DOI: 10.1111/prd.12279

REVIEW ARTICLE

Periodontology 2000 WILEY

Minimizing risk of bias in clinical implant research study design

Georgios A. Kotsakis

Department of Periodontics, University of Texas Health Science Center, San Antonio, Texas, USA

Correspondence
Georgios A. Kotsakis, Department of Periodontics, University of Washington, Seattle, WA, USA.
Email: kotsakis@uw.edu

1 | INTRODUCTION

The field of clinical implant research has been flourishing in recent years. The number of related publications has been on the rise, as has the number of dental journals. In fact, a 2015 bibliometric survey reported that the number of dental journals in publication doubled from 2003 to 2012.[1] During these humming times, the oral health research arena is being reformed to incorporate the prevailing concept of "levels of evidence" and the grades of recommendation that derive from the strength of this evidence.[2] Journals that focus on clinical research and chairside translation are dethroning basic oral science journals from the top of the impact factor ladder. The impact factor is considered, by many, as a touchstone of a journal's scientific impact in terms of frequency of citations per article. As an example, from 2003 to 2012, the *Journal of Clinical Periodontology* recorded the highest annual impact factor rise, whereas the *Journal of Dental Research* noted the smallest impact factor increase among the highest-impact journals, according to Jayaratne & Zwahlen.[1] Undeniably, more periodontal and implant research than ever is being published. This quantitative increase, however, seems to have had a wearing effect on the quality of articles published.

Information gathered from surveillance studies on the quality of reporting and methodologies used in published studies is unsettling.[3-5] In addition, the quality of evidence in published oral health systematic reviews that guide clinical recommendations is "predominantly low."[6] It becomes evident that together with the steep quantitative increase in the number of published studies, the need for an increase in the methodological robustness of published work is heightened. The present article will review the key methodological components of clinical studies and discuss study design and research methodology in the context of implant research.

2 | CLINICAL IMPLANT RESEARCH: THE "IMPLANT SURFACE" FACTOR

An important challenge faced by researchers endeavoring to conduct implant studies relates to the unique properties of metal oxide implant surfaces. Until the 1980s, much of the dental literature was centered around the study of pathology affecting living tissues composed of organic matrices with varying levels of mineralization (eg, periodontal ligament and cementum in periodontal research, and dentin in caries research). Given the considerable experience gained in research methodology at that time, existing dental study-design methodologies were directly applied to implant research without much consideration for the fundamental differences between implants and teeth. One such example relates to differences in the surface properties of titanium and dentin that affect bacterial adhesion and microbiome composition. As a result, the targeted search for known periodontal pathogens (eg, red complex bacteria) in peri-implant sulci when peri-implantitis was first recognized led to the belief that peri-implantitis was periodontitis around implants. It was not until much later, when more refined microbiologic methods with a broader scope of microbial identification were applied in peri-implantitis and unveiled, that the peri-implant and periodontal microbiomes were found to be fundamentally different.[7] Further discussion on this topic can be found in this volume of *Periodontology 2000* by Daubert & Weinstein.

Similarly, the application of tooth-designed approaches for biofilm removal from cementum and dentin surfaces led to major setbacks in developing efficacious treatment modalities for peri-implantitis. In fact, it was only recently demonstrated that cleaning devices suitable for use around teeth, such as ultrasonic metal tips, can have a hazardous effect when used on titanium implant surfaces.[8,9] The list of such examples is quite extensive; for instance, chemotherapeutic agents that are useful in the treatment of inflammation around teeth become adsorbed onto implant surfaces and exhibit

protracted cytotoxicity, as in the case of chlorhexidine,[10] whereas fluoride helps to prevent dental caries but leads to accelerated corrosion of titanium implants.[11] These examples serve to underline the fundamental differences that exist between teeth and implants, and highlight the implant-related variables that need to be considered in research design.

3 | CLINICAL RESEARCH DESIGN

Any research study is designed with the main goal of discovering the scientific truth regarding an important research question. In the context of implant and peri-implant research, clinical research studies are basically designed with the intent to study the association between an exposure and an implant-related outcome. Better said, "the essence in clinical research is to relate measures of disease occurrence to suspected causes or interventions."[12] For example, a study may be designed to seek if an association between diabetic status and implant osseointegration success exists. When planning such a study, many study designs can be employed to answer the research question. A critical first step is to determine the exposure and the outcome. In this instance, the exposure would be diabetic status and the outcome would be implant success. Study design can vary widely, but a basic distinction should be made between interventional and observational studies. As implied by the name of each category, the categorization is based on whether the researchers allocate the exposure, in which case it is referred to as an intervention, or if they merely observe and measure the exposure.

Selection of an appropriate study design to address a specific research question is dependent on various factors, often including practicality as one. In fact, clinical researchers view the selection of studies through the prism of the strength of their resulting conclusions. The dental cosmos is becoming increasingly familiarized with the buzzword "evidence-based".[2] As practitioners seek high-quality evidence to determine which therapies will consistently yield successful outcomes for the patients in their chairs, it becomes increasingly important to determine which studies are more likely to yield reproducible results (ie, are reliable).

The level-of-evidence pyramid has been constructed to rank clinical study designs according to their reliability.[13] Randomized clinical trials are regarded as the most reliable primary research design for studying the prevention and therapy of diseases. One of the main reasons why the strength of evidence derived from well-conducted randomized clinical trials is considered to be strong is that the randomization process ensures comparability between groups and minimizes selection bias. One example would be the use of a randomization process to ensure comparability of smokers between intervention groups in a two-arm randomized clinical trial comparing the effect of a new antimicrobial gel for the prevention of peri-implantitis. Consider a very small randomized clinical trial with a total of 4 participants, with 2 being smokers. The investigators decide to flip a coin to determine group allocation and it so happens that the flip of the coin assigns both smokers to the control group. Having the knowledge that smoking is a risk factor for peri-implantitis sheds doubt on the results of a positive effect of the mouthwash that are subsequently recorded in the randomized clinical trial. If, however, the same study was re-run with a sample size of 500 participants, half of whom were smokers, the chance of randomly allocating a comparable number of smokers into each group is very high. In fact, theoretically, as the sample size converges to infinity, the percentage of smokers per group would be identical. It becomes apparent that for randomization to serve its purpose, a reasonable sample size is necessary, which is, of course, also important for making inferences about a population mean based on the sample mean. Furthermore, more refined randomization strategies can be utilized in lieu of simple randomization to force balance in covariables of interest between groups. This is one of the many reasons why involvement of a biostatistician early in the design phase of a clinical study is critical.

