Published in final edited form as:

Oral Dis. 2016 March ; 22(2): 87–92. doi:10.1111/odi.12385.

Design and interpretation of clinical research studies in oral medicine: a brief review

JC Atkinson¹ and DB Clark²

¹Center for Clinical Research, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA

²Behavioral and Social Sciences Research Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA

Abstract

The objective of this short review is to help researchers improve the designs of their clinical studies. Also included is a discussion of the level of evidence provided by the various clinical research study designs.

Keywords

clinical research; epidemiology; oral medicine

Introduction

In the last 30 years, the number of publications related to oral diseases has increased tremendously. The majority describe the clinical presentation of a disease such as oral lichen planus, biomarkers related to and risk factors for the disease, and the results of small trials testing therapies for treatment. To manage the increasing volume of publications, systematic reviews are conducted to help clinicians and researchers identify the best research to advance patient care. While these reviews first evaluated results from clinical trials, the systematic review process has been extended to observational studies. An example would be the review of the studies investigating the association between periodontal disease and cardiovascular disease conducted by the American Heart Association (Lockhart *et al*, 2012).

When one reads a systematic review, one cannot help but note that many publications are excluded. Often the reasons for exclusion relate to research design and study implementation methods (Atkins *et al*, 2004; The Cochrane Collaboration, 2011). This short review will discuss common weaknesses found in many clinical studies and provide suggestions for their improvement (Table 1). We use the word ‘design’ to include more than just the classification of study type, such as cohort design, clinical trial, and case series. Design in

Correspondence: Jane C. Atkinson, Center for Clinical Research, National Institute of Dental and Craniofacial Research, National Institutes of Health, 6701 Democracy Blvd., Room 634, MSC 4878, Bethesda, MD 20892-4878, USA. Tel: 301 435 7908, Fax: 301 480 5912, jatinso@mail.nih.gov.

Author contributions

The authors both contributed to the design, content and writing of this review.

this paper includes study methodology, such as how disease is measured, how subjects are selected, and how the data are analyzed. Implementation will include issues about how study data are collected, how outcomes are judged, and how well subjects enrolled in the study are retained. The good news is that many studies can be improved without expensive new technologies. We hope that readers of this review will realize that the most important ingredient in a successful clinical study is adequate planning.

Design selection

'Clinical research' is defined broadly as patient-oriented research that usually involves interaction with research subjects to collect data about their health or disease. If the data are collected from existing records, it is a retrospective study. Prospective studies involve collection of data as time moves forward. The research design generally dictates the level of clinical evidence that is the product of the study (Hujoel, 2009). For example, a case series describing the successful treatment of 20 subjects with disease X with a drug approved for another use should not be interpreted as evidence of the drug's efficacy for treatment of disease X. A randomized controlled trial is almost always needed to make that conclusion. Therefore, an appropriate design for a clinical research study should be selected after a research question is conceived. A brief overview of designs and factors to consider with each type is presented below (Manolio, 2002; Gordis, 2004).

Case report and case series

A case report or case series provides a careful description of a patient or several patients with a common disease or syndrome. Examples include descriptions of a patient's clinical course with an unusual facial infection, the presentations of cases with oral tuberculosis, a group of patients with fluconazole-resistant candidiasis or an unusual presentation of allergy in the oral cavity. The description should be complete and include other characteristics that may have influenced disease severity or symptoms. A case report is meant for use by another clinician who may evaluate a similar case.

Case reports and series must be viewed only as hypothesis generating, such as calling attention to a new clinical finding. There are no controls, so any descriptions of therapeutic outcomes must be viewed cautiously. Importantly, associated clinical data are usually gathered from clinical records rather than from a research form completed during a research visit. Therefore, there may be incomplete data in the report, and the data are not collected systematically. However, initial presentations of important diseases such as HIV/AIDS (Gottlieb *et al*, 1981) or bisphosphonate-related osteonecrosis of the jaw (Ruggiero *et al*, 2004) were first reported as case series, so there is value in this type of publication.

