

SECTION



Epidemiologic Concepts in Oral Health

1

Definition and Background

Oral health is defined as “being free of chronic mouth and facial pain, oral and throat cancer, oral sores, birth defects such as cleft lip and palate, periodontal (gum) disease, tooth decay and tooth loss, and other diseases and disorders that affect the mouth and oral cavity” (World Health Organization [WHO], 2008a). Although the definition of epidemiology has undergone changes over time, the current and most useful definition is provided by John M. Last: “epidemiology is the study of distribution and determinants of health-related states or events in specified populations, and the application of this study to control of health problems” (Last, 2001). By extension, oral epidemiology can be defined as the study of distribution and determinants of oral health-related states or events in specified populations, and the application of this study to control of oral health problems. For convenience, we use the term *disease* to imply all impairments of health or conditions of abnormal functioning in its broadest application, including illness, sickness, abnormal conditions or states, and injuries.

Within the field of epidemiology, oral epidemiology is the only subdiscipline that is defined according to an anatomic section of the body. Other subdisciplines are either defined by types of diseases or by pathophysiologic or other processes. For example, epidemiology may be defined according to disease or outcome such as infectious disease epidemiology, chronic disease epidemiology, cardiovascular disease epidemiology, cancer epidemiology, injury epidemiology, reproductive epidemiology, and so on. Alternatively, epidemiology may be subdivided by type of application or exposure such as: environmental epidemiology, occupational epidemiology, nutritional epidemiology, behavioral epidemiology, epidemiology of medical care and pharmacoepidemiology, among others. Scientific and socioeconomic-political developments have established several more areas of epidemiology

such as epidemiology of aging, genetic epidemiology, molecular epidemiology, epidemiology of war or disaster, climate change epidemiology, and several more. In its entire vision and scope, epidemiology has become established as a truly interdisciplinary science. Oral epidemiology, based on an anatomical definition, therefore encompasses all other subdisciplines of epidemiology as applied to the orofacial region.

Epidemiology embraces the central dogma of science: that the universe is understandable, and it involves a central assumption that *diseases do not occur at random*. Epidemiology presupposes that there exist causal, enabling, contributing, and preventive factors that protect or predispose populations to diseases. Following the central dogma of science, epidemiology assumes that factors affecting diseases can be identified through systematic investigations and manipulated by human agency.

Essentially, epidemiology examines interplay of three fundamental aspects of diseases: person, place, and time. Therefore, distribution of disease is described by answering the questions: who, where, and when? Overall, determinants of diseases are characteristics that influence occurrences or propagation of diseases, which have classically been described to form three angles of a triangle contributed by the host, the agent, and the environment. Determinants of diseases are many, and these may exhibit complex interplay among each other; depending upon the type of role they play in the disease process, these may be named or classified differently. The same factor may have different roles in different diseases. In general, epidemiology views a disease as an outcome of a series of interacting chain of events. By understanding the mechanisms involved in this chain of events, epidemiology aims to eventually prevent occurrences of diseases, or at least, to improve disease outcomes. Specifically, epidemiology aims to find etiology (cause) of disease, define the extent of disease occurrences (burden of disease), study the natural history (progress) of diseases, assess therapeutic interventions and policies, and identify modifiable factors that can impact disease occurrences in some meaningful way by providing a strong foundation on which better health policies can be built. Advanced understanding has modified the classical epidemiological triangle to incorporate other factors and rename the angles of the classical triad (see Figure 1.1).

There exists no single “theory of epidemiology.” Models of disease causation based on principles from all branches of science are generally used as guides to the practice of epidemiology. Epidemiology uses methods of experimentation and analyses borrowed from different fields toward its overarching goal of examining distribution and determinants of diseases in populations.

Epidemiology is to population what clinical medicine is to the individual. Epidemiology differs from basic sciences in that basic sciences are involved with the fundamental mechanisms of the disease process, whereas epidemiology is involved with disease mechanisms at the population level.