The benefits of randomization are absent, by definition, in observational studies (in which the effect of an exposure is monitored by the researcher); this contrasts intervention studies, in which the

TABLE 1 Examples of study designs for examining the effects of diabetic status on implant osseointegration success

Clinical Question: Does diabetes mellitus affect implant success?	
Study type	**Example**
Retrospective cohort study	Researchers use an electronic dental record query to identify people with a history of diabetes mellitus and normoglycemic people (the control population) who presented for treatment in an implant clinic from 1990 to 2005. Subsequently, recall examination entries for the same people were analyzed from 2000 to 2015 to assess whether these implants were successful 10 years post-placement. The analysis reports the risk of implant failure for participants with diabetes mellitus compared with normoglycemic participants
Prospective cohort study	Researchers recruit edentulous people, with or without diabetes mellitus, and conduct implant placement. The participants are followed prospectively until implant success is assessed. The researchers then determine if participants with diabetes mellitus are at increased risk for implant failure compared with normoglycemic participants
Case-control study	People receiving care at a university clinic and who have successfully integrated implants (controls) or implants that failed to integrate (cases) are enrolled in the study. The researchers then review the medical records of participamts to identify any history of physician-diagnosed diabetes mellitus and thus determine whether participants with failed implants are more often diagnosed with diabetes mellitus than participants with successful implants

intervention is randomly assigned and monitored by investigators. Several observational studies may arise depending on the temporal association between ascertainment of exposure and outcome. Utilizing the example, discussed above, of implant success and diabetic status, examples of observational study designs are presented in Table 1.

As previously mentioned, observational studies are defined by the absence of researchers' influence on exposure allocation. In the examples presented in Table 1 it becomes obvious that an investigator could not ethically expose study participants to medications used to induce diabetes to then test its effects on implant success, in contrast to the induction of diabetes by streptozotocin injection in animal models. Consider the differences between people with diabetes and normoglycemic people (eg, age, metabolic status, comorbidities) and their effect on implant success. Variables that are associated both with the exposure variable and the disease variable are referred to as "confounders". When confounding is present it may lead to biased estimates for the association between exposure and outcome. Statistical approaches for obtaining unbiased estimates in the presence of extraneous variables are beyond the scope of this article. Therefore, we will discuss strategies to minimize confounding that can be incorporated into the study design.

Moreover, there are various other sources of bias that can affect the validity of a clinical study. Minimizing these sources of bias makes it more likely for the given study to record true research findings in the observed study population without overestimating or underestimating these findings as a result of bias in the design, measurement, analysis, or presentation of the data.[14] An example of a form of bias is "detection bias". Consider an implant surgeon who prefers to place standard-length implants in the posterior maxilla, in conjunction with sinus elevation procedures, to avoid placing short implants. The sinus augmentation plus implant placement procedure performs well in their hands and they subsequently decide to join a multicenter randomized clinical trial in which participants are randomized to either implant placement simultaneously with sinus augmentation surgery or short implant placement in native bone. The implant surgeon places the implants as well as functions as an examiner to assess participants' postoperative pain and clinical signs of edema over the first 2 weeks after the operation. As this researcher was involved in the surgery and therefore has prior knowledge of the randomized intervention, effective masking of the assessor is not possible and it is likely that their prior knowledge may influence their assessment of these subjective outcomes.[15] In fact, Juni et al[15] used meta-data from trials comparing low-molecular-weight heparin with standard heparin for the prevention of postoperative thrombosis and showed that unmasked assessment of outcome inflated the effect of low-molecular-weight heparin by 35% (95% confidence interval: 1%-57%).

In the next section, the main types of study design that are applicable to implant research will be discussed, together with examples to highlight how to minimize risk of bias.

3.1 | Case reports and case series

Case reports and case series are very often utilized in all aspects of medical and dental science. They are valuable in providing proof-of-principle data, challenging prevailing concepts, and reporting disease symptoms or novel therapeutic approaches. The value of such studies in clinical education in an application-driven field, such as periodontology and implantology, should not be disregarded. A high-quality case report (also referred to as a case study) is informative and maintains its scientific rigor. Even the most prestigious medical journals accept and publish informative case reports. For instance, The Lancet summarizes the values of a quality case report as follows: "The ideal Lancet Case Report is of general, not specialist interest. It tells a clinical story of a difficult differential diagnosis in an engaging and concise manner, while respecting the dignity of the patient. Novelty is not essential, but at least one broadly useful learning point is".[16] One example is a case study of a young woman experiencing premature exfoliation of deciduous teeth and alveolar bone loss with a medical history of positive upper respiratory tract infections, tonsillitis, and recurrent otitis media.[17] The authors presented a structured diagnostic approach to construct a differential diagnosis list, together with results from targeted clinical, radiographic, immunologic, and microbiologic assays that supported the diagnosis of periodontitis as a manifestation of chronic idiopathic neutropenia.[17,18] The conclusion of the same case report followed the patient post-treatment and provided results of the proposed therapeutic protocol.[17] The contribution of case series and case reports is particularly impactful in describing findings from sample populations with rare diseases as it may be challenging to recruit such populations in controlled studies. However, moderation is the key; case reports should not be viewed as an easy-to-publish study type, but rather as useful tools with limited applications.

Another example demonstrating the benefits of a case study is the report of novel therapeutic protocols in the form of proof-of-principle studies.[19] Such feasibility studies gather preliminary clinical data to support that a biomaterial or a therapeutic protocol truly demonstrates the desired effect on a pathophysiologic mechanism that has been postulated in preclinical experiments. One such example is a case study describing the acceleration of post-extraction socket healing and minimization of overall treatment time following delivery of autologous growth factors in a fibrin matrix.[20] The clinical steps were illustrated for researchers to be able to reproduce the therapeutic protocol, and histologic data that supported a prior biologic hypothesis were gathered.[20] This type of study is of limited use and can provide proof-of-principle data in support of the feasibility of this therapeutic protocol, but by no means does it yield any information on the reproducibility of this therapeutic approach. As noted by the authors in the conclusions of the aforementioned study, higher-level evidence is warranted to assess the reproducibility of this treatment concept and to identify specific indications where its use may lead to significant clinical results.[20] Obviously, the publication of a similar case report after the publication of the first would be morbidly repetitious.

A major shortcoming of case reports and case series is the lack of reproducibility of their findings. For instance, a series of N = 10 nonconsecutively treated cases by means of a novel therapeutic intervention yields weak evidence on how consistently such favorable

outcomes can be achieved in routine practice by clinicians other than the authors. Consistency is, of course, the key; as biomaterials science and clinical techniques advance, it is beyond any doubt that thaumaturgy often takes place in implant practice and unexpectedly favorable results are much more likely to be published. What matters is how predictably favorable outcomes can be accomplished compared with an established treatment alternative. The lack of comparators in case series hampers the feasibility of determining that the observed effect was not the result of an extraneous exposure.[12] Higher-level studies that employ comparator arms and controlled interventions are necessary to justify clinical recommendations.

