2 Study Design

Designing a research study is not a simple task. Just as the key elements and determinants of outcomes of war are fixed even before it is fought (i.e., at its planning stages), the success of a study is also largely determined at its initial planning stage. Successful studies need to address two important dimensions: reliability and validity. A reliable study should be replicable, providing similar results if the same study parameters are applied. Validity is concerned with the ability of the study to correctly answer the question it asks. Internal validity deals with the ability of the study to correctly infer about the relationship between the independent variables and the outcome(s) being studied. External validity deals with application of the findings to other observations, samples, or populations and implies generalizability of the study results. The need for reliability and validity of studies dictates sound study designs. Box 2.1 outlines some of the important steps in designing a study, as well as some possible problems and solutions to them.

A common misunderstanding about studies comes into focus in the discussion section of study reports where most associations are simply *assumed* to be causal and investigators then proceed to interpret the "causal" association. This error is easily avoided if the objectives of a study are clear before designing it; that is, the investigators should be clear whether their study and involved analyses address a causal (or other) "hypothesis testing" situation of a "hypothesis generating" situation. A *hypothesis testing* study has to explicitly state the involved hypothesis a-priori, and the study should be designed and powered to address that question. Therefore, post hoc, secondary exploratory data analyses would not be a part of the hypothesis testing situation unless special analytical arrangements are made (e.g., adjusting for multiple testing). There is only a limited set of opportunities to find significant and meaningful results from data already collected—if it is found

BOX 2.1 A Short Checklist for Designing a Study

Steps Involved in Designing a Study

- Outline of study
- Title of study
- Research question (what is the overall, broad question?)
- Research hypotheses (specific clear and sequential sets of questions to be addressed)
- Significance of the study—new data/confirmatory, and why is it important?
- Why should anyone other than the interested investigator spend time or resources in this proposal?
- Comprehensive description of study design
- Inclusion/exclusion criteria; how will participants be recruited?
- What outcomes and other factors will be measured—why and how?
- How will data be analyzed—describe how each variable will be handled; how missing information will be handled; sample size and power? What is an important difference to detect (10%, 25%, 5 inches, 10mm of something . . .)?
- How will the data be safeguarded (HIPAA compliance)?
- Potential problem areas—logistics (e.g., power outage—will biological samples be lost? safeguards?)
- Limitations of study and how can these be corrected?
- Is the study proposal approved (approval pending) by institutional IRB?

Potential Problems and Solutions in Designing Studies

- Research question is too broad/general (*Narrow the question; use smaller set of variables*)
- Not enough subjects available (*Broaden inclusion criteria; increase time*)
- Methods beyond investigator's skill level (*Collaborate, consult, learn*)
- Too expensive (*Consider less expensive alternatives, smaller study, less follow-up*)
- Question not interesting/novel enough (*Consult and discuss with mentors/peers; modify research question*)
- Uncertain ethical suitability (*Consult with IRB; modify research question*)
- Vague study plan (*Write proposal early and revise several times*)
- Proposal confusing/unclear (*Write proposal in point-by-point manner for specificity*)

that more new information is needed, the testing mechanism cannot go any farther. However, post hoc analyses is not a condition sine qua non for data already collected. The recognition that the study was not designed for testing the involved questions in post hoc analyses clearly indicates that although such analyses may be carried out, they can, at best, be used only as indicators of potentially new information; that is, viewed as hypothesis-generating analyses. Studies carried out in such manner are correctly classified as *hypothesis-generating studies*. Therefore, once hypotheses are generated, these can then be tested in new, specifically designed studies.

Studies may be "unrepeated" measure studies where measurements are taken at one time only, and two or more groups are compared for outcomes. However, studies may also be "repeat-measure" studies where the unit of observation is measured more than once (i.e., the measurements are repeated). In repeat-measure designs, the second and subsequent measures from a subject are considered to be correlated with the first measurement because the basic biological and sociocultural fundamentals of the individual remain the same across all of the observations. Such study designs need special analytical handling, which is discussed in Chapter 8. The analyses used for repeat-measure designs are not specific to studies that incorporate measuring the same person multiple times, but are analyses that address any correlated data. Therefore, from an analytical standpoint, repeat-measure analysis may be considered a special case of correlated data analysis. However, whether one conducts repeat-measure studies or not, in terms of comparison of groups the architecture of studies follow the same paths as discussed below.

Studies may be classified as observational or experimental. In both, the effects of causes may be assessed. The effects are the outcomes, whereas the causes are usually generically called "exposures." The associations examined by the studies are exposure–outcome associations, which may be measured in different ways. The key difference between an observational study and an experimental study is the control of the exposure. In *experimental studies*, the investigator controls the exposure and determines who gets the exposure and how much exposure one gets. For example, in a phase III clinical trial, the investigator decides the dosage and allocation of the drug to the participants through a *randomization* process. In some experimental studies, the investigator is able to control allocation of exposure but is not able to ascertain randomization; that is, participants are not randomly assigned to exposure groups. Such studies are generally referred to as *quasi-experimental studies*. Compared to randomized designs, quasi-experimental studies have poorer internal validity, but they are more easily (and frequently) conducted than randomized studies.

In contrast, in *observational studies*, the investigator does not get to control the exposure but classifies the participants based on their preexisting exposure status. The determinants of the exposure lie in the population, or

otherwise outside the control of the investigator. For example, in examining the association of smoking and alcohol with oral cancer, the investigator does not get to choose who will smoke and drink or how much. The investigator is a passive observer of the exposure but is able to classify people based on their exposure status. Observational studies may be conducted with different designs. Studies that follow participants over time into the future are generally classified as *prospective studies*, whereas studies that look at information already collected before the start of the study are generally called *retrospective studies*. Sometimes, the investigator collects all information about the exposure and outcomes at the same point in time—such studies are called *cross-sectional studies*.

Types of Study Designs: Observational Studies

Case-Control Study

Case-control studies are common study designs used in oral health research. Essentially, the investigator selects a group of persons with a disease of interest (cases), and then selects a group of persons who do not have the disease, called controls, and compares the exposures of interest to find out which exposures are associated with the cases more than the controls. Case-control studies ask the question: What are the determinants of this disease? Box 2.2 enumerates the main properties of case-control studies. Some key points about this study design are discussed below.

Retrospective vs Prospective Case-Control Study

Mostly, case-control studies are viewed to be retrospective in nature. The common view is that because the investigator asks the participant about his or her history of exposure, which occurred sometime in the past, the design is "retrospective." Such a design is retrospective only with respect to the starting time of the study. In traditional case-control studies, prevalent cases were generally picked up, and their exposure history was obtained. However, as study designs progressed, and differences between prevalent and incident cases became clearer, the scientific community made greater efforts to examine incident diseases. The fundamental change in this paradigm shift was that a "new" disease event could be demonstrated in studies examining incident diseases as opposed to prevalent diseases. Therefore, logically, if exposure history could be *definitively ascertained* to have occurred *prior* to disease occurrence, the case for cause–effect association would be stronger, which was not possible in traditional case-control studies examining prevalent diseases. This awareness delivered the possibility of designing "prospective" case-control studies.

