequivalent group), concurrent studies the subjects are allocated to groups according to the decision of the subjects and the practitioner. Patients are often eager to participate in this type of trial, because they receive the therapy that they desire. However, if a particular therapy is viewed to be more sophisticated or in vogue, it may be difficult to obtain subjects for the other therapy (control group). The assignment of the maneuver is not under the control of the investigator through random allocation. However, the outcome has yet to occur, so subjects can be followed prospectively with a standardized protocol of outcome assessment. If the 2 groups are treated concurrently, the maneuver protocol can also be standardized. However, if the 2 test maneuvers have already taken place and the subjects were gathered after the fact and asked to participate in the follow-up, the design is weakened because the maneuvers were likely delivered with minimal standardization within or between the groups. In addition to group assignment based on the 2 maneuvers, individual assignments to the 2 groups are "matched" based on critical known prognostic factors (known variables) that may influence the outcome of interest in the study. There is a likelihood that unknown variables that may effect the outcome are not matched, and are not distributed evenly between the groups. Prospective outcome assessment allows standardization. However, if there is no attempt to blind the data gatherers to the therapy, there may be bias in outcome measurements between the 2 groups. Studies in which patients are coming from the same patient pool, usually have all subjects coming from the same institution or practice. When the groups come from different patient pools, one test maneuver is administered at one institution and the other test maneuver is administered at another institution. When the 2 groups come from 2 patient pools, there is the additional risk that co-interventions or patient-directed maneuvers that influence outcome might occur in one group and not the other (Table IV).

Before and after study

Question: *Do subjects who have previously been educated by dental personnel to perform oral hygiene, have better oral outcomes of gingival health, (defined as having minimal or no bleeding on probing) after receiving additional hygiene education of viewing an instructional video supplemented with personal hygiene instruction?*

Study design. All patients with at least 10 teeth per arch, who had previously been taught oral hygiene measures by personal instruction from their hygienists were examined for bleeding on probing. Those who were still found to have bleeding on probing in at least 3 tooth surfaces per quadrant, viewed an oral hygiene instructional video and had additional personal oral hygiene instruction. Subjects were educated using the same methods on each follow-up visit. They were followed prospectively at 2-month intervals, for up to 6 months, to evaluate the presence or absence of any bleeding. The rate of subjects (with no bleeding or less than 3 bleeding surfaces per quadrant on probing) in the before group was compared with the after group.

The before and after study allows the assessment of the outcomes (1) before the exposure or maneuver occurs, and (2) after the maneuver occurs. When the same patients are used for the before and after periods of the study, the outcome is assessed after the administration of the control therapy (which is usually a placebo or previous standard of care). This is the "Before-period," because it occurs before the experimental therapy is offered. Often the initiation of this study occurs when "before-therapy" has already been completed and the outcomes have occurred, because it is the experience of negative outcomes with the first (before) therapy that causes the practitioner to initiate

Table IV. Ideal design features of a cohort (nonequivalent group) concurrent study

	Cohort (nonequivalent group) concurrent study	
Ideal design features	**Advantages**	**Disadvantages**
Feasibility		
- Manageable costs		- Expensive in terms of time, money, people
- Manageable time		- Long time for outcome to occur
- Available subjects		- May be difficult to accrue patients if one intervention is in vogue
Sampling direction		
- Outcome absent at sampling	- Absence of outcome can be easily assessed at beginning of study	
- Standardized eligibility criteria	- Can be assessed easily and aids in generalizability of study	
Equivalence of groups		
- Concurrent control groups	- Concurrent control group aids standardization of methods	- Exposure to intervention or causal putative factor is determined by unknown factors (determined by patient), therefore it is difficult to determine whether an unknown variable is responsible for maneuver and the outcome (volunteers tend to have better outcomes than nonvolunteers)
- Random allocation		
- Prognostic stratification	- Patients can be matched on variables known to influence outcome (confounders)	
Maneuver		
- Standardized protocol for administration	- Standardized protocol allows same delivery of therapy	- Contamination with other exposures that might influence outcomes are difficult to minimize when subjects are from different patient pools
- Control of co-intervention or other exposures	- Contamination with other exposures that might influence outcomes can be minimized with subjects from same patient pool	
Outcomes		
- Standardized criteria for measurement	- Prospective design allows standardized measurement criteria	- Blindness to intervention difficult to achieve for clinician or subject creating likelihood of bias (increased attention to experimental group, or preferential search for outcome in experimental group, and placebo effect of intervention itself)
- Blind assessments		
Data analysis	Statistically efficient when equal numbers of subjects in both test groups	Potential to violate statistical assumption of randomized allocation
Ethics	Avoids having to withhold treatment from anyone who wishes to receive it	
Other		

the second (after) therapy. If the before-outcome is not reversible, the same subject cannot be used in the after study.