A good case series should include the fullest spectrum of the disease as possible or a unique trend not previously noted. A strategy could be systematically extracting hospital records for a defined period of time and including all applicable patients with the same diagnosis in the report. It is important to avoid the temptation to only include the most dramatic cases in the report.

Cross-sectional studies

Cross-sectional studies involve examination of a population for a disease or outcome and determining its prevalence and factors (exposures) associated with its prevalence. Like a case series, the disease of interest or exposure has already occurred, so the data collected about past events such as the time of presentation or exposures that may cause the disease are retrospective and may not be accurate. Cross-sectional studies using representative country populations, such as National Health and Nutrition Examination Survey (NHANES) (Dye *et al.*, 2014), are very useful in determining the prevalence of more common diseases in the population of interest and to indirectly assess the value of public health measures such as water fluoridation, human papilloma virus (HPV) vaccination, or smoking cessation programs on disease prevalence. However, one must not infer causality from these studies. For example, while caries prevalence may decrease in a municipality that began adding fluoride to water 5 years previously, other factors, such as increasing access to dental care, improved diet and increased the use of fluoridated toothpaste may have equally contributed to caries decrease.

The best cross-sectional studies evaluate outcomes using standardized methods and/or validated instruments (Table 1). Ideally, complete data need to be collected from every participant enrolled in the study. Those evaluating clinical outcomes such as caries should be calibrated, and images such as radiographs or pathology slides should be judged by individuals using predetermined outcomes, with a system to classify abnormalities. Those judging images or pathological slides should not be aware of a participant's diagnosis. The overall goal is to consistently and independently collect and analyze the research data. Investigators should be careful to decide what information they can collect given their time and resources. It is impossible to collect everything; instead, focus on collecting the most valuable data using robust methods and collecting complete data from all research participants.

Case-control studies

A case-control study is used to determine whether an exposure might be related to the development of a certain disease. Cases are individuals with the disease; controls are those without the disease who are carefully matched and drawn from the same population as the cases (Table 1). The goal is to determine whether an exposure (such as smoking or past infection with mumps virus) is more prevalent in the cases than the controls to the extent that the exposure can be considered a risk factor or protective factor for the development of the disease.

Case-control studies are often used to generate hypotheses about causality, but should not be interpreted as demonstrating that the risk factor causes the disease. Case-control studies often collect retrospective data about an exposure (such as past exposure to benzene) and a one-time assessment of disease severity (such as a complete blood count) and current exposure levels. Other factors to consider when designing a case-control study include the numbers and appropriateness of controls. Minimally, there should be at least the same number of controls as cases, cases and controls must be evaluated using the same methods, and the cases and controls should come from the same populations (Wacholder *et al.*,

1992a,b,c). Case-control studies are often used to study rarer diseases that would not occur frequently in a large population-based cross-sectional study. A good example is head and neck cancer. Case-control designs are used to establish that smoking and alcohol consumption are major risk factors for this malignancy (Day *et al.*, 1993; Conway *et al.*, 2015). Case-control studies typically report associations between risk factors and the outcome of interest as odds ratios, which are close to relative risk for rare diseases.

Electronic medical records (EMRs) can be used to enhance case-control studies. Past patient data that are not subject to recall bias, such as total head and neck radiation dose, may be available from the medical record, and control data from the same population may be extractable. However, there may be missing or incomplete data in the EMR and the data are typically not collected systematically, so one must be cautious not to overinterpret results. Another limitation of a case-control study is that those with the most serious cases may have died from their disease and are not available for evaluation (Manolio, 2002). If an investigator was determining the long-term consequences of head and neck radiation on adolescents who received radiation as part of the conditioning regime for stem cell transplant, a limitation of this research study would be that children who did not survive the transplant would not be included in the study population. This may be an important consideration depending on the research questions being asked.