BOX 2.2 General Properties of Case-Control Studies

Definition: Case-control studies compare cases and disease-free controls for their exposure status and compare the risk of exposure in cases and controls. Usually, cases are people with disease, but treatment outcomes or other criteria can be used to define a "case." These studies may be incorporated inside a cohort and are called nested case-control studies.

Assumptions

- Disease prevalence is low.
- Cases and controls are representative of the population.
- Relative risk cannot be directly calculated.

Case Selection

- Clear basis of case definition needed: sign/symptoms; clinical examination; diagnostic tests; confirmatory tests.
- It is better to err on the side of restriction rather than inclusion in doubtful cases.
- Cases should have the disease.
- Use incident cases rather than prevalent cases, if possible.
- Cases may be identified from clinic rosters, death certificates, disease/outcome registries, surveys, administrative databases in some situations, adverse drug reactions (ADR) databases.

Control Selection

- Controls should represent the same population from which cases arise; that is, if the control group members previously had the disease, they would have become cases.
- Controls should have the same probability of getting disease as the cases.
- Controls should provide exposure information about the population.
- Controls should be selected independent of their exposure status; that is, selection probability and sampling fraction of exposed and unexposed controls should be the same.

Types and Source of Controls, and Sampling Frame

- *General living population*—telephone directory, vital records, voter list, tax list, driver license roster, pharmacy roster, employment roster, insurance roster, professional roster.
- *Random-digit dialing*—can approach all households in a designated area; however, by definition assumes that all households have a telephone connection and will miss households without telephone connections that are likely to be the poorest and with the most burden of disease.

BOX 2.2 General Properties of Case-Control Studies *(Continued)*

- *Hospital/clinics*—hospital and clinic attendees; their disease status should, however, not be related to the disease being studied, and referral pattern should be similar to that of cases.
- *Dead persons*—dead persons may be used in special situations when selected cases are also dead.
- *Neighborhood controls*—May share environmental exposure with the cases and may be useful in environmental exposure-based studies such as cancer cluster studies, fluorosis studies, etc.
- *Family/friends*—They may share important characteristics with cases such as environment, socioeconomic position, race/ethnicity, etc.

Advantages: Useful for studying uncommon diseases; less expensive (prospective case-control study can be more expensive and logistically difficult); short duration studies, logistically easy; yields a reasonable estimate of risk ratio (odds ratio)—if prevalence is low, then odds ratio approximates relative risk well.

Disadvantages: Temporal relationship between exposure and outcome cannot be examined; subject to substantial selection, survivor, and recall bias; one outcome can be studied at a time; does not provide prevalence, incidence, and excess risk.

When to Do Case-Control Study: When disease prevalence is low; when not much information is available about the disease; when population dynamics do not permit longer studies; when obtaining exposure data is difficult/expensive; when disease has very long induction period/latency.

Once cases and controls are selected, there are only four possibilities that exist related to their exposure status:

1. Cases and controls were not exposed (unexposed).
2. Cases and controls were exposed before the disease occurred.
3. Cases and controls were exposed after the disease occurred, but before their clinical detection.
4. Cases were exposed after disease occurred, and about the time when the disease was detected.

Most often, participants are asked about their exposure, and this information is recorded, but no documentary evidence is required or produced to prove that the exposure actually occurred, or when the exposure occurred or started. There are situations when the exposure status of participants can be *proven* to have occurred *prior* to disease occurrence. For example, in a hos-

pital-based case-control study of adverse drug reactions, exposure details of relevant drugs would be clearly mentioned in the medical charts, which can be captured by chart abstraction. In such situations, there would be documented evidence that exposure preceded the disease, thereby establishing the temporal sequence between them. If such studies can be refined a little more by including only incident cases occurring *after* the proven exposure, then causal arguments can be further strengthened. Such studies will be *prospective case-control studies*. Studies of dental sealant failure, secondary caries, and several such outcomes can be conducted as prospective case-control studies. The key factor that distinguishes a retrospective case-control study from a prospective case-control study is the sequence of the recording time of disease occurrence. Rothman, Greenland, and Lash (2008) recommend that the terms *retrospective* and *prospective* be used in relation to study designs *only* in the context of whether the disease could influence the exposure information.

Kirakozova and Caplan (2006) conducted a case-control study in which they used a university hospital computerized treatment database to identify all patients receiving a single-unit crown on a nonendodontically treated permanent tooth over a 5-year period. They classified these patients as cases if their crowned teeth received root canal therapy and as controls if their crowned teeth did not receive root canal therapy before the study cut-off date. The authors wanted to identify variables predictive of *subsequent* root canal therapy in teeth receiving full coverage restorations. In this study, among other exposures determined from chart entries and radiographs, the key exposure was defined as the extent of coronal and root destruction at the time of receiving the crown. The exposure was therefore *definitively* recorded *prior* to becoming a case. The study concluded that younger age and greater extent of coronal and root destruction were important predictors of receiving full coverage restorations.

Perhaps it is true that usage of the terms *retrospective* and *prospective* are redundant, and are relevant only in historical context because health research literature is replete with study reports self-classified that way. Looking to the future, it might be more informative to drop these terms from regular usage unless it becomes critical to describing a study such as "retrospective cohort"—and even then, to avoid confusion, "nonconcurrent" cohort might be the more appropriate terminology to use.

Population and Control Selection

Defining the population and selection of appropriate controls are the two most important factors that determine the success of case-control studies in terms of arriving at correct inferences about strength, direction, and the importance of association of study factors and outcome (see also the section on counterfactual concept in Chapter 3). Essentially, definition of the source

population determines the population from which controls are sampled (Rothman et al., 2008). *Source population* is the population from which the cases and controls arise. Generally, cases in a case-control study should represent all the cases in the source population, and controls should arise from the source population from which the cases arise. Ideally, controls should be selected using a random sample from the source population. Sometimes, studies are restricted to a selected and restricted population of subjects and are not generalizable to the whole population. Such restricted populations on which studies are focused are *target populations* and may prevent comprehensive generalizability of study results to the whole population. Such restricted target populations may serve the purpose of program planning and service provision to a select group, but in most epidemiologic studies, they are more of a handicap because answers that are generalizable to the source population are the ones that are needed most often. The following four population-related terms are often used in epidemiology:

- *Target population*: The population about which we *intend* to make estimates and inferences.
- *Source population*: The population from which cases, controls, and samples arise.
- *Actual population*: The population to which our estimates and inferences *actually* apply.
- *Study population*: The subjects included in all phases of the study.

