Oral Health Epidemiology

PRINCIPLES AND PRACTICE

Amit Chattopadhyay

Oral Health Epidemiology:

Principles and Practice

Amit Chattopadhyay

PhD, FFPH-RCPUK, MPH, CPH, MDS, BDS, DipGlobalEthics, GCertID, GCertPHO, DipJourn, DcFM, MSASMS, MACE
Diplomate: American National Board of Public Health Examiners
Diplomate: American Board of Dental Public Health Epidemiologist
Office of Science Policy and Analysis
National Institutes of Health—NIDCR
Bethesda, MD

JONES AND BARTLETT PUBLISHERS
Sudbury, Massachusetts
BOSTON TORONTO LONDON SINGAPORE

World Headquarters
Jones and Bartlett Publishers
40 Tall Pine Drive
Sudbury, MA 01776
978-443-5000
info@jbpub.com
www.jbpub.com

Jones and Bartlett Publishers
Canada
6339 Ormindale Way
Mississauga, Ontario L5V 1J2
Canada

Jones and Bartlett Publishers
International
Barb House, Barb Mews
London W6 7PA
United Kingdom

Jones and Bartlett's books and products are available through most bookstores and online booksellers. To contact Jones and Bartlett Publishers directly, call 800-832-0034, fax 978-443-8000, or visit our website, www.jbpub.com.

Substantial discounts on bulk quantities of Jones and Bartlett's publications are available to corporations, professional associations, and other qualified organizations. For details and specific discount information, contact the special sales department at Jones and Bartlett via the above contact information or send an email to specialsales@jbpub.com.

This publication is designed to provide accurate and authoritative information in regard to the Subject Matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the service of a competent professional person should be sought.

Dr. Chattopadhyay has contributed to this book in a personal capacity. The views expressed in this book are his own and do not necessarily represent the views of the National Institutes of Health or the United States Government.

Production Credits
Publisher: Michael Brown
Editorial Assistant: Catie Heverling
Editorial Assistant: Teresa Reilly
Senior Production Editor: Tracey Chapman
Associate Production Editor: Kate Stein
Senior Marketing Manager: Sophie Fleck
Manufacturing and Inventory Control Supervisor: Amy Bacus
Composition: Publishers' Design and Publishers Services, Inc.
Cover Design: Scott Moden
Cover Image: © Jan Carbol/ Dreamstime.com
Printing and Binding: Malloy, Inc.
Cover Printing: Malloy, Inc.

Library of Congress Cataloging-in-Publication Data
Chattopadhyay, Amit, 1963–
Oral health epidemiology : principles and practice / Amit Chattopadhyay.
p. ; cm.
Includes bibliographical references and index.
ISBN-13: 978-0-7637-5409-9 (hardcover)
ISBN-10: 0-7637-5409-9 (hardcover)
1. Mouth—Diseases—Epidemiology. I. Title.
[DNLM: 1. Mouth Diseases—epidemiology. 2. Epidemiologic Methods. 3. Oral Health. WU 140 C495o 2010]
RC815.C495 2010
617.5'22—dc22

2009029052

6048

Printed in the United States of America
13 12 11 10 09 10 9 8 7 6 5 4 3 2 1

Dedicated to:

The fearless spirit of the inquisitive mind in search of truth to unravel the mysteries of life and the Universe

Contents

Preface

Sound substantive knowledge about questions and methods used to address those questions, tempered by reasoned analysis and considered inferences are the responsibility of epidemiologists as a service to science. Epidemiology draws important understanding from other disciplines to study the distribution and determinants of health-related states and their outcomes in populations. The oral epidemiologists' charge is to adapt epidemiological techniques for answering questions related to health states in the oral and craniofacial region of the body. In the following chapters, we will traverse several exciting and varied terrains on this fascinating journey.

It may be argued that because, by definition, epidemiology is a study of health states in populations, it has little role in basic sciences, which are mostly concerned with narrowly focused laboratory-based projects assessing causal relationships. However, epidemiological concepts have major applications in basic science, general clinical research, and public health as well. The phenomenal growth of molecular applications in everyday health care, genome- and stem cell-based reorientation of medical applications and developments in neurobiology to name a few events, have changed the paradigm of epidemiology from being viewed as being remote to basic science. In current day practice, scientific studies, whether those are laboratory-aided computational biology studies, clinical trials, population-based observational studies, hospital-based clinical studies, or computer-based in-silico simulation studies, all use epidemiological principles.

The range of disciplines that oral epidemiology covers is vast, which makes it a truly inter-/multi-/trans-disciplinary subject. Although several excellent books exist that discuss epidemiologic methods and analysis, none deal with oral health issues to serve as a ready-reckoner and quick general reference for oral health students, practitioners, and researchers to help assimilate and organize scientific information for drawing logical inferences. This book was conceived to fill this lacuna in this growing field. It is expected to stir interest among a variety of persons associated with oral and craniofacial health research who think of epidemiology holistically, as a basic and necessary science of not only public health, but for all clinical and

basic science research. I am hoping that this book will appeal to a wide readership.

This book, in general, refrains from presenting descriptive data about disease burden that are found in several other books and reports dealing with various aspects of oral and craniofacial health. Instead, this book emphasizes the application of epidemiological principles in oral health studies, and aims to encourage the reader to think critically about different aspects of studies that may impact their results and interpretation by approaching application of epidemiological principles in oral and craniofacial health research from a conceptual stand point. Although mathematics is the language of science, there exists a serious risk of getting lost in the myriad formulas and symbolism, especially for those who have left the subject behind and spent a substantial part of their scientific lives in applied biological and clinical fields and in the company of peri-basic mathematical skills only. This need not necessarily be viewed as an unbreakable wall because I believe, as Kurt Gödel said "either mathematics is too big for the human mind or the human mind is more than a machine," leaning towards the second of the two options of Gödel. Therefore, this book has kept mathematical formulas to a minimum, and tried to explain the concepts and implications of those formulae in a way that may be easily decipherable for the non-mathematically oriented intelligent professional. The key to epidemiology is logical and critical thinking—the complex analytical tools come in as important support systems for good epidemiological practice.