A high-quality case-control study will have the following features (Table 1). First, everyone in the study is evaluated using the same methods or instruments. Secondly, if possible, those judging clinical outcomes should not know whether a subject is a case or control. This is important, as investigators may be more likely to judge a certain condition such as mucosal erythema or gingival inflammation as present if they know the subject is a case. A good example of a high-quality case-control study was one started soon after HIV began to present in children. Six hundred children born to HIV-positive mothers were evaluated from birth longitudinally by examiners who initially did not know the children's HIV status (European Collaborative Study). In some cases, both infected and uninfected infants had the same mother. This design allowed more accurate estimates of the prevalence of HIV complications in children, such as oral candidiasis and parotid enlargement.

Case-control studies are subject to many types of bias, including exposure suspicion bias and recall bias. Enrolled participants with the disease (cases) may be seeking a reason for developing the disease, and controls may be less familiar with their medical histories. To minimize the potential for bias, the case and control groups must be carefully defined and selected.

Longitudinal cohort studies

A longitudinal cohort study is used to collect data on a certain population prospectively, usually to document disease course or to determine the incidence of a disease and factors related to its development. They are time and resource intensive, but generate higher quality evidence about exposures and disease development than a case-control or cross-sectional study. For example, the metabolic syndrome and cancer project, which includes seven population-based cohorts from Norway, Austria, and Sweden, enrolled 577 799 adults who were followed on average for 12 years. Longitudinal cohort studies may be the only method

to determine whether reduction or continued exposure to a risk factor, such as smoking or alcohol, affects the incidence or severity of a disease such as cancer (Stocks *et al*, 2012). In this situation, it is unethical to randomize individuals in a clinical trial to an arm that would deliberately expose the research participant to agents known to cause serious adverse outcomes.

The key elements of a good longitudinal study include the following:

- Having established intervals for follow-up and evaluating subjects within a defined interval, such as yearly or every 3 months;
- Having sufficient follow-up time to detect changes in the disease of interest or onset of new disease of interest; and
- Retention of a high percentage of those enrolled initially.

On occasion, a cross-sectional study is continued as a longitudinal study using a panel design. The same subjects are followed over time, collecting the same measures at specified intervals. An example is the Medical Expenditure Panel Survey (MEPS) (U.S. Department of Health and Human Services, August 21, 2009) that follows a random subset of NHANES participants for 2 years after completion of the NHANES examination. Following these participants for 2 years and collecting data at 6-month intervals allows for a snapshot of medical spending by the population with a baseline understanding of their overall health. Despite the prospective data collection involved, caution should be taken to avoid drawing causal inference from panel surveys.

Interpretation of longitudinal studies

Causality should not be assumed when two conditions are associated. An example would be cardiovascular disease and periodontal disease. While periodontal disease may be present more frequently in those with cardiovascular disease that is not sufficient evidence that periodontal disease causes cardiovascular disease. As the exposure precedes the development of disease, longitudinal studies provide additional evidence that the exposure (such as periodontal disease) causes the disease. However, one should be careful about interpretation of this evidence, as temporal sequence is not the only criterion required to establish causation. The case for causality is strengthened in an association study if the incidence of disease increases or decreases relative to the dose of the exposure (Gordis, 2004). However, changes in disease incidence that are correlated with exposure could still be explained by unidentified causal factors. Also, while longitudinal studies follow individuals prospectively and collect data from baseline forward using standardized methods, some of the data may be retrospective in nature. If a longitudinal study followed a group of adolescents with and without cleft palate for 5 years to determine changes in body image and other patient-related measures, information about past medical treatments for both groups would likely be collected from medical records that may be semi-complete (Manolio, 2002).

Clinical trials

If a research study is testing an intervention as a treatment or an improved diagnostic for disease management, the study is a clinical trial. 'Intervention' includes anything that can alter the course of a disease, such as a pharmaceutical agent, a medical device, a surgical technique, a behavioral intervention, or a public health program. Clinical trials are a subset of clinical research, and randomized controlled trials provide the strongest evidence for the causal nature of a modifiable factor and a disease outcome. Examples would include determining whether an anti-inflammatory agent that decreased autoantibody levels and other measures of inflammation increased the salivary flow rates of individuals with Sjögren's syndrome, or whether blocking osteoclast activity with a bisphosphonate drug prevented osteoporosis progression.