When cases and controls are sampled directly from the source population, the study is called a *population-based* study, also known as *primary-based study*. Such sampling occurs before the cases are identified. Sometimes, direct identification of source population may not be possible, and cases are identified without specifying the source population. In such situations, before controls are selected, it is important to identify and define the *study base* (source population for cases) so that controls can be selected from that study base using appropriate sampling methods. Such studies are called *secondary base studies* because controls come from a secondary source population. The term *case-based case-control study* usually implies a study in which cases are identified from a hospital and controls are identified from the community served by that hospital that did not have the disease under investigation.

Box 2.3 describes types of controls that are commonly used in epidemiologic studies. In general, Rothman et al. (2008) suggest that control selection should follow the following two important guidelines:

1. Controls should be selected from the same source population from which cases arise.
2. Within the stratification factors, controls should be selected independently of their exposure status.

BOX 2.3 Types of Controls

Types and Sources of Controls

Population Controls: The controls are selected from the same precise population from which the cases arise. The sampling of controls can be done randomly, or an incidence-density sampling can be done.

Neighborhood Controls: Controls are selected from neighborhood residencies of the cases in a systematic way with or without matching.

Random-Digit Dialing: Controls are selected randomly by calling numbers from a telephone book. It is assumed that cases arise from the population represented by the telephone book. It is easy to conduct. However, the number of people representing each telephone number may vary as different households have different numbers of residents and different numbers of telephone lines. Logistic issues might arise because members may not be home; permission to contact may be needed if the numbers are in do-not-call registries; commercial, residential, and cell phone numbers may be difficult to distinguish; Internet telephony may not necessarily associate a telephone number with the actual residence of the holder; and call screening and answering machines may be difficult to bypass.

Hospital/Clinic-Based Controls: Often used in hospital/clinic-based studies. The source population may be people treated in the hospital, and controls may be selected form the same source population. The catchment area of the hospital/clinic may be ill-defined, thus compromising defining the source population properly. It may be possible to match cases and controls on disease criteria, but healthy controls may be difficult to obtain. Controls cannot be selected randomly, and exposure–disease associations may bias the studies, e.g., Berkson's bias.

Dead Controls: Sometimes it may be possible to use dead people as controls if their exposure history prior to death can be ascertained. Such controls might be a useful strategy if the cases are already dead. However, if the cases are living, then the dead controls do not exist in the source population. Information for dead controls may be elicited from their medical records, vital registries, and/or proxy persons who knew the dead person well.

Sibling Controls: In some studies, siblings of cases may be used as controls as they share similar family characteristics, neighborhood, socio-economic characteristics, etc. Overmatching may be a problem. All cases may not have siblings.

Friend Controls: Friends of cases may be used as controls. Cases are asked to provide names of friends of same gender and age group for this purpose. Overmatching may be a problem. Having friends is a function of sociability which may vary between cases. Furthermore, cases may refer only certain friends based on some criteria that may introduce bias.

BOX 2.3 Types of Controls *(Continued)*

Matching

- Matching is the process of equating the groups being compared (e.g., cases and controls) on one or more factors so that whatever differences are noticed between the two groups would not be attributable to the factors on which they were matched. For example, if cases and controls were matched by age group and gender, then the differences between them would not be due to gender and age.
- May be done as individual matching, category matching, caliber matching, and frequency matching.
- Matching increases study efficiency in case-control studies by using fewer numbers of factors and variables, and may improve validity of cohort and experimental studies.
- Effects of factors used in matching cannot be examined in the matched study.
- May introduce selection bias in the study.
- Matching results in paired data that must be handled especially for analysis—correlated data analysis methods must be used.
- Over- and under-matching can create various problems in the study.

Case to Control Ratio

Number of controls per case (control to case ratio = r) in a case-control study has been actively discussed. The general paradigm is to use one control per case. However, this may limit the power of the study and precision of the effect estimates. Furthermore, studies are also limited by the cost incurred in the conduct of the study.

> If C denotes the ratio of cost of study of a member of a study group to the correspondent cost for the member of the referent group, then the optimal value of r for fixed total costs is approximately the square root of C. Thus, if the costs of studying the two types of individuals are equal, then the optimal strategy is to select equal numbers [of cases and controls]. In contrast, if the cost of studying a person in the study group is twice as great as the cost of studying a person in the referent group, then the optimal value for r is approximately $\sqrt{2} = 1.41$. (Kelsey, Whittemore, Evans, & Thompson, 1996)

Increasing the sample size in a study increases its power and the precision of the estimates derived from it. Substantial gain may be achieved by raising r to 4 or 5 (i.e., 4 or 5 controls per case). However, increasing the value of r to beyond 4 or 5 does not lead to further significant gain in power or precision of effect estimates. In matched case-control studies, meaningful

increases in statistical power can be obtained by increasing *r* above 5 "when there is a high (but plausible) correlation in exposure status between cases and matched controls, or when there is a low prevalence of exposure among controls" (Hennessy, Bilker, Berlin, & Strom, 1999).

In some special situations, more than one control group might be selected because individually none of the control groups truly represent the source population, and between them have certain advantages that other control groups may not have. For example, a case-control study of oral cancer may use hospital controls that may not be representative of the source population from which oral cancer cases arise and may therefore decide to use a second control group comprised of friends of the oral cancer cases. Although the friend-control group will be different from hospital controls, neither represents the source population. In such situations, both control groups should be compared to each other. Usually such issues can be logically resolved using only one control group. However, if *multiple control groups* are used, and exposure associations differ between different case-control comparison combinations, it would be difficult to pinpoint the true association. Unless an explicitly established reason exists and a clear a-priori decision about interpreting conflicting outcomes is laid down, multiple control groups may not offer any advantage, but may lead to logistic and budgetary issues and create confusion about interpreting results.

Power of Study and Sample Size

If an association between exposure and outcome exists in reality and is also detected by the study, or if an association does not exist in reality and the study detects the absence of an association, then a correct decision has been made. However, if there is no real association but the study detects an association, then the situation is similar to a false–positive test, and such errors are called *Type-I errors* or alpha errors (finding a difference that does not exist). Alternately, if a study fails to detect a true association, then the situation is similar to a false negative test and such errors are called *Type-II errors* or beta errors (not finding a difference that exists). Studies need to guard against and minimize both these types of errors.