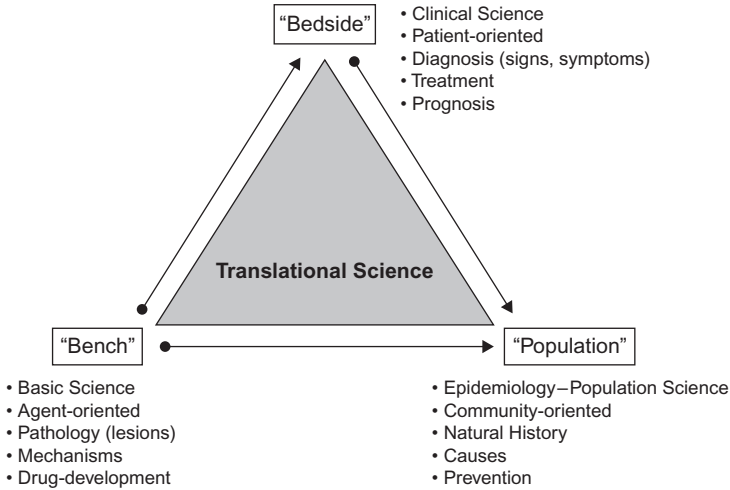


FIGURE 1.1 Triad of Biomedical Science

An analogy may be drawn by comparing *efficacy*, which represents how well drugs work under experimental conditions, with *effectiveness*, which relates to how well drugs work in real-life situations. These two attributes may differ because even though a drug is very efficacious, if it has inconvenient dosages or problematic side effects, patients may not comply with it, thereby reducing its effect in real-life situations (effectiveness). Another example of different concepts between clinical medicine and epidemiology can be demonstrated by herd immunity. Whereas active or passive immunity imparts resistance to disease to the individual, herd immunity aims to restrict propagation of the communicable disease in the population. By immunizing most (not all) people in the population, the means of propagation of the disease is disrupted, so that the disease may eventually disappear in the population, or remain in controlled, manageable proportions even if all people in the population may not be disease-free.

Epidemiologic studies generally follow a series of steps that are called the “epidemiologic sequence”—a misnomer because the sequence is often disrupted. This “sequence” includes observing by counting cases and events, relating cases and events to the population at risk, making comparisons, developing hypotheses, testing hypotheses, making scientific inferences, conducting experimental studies, intervening, and evaluating. Ultimately, epidemiology examines the associations between sets of events, defined as outcomes, and determinants of those outcomes. An outcome in one study may be a determinant in another study. Similarly, a disease may

be an outcome in one study but an exposure in another study. Several associations may exist between factors, but not all associations are causally linked. The key questions that epidemiology tries to answer are (1) is the observed association real or spurious? and (2) is the association causal (i.e., exhibiting a cause–effect relationship)? Thereafter, epidemiology tries to establish whether the determinants of outcomes are independent. In epidemiology, determinants of diseases are often called “exposures,” which may be causative factors or protective factors for diseases.

Koch’s postulates mandated one organism for a disease and Sir Austin Bradford Hill’s (1965) causal criteria suggested one cause for a disease. However, it is generally seen that although certain diseases may have a single cause, most diseases are outcomes of a complex interplay of several factors in different ways under a variety of environments and conditions. These observations have also instituted a paradigm shift in thinking about disease causality. Current understanding has deviated from traditional “one-cause, one-disease” paradigm toward interaction of multiple causes classified in several different ways: sufficient cause, necessary cause, or component cause; or causes that may be modifiable or nonmodifiable, acting at the same or different levels of exposure.

Observational vs Experimental Epidemiology

Observational epidemiology includes observing the effects of exposures on populations. In this situation, the exposure is not in the control of the observer (investigator), and the investigator merely observes the effects of prevailing exposures. For example, the investigator examines HIV-1–positive patients and notes their oral diseases and compares these with those who are HIV-1 negative. In this example, the observer did not have any role in the patients being exposed to and infected with HIV-1. Similarly, the observer may compare the results of different treatments carried out in a hospital—although the patients were treated by a clinician the observing investigator played no role in the treatment.

On the contrary, in experimental epidemiology, the exposure is under purposeful control of the investigator. For example, the investigator may treat one group of partially edentulous persons by providing them with removable partial dentures and providing another similar group of patients with implants, then compare chewing efficiency and patient satisfaction with the two rehabilitation schemes. In this situation, the investigator chose which kind of exposure (partial denture vs implants) was provided to which group, thereby conducting an experimental study. Random assignment of the exposure is a hallmark of true experimental study designs. Clinical trials of drugs and devices are experimental studies. However, in some situa-

tions, although the exposure is under the investigator's control, random assignment may not be possible. Such studies are generally classified as *quasi-experimental studies*.

Descriptive Epidemiology

Descriptive epidemiology provides a general description of the distribution of disease and/or factors associated with disease in a population in terms of person, place, and time. Such description can be obtained from new data or preexisting data. Descriptive epidemiology may be viewed as the first step in examining a disease and/or exposure, and is useful in generating hypotheses about exposure and outcome. Systematic differences in the distribution of disease/exposure can provide major insights into disease occurrence, etiology, and mechanisms. For example, a cancer cluster immediately alerts the investigator to look for possible local environmental exposures that may be linked to the cancers. Similarly, an outbreak of infectious gastrointestinal disease requires tedious description of the affected persons and the food items they might have eaten during the purported exposure period.

Disease surveillance systems rely on descriptive epidemiology. For example, the National Oral Health Surveillance System (NOHSS; Centers for Disease Control [CDC], 2006) uses descriptive epidemiology to disseminate important oral health information. The NOHSS was established in 2001 as a collaborative effort between the CDC Division of Oral Health and the Association of State and Territorial Dental Directors (ASTDD). The NOHSS is designed to monitor the burden of oral disease, use of the oral healthcare delivery system, and the status of community water fluoridation on both a national and state level. It includes eight indicators of oral health, information on state dental programs, and links to other important sources of oral health information.