Clinical trials typically are classified into phases (Phase I, II, III, or IV) that indicate their size and stage of development. A Phase I study is used to establish dose and safety of a drug or intervention. They are not randomized and treatment allocation is not concealed (i.e., no one is blinded or masked to the agent the participant is receiving), and research subjects are given escalating doses of the intervention until the predefined dose limiting toxicity (DLT) is achieved (Johnson *et al*, 2007). A Phase I study could also be used to enroll a small number of subjects with a condition (such as ONJ) being treated with a new surgical technique to determine essential parameters for a future randomized trial. This could include determining an expected healing time, the postoperative complications, and eligibility criteria for a subsequent study. Once the dose and other parameters are established, the next phase (Phase II) begins.

Many designs are employed for Phase II trials, which are conducted to determine initial efficacy of a new intervention. The most straightforward design is a randomized controlled trial (RCT) that randomly allocates subjects to a placebo arm or active therapy arm. Ideally, the treatment assignment is concealed from the subjects enrolled in the study, the investigators conducting the study, and those conducting the statistical analyses until the final results are determined. The goal is to find the superior treatment and determine whether a future, larger trial is warranted. Phase II trials may be conducted at a single center or only enroll a population from one geographic area. They lack generalizability to a larger population of individuals with a disease or disorder.

The strength of the RCT methodology is the randomization of subjects to compare interventions to each other and/or to a non-intervention control group. The randomization procedure presumably controls for hidden factors that could influence outcomes. For example, if a novel treatment for periodontitis was being tested and smokers were eligible, randomization should allow for subjects who smoke to be equally distributed between study arms. This greatly reduces the chance that smoking status will influence any difference that is observed between study arms.

A Phase III study is necessary to establish the *efficacy* of an agent and sometimes is referred to as a pivotal trial. Large Phase III trials usually enroll hundreds to thousands of subjects at several clinical centers. While the inclusion/exclusion criteria select a uniform subset of individuals with the disease of interest, conducting the trial at several centers should enroll a

more diverse array of individuals. A Phase III trial has different teams of individuals delivering the therapy and judging outcomes. While the methods for delivering therapies and judging outcomes are standardized, there is always some variation between centers. Therefore, if the new agent is determined to have a positive effect on the disease across centers, it is more likely that the agent will be beneficial in practice when prescribed by many different practitioners. All of these characteristics of Phase III multicenter trials make results more generalizable to the population at large.

A Phase IV trial determines how well an efficacious treatment works in practice. This could include assessing how well the community of patients can comply with a drug schedule, such as taking a drug three times per day, and the frequency and severity of side effects associated with long-term use of a drug. These studies determine the *effectiveness* of a therapy.

Investigators often fail to consider that establishing evidence for a new treatment requires a series of studies. Circumventing those studies and going straight to an advanced phase clinical trial can increase the probability of a failed trial. For example, the correct dose for a new agent should be established with a smaller study that tests different doses and different lengths of exposure. A drug could have a slowly escalating positive effect that might be missed if subjects were not observed long enough in the trial.

One should also be cautious when reading the results of a Phase II or small clinical trial, even if a statistical difference is detected. Small trials that are published typically overestimate treatment effects (Hennekens and Demets, 2009), and small trials that are negative may not be published (sometimes referred to as publication bias) (Califf, 2007). A meta-analysis that combines results from several small clinical trials does not have the generalizability of a large Phase III trial and may inflate the therapeutic effect in part because negative trials are not included (LeLorier *et al*, 1997).