Power of a study is the ability of the study to correctly detect an association when such an association truly exists. Numerically, power = 1 – Type-II error rate (i.e., 100% – Type-II error rate %). Power is directly proportional to effect size, sample size, variance, inverse of Type-II error, and the statistical significance level. It also depends on disease prevalence, exposure prevalence, study design, and sampling scheme. Of these, only the sample size is controllable by the investigator, although the other factors, including the study design, may be varied in different ways. *Effect size* is the difference between the magnitude of the observed association in the study and the hypothesized/true association (i.e., it is a measure of the magnitude of effect under the alternate hypothesis) and represents the smallest difference that

would be clinically important or substantively significant to detect. If an investigator wants to detect large differences, then a small sample size would suffice, but if smaller differences need to be detected, sample size must be increased. Small decreases in effect sizes may result in large increases in required sample size. In situations where a study has a predefined, fixed sample size, the investigator may calculate power of the study and calculate the minimum difference that can be detected given the fixed sample size. If it is possible to select a sample, then the investigator usually defines a level of power and then calculates the sample size required for that power for the given effect size. Studies with power below 0.8 (80%) are generally not considered to be useful. Studies targeting at least 90% power are becoming common.

Sample size requirement of a study is determined by the power level required, the effect size to be detected, and the type of analyses to be performed. The sample size for a simple comparison of two groups may be modest. But if the investigator wishes to compare several groups and conduct subgroup analyses with a-priori hypotheses (i.e., conduct a detailed hypothesis-testing study), then allowances need to be made for the number of comparisons to be conducted. A simple way is to proportionately reduce the level of statistical significance (normally 0.05) by the number of planned comparisons. Such adjustments are called *Bonferroni corrections*. For example, if five comparisons are planned, the statistical significance level can be redefined at $0.05/5 = 0.01$. Such corrections will adjust for Type-I errors. However, for more complex analyses, sample size calculations will need more complex handling. Power and sample size calculations for detecting differences between means and differences in proportions under different disease/exposure prevalence rates are varied, as are the calculations for different types of analyses. Two-tailed tests are more conservative than one-tailed tests. However, unless the investigator clearly knows and can establish and support a one-tailed hypothesis, two-tailed tests are the norm. Also, the paradigm "when in doubt, use two-tailed tests" is held universally.

In some situations, investigators may not only be interested in demonstrating differences between groups using hypothesis testing, but may also have ancillary aims such as demonstrating a certain magnitude of effect (e.g., improvement of biochemical parameters by 20%, or 100% change in relative risk). In such situations, although the central goal of getting an adequate sample size for the power level relevant to hypothesis is important, the study may focus on obtaining a predefined precision about the magnitude of effect according to the study goals. The ability to precisely report the magnitude of effect of interest depends on sample size, confidence interval, and outcome variance.

There exist several commercial software programs that conduct sample size, power, and precision calculations. However, if an investigator has re-

strictions to accessing these software applications, there are several freeware and shareware solutions easily downloadable from the Internet. Furthermore, several university and other sites maintain freely accessible sample size/power calculation web-based applications that are fairly reliable; for example: University of California at San Francisco sample size web page (http://www.biostat.ucsf.edu/sampsize.html), StatPages.net (http://statpages.org/index.html), Open Epi: Open Source Epidemiologic Statistics for Public Health (http://www.openepi.com/Menu/OpenEpiMenu.htm), and Creative Research Systems (http://www.surveysystem.com/sscalc.htm).

Prevalent vs Incident Cases

The anecdotal notion about case-control study is that it is utilized only with prevalent cases, and this notion works as a self-fulfilling prophecy. However, in many situations, a mixture of incident and prevalent cases or only prevalent cases may need to be used for logistic, financial, or other compelling reasons. Ideally, incident cases should be used in a case-control study. Why?

If incidence rate ratio is equal to the prevalence odds ratio measured in a case-control study, then prevalence odds ratio may be a good and unbiased estimate of the incident rate ratio (for definitions, see Chapter 4). However, in most oral diseases, such is not the case. For prevalence odds ratio to approximate the incident rate ratio, incidence and prevalence need to be in balance; that is, occurrences of new cases in a population should be balanced by removal of cases from the prevalence pool by treatment or other modes of emigration. If removal from the prevalence pool is slow, then the rate of growth of the prevalence pool is faster than the case generative rate of incidence rate, and the prevalence odds ratio will overestimate the incident rate ratio. On the contrary, if removal from the prevalence pool is faster than the incidence rate, then the prevalence pool will shrink and the prevalence odds ratio will underestimate the incidence rate ratio.

Prevalence (P) is related to incidence (I) by the duration (D) of disease (i.e., $P = I \times D$). If exposure is associated with disease duration or rate at which the prevalence pool is depleted, then measures based on prevalent diseases will not be able to differentiate the effects of association of exposure with the disease occurrence from the effects of association of the exposure with disease duration or prevalence pool depletion. In such situations, only the use of incident disease can remove the effect of association of exposure with prevalence pool size and growth rate. Furthermore, if the size of the exposed and unexposed population also changes over time, prevalence odds ratio becomes a yet more unreliable measure of association between exposure and disease.

The two most common dental diseases, dental caries and periodontal disease, are both measured in a cumulative way using the decayed, missing,

or filled teeth (DMFT) index and periodontal attachment loss, respectively and these are both irreversible in nature. Therefore, the only effect of these measures of disease in a population is to increase the prevalence pool over long periods because teeth are maintained over long periods of time—removal from the prevalence pool is slow, and occurs when a tooth is lost (for periodontal disease only but not caries), or upon physical removal of the person from the population. Because the times of exposure onset for these diseases are ill-defined and difficult to pinpoint, true population risk may be estimated by studying incident disease rather than prevalent diseases.

Case-Control Study Within a Defined Cohort

It may be possible to conduct a case-control study within defined cohorts. Such studies are called *hybrid* or *ambidirectional* studies. Cases are composed of all cases in the cohort, whereas controls are selected either at baseline or from a nondiseased group at the same time when a case arises. Two types of such studies are recognized: (1) nested case-control study and (2) case cohort study.

Nested Case-Control Studies

In nested case-control studies, cases are compared with a sample of the nondiseased members of the cohort who serve as controls and are selected at the same time the cases arise. The sampling in this design is called *incidence density sampling*. In this sampling method, controls are sampled throughout the study period—one waits for a case to arise and as soon as that happens, a time-matched control subject is selected from the same cohort. Therefore, because the control is selected after the start time of the study, controls contribute person–time to the study, which is why the sampling method is called incidence density sampling. In this type of design, it may be possible that a control selected at a certain point in time may itself become a case in the future. In such situations, these "future cases" are permitted to be controls for other cases. If the disease being studied is rare, then the probability of a control becoming a case in the future is very low. Although analyses of nested case-control studies may follow routine procedures, the effect measure is not an odds ratio, but is a *rate ratio* (or a density ratio). If cases are removed from the control group, then the effect measure is a *density odds ratio* (for definitions, see Chapter 4).