3.2 | Observational studies

In observational studies, investigators seek to identify associations between exposures (ie, factors that are thought to affect outcomes of interest), and disease variables. The term "disease variable" may not always be the ideal term to describe the outcome variable in implant research, as this variable may be related to the opposite of disease. Specifically, in the majority of published implant studies, investigators seek to determine implant success, which is a state of health. One would think that regardless of whether the outcome of an implant study is seeking to determine states of health (eg, implant success) or disease (eg, peri-implantitis), the definitions of health and disease should be concrete. However, this is nearly the case in implant research.

When root-form implants became mainstream following the Toronto conference, the primary consideration was whether they could successfully integrate and retain prostheses without clinical mobility. This led to a set of implant survival criteria that called for implant stability followed by a slow progression of bone loss around the implant-abutment connection as sufficient criteria to determine implant survival.[21] These criteria for implant survival were authored in the 1980s and are still in use today. Although very useful for establishing implant therapy efficacy 40 years ago, one would think that advancements in implant dentistry warrant further scrutiny regarding the clinical outcomes of implants. For instance, advances in implant design[13] and evolution of clinical protocols[22] may limit the expected remodeling around implants that would be considered as within normal limits in 1986. In fact, more strict criteria for implant success, rather than mere implant "survival", have been established.[23] It is quintessential for the dissemination of clinically relevant information for researchers to use relevant success criteria, the selection of which is related to the subject of research. For example, implant placement in the esthetic region requires consideration of criteria of esthetic implant success.[16,24] Furthermore, patient-perceived

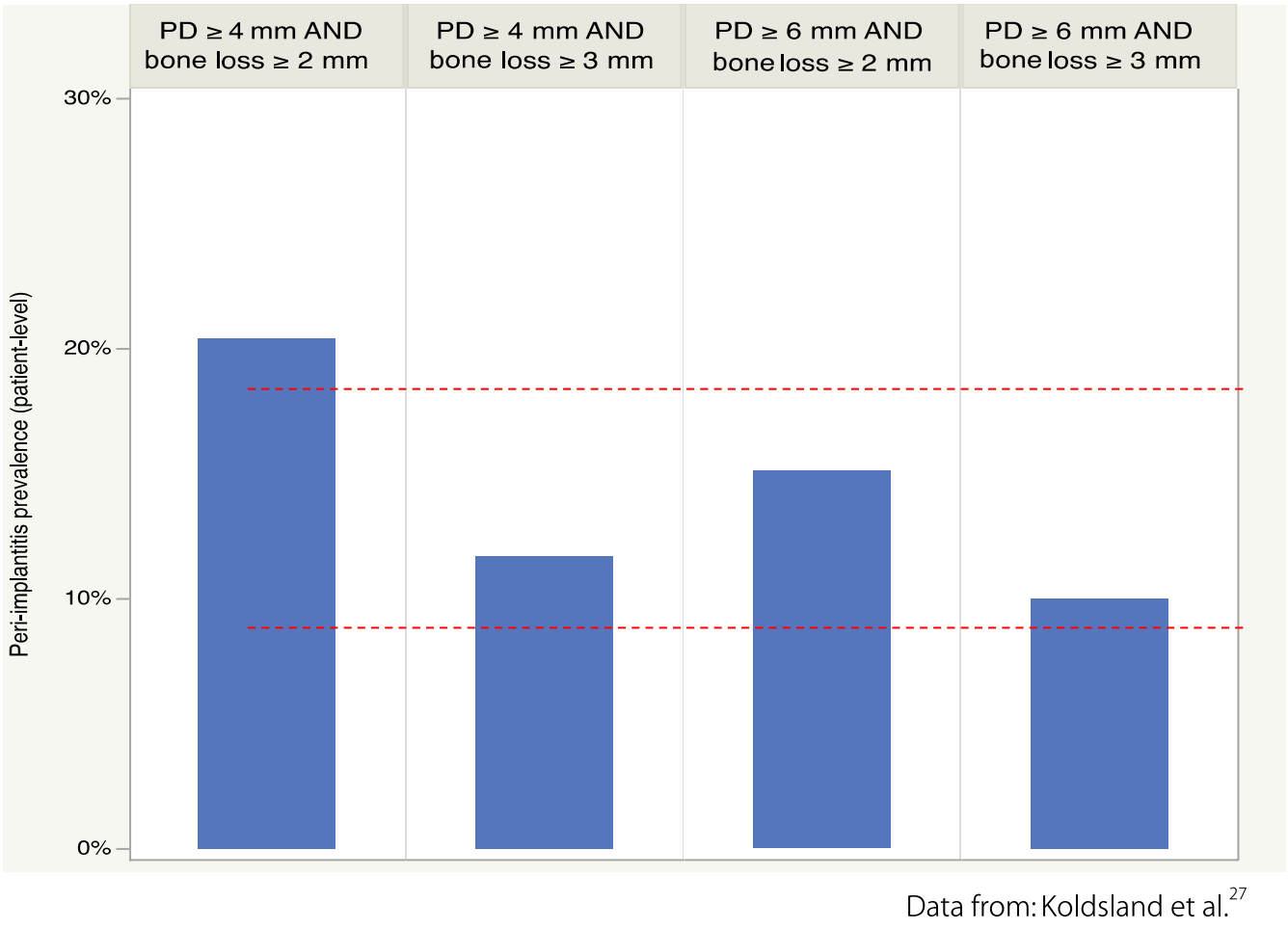

FIGURE 1 Prevalence estimates demonstrate volatility that is dependent on disease case definitions. PD, probing depth

outcomes should be taken into account in conjunction with clinical outcomes.[25,26]

Another important consideration for clinical research design in implant dentistry is selection of case definitions that accurately capture disease experience and provide comparability across studies. This is particularly challenging in peri-implantitis research because there is great heterogeneity in case definitions because of the absence of a clear consensus. Case definitions of peri-implant health and disease are not only important for recruitment of people into interventional studies but are also fundamental in observational studies for determining the disease experience of sample populations. A well-known example of the challenges associated with the heterogeneity in case definitions of peri-implantitis was published by Koldsland et al.[27] In their study, the authors employed 4 distinct definitions of peri-implantitis that differed based on the probing depth and bone loss thresholds used to define disease. The results of this study showed tremendous variability in prevalence estimates across the different definitions, which would have been of sufficient magnitude to affect public health policies (Figure 1).

Currently, the definition of peri-implantitis published in the Proceedings of the 8th European Workshop of Periodontology is considered to be well accepted and can serve as a reference point for comparability among studies. Based on this definition, "changes in the level of crestal bone, presence of bleeding on probing and/or suppuration with or without concomitant deepening of peri-implant pockets" define peri-implantitis,[28] with a threshold of 2 mm of bone loss being considered diagnostic in the absence of baseline radiographic records.