Questions that one should consider when designing a clinical trial include the following:

- Is there enough good evidence from previous studies to design a clinical trial? Has the natural history, including the variability of disease severity across time been established? Diseases such as oral lichen planus will have periods of quiescence or periods with little pain in the absence of therapy. This variability must be known to adequately design an interventional study, and can be established with a well-designed longitudinal (natural history) cohort study.
- Was a previous study conducted to establish a dose and duration of therapy? When considering the existing evidence, what should be the appropriate phase or design for the next research question(s)?
- Can subjects be consistently diagnosed with a disease or condition using accepted, validated criteria?
- Are there existing validated measures to determine the outcome variables of interest and can they be applied consistently by all individuals conducting the evaluations?

- Does the proposed change of disease activity, such as a decrease in oral mucosal pain or increase in the measurement of the oral aperture, translate into a meaningful change for the patient? Statistical significance does not necessarily mean clinical significance.
- Can the study protocol be implemented consistently across all the sites so that each research participant is treated and evaluated in a similar manner?

Analysis of clinical studies

An analysis plan must be developed before the data from a clinical study are analyzed, regardless of the clinical research design (Table 1). This helps assure that data are analyzed independently and minimizes bias. The plan must take into account many factors such as the most important variables, the distribution of the data, the completeness of the data, and the conclusions that can be made given the design of the study. *Post hoc* data analysis (analyses performed after the original analyses and after investigators know results of the first analyses) must be viewed with caution, especially as the *P*-values for detecting significance are often not adjusted for multiple tests and there is a high potential for investigator bias.

Another challenging issue in clinical studies is poor retention. If the study intervention is too taxing for participants or the primary outcome is collected 2–3 years after randomization or a baseline visit, study subjects may drop out of the trial. This threatens the validity and power of the study. Investigators must consider their design carefully and realistically to retain as many subjects as possible for the entire study and have a retention plan in place from study outset.

This very brief review highlights important factors to consider when designing a clinical research study. More guidance about the design and implementation of clinical studies can be found in the references.

References

- Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004; 328:1490. [PubMed: 15205295]
- Califf, RM. Large clinical trials and registries: clinical research institutes. In: Gallin, JL; Ognibene, FP., editors. *Principles and Practice of Clinical Research*. 2nd. Burlington, MA: Academic Press; 2007. p. 237-263.
- Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0. 2011
- Conway DI, Brenner DR, McMahon AD, et al. Estimating and explaining the effect of education and income on head and neck cancer risk: INHANCE consortium pooled analysis of 31 case-control studies from 27 countries. *Int J Cancer*. 2015; 136:1125–1139. [PubMed: 24996155]
- Day GL, Blot WJ, Austin DF, et al. Racial differences in risk of oral and pharyngeal cancer: alcohol, tobacco, and other determinants. *J Natl Cancer Inst*. 1993; 85:465–473. [PubMed: 8445674]
- Dye BA, Li X, Lewis BG, Iafolla T, Beltran-Aguilar ED, Eke PI. Overview and quality assurance for the oral health component of the National Health and Nutrition Examination Survey (NHANES), 2009–2010. *J Public Health Dent*. 2014; 74:248–256. [PubMed: 24849242]
- European Collaborative Study. Children born to women with HIV-1 infection: natural history and risk of transmission. European Collaborative Study. *Lancet*. 1991; 337:253–260. [PubMed: 1671109]
- Gordis, L. *Epidemiology*. Philadelphia: Saunders; 2004.