Case Cohort Study

In case cohort studies, controls are selected as a random sample at the baseline. Every person in the cohort has an equal probability of selection as a control regardless of the time contributed. It may thus be possible that a participant selected as a control at baseline may develop disease and become a

future case. The same control group selected at baseline may be compared with different disease sets that arise in the future. In contrast, if the controls are selected only from those who remain free of disease throughout the study period, the sampling method is called *cumulative incidence sampling*. The incidence odds ratio calculated from these studies approximates the incidence proportion if the disease is rare. Analysis for case cohort studies requires refinements that are addressed by survival analysis methods. When controls are selected from the total cohort at baseline, it yields a cumulative incidence ratio (relative risk), whereas when cases are excluded from the control group, it yields an odds ratio as a measure of risk.

Case Crossover Study

Classic crossover studies involve a sequence of exposures interspersed with washout periods and are discussed later in the experimental study section. Case crossover studies have been described as the case-control version of a crossover study. In a case crossover study, one person acts as his or her control. However, the key issue is the time point of exposure. Once a case is identified, and exposure is ascertained for the case, defined time periods are identified either *before* or *after* the disease onset or remission. These time periods are called control periods. The exposure status of the control is determined in the control periods. Thus, in this design, either the control crosses over to become a case or a case crosses over to become a control at a different time point. All diseases may not be amenable to case crossover design. This design also assumes that although the exposure must vary over time, the exposures and confounders do not change over time in *systematic ways*—at least not in the "control" periods. Exposures that remain constant over time cannot be used in this study design. To be amenable for this design, the exposure must be short lasting, have a short induction time, and have a short-lasting effect. Short-lasting diseases with or without episodic nature, and exposures that come in intermittent fashion, are suitable for case crossover design-based investigation.

Cohort Study

Cohort studies require that the investigators select two (or more) disease-free cohorts—one (or more) exposed, and the other unexposed—and follow these *study cohorts* over time to see how many cases develop in each group. The exposed group is called the *index cohort*, whereas the unexposed group is the *reference cohort*. The number of new cases and the rate at which cases develop are compared between the index and reference cohorts to yield the relative risk of disease given the exposure of interest. Cohort studies ask the question: What are the effects of this exposure? Box 2.4 enumerates the main

BOX 2.4 General Properties of Cohort Studies

Definition: Cohort studies compare exposed and exposure-free controls for their disease status after following them over time and compare the risk of disease in cases and controls. They can be designed as "prospective" or "retrospective studies." Cohort studies ensure that exposures occurred before disease outcomes. The term *inception cohort* is used to identify a group of persons who are aggregated together close to the onset of the disease (inception of disease).

Assumptions:

- Disease prevalence is not low.
- Exposed and unexposed are similar in all other respects and are comparable.

Case Definition:

- Case definition clearly defined a-priori because with new research information and understanding of disease, working definitions may change over time.
- Periodic follow-up should be planned in a way that new disease could be identified before its remission.

Comparison Groups:

- *Internal comparison*—A cohort of people may be followed, and the unexposed subsection of the cohort serves as the comparison group.
- *General living population*—Comparison cohort from another population may be used; available data on disease occurrence may be used.

Exposure Definition and Measurement:

- Exposure definition and measurement is critical to the study and it must be carefully defined and measured.

Advantages: Allows causal interpretation; can study multiple outcomes of an exposure; provides the real measure of risk of disease–relative risk; yields excess risk; can ascertain disease incidence (cumulative incidence and incidence density); can incorporate information about changing patterns of risk factors; can authoritatively assess dose–response relationship; can examine multiple outcomes of the same exposure; allows more control over subject selection, measurement, and control of measurement bias; multiple cohorts can be studied; and smaller case-control studies can be nested within cohort studies to make studies more resource efficient.

Disadvantages: Usually requires large sample size; usually expensive (retrospective cohort can be less expensive and logistically simpler, with shorter duration); takes long time for completion; difficult to conduct for rare diseases; logistics-intensive; subject to loss due to follow-up bias, healthy

BOX 2.4 General Properties of Cohort Studies *(Continued)*

worker effect, and so on; "case definition" may change in the future with new research; new diagnostic techniques may come out in the future compromising the study; may be affected by secular trends; and if multiple cohorts are studied, then probability of selection bias increases.

When To Do Cohort Study: When disease prevalence is high, when causal-association is being tested, when true relative risk is sought, when disease prevention methods are being tested, when disease does not have a very long induction period/latency, when etiological mechanisms are being assessed, and when multiple effects of exposures are being examined.

properties of cohort studies. Some key points about this study design are discussed next.

Population and Cohort Types

Study population in cohort studies can be defined in different ways. For example, ongoing cohorts may sample participants from within those living in a certain geographic area (the Florida Dental Care Study: http://nersp.osg.ufl.edu/~gilbert/); from those at high risk for a certain disease (Women's Interagency HIV Study: http://statepiaps.jhsph.edu/wihs/); from selected sample to investigate etiology and natural history of disease (Atherosclerosis Risk in Communities Study, [ARIC]: http://www.cscc.unc.edu/aric/); or from convenience cohorts formed of willing participants or for logistically convenient reasons (Nurses' Health Study: http://www.brighamandwomens.org/publicaffairs/NursesHealthStudy.aspx?subID=submenu5). Alternately, cohort studies that are more focused, smaller in scope, and shorter in duration may be conducted, such as HIV-Associated Oral Disease Study in North Carolina (Patton, McKaig, Strauss, & Eron, 1998), Tooth-Loss Risk Factor Study in Michigan (Burt, Ismail, Morrison, & Beltran, 1990), and the Multicenter Clinical Trial: Obstetrics and Periodontal Therapy (OPT) Study (National Institutes of Health [NIH], 2008).

Study groups in cohort studies can be of different types depending upon the amount of time their members contribute to the study. Sometimes, in a cohort study with multiple index cohorts, the index cohorts may be defined by levels/doses of the same exposure. As participants' exposure levels change over time, they may be moved to a different exposure index cohorts in the study. Studies that allow such movement of participants between index cohorts are called *open cohorts* or *dynamic cohorts*. This is sometimes confused with *open population*, which is defined as a population whose membership is changeable over time; that is, people are free to join or leave

the population. An example of open population is a state cancer registry that keeps track of oral cancer cases. Because people are free to move between states, their residencies may determine their membership in the cancer registry. However, cohorts are usually defined as a fixed group of persons sharing a common experience. Participants in an open cohort may be free to move *within* the cohort *between* exposure groups, but their handling is different when they leave the cohort study itself.