3.2.1 | Cross-sectional studies

Cross-sectional studies, also called prevalence studies, are primarily descriptive, meaning that they provide estimates of disease frequency according to time and place. Prevalence is a snapshot capturing the number of people having a condition at a given point in time (ie cases) proportionate to the size of the sample population. For instance, a recent cross-sectional study reported that the prevalence of peri-implantitis was 26% among people with implant-supported restorations visiting the periodontology clinic at the University of Washington between September 2011 and March 2013.[29] The distinct components of a cross-sectional study were as reported in the previous sentence: definition of a sample population (people with implant-supported restorations visiting the periodontology clinic at the University of Washington); time period of assessment (September 2011 to March 2013); and outcome of interest (peri-implantitis). Prevalence studies are important for determining the magnitude of a characteristic of interest (eg, peri-implantitis prevalence) at a population level. An important source of bias that has to be minimized in order to obtain realistic prevalence estimates in descriptive cross-sectional studies is survival bias.[30] The term "survival" highlights the fact that cases which did not survive (eg, failed and explanted implants) are not captured in this cross-sectional assessment and may lead to biased prevalence estimates (eg, underestimation of the prevalence of peri-implantitis).

Cross-sectional studies also allow the opportunity for limited analytic study of associations between a putative risk factor and an outcome, within certain limitations, thus presenting a step-forward to preclinical research. Preclinical data or expert opinions may purportedly suggest associations that cannot often be replicated in clinical studies. Cross-sectional studies offer a convenient and resource-efficient tool to assess such associations under clinical conditions before devoting the effort and funds required for a longitudinal study. For example, preclinical data in implant dentistry have shown that dental cements attract bacteria which may be implicated in peri-implantitis.[31] This information was further corroborated by mechanistic or proof-of-concept studies, and was often interpreted as guidelines to favor screw-retained restorations for implants. Although valuable in gathering preliminary data, such studies cannot, and should not, be used to guide clinical recommendations. Obviously, in human peri-implant tissues, the interplay among immune cells, host enzymes, bacteria, and environmental factors cannot be replicated even in the finest model. Then come the cross-sectional studies to put preclinical findings to the test as the initial form of clinical investigation. Such studies come to rescue researchers from the exhaustive, and possibly vain, efforts to put together a randomized clinical trial that will point to the null.

In the case of cement and peri-implantitis, cross-sectional studies consistently pointed to the direction that the culprit was not cement retention compared with screw retention, but rather a specific type of cement; methacrylate.[29,32,33] In fact, a subsequent interventional study was even successful in proving causality.[34] Researchers followed existing data indicating increased likelihood of excess cement when methacrylate cements were used, and showed that when excess methacrylate cement caused peri-implant inflammation, this could be treated, 77% of the time, by simply removing the excess cement and re-cementing with a zinc-based cement.[34] Then, clinical research passed the torch back to preclinical research to confirm the biologic plausibility and dissect the etiologic mechanisms of the observed clinical findings. Indeed, microbiome analyses revealed that, compared with zinc-based cements, methacrylate cements favored the growth of periodontal pathogens, therefore corroborating the clinical findings.[35,36] Thus, now the tools are in place and adequate preliminary data have been gathered to pursue the question of whether resin cements are associated with peri-implantitis in longitudinal studies; a question considering prior information without confirmation bias and has a better chance to identify a true association.

A common way to determine analytic characteristics in the sample population in a cross-sectional study is the use of odds ratios (ORs). Crude ORs are easily computed from contingency tables. They are defined as the quotient of the odds of being diseased vs not being diseased among the exposed population divided by the odds of being diseased vs healthy in the nonexposed population. A typical example is found in large population surveys that aim to identify associations between various variables and disease.

TABLE 2 Core of the analytical component of a cross-sectional study

	Severe periodontitis	Periodontal health
Low-dose aspirin intake	Pa	Ha
No aspirin use	Pn	Hn

Ha, periodontally healthy among aspirin users; Pn, periodontally healthy among nonusers of aspirin; Pa, periodontitis cases among aspirin users; Pn, periodontitis cases among nonusers of aspirin.

For instance, a recent exploratory analysis of the National Health and Nutrition Examination Survey data aimed to determine if an association exists between low-dose aspirin intake and prevalent periodontitis.[37]

From the information shown in Table 2, the OR is defined as (periodontitis cases among aspirin users/periodontally healthy among aspirin users)/(periodontitis cases among nonusers of aspirin/periodontally healthy among nonusers of aspirin). In this specific example from an analysis of the 2011-2012 National Health and Nutrition Examination Survey population, the crude OR was 1.27 (95% confidence interval: 0.99-1.63).[37] Considering its definition, an OR = 1 would imply identical odds for severe periodontitis in aspirin users and nonusers of aspirin. Thus, it becomes obvious that as the OR (0.99-1.63) was inclusive of the value 1, it is likely that the odds of having periodontitis were not associated with intake of low-dose aspirin. However, the fact that the lower extent of the 95% confidence interval is very close to "1" (in this case "0.99") may reveal a trend for an association in this sample population. Importantly, the influence of confounding on the observed estimates from a nonrandomized study needs to be considered carefully, as discussed below.

Limitations

Confounding. The OR reported for the periodontitis-aspirin association in the aforementioned examples was referred to as crude. The term "crude" contrasts to "adjusted", in which the effect estimates are adjusted to consider the effect of additional variables. Such confounding variables are defined as extraneous variables that can explain the relationship between an exposure and an outcome.

Consider, for instance, a cross-sectional study in which the investigators report an OR of 2.0 (95% confidence interval: 1.5-3.5) for periodontitis and emphysema, which is a subtype of chronic obstructive pulmonary disease. This translates into people with periodontitis being twice as likely as periodontally healthy people to have emphysema. A cursory review of this finding may lead to the conclusion that periodontitis may be a risk factor for emphysema, or vice versa. Nevertheless, as smoking is considered to be the leading cause of emphysema as well as a major risk factor for periodontitis, careful assessment of the sample population is needed. If the percentage of smokers in this hypothetical sample population was significantly greater among periodontitis patients vs nonperiodontitis patients, consideration of smoking in the analysis would negate the observed association.