During the process of developing this book, I changed jobs to join the NIDCR at the National Institutes of Health in Bethesda, MD, USA. I wish to point out that *I have contributed to this book in a personal capacity. The views expressed in this book are my own and do not necessarily represent the views of the National Institutes of Health or the United States government*. I must acknowledge that what is good in this book is a function of multi-factorial inputs from different persons at different times, and I alone bear the responsibility for errors, omissions, and for what may not be well developed in the book.

An individual usually does not accomplish much in isolation. I would like to extend my gratitude to all those persons whose influences have helped me put life in perspective of reality and kindled the spark of inquisitiveness, opening up the ability to delve into the reality and bask in the wonderful, consistently sustained "high" through science that no recreational drug can provide! Notable among these many guides are Drs. R.K. Bali, J.D. Beck, B. Bera, D.J. Caplan, M.P. Chayya, P.K. Dayal, Mr. R. Dhingra, Drs. B.B. Dutta, J.L. Ebersole, J.E. Eichner, J.S. Kaufman, J.V. Kumar, D.E. Parker, L.L. Patton, R.R. Paul, T.E. Rams, G.D. Slade, D.C. Shugars, D.W. Smith, R.G. Rozier, R.P. Strauss, and C. Poole. I have also learnt much about life, science, and art from long and meaningful discussions with family and friends notably: Subrata Bhattacherjee, Meeta Chatterjee; Oscar Arevalo, Martha Bardour, Steven Browning, Nancy Colvin, Pawan Gulati, Upma Gulati, Younis

Khan, Prashant V.K. Nair, Sarvesh Passi, Jay G. Ray, Amit Roy, Pramod Sinha, and Victor Voss. There are many others who have influenced my thinking in several ways. Some helped run fantasies to the ground which brought forth realities of life very vividly by defining intangibles; whereas others helped keep the dreams going—fuelling new visions and expanding horizons. Among the latter is my mother-in-law, Prity Chatterjee who never fails with her warmth and blessings. And, there has been Mike Brown, keeping a tight control on the book, bearing with the several delays, wielding the stick, and tolerating the reactions—to him my special thanks because without him and his hard working and tolerant team, this book would not have happened.

The obvious, due to its ubiquitous nature, often goes unrecognized—for me, the buried foundation bedrock of outcomes in my life has been the spark of curiosity and self-reliance ignited by my father Salil K. Chatterjee; the constant unfailing support from my mother Meera Chatterjee, and my wife Sharmila Chatterjee, who have borne all my unapologetic whims and oddities unconditionally—to them I remain indebted.

God does not play dice.

—Albert Einstein
(Nobel Laureate: Physics, 1921)

Reason and understanding concern two levels of concept. Dialectics and feelings are involved in reason.

Every error is caused by emotions and education (implicit and explicit); intellect by itself (not disturbed by anything outside) could not err.

Intuition is not proof; it is the opposite of proof. We do not analyze intuition to see a proof but by intuition we see something without a proof.

—Kurt Gödel
(Albert Einstein Award: Theoretical Physics, 1951)

যদি তোর ডাক শুনে কেউ না আসে তবে একলা চলো রে

If they answer not to thy call walk alone,
If they are afraid and cower mutely facing the wall,
O thou of evil luck,
Open thy mind and speak out alone.

—Rabindra Nath Tagore
(Nobel Laureate: Literature, 1913)

SECTION

I

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.

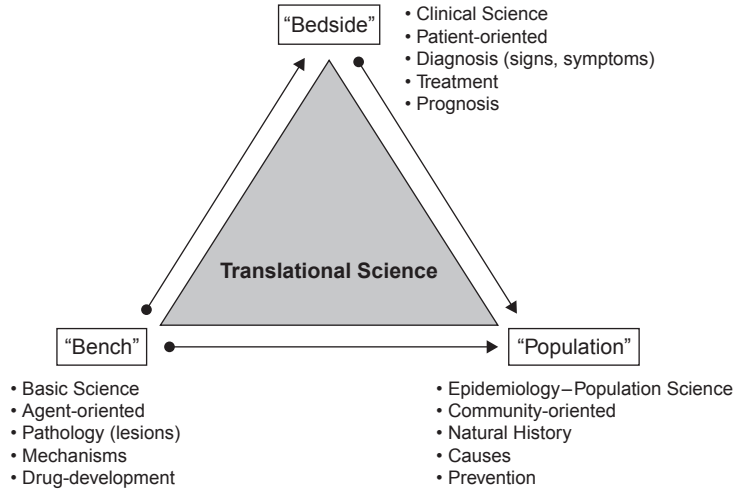


FIGURE 1.1 Triad of Biomedical Science

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 noncausally 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 (*etiologic 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 oropharayngeal 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 *premalignant 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.

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.

3 Associations and Causation

Two factors are said to be associated if a change in one is also manifested as some change in the other. For example, we can say that the education level of an individual is associated with the salary that will be earned in his or her lifetime. In broad terms, the above statement means that the salary of someone with a high school education is generally lower than someone with a bachelor's degree, who in turn earns a lower salary than someone with a graduate degree, and so on. In this example, as the education level increases, the salary increases. Such associations where increase in one factor also implies an increase in the another factor are called positive associations. On the contrary, if salary were to decrease with increasing education level, then that association would be described as negative; that is, the increase in one factor would imply a decrease in the other. This positive and negative quality of the association is called *directionality* of the association (i.e., the direction of association is positive or negative).