Fixed cohorts are groups within the cohort study where participants are not free to move between different exposure groups. In clinical trials, most participants are not free to move between treatment arms; the intent-to-treat analysis paradigm ensures that the treatment arm cohorts are fixed cohorts. A *closed cohort* is not the opposite of an open cohort. In a closed cohort, participants are usually free to drop out from a study if they choose, or they may be removed from a study if the conditions so demand. At the end of the study, if no participant has dropped out, then the cohort can be considered a *closed cohort* because the cohort composition remained the same without change for the entire study. This core element of fixed cohort is similar to the definition of a *closed population* where the population constituents do not change over time.

When cohort studies are set up in calendar time in such a way that the study starts in the present and continues into the future, it is called a *prospective cohort study*, *cohort study*, or *concurrent cohort study*. However, if a cohort study is defined in a database with data already collected such that the calendar time for the study to be completed has already passed before the formal study, it is called a *retrospective cohort study*, *nonconcurrent cohort study*, or a *historical cohort study* (see the case-control study design section earlier in this chapter about use of the terms *retrospective* and *prospective*). Because cohort studies start with a disease-free population divided into exposed and unexposed groups, the distinction between retrospective and prospective studies is simpler compared to case-control studies. Some cohort studies may be mixed, that is, partly retrospective and partly prospective. For example, index and reference cohorts may be identified from databases, and then followed up historically till present time as a retrospective cohort study. This follow-up may then be continued farther into the future, adding a prospective part to the whole study.

Exposure and Case Ascertainment

Case Ascertainment

Cohort studies start with a clear case definition and select participants who are disease-free at the starting point, which is called *baseline*. Their characteristics representing different factors under study are measured at baseline and then remeasured at predetermined follow-up intervals. Case ascertainment is simple because of the close follow-up that is maintained. The frequency of follow-up time must be carefully chosen depending upon the

characteristics of the disease. If the disease has a long induction phase and a long clinically detectable phase, then follow-ups may be longer. However, for diseases with a short induction period and a short clinically detectable phase, duration between follow-ups should be short. This becomes more problematic when the disease is transient and episodic.

For example, in a mixed cohort study of HIV-associated oral candidiasis, the follow-up periods were 6 months apart. The participants were clearly informed to perform self oral check-ups and report to the clinic immediately if they observed any whitish patch or streaks in their mouth. Because this was an HIV study, the compliance of participants with self oral check-up and immediate reporting was high and disease ascertainment could be carried out accurately and timely (Chattopadhyay, 2003; Chattopadhyay et al., 2005). However, for a person with borderline immune compromise, oral candidiasis could conceivably present as a mild, small nonfulminant episodic patch in a difficult to observe spot, or be wiped away by food or routine oral hygiene practices. Because the patient would perhaps not feel sick, the event may go unreported. The common clinical practice is to prescribe antifungal ointments for management of oral candidiasis. Patients are usually told to reapply the ointment if candidiasis recurs after initial remission. This poses a problem for studies that may want to record time between subsequent episodes of oral candidiasis if the duration between follow-ups is long because multiple episodes could be missed. Recognizing this problem, the primary goal in the studies mentioned above, was restricted to assessing the first episode of oral candidiasis.

Exposure Ascertainment

In most studies, it is generally assumed that exposures are chronic and constant. However, exposures may also be acute, transient, and may vary over time in terms of occurrence or dose. For example, the Stephan curve dictates that exposure to a low-pH environment will be shorter if frequency of sugar intake is less because in an hour's time, the pH is restored and further damage to enamel is minimized. However, if sugar intake is frequent, then exposure to low pH lasts longer. If this is taken to greater extremes and sugar is continuously present, then a chronic exposure to low pH occurs. Risk of dental caries increases directly with frequency of sugar consumption and is greatest in early childhood caries when sugar is habitually present in the mouth (Edgar & Higham, 1995).

Selecting appropriate exposure categories is important in cohort studies dependent upon the hypothesis being tested. In chronic exposures, such as sugar consumption, exposure assessment must distinguish between transient acute exposure and larger, chronic exposure while keeping in context the pathophysiology of the exposure–disease relationship. Chronic exposures also tend to be cumulative and may accelerate disease outcomes. If a child's exposure to a nursing bottle is measured as a dichotomous response

(yes/no), then we will not be able to understand the difference in early childhood caries risk between transient–acute exposure and chronic exposure related to nursing bottles. Therefore, careful ascertainment of exposure in terms of its start, duration, frequency, and periodicity will allow us to assess dose–response, dose threshold, and critical exposure duration. Such characterization of exposure is critical in making correct inferences about disease risk and etiological associations.

Loss to Follow-Up

Cohort studies assume that persons who are lost to follow-up are similar to those who remain in the study. Therefore, it is important to conduct a separate analysis assessing the dropout group to check the validity of the ascertainment in the index and the reference cohort groups. Differential loss to follow-up would induce a selection bias into the study. For example, in the HIV-associated oral candidiasis study mentioned above, if sicker participants (i.e., those with lower CD4+ cell counts and therefore at greater risk for oral candidiasis) started to miss follow-up periods, then it would be prudent to assume that they were perhaps hospitalized, and they should be followed-up there or their medical charts should be assessed. However, if the study ignored this systematic drop off of sicker participants, then the observed risk estimate in the study would be an underestimate of the true risk because individuals with stronger association between CD4+ cell counts and oral candidiasis would have been differentially eliminated from the study as dropouts. However, if loss to follow-up in index and reference cohorts is nondifferential, then these would cancel out when calculating the risk estimate, and the observed relative risk in the study would be the true risk. In this context, the risk estimate is biased only if there is a differential loss to follow-up between the index and reference cohorts. Investigators, however, cannot assume nondifferential loss to follow-up and therefore, the dropout phenomenon must be examined.

Cross-Sectional Study

Cross-sectional studies measure disease while exposure statuses are measured simultaneously in a given population. Cross-sectional studies ask the questions: "How common is the condition?" and "Are exposures and diseases associated?" These studies are sometimes thought of as "freeze-frame" of the population that provide a "snapshot" of the disease and exposure characteristics in a population at a particular point in time. The participants in a cross-sectional study are sampled from a population and then classified into disease and exposure categories that are then compared. Obviously, in such a study design there is no way to ascertain incident disease or whether the exposure came before or after the onset of disease. Cross-sectional studies that aim to assess disease prevalence are called *prevalence studies*. Analyses

of cross-sectional studies compare the point prevalence rates between the exposed and the unexposed group. These analyses assume that the data came from case-control studies and then follow the analytical paradigm of case-control studies. The main advantages of cross-sectional studies are that they may examine several exposures and outcomes at the same time; are quick, easy, and relatively inexpensive to conduct; provide prevalence and relative prevalence estimates; and may provide good insight in developing cohort studies. However, cross-sectional studies do not allow inferences related to temporal sequences between exposure and disease, cannot accommodate changes in exposure and outcome rates over time, are not efficient if disease or exposures are rare, and do not provide estimates of incidence rate or the relative risk (for definitions, see Chapter 4).