In interventional studies, appropriate randomization is expected to lead to equal distribution of patient characteristics in the groups. Nonetheless, in nonrandomized studies, disproportionate distribution of confounders in the study groups can significantly overestimate or underestimate the observed associations, or even negate them. Appropriate adjustment for measurable confounders is necessary in the analyses of such studies. In the periodontitis-aspirin example mentioned above, the trend for association observed in the crude analysis was further attenuated when adjusting for known confounders, such as age, gender, ethnicity, socioeconomic status, and smoking, with an adjusted odds ratio of OR = 1.06 (95% confidence interval: 0.74-1.50).[37] Even in the presence of adjusted analyses, the results of nonrandomized studies should be interpreted conservatively because model-based adjustment may not be sufficient to minimize confounding bias.[38] Furthermore, adjusted analyses are inherently limited to confounders observed during the study; thus, residual or unmeasured confounding may distort the findings and lead to spurious associations.[38,39]

Temporality. Sir Bradford Hill[40] suggested 9 key criteria for determining whether the information available is of adequate strength to establish a causal association between an exposure and an outcome. One of the key criteria for ascertaining causality namely, temporality, requires establishing that the exposure precedes the outcome.[41] For example, consider a cross-sectional study assessing people with a lack of attached gingiva in their mandibular anterior dentition that identifies a significant association between increased plaque index and lack of attached gingiva. Consideration of plaque as a risk factor for a lack of attached gingiva is biologically plausible. However, and given that the data arise from a snapshot in time, it is impossible to determine if plaque accumulation indeed preceded the loss of attached gingiva. It is equally likely that the lack of attached gingiva occurred for reasons other than plaque accumulation, and that the, affected people altered their oral hygiene habits accordingly as a result of the lack of attached gingiva. Sir Bradford Hill, himself, initially seemed to underestimate the importance of temporal precedence, stating that "This temporal problem may not arise often, but it certainly needs to be remembered…".[40] The issue of temporal precedence inarguably remains the most elusive among the components necessary to determine a cause-and-effect relationship since Aristotle's chicken or the egg dilemma.

3.2.2 | Cohort studies

Cohort studies are designed based on the hypothesis that an exposure is associated with the disease of interest. Participants are stratified into groups ("cohorts") based on the exposure of interest (eg, smoking) and are then followed over time to determine if the risk for developing the disease of interest (eg, peri-implantitis) is increased in the presence (or absence if the exposure is protective, such as vitamin D supplementation) of the exposure. The temporal component of these studies allows calculation of the incidence of disease, which

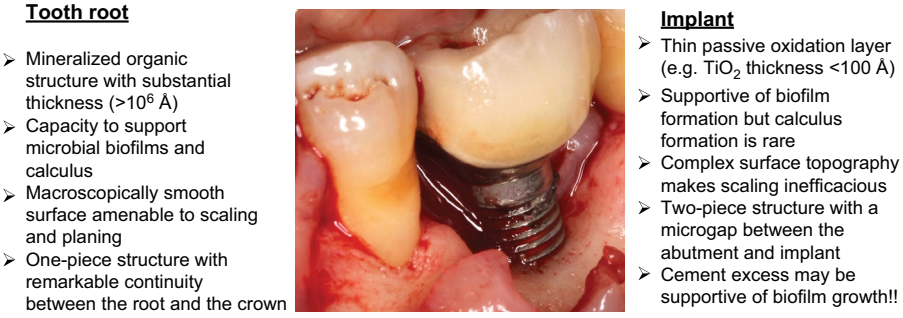

FIGURE 2 Differences between teeth and implants that need to be considered when designing the study

is defined as the number of incident cases/number of people initially at risk. Incidence is dynamic; it is defined as the fraction of people who develop the disease of interest over a given time period. This contrasts with prevalence, which was discussed in the context of cross-sectional studies and describes the number of people having the disease of interest at a given point in time.

Cohort studies are among the most powerful observational study designs, specifically because of that temporal component which can provide cues for causal associations between exposure and outcome. However, the strength of the evidence depends on how comparable/exchangeable the groups are. For example, consider two conflicting studies, with one identifying a significant association between loss of implants due to peri-implantitis in persons with a history of periodontitis, whereas a second study found no association. If the 2 groups were imbalanced for variables that may have affected the outcome, then the differences in the observed associations may be explained by these factors. When determining these variables in implant research, it becomes increasingly important to consider the intricacies related to implants as compared with teeth, as depicted in Figure 2. Risk factors intrinsic to peri-implantitis must be considered, and their allocation between groups of interest should be balanced. Examples of such factors include the type of prosthesis retention (i.e cement-retained vs screw-retained restorations) and the type of implant surface (e.g hydroxyapatite-coated vs machined titanium surface).

Limitations

In cohort studies, the allocation of exposure is beyond the investigators' control. Therefore, increased vigilance is necessary to ascertain that a balance between the 2 groups exists. The optimal time to consider this balance is when the study is being designed. Investigators should make sure that no selection bias occurs due to enrollment strategies. One example would be a study in which investigators want to know if diabetes mellitus is associated with increased risk for peri-implantitis and plan to enroll participants receiving implants in 3 dental practices. Two of the practices place implants that receive screw-retained restorations, whereas the third places implants that receive cement-retained restorations. For the sake of this example, it is hypothesized that the first 2 practices serve populations with increased obesity burden compared with the third practice, and indeed an increased association was identified between diabetes and risk for peri-implantitis after 5 years of implant placement. How could one be certain that the increased risk for peri-implantitis is truly due to diabetic status and not because of screw retention? There are 2 strategies to address this imbalance in observational studies. The first would be to make an effort to enroll participants from clinical centers that perform as comparable treatment as possible with the hope that this would lead to balance between implant variables that can be assessed. The second strategy is to address imbalances that may arise at the data-analysis stage, as described above, under "cross-sectional studies". Such adjusted analyses are most often necessary in observational studies because achieving balance between the groups is practically very challenging, despite careful enrollment strategies.

3.2.3 | Case-control studies

In the previous section it was discussed that cohort studies are designed based on the hypothesis that an exposure is associated with the disease of interest. Cohorts are defined based on exposure and are followed over time to determine the risk for disease occurrence. Case-control studies, on the other hand, are grounded upon categorization of the sample population into 2 groups based on their disease status: participants who have the disease/condition compose the "cases", whereas participants who are at risk for the disease, but do not have the disease, are the "controls" (Figure 3).[42]

The focus of a case-control study is to determine whether prior exposure to suspected causal factors differs between cases and controls. An example from the oral health literature is a case-control study that sampled cases and controls from a practice-based research network to identify risk factors for osteonecrosis of the jaw.[43] The researchers selected samples from various clinical centers and identified 3 controls per case from the same clinical centers. In case-control studies, researchers often elect to sample more controls per case (even when investigating relatively rare diseases) as a convenient way to increase the study's power. To minimize selection bias it is important that both the control and the case participants are selected from the population at risk for the disease.[44] A reasonable question to ask at the time of designing a case-control study is: "had a control hypothetically developed the disease, would it have been selected as a case?"