In the hygiene scenario, subjects who achieved no bleeding in the before group were unable to enter the after group, so the 2 groups are not equivalent groups. An example of a situation in which a maneuver might reverse itself would be one in which the clinician asks patients if any of them experienced difficulty in jaw range of motion (ROM) after a particular dental injection technique. The presence or absence of this ROM outcome serves as the "before-period" data. Once the outcome has reversed itself (normal ROM has resumed), all the same patients can receive a different injection technique that will supply the "after-period" data. A major problem with the "before-period" is that the before data gathering is retrospective; the before-maneuver was likely not standardized, there may have been co-interventions initiated by the patient, and the clinical records or patient memory of the before-outcome are often incomplete or inaccurate. This before and after study differs from the randomized before and after study, in that the before maneuvers and co-interventions are not under the control of the investigator, the treatment order bias is not controlled, and the outcomes are not prospective in the before therapy.

When the before and after study involves different patients receiving the before-intervention and the after-intervention, the treatments are frequently offered at the same institution, but perhaps by different practitioners. Practically speaking, patients with a particular malady received a particular therapy (the previous standard of care), but patients currently coming to that institution with the same malady are likely to receive a different therapy (the experimental therapy or "new standard of care"). Some researchers have called the

Table V. Ideal design features of a before-and-after study

	Before and after study	
Ideal design features	**Advantages**	**Disadvantages**
Feasibility - Manageable costs - Manageable time - Available subjects	- Can be inexpensive and performed in short time - Subjects have already received "before treatment"	
Sampling direction - Outcome absent at sampling - Standardized eligibility criteria	- Able to exclude patients with outcome of interest from the after group in the different patient pools - May be possible to achieve standardized eligibility criteria	- Assessment of eligibility, facts of before- therapy, and outcome of before-therapy collected in a retrospective fashion and may be incomplete and inaccurate
Equivalence of groups - Concurrent control groups - Random allocation - Prognostic stratification	Patients act as their own controls in the same patient pool, but is not a concurrent control group	- No concurrent control group in the different patient pool - No random allocation and very difficult to match or equalize prognostic variables between groups in different patient pool - Disease severity or other important prognostic factors may differ between the same patient pool because of the lapsed time between maneuvers - There may be coexistent disease affecting outcome between the before and after periods
Maneuver - Standardized protocol for administration - Control of co-intervention or other exposures	- Might be possible to standardize maneuver - Contamination with other exposures that might influence outcomes can be minimized with subjects from same patient pool	- Contamination with other exposures that might influence outcomes are difficult to minimize when subjects are from different patient pool or from same patient pool, due to time lapse between therapies, probable lack of standardized maneuver protocol in "before-period," and additional patient selected remedies
Outcomes - Standardized criteria for measurement - Blind assessments		- Often this study is implemented when new therapy comes along, and outcomes assessment for "before-period" is reliant on patient charts and patients perception or memory of outcome - Blindness to intervention is unlikely
Data analysis	In same patient pool decreased subject variation	Passage of time may be a cofactor in outcome, irrespective of maneuver
Ethics	- Avoids having to withhold therapy - Patients view maneuvers as routine therapy. May lessen volunteer bias	
Other		

group treated in the "before-period," the "historical control group." Except for the fact that all subjects in each group received similar therapy, there is little likelihood that the groups would be equivalent in known and unknown variables that effect outcome. Blindness of the subjects or investigators is unlikely, and although standardization of outcome assessment and maneuver are possible in the "after-period," it is unlikely in the "before-period." Patients have few qualms about enrolling in this type of study, as they are less likely to view this type of study as "experimental." They tend to view the therapy of the "after-period" as the new standard of care (Table V).