Despite their limitations, cross-sectional studies can be very useful, for example, in genetic epidemiology because they can provide an estimate of genotype frequencies, allele frequencies, and population exposure levels. These studies can also provide an assessment of relationships between genotypes, genetic variants, phenotypes, and population-level environmental exposures in a relatively short time with little expense. Such information may be useful in policy formulation, designing hypothesis-driven studies, and in interventions. Cross-sectional studies are also useful for surveillance; for example, cross-sectional surveys such as National Health and Nutrition Examination Survey (NHANES), Behavioral Risk Factor Surveillance System (BRFSS), Medical Expenditure Panel Survey (MEPS), National Survey of America's Families (NSAF), and several others provide data into national surveillance systems. Cross-sectional studies are, however, not suited for etiologic research or causal analyses because they cannot identify incident disease, cannot establish temporal sequence of exposure of disease, and are subject to length bias (a function of duration of disease).

Types of Study Designs: Experimental Studies

Experimental studies are those in which the investigators control the exposure. Experimental studies generally ask the questions: "What are the effects of this change of conditions?" and "What are the effects of this intervention?" Clinical trials are completely randomized experimental study designs. These very important experimental studies are discussed in Chapter 11 under pharmacoepidemiology. Community trials are experimental studies that are carried out at the community level and follow the same general methods as in clinical trials except that such trials have to accommodate the potential for exposure modification from community-level factors, competing exposures, and changes in secular trends. In this section, we discuss other experimental study designs such as randomized block designs, stratified designs, split-plot designs, crossover designs, and factorial designs.

Crossover Study

In crossover studies (common in oral disease-related experimental designs), the same person (i.e., the case) receives two (or more) types of exposures in sequence. After the first exposure/treatment is delivered, outcomes are noted, and the person is allowed an exposure-free period called the "washout period" when the exposure and its effect are allowed to be completely eliminated. Thereafter, the subsequent exposure is applied and outcomes are noted. The results of the two exposures are then compared. The key advantage of such a design is that the same person acts as his or her own control or comparison group, thereby effectively matching for all person-related factors. The disadvantage is the possibility that the first exposure may in some way influence the responses from the second factor, which may or may not be independent of the adequacy of the washout factor. In some situations, it may be possible that the effects of the first exposure persist for a long time, which is called a *carryover effect* from the first exposure. Carryover effects may impact the outcomes from the second exposure. In education-related intervention, the possibility of learning from the first exposure can significantly impact learning outcomes of the second exposure if substantive learning of the first exposure is cumulative and correlated with the learning capacity from the second exposure. The washout period is generally designed to dilute this learning effect. This design should not be confused with pretest–posttest designs because in pretest–posttest designs, the effect of *one exposure* is assessed over two time periods, whereas in crossover studies, *two (or more) exposures* are assessed sequentially.

For example, a recent study compared the antimicrobial effects of a new 1% zinc citrate dentifrice with a control formulation (Sreenivasan et al., 2008). The investigators collected baseline (and subsequent) samples of dental plaque, buccal mucosa, tongue, saliva, and plaque. Thereafter, a washout phase was instituted and then a test dentifrice was randomly assigned to the participants to use for the next 13 days. This was followed by a second washout period after which the study was repeated with the alternate dentifrice. Because people use dentifrices daily, in this study, the first washout period was necessary to remove the carryover effects from their regular dentifrice. Thus, to remove the carryover effects of the assigned dentifrice, the second washout period was needed.

Split-Plot Design

The split-plot design uses one side of the person as a case and the other side as a control. For example, in a person with generalized periodontitis, one side (left or right) may be treated with a surgical or medication intervention and the results may be compared with the untreated other side of the

mouth. In general, such designs are called split-plot designs, although oral/dental studies using this design prefer to label them as *split-mouth designs*. Split-plot designs may or may not add a crossover design component to them. Split-plot designs are the only "pure" repeat measure design where *exactly* the same person (experimental unit) is observed under more than one set of conditions at only one measurement location to yield two (or more, if more than two splits are designed) sets of measurements at the same time.

In this design, essentially, the unit of the experiment—the physical entity is considered as a "plot," and is split into several locations—is used for carrying out the experiment. The plot is considered a uniform entity, and any observed differences in the experiment can be attributed to factors other than the plot. The key reason for using such a design is to minimize the variability due to responses between comparison groups. Different physiological processes in the body can be viewed as being *nested* within the body just as a nucleus is nested inside a cell. Although these nested entities work in coordination, changes in parameters common to these nested entities will affect them somewhat differently compared to changes in another nested set of entities (e.g., another person). The split-plot design tries to minimize *nested variation*.

This design is especially useful when the investigator wants to control for factors that are very difficult to control, such as the genotype, salivary flow rate, and immunological response among others. Therefore, in this design, the control is exactly matched to the case for all measured and unmeasured sources of variation (see the section on counterfactual concept in Chapter 3 for a discussion on exact matching). For example, a recent study reported a randomized controlled clinical trial using a split-mouth design to evaluate the clinical performance of a plasma arc light against conventional tungsten-quartz halogen curing light for direct orthodontic bonding. The authors divided the mouth into quadrants that were randomly assigned to treatment groups (Russell, Littlewood, Blance, & Mitchell, 2008). They found that bracket survival, patient sensitivity, discomfort, and rebond times using the plasma arc light and conventional halogen light were similar, but the bond-up times were typically reduced by 204 seconds per patient with the plasma arc light.

Randomized Block Design and Other Designs

Blocks are collections of experimental units (e.g., participants in a study) that are similar to one another. For example, participants may be grouped by blocks of sex (two blocks) or race/ethnicity (multiple blocks). The participants in each block are then randomized to treatment groups. The blocks should be as homogenous as possible. Factors on which blocks are created are generally considered "nuisance" factors; the design "removes" the effect

of the blocked factor. For example, if the investigator thinks that race/ethnicity impacts the outcome of a drug trial, then participants can be "blocked" by race/ethnicity. The drug can be randomized within each race/ethnicity block to "remove" the effect of race/ethnicity and the investigator can then assess the effect of the drug on clinical outcomes within each race/ethnicity block. The general rule of thumb often used in controlling nuisance factors is to "block what you can, and randomize what you cannot."