As in cross-sectional and cohort studies, a consideration that arises in case-control studies is the comparability of baseline characteristics between controls and cases. Comparability of cases and

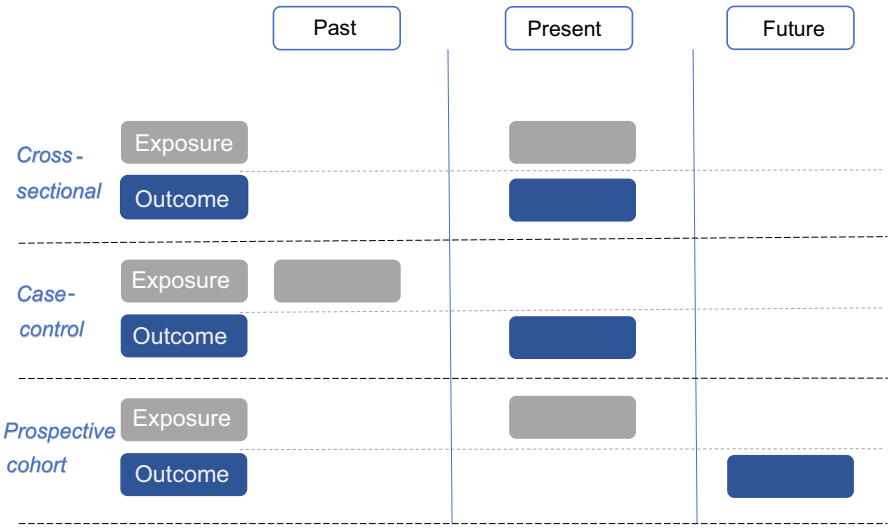

FIGURE 3 Differences in temporality among observational study designs

controls should be carefully assessed, either at the level of participant selection or at the level of data analysis. A pre-hoc approach to ascertain comparability is matching of cases and controls. For instance, selection of control participants in a case-control study evaluating a periodontitis risk factor could be performed with the intent to match them to cases according to age, race, smoking status, and diabetic status, all of which are known risk factors for periodontitis. Nonetheless, this approach can be unwieldy and lengthen the sample recruitment period with a nonnegligible cost in funds and resources. An alternative approach is post-hoc balancing of study characteristics among cases and controls. In the osteonecrosis case-control study, an analytic approach that is known as propensity score matching was employed to determine whether imbalance in the distribution of conditions for which bisphosphonates are prescribed affected the observed association between bisphosphonates and osteonecrosis.[43]

Limitations

Case-control studies are attractive studies that offer the opportunity to include a comparator group and observe chronic diseases, but not without the cost. They warrant due attention at the time of study design and exposure assessment to minimize bias. Selection bias has already been discussed above. Moreover, the retrospective assessment of risk factors that is interwoven in a case-control design limits their reliability. This is particularly true when the assessment of prior exposure is based on participant recollection.[12] The "recall bias" refers to the fact that patients affected by the disease (ie, cases) are more likely to recall past exposures than controls. Minimization of recall bias requires careful assessment of exposure and cross-validation of participant-reported exposure using medical or dental records. Due to these limitations, case-control studies, albeit valuable, are considered to yield low-to-moderate evidence.

3.3 | Interventional studies

The term "clinical trial" can refer to either one of two study designs: nonrandomized clinical trials or randomized controlled clinical trials. Well-conducted randomized controlled clinical are considered to hold the highest level of evidence. The power to determine the interventions being tested in a randomized controlled clinical trial offers the opportunity to study a research question in a well-controlled environment. In addition, appropriate randomization ensures that comparators are well balanced for both known and unknown confounders. In certain instances, however, the use of randomization is not feasible or ethical. When the intervention is applied in groups that are not randomly allocated, use of these interventional studies is considered as lacking scientific rigor compared with randomized controlled clinical trials but more highly than observational studies.[45] In nonrandomized trials, the study design bears much resemblance to a cohort study, with the exception of the researchers introducing a predetermined intervention. The major shortcoming of nonrandomized studies is the potential for selection bias.[45] One such example would be a public media campaign to promote annual dental check-ups being tested as a public health intervention for oral disease prevention in various states in the USA. Researchers interested in the effect of this intervention could recruit people from the states that were included in the campaign in the test group and compare them with a control group composed of residents of states not included in the campaign against relevant oral health outcomes. The group allocation is determined according to state residency and cannot be determined by chance in this hypothetical example.

A well-designed and adequately powered randomized controlled clinical trial is considered to be the most authoritative study type because appropriate randomization minimizes the risk for selection and confounding biases.[45] Randomized controlled clinical trials are employed to gather definitive data on the effect of an intervention. There are many instances in which a clinical trial has negated findings observed in preclinical studies or studies using unsuccessful surrogate outcomes that had actually already been used as clinical recommendations despite their weak strength. For example, ample in vitro and in vivo studies had provided information suggesting that chlorhexidine products for oral use were efficacious against *Streptococcus*

mutans.[46] Such findings had been put to use prematurely as clinical recommendations for the use of chlorhexidine for caries prevention, without prior confirmation in high-level clinical studies. When a chlorhexidine mouthwash was assessed as an intervention to prevent tooth loss in people aged ≥60 years, the results negated the aforementioned preclinical findings; in fact, they showed that the use of chlorhexidine rinses did not have a substantial effect on decreasing the number of filled, decayed, or extracted teeth in older adults.[46]

4 | KEY PRINCIPLES OF RANDOMIZED STUDY DESIGN

Designing a clinical trial entails several key methodologic components. Following inception of the research hypothesis, 2 core variables have to be determined: the intervention; and the outcome. In a classic 2-arm randomized controlled clinical trial, the investigators select an intervention and an appropriate control, with the premise that their head-to-head comparison will definitively answer the research question. The ideal control would be identical to the intervention in every possible way with the exception of only having a placebo effect. However, ethical concerns or practical issues often hamper such an ideal scenario. One example of a placebo-controlled randomized controlled clinical trial in implant research would be assessment of the effect of systemic antibiotics on the treatment of peri-implantitis. In a placebo-controlled trial, the investigators would randomly assign participants to receive a 7-day course of opaque white capsules containing either a systemic antibiotic (intervention arm) or nonactive ingredients (eg, sugar pill; control arm); in both groups the capsules would be taken at the same frequency. Such an approach would, in fact, allow for 2 levels of masking, namely at participant level and at assessor level.