Person-level factors that are often assessed in epidemiology include age, sex, race/ethnicity, individual behavior/life style, cultural values, education, family size, employment, income, presence of insurance, stage in life (e.g., fetal, childhood, youth, adolescence, adulthood, old age, etc.). Sometimes, a distinction is made between use of the term *sex* and *gender* to define biological sex of the individual. Efforts for political correctness nudge us to use the word *gender* to define biological sex so that insinuation to "sexual act" is avoided. This works very well for the biological–medical model. However, with the paradigm shift from biological to sociobiological models of disease causation, the meaning of the word *gender* has become more important. Sociologically speaking, the world divides humans into two genders, male and female, based on the types of work one performs. Under this concept, a stay-at-home father assumes a "female" gender role, and a professionally occupied mother assumes a "male" gender role. In examining

associations of parental influence on specific behavior attributes of children, merely classifying parents by biological sex while disregarding the “changed” gender roles may lead to misclassification of exposures. As transsexualism and gender reassignment surgery becomes more availed, gender–sex related issues will become important, more so because of involved legal, ethical, and moral challenges to the society (Sharma, 2007).

Place-level descriptors used in epidemiology include definitions of clusters, geographic zone of the spread of disease/exposure, climate, rural/urban infrastructure, location of factories, workplace environment and other shared environments, sanitary conditions, and common sources of infection or disease propagation, among others. Accurate description of place-related factors becomes important in most epidemiological work, especially in war- or disaster-affected areas or places with special characteristics. For example, disease patterns in correctional facilities may vary substantially compared to the “outside world”—a study demonstrated recently that the oral and general health of remand prisoners was severely compromised compared to the general population in the United Kingdom (Heidari, Dickinson, Wilson, & Fiske, 2007). Similar observations have been reported in South Africa (Naidoo, Yengopal, & Cohen, 2005) and the United States (Heng & Morse, 2002).

Time is the most difficult of all concepts to address in epidemiology. Descriptive epidemiology incorporates time as a calendar-year-based entity, and describes disease/exposure distribution in blocks of time period. The selection of time period chosen for describing disease is arbitrary and generally attributed to conventional practice of convenience. However, certain disease may occur at different times in different manners, such as seasonal allergies and episodic infections. Secular trends are occurrences of disease and outcomes over time, most commonly described over years. For example, a recent study from Italy reported a reduction of upper arch width from the 1950s to 1990s (Defraia, Baroni, & Marinelli, 2006), and another study described changes in transverse dental and dental arch depth dimensions among Norwegian children from the 14th to the 19th century (Lindsten, Ogaard, Larsson, & Bjerklin, 2002). Interpreting secular trends needs care. Because outcomes are compared over several years or decades (even centuries), such observations are especially susceptible to biased overinterpretation as functions of new knowledge. Threats to correct interpretation of secular trends include changes in disease definitions, altered categorization of diseases, establishment of new disease entities, changes in disease outcomes, newer and more accurate diagnostic techniques, updated understanding of disease etiology, “new”/evolved mechanisms of diseases, demographic changes in a locality, changes in living conditions, lifestyle changes, landscape changes, catastrophes, and migration.

Epidemiological transition is a change in patterns of diseases in society that occur regularly. Such shifts may manifest in different ways such as al-

tering of disease pattern in a population from primarily acute-infectious in nature to a mainly “chronic” type of disease. An example of such a transition in contemporary times is HIV/AIDS. In the early 1980s, HIV/AIDS was essentially an infectious disease with fulminant upswing in its population dynamics. However, in the developed world, with successful highly active antiretroviral therapy (HAART), HIV/AIDS has turned into a stable chronic disease with much more controllable dynamics, and apparently, this stability can be maintained as long as HAART remains effective. Diseases considered to be eradicated have often reemerged in a modified form; that is, newly emerging and reemerging infections also contribute to epidemiologic transition. Direction of epidemiologic transition need not necessarily be from infectious toward chronic disease. At any one time, epidemiological transition of several different types may coexist. For example, transition of HIV/AIDS, emergence of multidrug-resistant tuberculosis, occurrences of prion diseases, severe acute respiratory syndrome (SARS), and increased occurrences of carpal tunnel syndrome have existed together globally in recent decades. Several mechanisms and factors may be involved in contributing to epidemiologic transition such as demographic changes; risk factor changes; biologic phenomena such as antigenic drift and shift; drug resistant strains; social, cultural, and environmental factors; increased travel and migration; increased stress levels; bioterrorism, wars, and disasters; iatrogenic factors; and advances in medical science and technology.

Analytic Epidemiology

Analytic epidemiology provides systematic assessment of relationships and hypotheses. These studies primarily test specific hypotheses. Although other hypotheses may be generated as an outcome of analytic epidemiological studies, the primary goal of analytic epidemiology is to analyze data and test hypotheses.

Analytic epidemiology opens up several prospects for assessing associations between exposures and outcomes and series of factors that may cloud these associations or may impart different associations in different categories of certain important factors. These associations are expressed through mathematical models. If a factor can be divided into two categories, it is called dichotomous, whereas several levels of the factor make it polytomous or simple multilevel factor. Most models take the form of an equation with the outcome factor (*dependent variable* because its value is dependent on several other factors) on the left-hand side and the explanatory factors (*independent factors* because in the equation these variables can take any value independent of the outcome or other factors) on the right-hand side. Depending on the nature of the data, both dependent and independent variables may be continuous and/or categorical, and they may be single or multiple.