The association between two factors also can be described based on their *strength*. Let us assume that for every degree obtained after a high school education, the annual salary of a person increases by $20,000. Based on this assumption, if someone with a high school education earns an annual salary of $30,000, then a person with a bachelor's degree will earn $50,000 and a person with a graduate degree will earn $70,000. On the other hand, we may find that for every 5 years of experience in a job, the salary of a person jumps by $3000. How do we compare the two associations? First, we see that in both cases, the direction of salary and the other factor (i.e., education level or experience) is positive. However, if someone with a high school education works for 10 years, then his or her salary at the end of tenth year is $50,000. If the same person spends the next 10 years in college to complete a graduate degree, the person will then earn a salary of $70,000 in the first year out of college (which is the same amount of time as the 10

years spent by the other person at work). Therefore, the graduate degree holder gets a larger return on investment (investing in education) than the person with a high school education (investing in work experience only). In this example, we can conclude that the association between salaries and education level is stronger than that between salaries and work experience alone, even though in both cases the association is positive. Associations between two factors can thus be described by both their strength and direction—making associations vector quantities rather than scalar quantities (i.e., having direction and strength rather than strength alone). Associations can be measured in different ways as discussed in Chapter 4.

Whenever we study associations, we are also interested in the change in one factor that is *caused* by the change in the other factor that is associated with it. Therefore, in the situation discussed above, the question is whether the salary increase was *caused* by obtaining a higher degree or by working for a greater number of years.

What Does an Association Mean?

Mostly, when a study finds an association, investigators generally start thinking of a causal link to explain the association. However, the truth is that associations are many; not all associations are causal associations. It is generally prudent to first think of demonstrated associations as spurious rather than real. Therefore, explanations such as random error, bias, and confounding should be sought first, and only after ascertaining that the observed association is not explained by these distractions should causality be considered, unless there is a lot of strong support already existing that points toward a causal association. If much supporting evidence exists for causal associations, then perhaps studies should be designed specifically to test for those hypothesized causal associations.

Precision and Accuracy

Once an association is established, its quality has to be assessed. This may be done by assessing precision and accuracy of the association. *Precision* is the degree to which a variable has nearly the same value when measured several times. Normally the estimates provided by different studies vary from each other. If the range of the estimates from different studies is close, the estimates are generally considered to be precise and vice versa. For example, to correctly measure the decayed, missing, or filled teeth (DMFT) of a person, investigators may decide to use several examiners and take the average of their individual scores to be the correct DMFT for the person. Let us consider that one group of five examiners reported the scores as 2, 4, 1, 1, 4 (average = 2.4). A second group of examiners found the following scores for the same person immediately after the first group of examiners had completed their examination: 1, 2, 1, 1, 1 (average = 1.2). We can see that the first group of examiners

reported a wider range of scores (1–4 for the first group compared to 1–2 for the second group). Even though the averages were different in the two studies, the second group of examiners were more precise because they were closer to their average score. The second group of examiners estimated the DMFT score more precisely. A third group of examiners gave the following result: 2, 2, 2, 2, 2; they were more precise than the first two groups because there was absolutely no variation in their reported scores (average = 2.0).

Accuracy, on the other hand, assesses the degree to which a variable actually represents what it is supposed to represent. Suppose we now reveal that the correct DMFT for the person examined above was 2.0. Then we immediately notice that the average of the first group of examiners was closer to the true value than the second group. Both the first and second groups of examiners, on average, did not get the correct score, unlike the third group. Therefore, the third group of examiners was accurate while the first two were not. Overall, we can say that the first group was neither precise nor accurate, the second group was more precise than the first, but was also inaccurate. The third group was precise and accurate. Similarly, another combination of imprecise, but accurate data can be derived (e.g., 4, 0, 1, 3, 2; average = 2.0). Figure 3.1 demonstrates the concept of precision and accuracy.

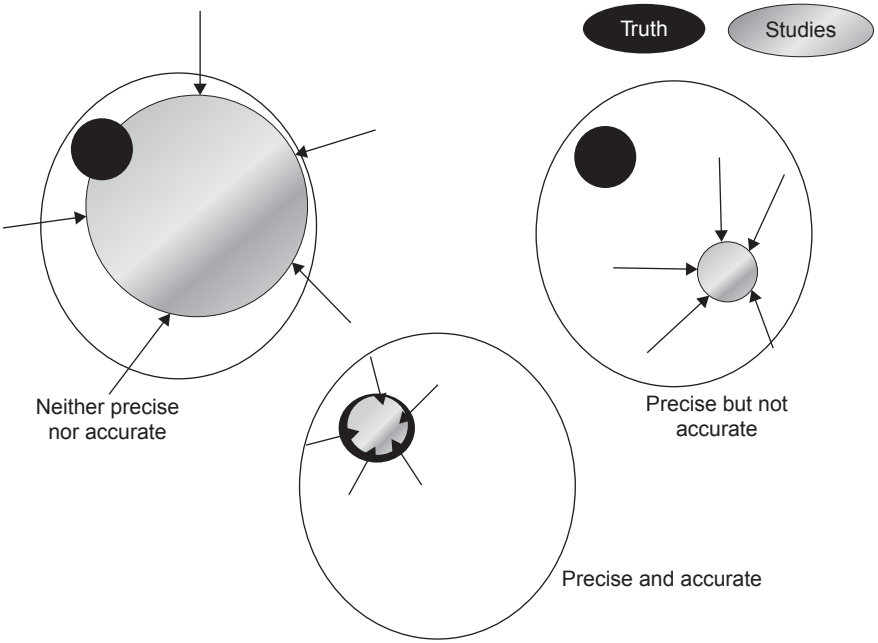


FIGURE 3.1 Precision and Accuracy Different studies (grey shaded circles) try to capture the position and width of the truth (dark circle). Wider circles are imprecise compared to narrower circles.