- Gottlieb MS, Schroff R, Schanker HM, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med. 1981; 305:1425–1431. [PubMed: 6272109]
- Hennekens CH, Demets D. The need for large-scale randomized evidence without undue emphasis on small trials, meta-analyses, or subgroup analyses. JAMA. 2009; 302:2361–2362. [PubMed: 19952322]
- Hujoel P. Grading the evidence: the core of EBD. J Evid Based Dent Pract. 2009; 9:122–124. [PubMed: 19737664]
- Johnson, LL.; Borkowf, CB.; Albert, PS. An introduction to biostatistics: randomization, hypothesis testing and sample size estimation. In: Gallin, JL.; Ognibene, FP., editors. Principles and Practice of Clinical Research. 2nd. Burlington, MA: Academic Press; 2007. p. 165-195.
- LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med. 1997; 337:536–542. [PubMed: 9262498]
- Lockhart PB, Bolger AF, Papapanou PN, et al. Periodontal disease and atherosclerotic vascular disease: does the evidence support an independent association?: a scientific statement from the American Heart Association. Circulation. 2012; 125:2520–2544. [PubMed: 22514251]
- Manolio, TA. Design and conduct of observational studies and clinical trials. In: Gallin, JL., editor. Principles and Practice of Clinical Research. 1st. San Diego: Academic Press; 2002. p. 187-206.
- Ruggiero SL, Mehrotra B, Rosenberg TJ, Engroff SL. Osteonecrosis of the jaws associated with the use of bisphosphonates: a review of 63 cases. J Oral Maxillofac Surg. 2004; 62:527–534. [PubMed: 15122554]
- Stocks T, van Hemelrijck M, Manjer J, et al. Blood pressure and risk of cancer incidence and mortality in the Metabolic Syndrome and Cancer Project. Hypertension. 2012; 59:802–810. [PubMed: 22353615]
- U.S. Department of Health and Human Services, A.F.H.R.A.Q. [accessed on 1 January 2015] Medical Expenditure Panel Survey Background [Online]. 2009 Aug 21. Available: http://mepsweb/about_meps/survey_back.jsp
- Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. Am J Epidemiol. 1992a; 135:1019–1028. [PubMed: 1595688]
- Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. II. Types of controls. Am J Epidemiol. 1992b; 135:1029–1041. [PubMed: 1595689]
- Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. III. Design options. Am J Epidemiol. 1992c; 135:1042–1050. [PubMed: 1595690]

Table 1

Common problems with clinical studies

Study factor	Problem	Solution
Population selection	Subjects do not represent the full spectrum of persons with the disease, disorder or syndrome	Enroll consecutive patients in a cohort study or use data from consecutive patients in a case series
	Referral center bias because academic centers treat patients with more severe disease	If the disease of interest is common, one may wish to use methods to select a random sample
		Include community-based centers that treat patients with better managed disease
Inclusion/exclusion criteria	The inclusion/exclusion criteria are overly restrictive or poorly defined, yielding results that are not generalizable to the entire population of individuals with the disease or cannot be reproduced	Criteria should be defined clearly and included in the publication Criteria should be as broad as possible to not exclude those most likely to benefit from a new therapy, such as the elderly
Subject selection	Subjects not selected with validated methods.	Subjects should be selected using validated diagnostic methods, realizing that diagnostic criteria are refined over time
Control selection	Controls are not well matched	Controls should be matched to controls carefully, such as by age, SES status and sex/gender Multiple control groups may be needed, such as a healthy control and diseased control group
Objective study outcomes	Methods for judging study outcomes are not validated	The methods for measuring outcomes should be validated before enrolling subjects. This is critical to generating clinical science that is reproducible
	Study outcomes judged with knowledge of group assignment. This includes observational studies and interventional studies	Both clinical outcomes and biological specimen evaluation should be performed by examiners who have no knowledge of group assignment or the diagnosis of a subject
Subjective study outcomes	Patient-reported outcomes (PRO) are not considered. A statistically significant reduction in the surface area affected by a mucosal disease may not represent a meaningful change to patients	When measuring PROs, it is important to use questionnaires that are validated for the same reasons it is important to use validated clinical assessment tools. A series of validated PRO questionnaires is available from http://www.nihpromis.org
Analytical methods	Post hoc analyses used to analyze data	The analytical plan should be established before examining the data. Post hoc analyses can yield biased, non-reproducible results
	Insufficient numbers of events to test study hypotheses	Extend observation to assess the disease or condition outcome
	Poor retention	Develop a good retention plan at the outset of the study