Describing the statistical details of a single variable under study is usually referred to as univariate analysis, whereas assessing the relationship of two variables is called bivariate analysis. There exists some terminology-related ambiguity in epidemiology and biostatistics literature when multiple variables are assessed together in statistical models. To describe models, the terms *multivariable* and *multivariate* are often used interchangeably. This was probably acceptable when most analyses involved a single dependent variable. However, with more advanced techniques being available in the epidemiology repertoire, modeling of multiple dependent variables has become commonplace these days. In this context, and to avoid ambiguity when reading and comparing literature, good analytic epidemiological practice dictates using the term *multivariable model/analysis* for those analyses that have a single dependent variable (the model may have several independent variables). For example, modeling the decayed, missing, or filled teeth (DMFT) score as an outcome or modeling the odds of presence/absence of a disease using multiple independent variables would be a multivariable analysis.

In contrast, the term *multivariate model/analysis* should be used for those models that have multiple dependent variables (the model may have several independent variables). For example, modeling occurrences of four different disease outcomes such as oral candidiasis (OC) only; oral hairy leukoplakia (OHL) only; both OC and OHL together; and all other HIV-associated oral diseases in HIV-1 infected persons, using several independent variables in the same model, would be an example of multivariate analysis. Although multivariate analyses are not yet very common in oral epidemiology, such analyses will be used more frequently in the future along with several other types of analytical methods that are uncommon today, including imputation methods, cluster analyses, nested models, different Monte Carlo methods, Bayesian models, multilevel models, multilinear methods, and several others.

Assessing Association

The practice of epidemiology bestows important responsibility on its practitioners. The need for information and diagnostic certainty and correctness of conclusions depends upon the penalty for being wrong about the true state of the population and the patient. The chances we are willing to take determine the burden of mortality and morbidity in the society, of which we are ourselves a contributing part. Epidemiology identifies and assesses associations between outcomes and determinants. One of the major charges in this exercise is to ascertain causation or establish a causal association. Epidemiological paradigm suggests that associations may be many and not all are causal—just as wisdom suggests: All is not gold that glitters!

Ambiguity in usage of terms is common in causal research. When trying to distinguish between causal associations and noncausal associations the term *risk factor* is used indiscriminately for all factors associated with the outcome, whether causal or not. The term *risk factor* should be reserved only for those factors that are causally associated with the outcome. The non-causally associated factors that may serve to indicate disease or its outcomes should be called *risk indicator*, *risk determinant*, or by other terms (Beck, 1998; Burt, 2001). For example, high sugar consumption is a risk factor for dental caries, but minority status in a society may only be a risk indicator for dental caries.

Per se, epidemiology is a population science, and causal associations are interpreted at the population level. However, epidemiological principles can be used in different settings and causal analyses can be conducted specific to that level. It is important to be constantly aware that the *unit of exposure* (that for which exposure has been measured) and *unit of analysis* (that entity about which analysis is being performed) are congruent for logical inferential conclusions. If the exposure is measured at a different level that does not correctly represent a person-level exposure, but the outcome is measured for the person and a purportedly causal association is inferred, then the causal conclusion is misplaced. For example, a retrospective cohort study report concluded that fluoride in water increases the risk of hip fractures among women (Kurttio, Gustavsson, Vartiainen, & Pekkanen, 1999). Whereas hip fracture was measured at an individual's level (person with hip fracture), fluoride levels were based on smoothed data from the fluoride registry averaged for the place where the women lived, and not upon actual measurement of individual fluoride consumption/biological ascertainment. Such a conclusion is called *ecological fallacy* because the outcome and the exposure were not measured on the unit of analysis (i.e., unit of measurement and unit of analyses were different).

Causal associations have more threats. Let us consider a hypothetical example. From a multivariable analysis, it was found that regular sugary hard candy consumption was associated with the decayed, missing, or filled tooth surface (DMFS) score of children in a study sample. The model had several factors included, among which was a significant variable—parental income. The report mentioned that candy was a risk factor for dental caries, whereas parental income was a risk indicator. The justification for this conclusion outlined the etiopathology of caries and the central role of glucose in the process, and explained that because parental income was not involved in the biological etiological pathway of caries, it could not be a risk factor and was therefore classified as a risk indicator. When this manuscript was sent for publication, a peer reviewer turned the argument around saying that low parental income would lead parents to handle multiple jobs leaving little time for their household chores and attending to children.

Therefore, parents would keep giving candies to their children to keep them satisfied and silent. Furthermore, low-income families would also compromise on oral hygiene measures and contribute further to occurrence of caries. According to this line of argument, because the income caused the increase in candy consumption and reduced oral hygiene maintenance, parental income is a causative factor, and not just a risk indicator! It is easy to understand that the author of the manuscript was concerned only with *biological causation*, whereas the reviewer brought in the concept of *social causation of disease*. Therefore, causal inference may vary depending upon the type of disease model being followed. However, parental/household income is not a child's individual-level exposure (a child's claim to returns from such income varies with parents' assessment of the importance of the issue in question, the child's age, personality, number of siblings, and the seriousness of other pressing needs the family may face). If parental income is included in the model as a "causal" factor, the most appropriate way to use it would be to define it as a higher-level variable in appropriate modeling techniques such as multilevel analysis to avoid ecological fallacy.