Precision is threatened by within-observer variability, between-observer variability, within-instrument variability, and between-instrument variability. The reproducibility of continuous variables can be expressed as the within-subject standard deviation. If the "Bland–Altman" plot of the within-subject standard deviation versus the mean shows a linear relationship, then it is better to use the coefficient of variation to assess the variability. A Bland–Altman plot (see Figure 3.2) plots data assessing the agreement between two variables. The *y*-axis is assigned the difference between two variables and the *x*-axis is assigned the mean of two variables. If the plotted data points are spread around, the correlation between the two variables would not be high, whereas a linear distribution would indicate a good correlation between the two. For categorical variables, percent agreements are inferior to kappa scores because observers may show 100% agreement by chance. Kappa statistics adjust for the by-chance agreement. Therefore, good kappa scores discount the occurrence of agreement by chance. Precision enhancement can be done by standardizing measurement, training and certifying observers, refining instruments, automating instruments, and using repetition and quality control in measurements (especially important in biological assays).

Accuracy is threatened by observer bias, subject bias, and instrument bias. The accuracy of a measurement is best assessed by comparing it to a "gold standard." For continuous measures, mean difference between gold standard and new measurement can be examined. For categorical variables, accuracy compared to gold standard can be measured by sensitivity and specificity. Accuracy may be enhanced by standardizing measurement methods, training and certifying observers, refining instruments, automating instruments, making unobtrusive measurements, blinding, and calibrating the measurement instrument. The decision of the extent to which to pursue these categories depends upon importance of the variable, potential impact of the inaccuracy, feasibility of enhancement measures, and costs.

Nature of Cause

The basic question about understanding associations is: What is a cause? At a fundamental level, we can define a cause as: (1) an event or events that provide the generative force that is the origin of something or (2) a series of actions advancing a principle or tending toward a particular end. From an epidemiological standpoint, overall, a cause is an antecedent event, condition, or characteristic that was necessary for the occurrence of the disease at the moment it occurred, given that other conditions are fixed. If factor A causes factor B, then we call factor A the causative factor and factor B the outcome. Primarily the cause or the causative factor must *precede* the outcome in time; that is, in a time-based sequence of occurrences the causative factors come first and the outcome follows. A reverse sequence of events

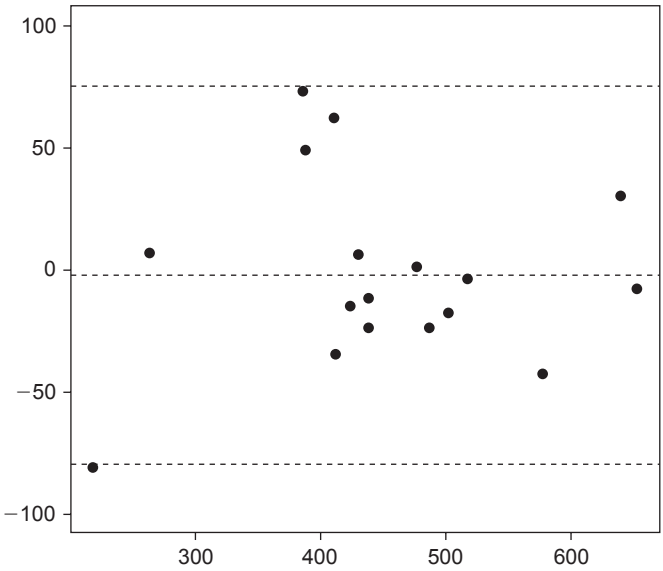


FIGURE 3.2 Bland–Altman Plot The spread of data points indicates correlations between the variables.

cannot occur in time. Although the treatment of time as related to causation is a subject of much study and philosophical interpretation (including manipulation in sci-fi literature), we interpret time as that which is measured by the clock. We consider it to be one-dimensional with a fixed forward flow like an arrow from the past to the future—the most common interpretation of time—and in this text we assume the measurement of time units in seconds, minutes, hours, days, and so on. There are several very fascinating and intellectually stimulating interpretations of time (for example, block time, which is static), but we will not delve into those issues in this text.

Any scientific research or epidemiological study must deal with the entity of time. In epidemiology, we generally either ignore time altogether (sometimes knowingly and sometimes unknowingly) or try to deal with time as a factor in our analyses. It is the latter situation that is intellectually most stimulating, but also conceptually and analytically most challenging; it leads to answers that are closer to the truth. Even if we choose to ignore it, time is ingrained in the physical reality—it is up to us to incorporate its reality in our scientific quest. At this point, it may be prudent to remind ourselves that whether we address a social, biological, economic, or other situation, we are always dealing with science and its central dogma. However, the scientific method is rigorous and requires demonstrable evidence that should be reproducible. Sometimes, science is viewed as opposed to religion. Although these views of the universe differ, it is science that requires

demonstrable evidence, whereas religion and prejudice require only belief (making manipulation easy based on individual personal experiences and views). There is no inherent intractable conflict between the two views based on causal issues as long as the origin of the universe is left outside the purview of investigation (i.e., which in today's time means the nature of the big bang theory). Delving into this issue would take us far from our primary goal of discussing the epidemiologic method in oral diseases! So, we now turn back to our primary focus.

Types of Causes

The general paradigm that has been dominant in oral health research is the "one disease, one cause" idea. Under this construct, there can be only one cause for a disease to occur. Whereas this holds well for several diseases; for example, HIV/AIDS (mostly caused by HIV-1 virus) and Down's syndrome (Trisomy 21), the construct fails to work for opportunistic infections where a prior immune compromise (itself a disease) is required (e.g., oral candidiasis). Similarly, for example, adult periodontitis may be caused by several organisms such as *P. gingivalis*, *T. denticola*, or *B. forsythus*, individually or in conjunction with other organisms. Furthermore, the same disease may be caused by different organisms in different circumstances; for example, dental caries could be caused by *S. mutans*, *S. mitis*, or *L. acidophilus*. Therefore, the exclusive one organism (or one cause) and one disease concept does not necessarily work. For dental caries to occur, a susceptible tooth, the acid-producing organism, and a local environment containing fermentable carbohydrates and plaque must exist. In this context, each of these factors—organism, plaque, and carbohydrate—can be considered as causes of dental caries. Although a paradigm shift from one disease, one cause to one disease, multiple possible causes has been acknowledged in principle (and more comprehensively for some disease states), it is not clearly visible in most oral epidemiological studies.