Returning to the selection of the intervention and the outcome, it is important to discuss biases preceding study design.[14] As mentioned above, randomized controlled clinical trials are resource-intensive studies. Therefore, before initiating a randomized controlled clinical trial, one should consider whether the research hypothesis is significant, innovative, and worthwhile pursuing. If the clinical relevance or the clinical utility that is derived is minimal, then the allocation of resources to such a research hypothesis is probably not going to be rewarding. Another critical component of prestudy design that is lacking in the field of periodontal and implant research is failure to consider prior evidence.[14] Our research field is fundamentally different from the medical field regarding the resources allocated to clinical researchers and fund availability. Federal funds for periodontal and implant clinical research are very limited; pharmaceutical companies that are among the top-funders for medical research in most instances do not have a direct interest in periodontal research. Other subtleties of the field exist, such as shorter residency time for dental residents than for medical residents, who are often the powerhouses in clinical studies.

All of these pressing factors shape a resource-limited landscape, while the fast-paced periodontal and implant research field often forces researchers to generate publications in a fight for academic survival. Avoidance of repetitious or underpowered studies due to lack of consideration of previously reported effect sizes could be easily overcome with updated systematic reviews and meta-analyses of existing data. This is, however, hampered by the alarmingly increasing rate of systematic reviews vs original studies published in dental journals. Albeit oxymoronic at a first read, the large number of systematic reviews generated on topics related to periodontology and implantology muddies the water. A fraction of these review studies are often offshoots from research groups with expertise in the methodology of conducting secondary research but they lack the context expertise to make them meaningful. Undeniably, systematic reviews are powerful and much-needed studies, and as such their misuse may carry a disproportionate hazard.

How then can prior evidence be appropriately utilized to design randomized controlled clinical trials? Once a research question that has scientific merit has been formulated, the relevant literature is assessed to determine the most meaningful outcomes for the condition being assessed. The most clinically relevant outcome should be selected to be the primary outcome. Careful interpretation of the phrase "clinically relevant" is warranted. Often, outcomes that can be well defined and objectively assessed are considered to be clinically relevant. One such example is the use of histomorphometric analyses in implants research. Obtaining a tissue sample, staining it, and analyzing it under a microscope is a procedure amenable to a high level of standardization, as it is almost completely restricted to ex vivo analyses. Nevertheless, there is no clear correlation between histomorphometric measurements, e.g surface area of new bone formation, and successful functional rehabilitation of an edentulous site and/or increment in patient-perceived satisfaction. The consideration of inclusion of true outcomes (ie, outcomes that are tangible to the patient), in periodontal and implant research is still in its incipient stages and needs to be strongly supported.[47]

5 | CONSIDERATIONS REGARDING BIAS IN CLINICAL TRIALS

It is well accepted that clinical trials are intensive research designs that, to be conducted with credibility, require interdisciplinary teams of experienced investigators, support staff, coordination centers, and oversight and monitoring boards. As such, they are the most resource-demanding study designs. To put this into perspective, the National Institutes of Health offers grants that span over a year to provide the necessary resources for researchers to set up the logistics of a clinical trial. The field of implant dentistry is quite challenged though, as federal funding for implant research is scarce. As a result, researchers often rely on industry sponsors as a resource for clinical trial funding. The support of industry to clinical research is much needed and offers a unique opportunity to further our knowledge on the efficacy of implant therapy and to identify efficacious techniques or biomaterials that hold merit for clinical use.

Nonetheless, such support should be regulated. One example of the effect of sponsorship on reported outcomes is given by.[48] Popelut et al[48]; they reported that based on meta-data from 38 randomized controlled clinical trials, industry-associated trials demonstrated annual failure rates that were 80% lower than those reported in non-sponsored trials.

ACKNOWLEDGMENTS

The author would like to thank Ms Sophia Nguyen, Clinical Periodontal Research Laboratory, University of Washington, Seattle, WA, for her editorial assistance in preparing this manuscript.