The two perspectives described above have major implications. Those professing the sugar-caries causation perspective could call for policies that ban candies, whereas those professing parental income-caries causal association would argue for an increase in income opportunity, social equity, social justice, and improvement of dental insurance mechanisms. Depending upon the type of policy professed, the associated budgets and infrastructural support needed would also vary.

It is being increasingly recognized that to prevent disease, target risk factors must be modifiable. Furthermore, it is also known that a large burden of disease lies on those who need the most help and have minimal resources to address these needs. For example, a large proportion of the dental caries burden is concentrated among the poorest and most needy families (United States Department of Health and Human Services [USDHHS], 2000). Some of the socioeconomic factors may be more amenable to modification and have wider general impact over disease-specific preventive measures, making for prudent and more efficient policies. Similarly, with an increasingly global interaction among people and increasing migration, population dynamics are changing across nation states rapidly. These factors raise more challenges to disease prevention efforts. Understanding the dynamics of disease patterns requires better sociocultural understanding of people from diverse backgrounds. This need has opened up possibilities for social epidemiology in a big way.

Qualitative Research

While discussing analytic epidemiology, we concentrated on quantitative methods to draw conclusions from studies. However, there are several situ-

ations where quantitative methods are not applicable or do not work well. In many such situations, qualitative research methods are useful, especially in social epidemiology and some behavior research areas. *Qualitative research* has been characterized as “multi-method in focus, involving an interpretive, naturalistic approach to its subject matter” (Denzin & Lincoln, 1994). These methods also generate and analyze data, but use different techniques compared to the usual epidemiological quantitative methods, such as content analysis, grounded theory analysis, triangulation, and narrative data analysis. Data for such research may be generated from focus groups, cognitive thinking, semistructured or open-ended questionnaires, interviews, and narratives, and may lead to important insights and explanations of the impact of the social phenomenon on disease occurrences (Sisson, 2007).

Although proponents of qualitative and quantitative research seek exclusive sway over the practice of research methods professing the advantages of their favorite methods, in reality the two are not replacements for each other. They serve different territories, and there are situations where qualitative research is better suited over quantitative research and vice versa. For example, if one wishes to gather information about the types of barriers that a certain population faces for accessing the healthcare system, qualitative research would probably be the path to take. However, if one wishes to estimate how much each cited barrier contributes to the population’s overall healthcare system utilization, quantitative research would provide the answer. Depending on the type of research question one asks, both qualitative and quantitative research can be brought together to provide comprehensive answers. Such approaches are called “mixed-method” research. This should not be confused with a mixed-model analytical approach that implies multilevel modeling.

Health Outcomes Research

All actions have outcomes, which could be positive (as hypothesized), or negative (unlike as hypothesized). Therefore, whether we examine a program, a new device, a new drug, a new communication method, or a health promotion drive, these have to be assessed for the effects they produce. *Outcomes research* aims to understand and assess the end results of particular healthcare practices and interventions. The Agency for Healthcare Research and Quality (AHRQ, formerly the Agency for Health Care Policy and Research) emphasizes that outcomes research is the key to knowing not only what quality of care we can achieve, but how we can achieve it (AHRQ, 2000). Clinicians usually assess the efficacy or effectiveness of treatments by using measures of disease process through clinical examination or using biological specimens of such tests. From a biological–medical view, this paradigm tests whether the biological abnormality is no more de-

tectable without considering the patient as a whole human being, thus ignoring the patient's subjective feelings or emotional response to the treatment. Such patient-based outcomes may be assessed by measuring patients' satisfaction, health-related quality of life, health awareness, behavior patterns, and belief systems. These assessments can be made using qualitative, quantitative, or mixed-method techniques. Importance of outcomes research is underlined by recent developments—outcomes research has now become an integral part of clinical trials and highly encouraged by the Food and Drug Administration (FDA; Burke, 2001).

Reasoning and Logic in Epidemiology

Scientific systems run on logical reasoning, especially when investigating causation. Epidemiology is rooted in logical reasoning and often uses the language of mathematics to express the inferences. Logical reasoning involves an argument consisting of one or more *premises* (statement that is either true or false that is offered in support of a claim) and one *conclusion* (a sentence that is true or false). The conclusion should follow from the premises based on the claims. A fallacy is an argument where the conclusion does not follow from the premises based on the claims forwarded. In contrast, a factual error involves getting the facts wrong. Errors in logical reasoning leading to fallacious conclusions occur in health research and are a major threat to concluding causal association. Four common errors that we examine briefly are inductive argument, factual error, deductive fallacy, and inductive fallacy.

1. Inductive Argument

Premise 1: Most persons with oral candidiasis are HIV-1 positive.

Premise 2: Mr. AC has oral candidiasis.