Rothman and Greenland (2005) have developed the idea of "multiple causation" and used the metaphor of a "causal pie" to explain the concept. As seen in Figure 3.3, disease occurs only if a circle is completed by one or more component parts (e.g., a "pie" from a pie diagram); each of these "parts" or "pies" are cause for the disease. This concept indicates that it may be possible for a disease to occur by a combination of different factors in different conditions. Although there might be several diseases that must always require the same set of conditions (same component pies together to complete the circle), most disease can complete a circle using different pies—some of which may be present always, and others which may be replaceable by some other pies or combination of pies. In the concept of multiple causation of disease, a *necessary cause* is a causal factor whose presence is required for the occurrence of the effect. Such a causal "pie" must be present

in all completed circles for disease. If this necessary cause is missing, circles cannot be completed and disease cannot occur.

A *sufficient cause* is a complete causal mechanism that can be defined as a set of minimal conditions and events that inevitably produce disease. There may be several different independent sets of sufficient causes for a certain disease outcome. The term *minimal* implies that all the conditions or events are necessary. Thus, if a cause must exist for a disease to occur and the disease cannot occur in absence of this cause, then it is a necessary cause. Some causes may be necessary but not sufficient, whereas some may be sufficient but not necessary (Rothman & Greenland, 2005). Each cause that contributes a pie to the circle under the sufficient cause model is a *component cause*. Therefore, different component causes may contribute to disease causation, and it may be possible that for some diseases, there are no necessary causes even though several different sets of sufficient causes may exist. Some causes may be sufficient and not necessary, whereas some causes may be necessary, yet not sufficient. For example, although *Mycobacterium tuberculosis* is a necessary cause for tuberculosis, it is not sufficient because mere exposure to the organism will not produce disease; a component cause of compromised immunity is required. Together, these may be sufficient to cause the disease.

Because more than one cause may be in action to produce a disease, it is important to consider causal coaction of joint causes. The joint action of two component causes does not have to occur simultaneously: one component cause could act many years before the other, but it would have to leave some effect that interacts with the later-acting component cause. In order to create an intervention to change disease outcomes, we need to assess those

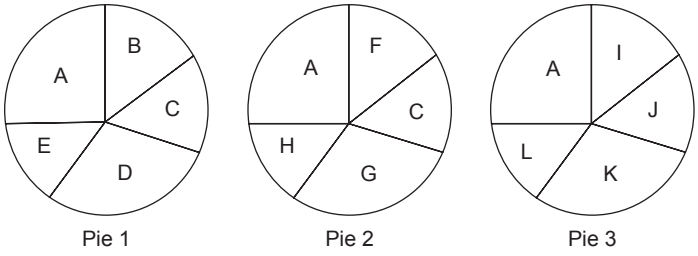


FIGURE 3.3 Multiple Causation and Causal Interaction: Sufficient Cause Model There are only three ways in which this outcome may occur, as represented by the three pies. Factor A is a necessary cause because it must be present in all the three situations. However, all other component causes (i.e., B, C, D, E, F, G, H, I, J, K, L) are not necessary causes, but may combine with Factor A to complete the pie. Each pie represents a combination of component causes grouping together as sufficient causes for the outcome.

causes that may be amenable to manipulation and change—such causes are *modifiable causes*. Even if we understand all causes and mechanisms of a disease very well, if no modifiable causes exist, disease prevention will not be possible.

Causal Inference

It is common in scientific oral health literature to see most associations be interpreted as causal associations. Even if no such direct claims are made, several studies interpret the associations as causal in the discussion of the articles. However, inferring causal association is a tedious and rigorous proposition and requires quality insight, sound judgment, and careful assessment of evidence. In making such inferences, design of the study must be considered as the limiting framework within which conclusions may be drawn. Errors of study design cannot generally be "fixed" posto facto. Similarly, the statistical analytical design is another factor that must be considered before making firm conclusions. Different types of analytical strategies and their limitations and appropriate use are key factors in correct inference making. After considering these limitations, the validity of the association should be considered; that is, whether the observed association is true or spurious. Several observed associations can be attributed to faulty study and analytical designs, chance, systematic error (bias), and confounding. Only if the association is considered to be true should the issue of potential causal association be explored.

How do we infer causality from studies? The U.S. Surgeon General's article on the association of smoking and health published in 1964 outlined five criteria for inferring causal association (Surgeon General, 1964). Adding four more criteria to these, in 1965, a nine-criteria general guide to making causal inference in epidemiology was provided by Sir Austin Bradford Hill, Professor Emeritus of Medical Statistics at the University of London, forming the cornerstone of epidemiological causal inference (Hill, 1965).

> I have no wish, nor the skill, to embark upon philosophical discussion of the meaning of 'causation.' The 'cause' of illness may be immediate and direct; it may be remote and indirect underlying the observed association. But with the aims of occupational and almost synonymous preventive medicine in mind, the decisive question is where the frequency of the undesirable event B will be influenced by a change in the environmental feature A. How such a change exerts that influence may call for a great deal of research. However, before deducing 'causation' and taking action we shall not invariably have to sit around awaiting the results of the research. The whole chain may have to be unraveled or a few links may suffice. It will depend upon circumstances . . . an association between two variables, perfectly clear-cut and beyond what we would care to attribute to the play of chance. What aspects of that association should we especially consider be-

> fore deciding that the most likely interpretation of it is causation? (Hill, 1965)

The nine criteria of Sir Bradford Hill (1965) mentioned below (in order) are adapted from his original article to dispel any ambiguity in the implied meanings.