Group C

Key features

- Exposure to intervention or putative causal factor not under control of the investigator

- Control group (may or may not be concurrent)
- Outcomes present at the time of enrollment in the study

Study designs

- (a) Cross-sectional survey
- (b) Ex post facto study
- (c) Case control study

Cross-sectional survey

Question: *Do subjects with poor oral hygiene have more bleeding on probing of their gingival collars than subjects with good oral hygiene?*

Study design. All subjects in a small rural town were asked to volunteer for a dental screening. Those with at least 10 teeth per quadrant, were screened for (1) their daily oral hygiene regimen and level of oral hygiene, and (2) any bleeding on probing of the gingival collars. The subjects were divided into groups of poor oral hygiene with and without bleeding on probing and good oral hygiene with and without bleeding on probing. A relationship between oral hygiene and bleeding on probing was examined.

The previous study design represents a cross-sectional survey. It is primarily used for epidemiologic causal relationships or associations between exposures and outcomes. The study requires examining a fixed population. Often the entire population cannot be examined, so a subset or cross-section of the population is selected. Selecting this subset should be quite rigorous and systematic to be certain that all demographics of the entire population are represented. A study might be able to evaluate all subjects in a small rural town, but a subset population would be more feasible in a metropolitan population of 4 million.

In the previous example, one is examining whether oral hygiene is associated with bleeding on probing of the gingival collars. The fixed population is examined at one point in time. The exposure or maneuver (in this case level of oral hygiene) and the outcome (bleeding on probing) are all ready present. We now know, after years of research, that there is a causal relationship between oral hygiene and changes in bleeding. Changes in bleeding on probing will occur very quickly if one's hygiene level changes, and because level of hygiene tends to remain constant on a daily basis, a "one point in time investigation" would likely assess the true level of "maneuver and outcome" in this population. This is not the case in all cross-sectional investigations, however. If one happens to be considering a relationship between an exposure and an outcome that is very delayed in onset, the one-point-in-time investigation may miss the outcome in a number of the exposed subjects. Likewise, if the outcome is short-lived and has subsided by the time the investigation occurs, the outcome may be missed.

In studies of exposure to causal agents, such as contaminated drinking water or chemical waste, the identification of subjects may come from property tax records, or other means to identify occupants living in an area. These initial records came from a fixed population, but if this study deals with a specific exposure incident that may have occurred several years earlier, persons may have moved away from the area and would not be available for interview. These unavailable subjects are lost to follow-up, and are not entered in the trial for outcomes assessment. It is almost impossible to estimate how the outcomes assessment in these lost subjects would influence the final study results.

The standardized criteria for eligibility in the study, and data recording of the maneuver / exposure, and outcome are critical. In the cases of exposure to causal agents, history taking may be the only means of verifying the exposure. Blinding to the maneuver or exposure is very difficult to achieve for the investigator and the subject. When the investigator or subjects have a preconceived notion as to the findings of the study, there is a risk for bias toward increased attention and more zealous data gathering in the experimental group.

Without the random allocation and standardized eligibility criteria, it is impossible to ensure that known confounding variables are evenly distributed in each group. Very sophisticated covariate statistical analyses are often used to examine associations of known confounding variables and their relationship to the particular outcome. The most common are age, gender, and comorbid conditions such as diabetes, obesity, and osteoporosis. However, of more concern are the unknown variables that may influence the outcome. If these variables are unknown, they will not be considered in the analysis. There is always the risk that some unidentified factor that is associated with the maneuver has the true causal relationship with the outcome. In this type of cross-sectional survey, it has been evidenced that subjects who volunteer for studies often have better outcomes than the nonvolunteers. There may be some unidentified factor associated with the characteristics that promote volunteerism that are responsible for better outcomes (Table VI).

Ex post facto study

Question: *Persons who have had oral hygiene instructions using video and personal instruction have a lifestyle change toward better outcomes of gingival health (defined as lower bleeding index) than subjects who have had personal oral hygiene instruction only*?

Study design. A group of patients who attend a federally funded health clinic during their pregnancy and were known to have had oral hygiene instruction of video and personal instruction, had an oral examination when they returned for their infants 1 year follow-up.