Conclusion: Mr. AC is HIV-1 positive.

Note: The conclusion does not follow from the two premises because premise 1 leaves room for some persons with oral candidiasis who are not HIV-1 positive.

2. Factual Error

Example: Candidiasis is caused by *Vibrio cholerae*.

Note: Candidiasis is caused by *Candida spp*.

3. Deductive Fallacy

Premise 1: If the dental pulp is alive and exposed, the tooth may be treated with root canal therapy.

Premise 2: The tooth was treated with root canal therapy.

Conclusion: The tooth was alive and exposed prior to root canal therapy.

Note: Live exposed pulp is one of the possible conditions under which root canal therapy may be performed.

4. **Inductive Fallacy**

Background: Dr. AC practices dentistry in a poor suburb in the United States.

Premise 1: In Dr. AC's practice, 90% of the elderly patients are totally edentulous.

Conclusion: Ninety percent of the U.S. elderly population is totally edentulous.

Note: Only 20.5% of adults aged 65+ years were totally edentulous in 2004 (CDC, 2008).

We classify events in life according to their time of occurrence; that is, in the past or the present, and we try to make allowances for the event happening in the future. Because the future will come only later and a decision is made in the present, we are never sure whether the decision will lead to the event we want to happen. Sometimes we are certain that a set of events will always follow a set of actions, but most often we are not sure. Therefore, our interest is to be as close to certainty as possible about future eventualities. This attempt is embodied in probabilistic thinking and reasoning. We generally express probabilistic reasoning by thinking about our chances (in percentages) for an outcome. We tend to choose the alternatives that have greater chances of being successful. For example, if we believe that pit-fissure sealants have a 90% chance of preventing dental caries compared to a 50% chance of prevention by using regular toothbrushing, we would decide to use pit-fissure sealants. If the above numbers were reversed, our choice would also reverse.

Probability is the positive counterpart of uncertainty. If we are highly uncertain about an event, our confidence about the event is low and vice versa. Therefore, if someone were to tell this author that investing in a particular stock is laden with major risk and it is highly likely that I'd lose my money, my confidence in investing in the stock will be low, and vice versa. However, if I had no knowledge about the stock in question, my confidence would be better than in the earlier situation. Earlier knowledge about an event modifies our thinking and action related to the event. Bayesian statistics incorporates changes due to experiences. Probability of events are easily understood in numerical terms (i.e., comparing 50% and 90% chances of success helps us decide better than comparing ambiguous statements such as moderate/high chance of success). Probability therefore is best estimated using mathematical operations. However, merely expressing probability as a number is not of much use to us. It must help us decide, and so it needs interpretation.

Mathematics is a popular language of science but it requires sound logical interpretation. Statistics is a mathematical science dealing with collec-

tion, analysis, interpretation or explanation, and presentation of data. Actual interpretation and methods of interpretation may vary, but these work very well within their reasoned logical frameworks. Statistical approaches have been viewed to belong to two major camps: the Frequentists and the Bayesians.

Frequentists view data as a collection of random variables that can be conditioned by probability distributions of the data or of functions of the data and are comfortable with considering data that are observed as well as data that are not actually observed. Their view suggests that one hypothesis is true and the rest are false. In contrast, Bayesians condition on the data actually observed and consider the probability distribution on the hypotheses (and not on the data). Therefore, Bayesians allow for choosing between several possible hypotheses. Bayesian statistics are influenced substantially by a-priori (prior) knowledge. They estimate a *prior-probability* for an event and may compare it with *posterior-probability* of the event. If there is no a-priori knowledge of an event on which to base a prior-probability, Bayesians will derive it using a set of assumptions. For example, let us consider the case of the well-known phrase “may he live in interesting times.” Evidence exists that Robert F. Kennedy, during a speech in Cape Town, South Africa, on June 7, 1966, cited this statement as an English version of an old Chinese proverb (JFK Presidential Library & Museum, 2009). It seems that no one has been able to find the “original” Chinese proverb until now. To solve this problem, Frequentists would view the question as: Did the Chinese say this first? Bayesians, in contrast, would frame the question somewhat differently: Who said this first—the Chinese, the Americans, or some others?

Disease Classification Systems

Diseases may be classified in several different ways based on their nature, etiology, progression, or numerical classification systems. Disease classification is done for our ease of organizing information about diseases. Although classification systems may use certain characteristics of diseases, it is not prudent to tie our inferences about a disease to its membership to a certain class in a group. A disease may belong to different groups depending upon the classification system used. For example, as mentioned earlier, HIV/AIDS is an infective disease that was historically not viewed as a chronic disease; but is slowly turning into a chronic disease in contemporary times. Therefore, if one attributes “acute disease” status to an infectious disease such as HIV/AIDS, then it would not reflect the true attribute of the disease as it stands today. Similarly, cancer has generally been thought to be a chronic disease, although research over the past several years has demonstrated several infective causes for many cancers.