1. **Strength:** The stronger the association, the stronger is the argument for *potential causal association*.
2. **Consistency:** Has the association been *repeatedly* observed by different persons, in different places, circumstances, and times?
3. **Specificity:** *If* the association is limited to specific exposure and to particular outcome, and there is no association between exposure and other diseases, then it is a strong argument in favor of causation.
4. **Temporality:** The cause must *precede* the outcome.
5. **Biological gradient:** If the association is one which can reveal a biological gradient, or dose–response curve, then *we should look most carefully* for such evidence. A stronger dose–response relationship strengthens the argument for causal association.
6. **Plausibility:** The association should be understandable in *plausible* terms. "It will be helpful if the causation we suspect is biologically plausible. But this is a feature I am convinced *we cannot demand*. What is biologically plausible depends upon the biological knowledge of the day."
7. **Coherence:** "The cause-and-effect interpretation of our data should not *seriously* conflict with the generally known facts of the natural history and biology of the disease."
8. **Experiment:** "Occasionally it is possible to appeal to experimental, or semi-experimental, evidence. . . . Here the strongest support for the causation hypothesis may be revealed."
9. **Analogy:** "*In some circumstances* it would be fair to judge by analogy. With the effects of thalidomide and rubella before us we would surely be ready to accept slighter but similar evidence with another drug or another viral disease in pregnancy." Although this criteria is generally considered to be weak, and sometimes superfluous, we need to recognize that phase IV clinical trials and post-marketing surveillance are important policy applications of analogy—just because some drugs have caused adverse events, it is assumed that others may also cause similar effects. Therefore, in contrast to anecdotal impressions of Hill's criteria, "analogy" is perhaps the most stringently applied.

The general idea of these guidelines for inferring causal association between two factors is that *causal inference is a process*, and not a binary yes/no phenomenon derived from one study. Hill (1965), in his original paper, went on to unambiguously impress this point and also qualified the use of statis-

tical significance for inferring causal associations. Several important sources of errors need to be addressed in inferring causal associations, such as overemphasizing statistical significance and overlooking biases. Emphasis on using Hill's criteria as a checklist to infer causal association is, although misplaced, a common practice.

> Here then are nine different viewpoints from all of which we should study association *before* we cry causation. I do not believe—and this has been suggested—that we can usefully lay down some hard and fast rules of evidence that must be obeyed before we can accept cause and effect. *None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non.* What they can do, with greater or less strength, is to help us to make up our minds on the fundamental questions: Is there any other way of explaining the set of facts before us? Is there any other answer equally or more likely than cause and effect? (Hill, 1965)

Tests of Significance

No formal tests of significance can answer those two fundamental questions. Such tests can, and should, remind us of the effects that the play of chance can create, and they will instruct us in the likely magnitude of those effects. Beyond that, they contribute nothing to the "proof" of our hypothesis (Hill, 1965).

Overall, Hill's criteria can be used in several ways, such as making distinctions between association and causation in epidemiologic research, critically assessing of epidemiologic studies, designing epidemiologic studies and setting up appropriate analytical schema, and appropriately interpretating of study results.

Causal Paradigms

Causes discussed thus far follow a "biomedical" causal paradigm and are fairly familiar to the clinician working in the clinic who interacts with patients on a one-to-one basis. However, once we consider disease prevention or occurrence in populations, there are several factors that are contributory to the occurrence of a disease, and some of these are potentially more easily, effectively, and efficiently amenable to modification for disease prevention or control in a population. Several of these *leverage points* for developing effective preventive methods lie outside the biomedical disease process framework. A good example is use of public water fluoridation to prevent dental caries, where the prevention control mechanism is essentially nonbiological, and its implementation lies in the sociopolitical arena. Similarly, causal association of smoking with oral cancer is well known, and the effec-

tiveness of smoking prevention is much more productive when implemented as a policy rather than individualized behavior modification.

Recognizing this link to factors outside the biomedical framework, a concept of "social causation" of disease has developed and is currently gaining ground. Examples of social causation of disease could be the prevalence of gunshot injuries, motor vehicle accidents leading to facial and dental trauma, higher dental caries prevalence in lower socioeconomic sections of society, and several more. However, it serves the population well to recognize that although causal paradigms may categorize chief causes within the biomedical or social frameworks, diseases occur independent of categories imposed by us—causes in both frameworks interact extensively. From a biomedical standpoint, social causation may appear as a set of facilitating factors that lie outside and are not modifiable in a clinical set-up, whereas from a social causation standpoint, biomedical causation may appear as not modifiable from a primary prevention standpoint.

Rigid adherence to a causal paradigm may sometimes lead to prevention and treatment-planning dilemmas, ineffective treatment outcomes, poor policies, and policy resistance. For example, a tooth that would have been otherwise restorable with root canal therapy followed by crowning, may have to be extracted (with or without replacement) if the patient is poor and cannot miss work due to an associated loss of wages. In such situations, the clinician, instead of following an ethical paradigm of "beneficence," is forced to follow the principle of "harm reduction." In this situation, subscribers to the biomedical causal paradigm would argue that dental caries were caused by well-described etiological factors, whereas those subscribing to the social causation paradigm would argue that poverty was the main cause because it disabled access to proper preventive and curative care, which was otherwise available. A counterargument would state that there exist hypersusceptible persons who would get dental caries despite access to and actual utilization of standard preventive care. To develop a workable prevention and treatment policy, it is necessary to address causal issues that are inclusive of all paradigms.

The Counterfactual Concept

In quantum mechanics, counterfactual definiteness is the ability to speak meaningfully about the definiteness of the results of measurements, even if those measurements were not actually performed. Epidemiology, however, does not insist on the definiteness of the counterfactual experiences. An ideal study would use counterfactual populations for comparison to derive correct population estimates. In order to attribute a disease to an exposure, ideally, the investigator should compare the outcomes in a group that is exposed to an *identical group* that is unexposed. If the only difference between these groups is the exposure, then the outcome definitely can be attributed

to the exposure. However, this would imply that every individual in the exposed group must also be present in the unexposed group at the same time, because only then can all attributes be exactly common between the two groups except the exposure under study. However, such exactly matching unexposed groups cannot be found, and are therefore called *counterfactuals* to the exposed group. Not only would the counterfactual control group arise from the same population as the cases, but they would be the same persons as the cases themselves without the exposure at the same time while being exposed! Ethical issues aside, if two identical human clones could be produced and subjected to randomized cohort studies at the same time under same environment, perhaps a counterfactual ideal study could be factually performed. In absence of that eventuality, crossover studies and studies with split-plot designs use the same person as the case and the control, and this comes closest to the counterfactual ideal comparison group. But the timing of exposure and control phases of the same person is different in a crossover study, and the "site-wise" status of exposure of the same person's case and control sections at the same time may vary in split-plot designs. Unless identical animal clones are used in studies, the goal of counterfactual ideal may be considered unachievable. However, growing use of experimental chimera models may, in due course, extend the use of counterfactual ideal comparison models in genetic epidemiology sooner than human counterfactuals.