Table VI. Ideal design features of a cross-sectional survey

	Cross-sectional survey	
Ideal design features	**Advantages**	**Disadvantages**
Feasibility - Manageable costs - Manageable time - Available subjects	- Relatively less expensive in terms of time, money and people - Do not need to keep subjects under study for any extended time period	- Are volunteer subjects and often volunteer subjects have better outcomes than nonvolunteers
Sampling direction - Outcome absent at sampling - Standardized eligibility criteria		- Outcomes present at time of sampling, not a prospective study - Difficult to standardize eligibility criteria - Way of selecting participants may not represent entire population
Equivalence of groups - Concurrent control groups - Random allocation - Prognostic stratification	Concurrent control group from fixed population	- Selection of participants may not represent entire population - Risk of known confounding variables not being equally distributed in each group - Risk of an unknown variable being responsible for exposure/maneuver and the outcome, namely, volunteers tend to have better outcomes than nonvolunteers
Maneuver - Standardized protocol for administration - Control of co-intervention or other exposures	Because a fixed population, co-intervention/ co-exposures likely to be same in both groups	Blindness to maneuver/exposure difficult to achieve, therefore bias likely to occur. Risk of increased attention to experimental group by investigator and placebo effect of the subjects
Outcomes - Standardized criteria for measurement - Blind assessments		- Examining before the late outcomes occur and examining after early outcomes have occurred = underreported outcomes - Difficult, if not impossible, to standardize outcomes assessment
Data analysis	Must use statistical analysis to examine known-confounding variables not evenly distributed in 2 groups	Unequal numbers of exposed and unexposed lessens statistical efficiency
Ethics	Because maneuver already rendered, avoids having to withhold treatment from anyone who wishes to receive it	
Other		

A second group of patients who attended a different federally funded clinic during their pregnancy and were known to have had oral hygiene instructions from dental personnel only, also had an oral examination when they returned for their infants 1-year follow-up examination. A bleeding index was determined for all subjects and a comparison of the bleeding index was made between the groups.

The ex post facto study is similar to the cross-sectional survey in that the outcomes have also occurred at the time of the investigation. The major difference between the 2 study designs is that in the ex post facto design, the subjects receiving / not receiving the maneuver or exposure are identified independently from different populations. In the cross-sectional survey, the 2 groups are derived from a single-fixed population.

The maneuver / exposure and the outcomes have already occurred. Eligibility criteria for receiving the maneuver or exposure are not within the investigator's control, therefore the maneuvers were likely carried out by a number of different, noncalibrated patients. There is little likelihood that the maneuvers would be standardized, even in the same clinic. Each group is assessed for the outcome of interest at one point in time. There are no follow-up therapies or examinations. There is the risk that, if the outcome attributed to the maneuver or exposure appears only after considerable time has passed, the outcome may

Table VII. Ideal design features of an ex post facto study

	Ex post facto study	
Ideal design features	**Advantages**	**Disadvantages**
Feasibility		
- Manageable costs - Manageable time - Available subjects	- Relatively less expensive in terms of time, money and people - Do not need to keep subjects under study for any extended time period	- Are volunteer subjects and often volunteer subjects have better outcomes than nonvolunteers
Sampling direction		
- Outcome absent at sampling - Standardized eligibility criteria		- Outcomes present at time of sampling, not a prospective study - Difficult to standardize eligibility criteria
Equivalence of Groups		
- Concurrent control groups - Random allocation - Prognostic stratification	Concurrent control group	- Risk of known confounding variables not being equally distributed in each group - Risk of an unknown variable being responsible for exposure/maneuver and the outcome, namely, volunteers tend to have better outcomes than nonvolunteers
Maneuver		
- Standardized protocol for administration - Control of co-intervention or other exposures		- Maneuver/exposure already occurred before study began, therefore no standardized protocol for administration - Risk of differing co-interventions because subjects from different settings
Outcomes		
- Standardized criteria for measurement - Blind assessments		- One-point-in-time examination gives risk of underreporting: examining before late outcomes occur and examining after early outcomes have occurred and subsided. May not recognize early stages of outcomes - Difficult if not impossible to standardize outcomes assessment - Blindness to maneuver/exposure difficult to achieve, therefore bias likely to occur. Risk of increased attention to experimental group by investigator and placebo effect of the subjects
Data analysis	- Must use statistical analysis to examine known-confounding variables not evenly distributed in 2 groups - Can sample groups of equal numbers, which increases statistical efficiency	
Ethics	No withholding of maneuver because it is rendered before study begins	

have yet to occur in some subjects. Likewise, if the outcome waxes and wanes, the clinician may miss the occurrence. Both of these could lead to underreporting of the outcome. In the case of the hygiene scenario, various circumstances could have led the treating clinician to send the patient for the hygiene instructions. These indications were likely different among clinicians and between the 2 clinics, making it difficult to have standardized eligibility criteria, and making it difficult to have the groups balanced for confounding variables or unknown variables.