For example, a recent study attempted to choose the best retraction agent by evaluating gingival inflammation related to three kinds of retraction agents (15.5% ferric sulfate, 25% aluminum chloride, and 0.1% epinephrine hydrochloride) and comparing to the control group (sodium chloride). The study used a randomized block design to allocate 40 maxillary premolars to the four treatment groups (Sun, Sun, & Xiao, 2008). Another recent study used a modified randomized block design in a two-arm randomized trial to evaluate the efficacy of a couples-based intervention designed for HIV-serodiscordant African American couples in four U.S. urban areas. The investigators used the Eban HIV/STD Risk Reduction Intervention as a treatment compared to the Eban Health Promotion Intervention as the control. The Eban HIV/STD Risk Reduction Intervention addresses multilevel individual-, interpersonal-, and community-level factors that contribute to HIV/STD transmission risk behaviors among heterosexual African American couples who are HIV serodiscordant. This study used the gender of the HIV-positive partner as the "blocking" factor to ensure that the distribution of HIV-positive men and women was equal across the interventions (National Institute of Mental Health [NIMH], 2008).

Randomized block designs differ from *stratified designs* where subjects are categorized into subpopulations called strata, and within each stratum, a completely randomized design is conducted. This is much like the blocked design, except there is only one sample, at least conceptually, from the strata. Examples might be litters of laboratory animals, surgical practices, or batches of a therapeutic agent. The desire is to make inferences about treatments in the population as a whole, not just in the strata that were actually sampled.

Factorial designs categorize interventions by two (or more) independent factors and randomize participants in the resulting groups. The advantage of this design is that it allows the investigator to simultaneously assess the effects of two (or more) independent factors on the outcome (main effects), and how these factors may modify each other (interaction effects). Obviously, this is a posttest-only design. The number of factors can be many. These designs are usually identified by the number of factors and their levels being examined. For example, a design with two factors having two levels each would be a 2X2 factorial design with four groups, whereas two factors each having three and five levels would be a 3X5 factorial design with 15 groups; a study with three factors having two, three, and four lev-

els, respectively, would be a 2X3X4 factorial design with 24 groups. Factorial designs are flexible because the design allows the investigator to examine the effects of treatment variations in an efficient manner compared to a series of independent studies assessing the effects of the factors concerned. Factorial designs also allow the investigator to examine the effect modification between two (or more) factors.

Table 2.1 provides a hypothetical example of a factorial design. In this example, the investigators are interested in examining the usefulness of allowing a trained public health nurse to apply fluoride varnish on children's teeth because the investigators argue that it would be a more cost-effective way to prevent early childhood caries (ECC) in public health programs. The investigators want to find if there are any differences in clinical outcomes when the varnish was applied by a pediatric dentist or a trained public health nurse. At the same time, the investigators want to examine whether a second follow-up application of fluoride varnish has an impact on significantly reducing ECC outcome (main effect) over a one-time application. They are also investigating whether outcomes vary with number of applications (by adding a second follow-up application) dependent upon application by pediatric dentist or nurse over one-time/follow-up application (interaction effect). The study could be designed as a 2X2 factorial design and randomly allocate children to the four treatment groups: (1) pediatric dentist, one-time application; (2) pediatric dentist, follow-up application; (3) nurse, one-time application; and (4) nurse, follow-up application, and the results would be assessed.

Types of Study Designs: Ecologic Studies

Ecologic studies measure factors at the group level and compare groups rather than individuals. Ideally, both outcomes and exposures should be measured at the group level. For example, the association of dental fluorosis incidence rate of a country with per-capita water consumption measured from public water supply is an ecologic study because the disease and exposure are both measured at the community level. However, measuring fluorosis at an individual level and measuring fluoride "exposure" from the water content of public water supply systems would imply that the exposure was measured at the ecological level, thus giving rise to ecological fallacy. To avoid this ecological bias, samples of individual exposure and confounder data within each area are required, which may be difficult to obtain.

Amstutz and Rozier (1995) conducted an ecologic study by examining factors associated with variation in dental caries prevalence in school classrooms, using classrooms as a surrogate for the larger community, in order to identify community risk indicators for dental caries. Although they measured DMFT and DMFS in children, they used only the average scores for

TABLE 2.1 Factorial Design

		Factor 1 with Two Levels	
		Fluoride varnish: one application	*Fluoride varnish: two applications*
Factor 2 with Two Levels	Dentist applied	Outcome group 1	Outcome group 3
	Nurse applied	Outcome group 2	Outcome group 4

There are two factors: (1) application time of fluoride varnish, and (2) type of professional applying fluoride varnish. Each factor has two levels.

classroom (group-level variable) among others, such as population density, parental education, coastal residence, age, and Medicaid expenditures, in their models. The investigators concluded that for population-level caries risk assessment, models based on community rather than individual variables were feasible, and suggested model refinement to further reveal factors useful in identifying high-risk communities.

Ecologic studies are usually relatively easy to conduct using routinely available data, are less expensive, and should be generally viewed as hypothesis-generating rather than hypothesis-testing studies. There are several disadvantages of ecologic studies, however, especially because they are subject to ecological fallacy. Ecological studies are often victims of misinterpretation and uninformed persons draw individual-level causal conclusions from the results of such studies, leading to disinformation among the lay public. Studies relating water fluoride level with skeletal fractures may fall victim to such misinterpretation if exposure ascertainment is done at the ecological level (see Chapter 17 for discussion on this topic).

Ecological inference is the process of extracting clues about individual behavior from information reported at the group level, and not about providing conclusive evidence about correctness of those clues. While collecting group-level data, individual-level information may be lost, often in a systematic way, thus introducing information bias. Recently, perhaps powered by the easy and large availability of data and the increase in computing power, interest in ecological inference has grown tremendously, which has resulted in improved methods for analyzing ecological data and providing useful results. However, ecological inference is not an easy process. Analyses of ecological data require special analytical skills. For example, Wakefield and Shaddick (2006) developed a model in a study of the association between mortality among the elderly and the previous year's environmental sulphur dioxide level in London. They showed that modeling the exposure surface and estimating exposures may lead to bias in estimation of health effects and developed statistical procedures to avoid ecological bias. They concluded that the use of their "proposed model can provide valid inference, but the use of estimated exposures should be carried out with great

caution" (2006). Using simulated data and a practical illustration through an analysis of lung cancer mortality and residential radon exposure in counties of Minnesota, Salway and Wakefield (2008) combined a Bayesian nonparametric Dirichlet process prior probability with an estimating functions' approach to develop a hybrid model for reducing ecological bias. For an ecological study to be feasible using only small samples of individual data, success of this model requires good quality prior information about exposure and confounder distributions and a large between-to-within-area variability ratio. This procedure was then extended to correlate ecological data with supplemental case-control data to develop a Bayesian spatial random effects model (Haneuse & Wakefield, 2008). These authors have suggested that their proposed design may be used to resolve the ecological fallacy.