Although we can never achieve the counterfactual ideal, we strive to come as close as possible to the ideal in the design of epidemiologic studies. Instead of comparing experience of an exposed group with its counterfactual ideal, we must compare that experience with that of a real unexposed population. The goal is to find an unexposed population that would give a result that is close, if not identical, to that from a counterfactual comparison. To achieve a valid substitution for the counterfactual experience, we resort to various design methods that promote comparability (Rothman, 2002).

Demonstrating causation requires careful assessment of study characteristics, especially selection of comparison groups and understanding the source population in light of counterfactual ideas. However, there is a tendency in research to focus on Hill's (1965) criteria as a checklist (which he himself refused to do). Counterfactuals, like Hill's criteria, promote scientific thinking and help gain insight into potential causal mechanisms that studies try to decipher.

Why do health researchers, seemingly much more than those in other fields, cling to rules for assessing causation to the point where we have several such lists, as well as a secondary literature, that try to assess and improve the rules?

Probably more importantly, the desire to find answers to countless different policy, social science, and biological questions creates the desire to study something once (in a particular population, at a particular time, with

particular variable definitions), declare an answer, and move forward. This does not provide much opportunity to actually test hypotheses. It encourages health researchers to conduct simplistic statistical calculations that are described in the language of hypothesis testing and mistake this for actually testing a worldly hypothesis. It discourages genuine hypothesis testing, along the lines of, "If we have observed a true causal relationship, then we would also expect to see. . . . Let's do more research to check that before reporting our result." We would certainly expect such testing from another science before it declared, say, the discovery of cold fusion or that unfettered free markets make people's lives better (poor examples, perhaps—let's call them exceptions that emphasize the value of the rule; Phillips & Goodman, 2006).

The Duhem–Quine thesis (or the Duhem–Quine problem) states that it is impossible to test a scientific hypothesis in isolation of its environment because background assumptions about associated issues form the basis of empirical tests (Gillies, 1998). Therefore, when a study is conducted under assumptions, the study results are also a statement on those assumptions themselves because all assumptions are defacto considered to be valid *prior* to the study. Therefore, the mere demonstration of an observed association cannot be taken as evidence for causality at face value. In oral health research, few studies are ever repeated under similar or varying circumstances (with or without correcting for the errors in an already published study), and often either publication bias prevents reporting of confirming results, study methodology is found wanting, or investigator bias steers one away from conducting a repeat of a published study in the interest of breaking "new ground."

Appropriate comparability of the exposed and unexposed group is the key determinant in etiologic studies. Therefore, selection of the control group is an extremely important issue. Strategies for obtaining such comparability include randomization, matching, restriction, and selecting controls with similar risk factors or community profiles and similar geographical/residential locations. Causal paradigms and criteria should be used as guideposts to better understand the associations observed in a study. Toward this goal, it is perhaps helpful if a causal pathway can be visualized. Several ways to draw visual graphs exist and are used in causal analysis. Application of such methods is discussed in Chapter 6.

4 Measuring Associations

Epidemiologic studies measure associations between exposures and disease in several ways. Table 4.1 describes measures of disease frequency and associations commonly used in epidemiology. The key definitions to keep in mind are: (1) prevalence measures all persons having an outcome as a proportion of all persons who *are at risk* of developing the outcome, and (2) incidence measures persons developing *new outcomes* as a proportion of all persons at risk for developing the outcome. The denominator data for incidence may or may not include an element of time, and is therefore named differently according to the construct it represents (see Table 4.1). Those who are not at risk for the outcome should not be included in the denominator. For example, when measuring prevalence of dental caries using decayed, missing, or filled teeth/surfaces (DMFT/S), all persons in the population/sample should be included even if they are completely edentulous. This is so because the "M" component of the index is represented among completely edentulous persons, as many of their teeth might have been extracted due to caries. On the contrary, if prevalence of active carious lesions is being assessed, then the edentulous persons are excluded from the denominator because by definition, they are not at risk for active carious lesion. Similarly, when incidence of caries is being measured, completely edentulous persons should not be included in the denominator because they are not at risk for *new* carious lesions.

Measures of disease frequency usually treat time as a nuisance factor and make assumptions about time in ways as if time is of no consequence to the disease process or its burden. When measuring prevalence of disease, point prevalence studies are conducted in very short time spans, hence the name—although it is impossible to conduct a study in an instantaneous point in time. These studies *assume* that they were actually conducted in such a short time frame. Period prevalence studies last longer, usually for a year or more. Depending upon disease duration, it may be possible that a

TABLE 4.1 Measures of Association of Exposures and Outcomes

Measure of:	*Measure Type*	*Measure Identity*	*Measure Calculation*	*Notes/Synonym(s)*
Frequency	Prevalence	Point prevalence	Number of cases/population at risk at a point in time	
		Period prevalence	Number of cases/ population at risk over a period of time	
	Incidence	Cumulative incidence	Number of new cases/population at risk	Incidence proportion, risk
		Incidence density	Number of new cases/(population at risk × time period under observation)	Incidence rate, hazard rate, person–time incidence
Probability	Prevalence	Prevalence odds	Prevalence %/(100–prevalence %)	
	Incidence	Incidence odds	Incidence %/(100–incidence %)	
Association	Absolute difference	Risk difference	Incidence (exp) – incidence (unexp)	Excess risk, absolute risk difference, attributable risk
	Ratio of risks (relative difference)	Risk ratio	Incidence (exp)/incidence (unexp)	Cumulative incidence ratio, relative risk
		Rate ratio	Incidence density (exp)/ incidence density (unexp)	Incidence density ratio
		Odds ratio	Odds (exp)/odds (unexp)	Relative odds
		Prevalence ratio	Prevalence (exp)/prevalence (unexp)	Prevalence odds ratio