Blinding to the maneuver / exposure may not be possible for the investigator or the subjects. In these trials, it is important that the investigator treat the outcome assessments similarly in each group. Over / under enthusiasm for history taking and the assessment could bias the results. Part of the history taking should include how the maneuver was delivered, or the intensity of exposure to the causal agent. If subjects are not blind to the maneuver / exposure, they may report biased history based on their preconceived notion of the outcomes.

In the hygiene scenario, there may be an unknown factor that influenced whether subjects were eligible for the subject. Patients must have kept their hygiene appointment at the clinic when they were referred. Patients who did not keep that appointment were not eligible for this ex post facto study. Patients who kept the hygiene instruction appointments were likely to be more motivated than those who did not keep the appointments; this motivation is likely related to the unknown variable that often affects outcomes, whereby volunteers have better outcomes than nonvolunteers (Table VII).

Case control study

Question: *Are college students who have poor gingival health less likely to have had professional oral hygiene instructions (defined as instruction by dental personnel using models, video, or "hands-on instruction") during childhood or adolescents, compared with students who have better gingival health?*

Study design. First-year military academy students underwent initial oral examinations in preparation for full dental treatment. All students who had poor gingival health (defined as bleeding on probing of more than 3 surfaces per quadrant) were determined by chart review and were later personally interviewed about their previous oral hygiene education. These students were entered in the "test" group. The control group students were obtained by reviewing the records of the student examined immediately after the "test" student. If that sequentially chosen record indicated that this student had less than 3 bleeding surfaces per quadrant on probing, this student was eligible for the control group. The control group students were personally interviewed about their previous oral hygiene education.

In this case control study, the maneuver/exposure and outcomes have already occurred. The subjects are chosen and assigned to the study groups, based on the presence or absence of the outcome. This is usually a fixed population and all subjects are examined for possible inclusion in the test group of the study. The selection of the control group is performed in a systematic fashion, but this method of subject enrollment is likely to allow uneven distribution of known or unknown variables between the 2 groups. The subjects are interviewed about their history related to the maneuver/exposure. Although one could examine the subjects to confirm the outcomes, as in the cross-sectional survey or ex post factor study, the primary means of identifying subjects for inclusion in the study, comes from previous health records. Many studies have shown that clinical disagreement occurs during routine medical or dental examinations, and omissions in chart documentation often occur. Clinical data recorded in the routine dental examinations that were used to identify the possible study population may be incomplete in the area of interest (bleeding scores), so those students with incomplete records could not be considered for the study. In the hygiene scenario, one could reexamine the students to confirm the bleeding outcomes and confirm which group the subject should be placed, but the outcome may have changed because of additional dental interventions occurring since the initial examination.

After inclusion in the study, the outcome assessment relies on the quality of the history given by the subject and method of "extracting" that history by the interviewer. The lack of blinding for investigator and subject, the difficulty in ensuring identical interviewing techniques, and the inability to ensure that the subject accurately remembers the maneuver/exposure are also weaknesses in this retrospective design (Table VIII).

Group D

Key features
- Exposure to intervention or putative causal factor not under control of investigator
- No control group
- Outcomes may be present at time of enrollment to study

Study designs
- Descriptive study
  - Assessment of exposure in patients with the outcome of interest
  - Assessment of outcome in patients with the exposure of interest
- Expert opinion

Descriptive study (subjects with outcome of interest)

Question: *What proportion of patients in a private practice setting who have good gingival health (defined as low bleeding index), have had professional oral hygiene instructions by trained dental personnel?*

Study design. All new patients, who came into a private practice in the past year, and had a low bleeding index (less than 3 bleeding surfaces per quadrant on probing) during initial examination were identified. A survey was conducted to determine what proportion of patients had had oral hygiene instruction, and what instruction techniques had the patient experienced.