Manifestation Criteria vs Etiological Criteria

Diseases may be classified according to their signs and symptoms or how they manifest themselves. Alternatively, they may be classified based on their causes (etiology). For example, ulcerative colitis, dental caries, temporomandibular dysfunction disorders (TMDD), leukoplakia, or vesiculobullous lesions (such as pemphigus and lupus erythematosus) are classified as such because of the way they present themselves (*manifestation criteria*). On the other hand, diseases such as tuberculosis, diphtheria, candidiasis, fluorosis, and berylliosis are classified according to their causative agents (*etiological criteria*). In oral epidemiology, it is important to recognize the difference between these classifications because disease measurement criteria may vary according to the classification/definition criteria used. For example, outcomes in an etiologic agent-based criteria may include demonstration of removal of the etiology (e.g., absence of the organism or appropriate reduction in appropriate antibody titers), whereas use of manifestation criteria may only need to demonstrate clinical remission. Yet, it may be possible to have clinical remission even though the etiologic agent may still be demonstrable and under control. Interpretation of the criteria for success and failure of treatments may differ depending upon the criteria used for defining outcomes of treatments. For example, one often-discussed situation is the measurement of success of root canal therapy. The point of consternation is how to define success—a clinically functional, treated tooth may have a short root canal filling. Therefore, if manifestation criteria are used, the outcome may be defined as a success, but if an etiologic type criteria is used (requiring hermetic seal of the apical one-third—the potential area that may lead to reinfection), then the same outcome may be classified as a failure.

Using etiologic criteria, diseases may be further subclassified as genetic or acquired, microbial (bacterial, fungal, viral, parasitic, prion based), autoimmune, iatrogenic, or diseases of unknown etiology. All these classifications focus on the biologic causation of disease. However, social issues also play a major role in disease occurrences and propagation at individual as well as population level. Therefore, social “causation” is often invoked in public health practice to understand factors that may be modified more easily and have a greater impact in disease mitigation at population level.

For example, smoking and alcohol consumption, individually and together, are important risk factors for oral cancer. Although there are proven biological mechanisms describing causation of oral cancer due to these deleterious habits, if we wish to reduce cancer incidence, the most effective step for oral cancer prevention perhaps lies at the social level and not at the biological level. Establishing programs to dissuade people from smoking and

drinking are perhaps more effective strategies than trying to “immunize” the population using a vaccine (if such an effective vaccine becomes available at a low cost). Therefore, it may be argued that even though the biological causation of oral cancer is linked to exposure to smoking and alcohol consumption, effective prevention lies not at biological level, but at a social level.

Infectious Disease vs Noninfectious Disease

Some of the diseases mentioned previously, such as tuberculosis and diphtheria, may be classified as infectious diseases because they are acquired as an infection, whereas others, such as TMDD and fluorosis, are not infectious diseases. Sometimes, the classification becomes ambiguous—dental caries is generally not thought to be an infectious disease although it is! There is substantial literature showing vertical and horizontal transmission of *streptococcus mutans* causing dental caries as an infectious disease (Caufield, Li, & Dasanayake, 2005).

Chronic Disease vs Acute Disease

A disease is usually classified as chronic if it has a lingering, persistent, and long-lasting course (such as cancer and diabetes mellitus), whereas it is classified as acute if the course of disease is short-lasting (such as influenza, mumps, and periapical abscess). However, several diseases may have a chronic course interrupted by periodic intensive acute phases (*acute exacerbation*). As we will see later, it is important to make these distinctions because the risk of a first occurrence of a disease may be substantially different than that of the risk of a subsequent occurrence. Analytical handling of these completely opposed outcomes needs to be different and requires an astute understanding of disease classification criteria to determine the case definition and outcome selection. A commonly stated result in several studies is “past disease predicts future disease.” Obviously, in this scenario, in order to predict future disease, past disease needs to have occurred first. However, for the first occurrence, because there was no past disease, the prediction criteria would necessarily be different. This fine point is often missed in most studies that seek to look for a prediction model.

Neoplastic (Benign vs Malignant) vs Nonneoplastic

A neoplasm is a new and abnormal growth in any part of the body. If this growth is uncontrolled, it is a *malignant* neoplasm; otherwise the tumor is

benign. The characteristic of a malignant tumor is its predilection to spread. A malignant tumor that spreads to distant parts of body from its main site of origin (primary tumor) is a *metastatic* tumor. Certain lesions that are space-occupying may increase in size, but are not tumors; that is, they are not characterized as new growth (e.g., cysts). In considering a tumor classification, especially when trying to examine its characteristics for making diagnoses, prognostication, or prediction of disease outcome, it is important to be able to correctly assess the nature of the malignant tumor. A common problem in the literature is clubbing all head and neck cancers together and viewing this disparate group of cancers as a single entity with common characteristics. Therefore, comparing the risk factors of oral cancers with oropharyngeal cancers or all head–neck cancers is clearly inappropriate. Furthermore, most oral cancers are squamous cell carcinomas (SCC; over 95%). Therefore, clubbing other histological cancer types along with SCCs should be considered poor case definition. At the same time, a deeper perspective suggests that although SCCs may be viewed as a homogenous group of cancers, their histologic nature and clinical manifestations differ depending upon the histologic differentiation of the cancer cells. Therefore, for certain outcomes, clubbing undifferentiated, moderately-differentiated, and poorly differentiated SCCs together may also give rise to erroneous conclusions. Certain disease entities fall between being nonneoplastic and neoplastic. Although technically these lesions (such as the clinical entity called leukoplakia) are nonneoplastic, their chances of converting to malignancy are substantially greater than several other lesions or normal tissue. Therefore, such lesions are usually classified as *pre-malignant lesions*. Furthermore, there are certain disease conditions, such as lichen planus or oral submucous fibrosis, which are not directly epithelial lesions themselves but create a condition as part of their natural course, so that the associated epithelium acquires greater probability of becoming cancerous. Such *preneoplastic conditions* should be studied as separate entities than precancerous lesions such as leukoplakia.