REFERENCES

1. Jayaratne YS, Zwahlen RA. The evolution of dental journals from 2003 to 2012: a bibliometric analysis. *PLoS One*. 2015;10(3):e0119503.
2. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926.
3. Al-Namankany AA, Ashley P, Moles DR, Parekh S. Assessment of the quality of reporting of randomized clinical trials in paediatric dentistry journals. *Int J Pediatr Dent*. 2009;19(5):318-324.
4. Savithra P, Nagesh LS. Have CONSORT guidelines improved the quality of reporting of randomised controlled trials published in public health dentistry journals? *Oral Health Prev Dent*. 2013;11(2):95-103.
5. Kotsakis GA. Mandatory trial registration in oral health research: long overdue? *J Am Dent Assoc*. 2017;148(6):353-356.
6. Pandis N, Fleming PS, Worthington H, Salanti G. The quality of the evidence according to GRADE is predominantly low or very low in oral health systematic reviews. *PLoS One*. 2015;10(7):e0131644.
7. Kumar PS, Mason MR, Brooker MR, O'Brien K. Pyrosequencing reveals unique microbial signatures associated with healthy and failing dental implants. *J Clin Periodontol*. 2012;39(5):425-433.
8. Eger M, Sterer N, Liron T, Kohavi D, Gabet Y. Scaling of titanium implants entrains inflammation-induced osteolysis. *Sci Rep*. 2017;7:39612.
9. Safioti LM, Kotsakis GA, Pozhitkov AE, Chung WO, Daubert DM. Increased levels of dissolved titanium are associated with peri-implantitis – a cross-sectional study. *J Periodontol*. 2017;88(5):436-442.
10. Kotsakis GA, Lan C, Barbosa J, et al. Antimicrobial agents used in the treatment of peri-implantitis alter the physicochemistry and cytocompatibility of titanium surfaces. *J Periodontol*. 2016;87(7):809-819.
11. Noguti J, de Oliveira F, Peres RC, Renno AC, Ribeiro DA. The role of fluoride on the process of titanium corrosion in oral cavity. *Biometals*. 2012;25(5):859-862.
12. Matthews DC, Hujoel PP. A practitioner's guide to developing critical appraisal skills: observational studies. *J Am Dent Assoc*. 2012;143(7):784-786.
13. Annibali S, Bignozzi I, Cristalli MP, Graziani F, La Monaca G, Polimeni A. Peri-implant marginal bone level: a systematic review and meta-analysis of studies comparing platform switching versus conventionally restored implants. *J Clin Periodontol*. 2012;39(11):1097-1113.
14. Ioannidis JP. Perfect study, poor evidence: interpretation of biases preceding study design. *Semin Hematol*. 2008;45(3):160-166.
15. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*. 1999;282(11):1054-1060.
16. Belser UC, Grutter L, Vailati F, Bornstein MM, Weber HP, Buser D. Outcome evaluation of early placed maxillary anterior single-tooth implants using objective esthetic criteria: a cross-sectional, retrospective study in 45 patients with a 2- to 4-year follow-up using pink and white esthetic scores. *J Periodontol*. 2009;80(1):140-151.
17. Kamma JJ, Lygidakis NA, Nakou M. Subgingival microflora and treatment in prepubertal periodontitis associated with chronic idiopathic neutropenia. *J Clin Periodontol*. 1998;25(9):759-765.
18. James E, Hinrichs GAK. Classification of diseases and conditions affecting the periodontium. In: Newman M, Takei H, Klokkevold P, Carranza F, eds. *Clinical Periodontology*, Chapter 3, 12th edn. Amsterdam: Elsevier; 2017.
19. Schmidt B. Proof of principle studies. *Epilepsy Res*. 2006;68(1):48-52.
20. Kotsakis GA, Boufidou F, Hinrichs JE, Prasad HS, Rohrer M, Tosios KI. Extraction socket management utilizing platelet rich fibrin: a proof-of-principle study of the "accelerated-early implant placement" concept. *J Oral Implantol*. 2016;42(2):164-168.
21. Albrektsson T, Jansson T, Lekholm U. Osseointegrated dental implants. *Dent Clin North Am*. 1986;30(1):151-174.
22. Koutouzis T, Gholami F, Reynolds J, Lundgren T, Kotsakis GA. Abutment disconnection/reconnection affects peri-implant marginal bone levels: a meta-analysis. *Int J Oral Maxillofac Implants*. 2017;32(3):575-581.
23. Karoussis IK, Bragger U, Salvi GE, Burgin W, Lang NP. Effect of implant design on survival and success rates of titanium oral implants: a 10-year prospective cohort study of the ITI dental implant system. *Clin Oral Implants Res*. 2004;15(1):8-17.
24. Konstantinidis IK, Siormpas KD, Kontsiotou-Siormpa E, Mitsias ME, Kotsakis GA. Long-term esthetic evaluation of the roll flap technique in the implant rehabilitation of patients with agenesis of maxillary lateral incisors: 10-year follow-up. *Int J Oral Maxillofac Implants*. 2016;31(4):820-826.
25. Babbush CA, Kanawati A, Kotsakis GA, Hinrichs JE. Patient-related and financial outcomes analysis of conventional full-arch rehabilitation versus the All-on-4 concept: a cohort study. *Implant Dent*. 2014;23(2):218-224.
26. Alzoubi F, Bedrossian E, Wong A, Ferrell D, Park C, Indresano T. Outcomes assessment of treating completely edentulous patients with a fixed implant-supported profile prosthesis utilizing a graftless approach. Part 2: patient-related outcomes. *Int J Oral Maxillofac Implants*. 2017;32(5):1080-1085.
27. Koldsland OC, Scheie AA, Aass AM. Prevalence of peri-implantitis related to severity of the disease with different degrees of bone loss. *J Periodontol*. 2010;81(2):231-238.
28. Sanz M, Chapple IL, Working Group 4 of the VEWoP. Clinical research on peri-implant diseases: consensus report of Working Group 4. *J Clin Periodontol*. 2012;39(Suppl. 12):202-206.
29. Daubert DM, Weinstein BF, Bordin S, Leroux BG, Flemming TF. Prevalence and predictive factors for peri-implant disease and implant failure: a cross-sectional analysis. *J Periodontol*. 2015;86(3):337-347.
30. Saracci R. Survival-related biases survive well. *Int J Epidemiol*. 2007;36(1):244-246.
31. Raval NC, Wadhwani CP, Jain S, Darveau RP. The interaction of implant luting cements and oral bacteria linked to peri-implant disease: an in vitro analysis of planktonic and biofilm growth – a preliminary study. *Clin Implant Dent Relat Res*. 2015;17(6):1029-1035.
32. Kotsakis GA, Zhang L, Gaillard P, Raedel M, Walter MH, Konstantinidis IK. Investigation of the association between cement retention and prevalent peri-implant diseases: a cross-sectional study. *J Periodontol*. 2016;87(3):212-220.
33. Korsch M, Walther W. Peri-implantitis associated with type of cement: a retrospective analysis of different types of cement and

their clinical correlation to the peri-implant tissue. *Clin Implant Dent Relat Res*. 2015;17(Suppl. 2):e434-e443.
34. Korsch M, Obst U, Walther W. Cement-associated peri-implantitis: a retrospective clinical observational study of fixed implant-supported restorations using a methacrylate cement. *Clin Oral Implants Res*. 2014;25(7):797-802.
35. Korsch M, Marten SM, Dotsch A, Jauregui R, Pieper DH, Obst U. Effect of dental cements on peri-implant microbial community: comparison of the microbial communities inhabiting the peri-implant tissue when using different luting cements. *Clin Oral Implants Res*. 2016;27(12):e161-e166.
36. Korsch M, Walther W, Marten SM, Obst U. Microbial analysis of biofilms on cement surfaces: an investigation in cement-associated peri-implantitis. *J Appl Biomater Funct Mater*. 2014;12(2):70-80.
37. Kotsakis GA, Thai A, Ioannou AL, Demmer RT, Michalowicz BS. Association between low-dose aspirin and periodontal disease: results from the continuous national health and nutrition examination survey (NHANES) 2011-2012. *J Clin Periodontol*. 2015;42(4):333-341.
38. Hujoel PP, Drangsholt M, Spiekerman C, DeRouen TA. Periodontitis-systemic disease associations in the presence of smoking – causal or coincidental? *Periodontol 2000*. 2002;30:51-60.
39. Hujoel PP, Cunha-Cruz J, Kressin NR. Spurious associations in oral epidemiological research: the case of dental flossing and obesity. *J Clin Periodontol*. 2006;33(8):520-523.
40. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58(5):295-300.
41. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerg Themes Epidemiol*. 2015;12:14.
42. Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet*. 2002;359(9304):431-434.
43. Barasch A, Cunha-Cruz J, Curro FA, et al. Risk factors for osteonecrosis of the jaws: a case-control study from the CONDOR dental PBRN. *J Dent Res*. 2011;90(4):439-444.
44. Lopez R, Scheutz F, Errboe M, Baelum V. Selection bias in case-control studies on periodontitis: a systematic review. *Eur J Oral Sci*. 2007;115(5):339-343.
45. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet*. 2002;359(9300):57-61.
46. Emilson CG. Potential efficacy of chlorhexidine against mutans streptococci and human dental caries. *J Dent Res*. 1994;73(3):682-691.
47. Cunha-Cruz J, Hujoel PP, Kressin NR. Oral health-related quality of life of periodontal patients. *J Periodontal Res*. 2007;42(2):169-176.
48. Popelut A, Valet F, Fromentin O, Thomas A, Bouchard P. Relationship between sponsorship and failure rate of dental implants: a systematic approach. *PLoS One*. 2010;5(4):e10274.

How to cite this article: Kotsakis GA. Minimizing risk of bias in clinical implant research study design. *Periodontol 2000*. 2019;81:18-28. https://doi.org/10.1111/prd.12279