In a descriptive study (subjects with the outcome of interest), the goal is to examine a fixed patient population in which the outcomes and maneuvers/exposures have all ready occurred. The data gathering is retrospective and is in a survey format. In the case of the hygiene scenario, the patients are identified from patient records made during previous initial oral examinations. Subjects with the outcome of interest (low bleeding index) were surveyed to determine whether they had had oral hygiene instructions at an earlier time. Oral hygiene instructions may have ranged from television shows, school teachers, dental

Table VIII. Ideal design features of a case control study

	Case control study	
Ideal design features	**Advantages**	**Disadvantages**
Feasibility - Manageable costs - Manageable time - Available subjects	- Relatively less expensive in terms of time, money and people - Do not need to keep subjects under study for any extended time period	
Sampling direction - Outcome absent at sampling - Standardized eligibility criteria		- Outcomes present and form basis of eligibility criteria, not a prospective study - Difficult to obtain accurate information on maneuver/exposure and confounding variables
Equivalence of groups - Concurrent control groups - Random allocation - Prognostic stratification	- Concurrent control group - All newly diagnosed cases are sampled in fixed population	- Risk of known confounding variables not being equally matched even though from same patient pool - Due to uneven distribution of all confounding variables, statistical adjustment may be needed
Maneuver - Standardized protocol for administration - Control of co-intervention or other exposures		- Maneuver/exposure already occurred before study began, therefore no standardized protocol for administration - Cannot control co-intervention and difficult to measure co-intervention or influence of other exposures or maneuvers
Outcomes - Standardized criteria for measurement - Blind assessments		- Difficult, if not impossible, to standardize outcomes assessment - Blindness to maneuver/exposure difficult to achieve, therefore bias likely to occur. Risk of increased attention to experimental group by investigator and placebo effect of the subjects
Data analysis	- must use statistical analysis to examine known-confounding variables not evenly distributed in two groups - can sample groups of equal numbers, which increases statistical efficiency	
Ethics	Because maneuver already rendered, avoids having to withhold treatment from anyone who wishes to receive it	
Other		

personnel, and parents. In this study, all subjects have experienced the outcome of interest, and the proportion of those who had dental personnel instruction versus no instruction can be determined. Because there may be many instruction methods, one could determine which methods were most or least common. There cannot be a determination about which of these methods is better than another, as all the subjects have achieved the outcome of interest (low bleeding index). Persons who had a higher bleeding index were not included in the study population.

Descriptive study (subjects with the exposure of interest)

Question: *Among high school seniors who have had "hands-on" oral hygiene instructions in their freshman health class, what proportion of them have good gingival health at their annual dental checkup?*

Study design. A group of high school seniors were identified, who were known to have had oral hygiene instructions in health class during their freshman year (which allowed them to physically practice brushing and flossing techniques). They were surveyed to deter-

mine their current hygiene practices and an oral examination was performed to determine what proportion of the seniors had less than 3 bleeding surfaces per quadrant on probing the gingival collars.

In this descriptive study (subjects with the exposure of interest), the maneuver / exposure of interest has occurred in a fixed population. The group is identified by the fact that the exposure is known to have occurred. Available subjects are then surveyed to determine the proportion of subjects who have had the outcome of interest. In the case of the hygiene scenario, the subjects were examined for the outcome of interest, but depending on the clinical question, the outcome and exposure of interest may be determined by a survey.

No control groups are used in either type of descriptive study, but when these trials are reported, they are often compared with other descriptive studies, which serve as "historical control groups." These historical control groups may come from the same institution at an earlier date, or from previous research reports. When considering descriptive studies, one relies on common knowledge or historical control groups to determine whether the results are better or worse than another maneuver / exposure. A fixed study population is identified for study eligibility; however, the exposure and outcome may have occurred long ago. The entire population may not be available for the one-point-in-time survey or second examination. This creates a problem of "lost-to-follow-up" subjects when determining the proportion of effected subjects. If there are a large number of lost-to-follow-up subjects who have died, moved away, or refused to participate, then the final proportion of the study is not a true representation of the population outcome. These biased final results may suggest a therapy is better or worse than it truly is.

Because the inclusion of subjects in the study is a function of "identifying" the subjects who have either had the exposure of interest or have had the outcome of interest, it is necessary to use existing records for identification. This information may not be documented in the records, therefore a number of the fixed population may be eliminated from the study because of lack of documentation. If it is a large number of subjects eliminated from analysis because of inadequate documentation, then the final proportion of affected subjects reported in the study is not a true representation of the population outcome. This number of omitted subjects may not be reported for scrutiny, therefore it can be difficult to assess the doubtful nature of these final proportions. The diligence with which the investigator searches for the available subjects and surveys the subjects that are found is a bias that affects this type of investigation.