Oral and Systemic Disease

The link between oral disease and systemic disease has been explored for many years and several such links have been established. For example, links have been found between periodontal disease and cardiovascular disease, cerebrovascular diseases, cancers, renal dysfunction, preeclampsia, pregnancy outcomes, low birth weight of newborn babies, and diabetes mellitus (Beck & Offenbacher, 2005; Joshipura, 2002; Kshirsagar, Offenbacher, Moss, Barros, & Beck, 2007; Lamster, Lalla, Borgnakke, & Taylor, 2008; Meyer, Joshipura, Giovannucci, & Michaud, 2008; Offenbacher et al., 2006; Pitiphat et al., 2008; Ruma et al., 2008; Xiong, Buekens, Fraser, Beck, & Offenbacher, 2006). Most of the evidence of such links has come from cross-sectional

studies, although several cohort studies are being conducted. However, the question arises about causal direction involved in these associations. For example, in assessing the association between periodontal disease and cardiovascular disease, it becomes difficult to establish whether the periodontal disease or the cardiovascular disease occurred first. In the former case, periodontal disease would be viewed as an exposure for cardiovascular outcomes, whereas in the latter scenario, periodontal disease could be an outcome of the cardiovascular disease. Such association studies may become more confusing if there are bidirectional associations such as those described between diabetes mellitus and periodontal diseases (Lamster et al., 2008). It may be possible that diabetes mellitus (through some biological mechanism) may impact periodontal disease occurrence and then periodontal disease in turn impacts occurrence or perpetuation of diabetes mellitus (or impacts its outcomes in different ways). Rhetorical needs may be satisfied by citing the association between oral and systemic diseases, but actual understanding of disease mechanisms and adoption of scientific evidence-based policies and practices for disease prevention and control will need clear elucidation of the causal mechanisms stemming from the directionality of the associations.

ICD-9 vs ICD-10

The International Classification of Diseases (ICD) system had its origin in an internationally applicable, uniform classification of causes of death at the first International Statistical Congress, held in Brussels in 1853. The first iteration of a disease classification that evolved into ICD-9 started as the International Classification of Causes of Sickness and Death in 1909. The ICD system is currently in its tenth iteration (ICD-10) and the next iteration of disease classification, the ICD-11, is planned for 2015 (WHO, 2008b). A brief history of development of the ICD systems can be found at the WHO website.

ICD-10 was endorsed by the 43rd World Health Assembly in May 1990 and came into use in WHO Member States in 1994. The classification is the latest in a series that has its origins in the 1850s. The first edition, known as the International List of Causes of Death, was adopted by the International Statistical Institute in 1893. WHO took over the responsibility for the ICD at its creation in 1948 when the sixth revision, which included causes of morbidity for the first time, was published (WHO, 2008b).

The ICD has become the international standard diagnostic classification for all general epidemiological and many health management purposes. These include the analysis of the general health situation of population groups and monitoring of the incidence and prevalence of diseases and other health problems in relation to other variables such as the characteristics and circumstances of the individuals affected.

The ICD is used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and hospital records. In addition to enabling the storage and retrieval of diagnostic information for clinical and epidemiological purposes, these records also provide the basis for the compilation of national mortality and morbidity statistics by WHO Member States (WHO, 2008b).

Although ICD-10 was established in 1990, its use came about slowly. Even now, use of the previous version, ICD-9, is common. An important reason for slow adoption of ICD-10 was that most diseases had already been classified using ICD-9, and the knowledge explosion in biology and medical sciences predated ICD-10. Therefore, almost all centers across the world had to migrate from ICD-9 to ICD-10, which required changes in database coding and also relearning new codes. Although the use of the ICD is generally claimed to be common, the codes are best suited for computer databases and are not very intuitive in regular clinical situations (the codes have to be memorized or recalled using computer systems). Therefore, in locations where computer systems are not available, or where advanced medical coding systems for diagnosis and billing have not yet reached, ICD codes are generally not used. Many locations may not yet have migrated from ICD-9 to ICD-10 due to a variety of reasons. Therefore, when conducting a study or examination, it is always prudent to inquire about the local coding practice (especially where global health outreach programs and studies are conducted).

In the final count, epidemiology is about identifying, understanding, and correctly addressing sources of variation in information; collection, assessment, analysis, and interpretation of data; and the development and application of solutions to health problems aimed at improving the health of populations.