Not every descriptive study is retrospective. Descriptive studies could include a defined series of subjects treated and accessed in a prospective manner. In this design, a standardized protocol could be established, and lost-to-follow-up subjects and incomplete documentation could be controlled. The reader must make a critical appraisal of these protocols, as bias can contaminate the data, despite the prospective design of the study. Possible bias could occur if outcome assessments are performed by the same clinicians that rendered treatment, if the assessors are not blind to the treatment, or if the assessors are not calibrated to the measurement techniques. The more bias is controlled, the more meaningful the results. These data can serve to inform an interested patient of the possibility of a particular outcome occurring, if he or she chooses to receive a particular therapy.

In retrospective, descriptive studies, the lack of standardization is a serious weakness; this includes eligibility criteria, delivering of the maneuver, assessing the outcomes, and assessing co-interventions. The lack of any standardization between the study group and the historical control group is a greater weakness. One is likely unable to draw correct conclusions concerning efficacy between therapies with a descriptive study. The only exception is the rare instance when the outcome is so universally poor, and the maneuver offers a dramatic improvement over any other known treatment (Table IX).

EXPERT OPINION, OR "IN MY EXPERIENCE"

Developing expertise in evaluating the literature is necessary to enable today's practitioner to provide the best treatment for his or her patients. It compliments another form of outcome information processing that occurs over the course of a practitioner's career. It is commonly described as clinical experience or intuition. Clinical experience involves the personal evaluation of our patient's conditions and responses to everyday treatments. Much of the literature in medicine and dentistry is devoted to such case study. These observations stemming from everyday treatment decisions and their results (without control treatments) must be understood for their strengths and weaknesses related to subsequent decision making.

Clinicians become more confident and refine their therapies through repetition and observation of clinical effects. This is especially true with therapeutic procedures. Poorly performing a superior procedure may not offer better clinical results than excellently performing a less efficacious therapy. The category of "expert opinion" includes data reported from case reports, undocumented case series, and recollections of treatment results. The media used to report these findings are often lectures, advertisements in dental circulars, and case reports in journals. Because these reports are not derived from bias-controlled maneuvers, clinicians should not consider this information sufficient to determine efficacy of therapies. A common expert opinion strategy is, "I perform the procedure that

Table IX. Ideal design features of a descriptive study

	Descriptive study	
Ideal design features	**Advantages**	**Disadvantages**
Feasibility - Manageable costs - Manageable time - Available subjects	- Data relatively easy to obtain - Do not need to keep subjects under study for any extended time period	
Sampling direction - Outcome absent at sampling - Standardized eligibility criteria		- Outcomes or exposure history form basis of eligibility criteria, not a prospective study - May be difficult to obtain accurate information on maneuver or exposure to select subjects for study - May have large number of subjects that qualify for study, but are not available to survey about the outcome or maneuver
Equivalence of groups - Concurrent control groups - Random allocation - Prognostic stratification		No control group, so must rely on historical control groups or clinical experience on whether the subject is better off than if had no treatment at all
Maneuver - Standardized protocol for administration - Control of co-intervention or other exposures		- Maneuver/exposure already occurred before study began, therefore no standardized protocol for administration - Cannot control co-intervention and difficult to measure co-intervention or influence of other exposures or maneuvers in this retrospective study
Outcomes - Standardized criteria for measurement - Blind assessments		- Difficult if not impossible to standardize outcomes assessment - Blindness to maneuver/exposure difficult to achieve, therefore bias likely to occur. Risk of increased attention to finding the "favorable" outcome subjects and including them in the proportions
Data analysis		Only appropriate analysis is use of proportions in a single population
Ethics	Because maneuver already rendered, avoids having to withhold treatment from anyone who wishes to receive it	
Other		

works best in my hands." This type of care delivery philosophy may offer excellent performance of individual procedures, but it is possible that patients with similar dental problems will be treated differently depending on whose office they find themselves. The addition of critical appraisal of the literature and evidenced-based dentistry to this strategy would be, "I search the literature for the best available evidence that supports a therapy, and then I become proficient in that therapy."

RCTS VERSUS ALTERNATE RESEARCH DESIGNS

If prospective, randomized-controlled trials offer the best possible evidence to determine the truth about a maneuver or exposure, why do other study designs permeate the dental literature? RCTs can be the most expensive trials to perform. Costs increase with the complexity of the maneuver, and the length of time patients must be followed to observe an outcome. Assuring standardization in all aspects of the trial is important to control bias, and adds to the expense of a trial. However, one should also consider the costs to the profession and patients (financially and personally) for the years of delivering care that might be less efficacious. RCTs are not categorically more expensive than studies of "lesser design" and results from well-designed RCTs can dramatically change the scope of clinical practice.

In the case of harmful exposures that occur because of life-style and accidents of industry or nature, ethical health care precludes willfully subjecting a person to a harmful exposure to study the exposure's possible deleterious outcome. (Harmful events may occur as side-effects of therapies. When a comparison of therapies is considered in an RCT, the patient is informed of the possible side-effects related to the therapy, and harmful events are monitored. When comparison of harmful events as well as comparison of efficacy between therapies is determined in the same RCT, the information derived is powerful in making a final treatment decision as to the therapeutic ratio of risk vs benefit for future patients.[3]) The prospective cohort trials (group B) for harmful exposures offers the ability to standardize the follow-up regimen and outcome assessment, thereby increasing the confidence in the study results. One can understand the compelling reason for studying subjects, where it is likely that the outcomes and harmful exposure are already present (group C). After all, the alternative is to wait for the harmful event to happen again in another patient population, and follow the subjects prospectively. It is obvious, however, that the risks of inappropriate conclusions are greater in retrospective group C studies. Group C study designs are easier to perform, and are often the only studies available for policy makers to use to establish guidelines for industry safety and prevention of personal injury. (Most of the data related to silicone breast implants and associated autoimmune and/or connective tissue abnormalities, have come from group C studies. This likely accounts for conflicting study results and the continued skepticism by patients and physicians on both sides of the argument.) When well-conducted, confirmatory groups C and B studies are available, clinicians can be more confident of the results. It should be noted that it took decades of case-control, cross-sectional, and cohort studies, from many different countries, before the harmful effects of smoking were documented to the satisfaction of science.

Unfortunately, dentistry has a preponderance of literature using groups B, C, and D study designs that espouse specific therapeutic regimens. The argument can be made that there are thousands of previously treated patients available for outcome assessment, and "it seems a waste to not make use of all of those data." When these research designs are examined in light of the advantages and disadvantages described in this article, it becomes obvious why there are still conflicting reports related to longevity of fixed partial dentures, the prosthodontic treatment of endodontically treated teeth, and the value of replacing dentition distal to premolars. The greatest value of reports from groups B, C, and D research is that they offer experience and knowledge on the maneuver and outcome to assist in designing a definitive RCT. By examining these reports, it will be possible for clinical investigators to maximize internal and external validity in a subsequent RCT. The previous studies will (1) give insight into eligibility criteria for enrollment, (2) delineate possible problems that might thwart standardization of the maneuvers and outcome assessments, (3) delineate time-to-observe a treatment effect, and (4) suggest co-interventions and confounding variables that are likely to be present in the study populations and the like.

SUMMARY

The purpose of this article is to highlight important features of research design that clinicians can use to determine which articles are useful when attempting to answer clinical questions. This article offers a systematic means of categorizing the quality of research reports for clinicians and clinical investigators. Clinical investigators are interested in the entire continuum of research reports, as they work to define the definitive research question and research method. Clinicians are interested in the highest quality research report available to determine the "best therapy" for their patients. This article will assist in framing the questions and categorizing the best available evidence.

REFERENCES

1. Bader JD, Shugars DA. Understanding dentists' restorative treatment decisions. J Pub Health Dent 1992;52:102-10.
2. Carr AB, McGivney G. Users' guides to the dental literature. How to get started. J Prosthet Dent 2000;83:13-20.
3. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston: Little, Brown and Company; 1991.
4. Tugwell P, Bennet K. Classification of design architecture. Evidence-based workshop handout. Hamilton, Ontario: McMaster University Department of Clinical Epidemiology; 1984.
5. Feinstein AR. Clinical epidemiology: the architecture of clinical research. Philadelphia: WB Saunders; 1985.
6. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials. 3rd ed. Baltimore: Williams & Wilkins; 1996.