(Continues)

TABLE 4.1 Measures of Association of Exposures and Outcomes *(Continued)*

Measure of:	*Measure Type*	*Measure Identity*	*Measure Calculation*	*Notes/Synonym(s)*
Impact	At individual level	Attributable risk	Incidence (exp) – incidence (unexp)	Excess risk
		Attributable risk percent	(Attributable risk/incidence [unexp]) × 100	Etiologic fraction among exposed, relative risk reduction, attributable fraction among exposed
	At population level	Population attributable risk	Incidence (total population)/ incidence (unexp)	
		Population attributable risk percent	(Population attributable risk × 100)/ incidence [total population]	Population attributable fraction
Benefit/ Harm	At study sample level	NNT: Number needed to treat	1/risk difference	Number of people who need to be treated to prevent one additional bad outcome
		NNH: Number needed to harm	1/risk difference	Number of people who need to be exposed to cause harm to one person who would otherwise not have been harmed

Note: exp = exposed; unexp = unexposed

disease measured in the prevalence pool in January was cured by March and the person was disease free until December, when others with prevalent disease were still being counted in the same study—furthermore, these latter persons may have been disease free in January. Period prevalence studies have an inbuilt assumption that they consider all diseases to have occurred at one point in time—the midpoint of the study period. Essentially all prevalence measures are period prevalence, although we choose to classify them as point and period prevalence based on arbitrary duration of time. Prevalence studies do not make etiological inferences and generally assess disease burden over time, so perhaps it does not matter much if these studies characterize time as a nuisance factor.

Studies examining disease incidence, however, have to account for time if etiological inferences are to be made or inferences about survival time, incubation period, time to recurrence, and similar events have to be made. Cumulative incidence looks only at proportions and ignores time, and is therefore not very useful in study situations mentioned above. However, cumulative incidence can answer questions about how many new cases are being generated over a period of time (study period), provide an assessment of the generative force of disease in populations, and has value in policy making. Time to event-type questions are answered by incidence density measures that include time contributed by every participant in the study, and try to incorporate a time factor in the study. By definition, *rate* is a change in the magnitude of a parameter per unit of time. Therefore, only incidence density can be correctly classified as a rate. It is common to find incidence proportion being referred to as incidence rate, which is an incorrect designation, just as is prevalence rate. As common examples, mortality rate, case-fatality rate, and birth rate come to mind as frequent misuses of the term *rate* because most of these terms indicate proportions. Interestingly, if any of these "rates," such as mortality, is assessed over a 1-year period, then one may argue that there would not be any numerical difference because the denominator would be multiplied by 1, and the term "rate" is justified. However, people die at different times and do not necessarily contribute a full year of observation period to the study before dying. Therefore, to be correctly measured, only the amount of time contributed into a study before the outcome (dying in this example) should be counted, which does not happen in cumulative incidence, in which all participants are assumed to have contributed time equivalent to the full study period even if they do not do so.

If an investigator was trying to assess the rate of occurrence or progression of disease following certain exposures, then careful time keeping is important and a person–time denominator is a must, and only under such situations can the term *rate* be truly applied. This is important in etiological research, and is the reason for the often-mentioned statement that incidence studies contribute to etiological research.

Measures of Association

Assessing relationships between two (or more) variables is a key feature of epidemiological studies. Association between interval (and ratio) scale variables are often assessed using their correlation coefficient, which as a concept is close to regression. Categorical variables are very common in epidemiology, and they are presented as proportions of participants falling in categories leading to a nonparametric correlation coefficient. If prevalence of disease in two groups is measured, then the ratio of prevalence between the groups is called *prevalence ratio*. Prevalence ratios are often used to describe the characteristics of disease in a population. In epidemiology, the term *risk* is usually associated with the measure that designates the proportion of persons actually having a disease out of the persons who may develop the disease (i.e., the population at risk of developing the disease). By itself, it is a measure of disease frequency. In a study, risk is measured as the disease frequency in the group arising over *a specified time period* divided by the population at risk over *the same time period*. In this operation, the numerator and the denominator contain the same time period, which gets cancelled by each other (magnitude as well as the units), and we are left only with the proportion of numbers of persons. However, if there are putative exposure factors, then the role of such factors can be assessed by comparing the risk of an outcome among those who are exposed and those who are not exposed. Such comparison can be done in two ways: (1) the arithmetic difference of the two risks can be calculated, and (2) the ratio of the two risks can be calculated. It is convenient to use notations to designate these risks.

Terminology for measuring association may get confusing. Currently, similar measures are known by several synonyms in epidemiology—although as the field develops, terminology is becoming standardized (see Table 4.1). Keeping in tune with this shift toward use of standard terminology and notation, we designate the risk among unexposed as R_0 and the risk among exposed as R_1. Therefore, the difference of risk between exposed and unexposed is called *risk difference* (RD) = $R_1 - R_0$; and the ratio of two risks is called *risk ratio* (RR) = R_1/R_0. Because both measurements *compare* the risks in two groups "relative" to each other, RD and RR are also sometimes referred to as *absolute difference* and *relative difference*, respectively (if abbreviations were also used for these terms, it would further compound the confusion!).

When we multiply (or take ratios of) two numbers, then the arithmetic operation is carried out on a multiplicative scale; when we take the difference of two numbers, the operation occurs on an additive scale. Therefore, RR is a measure on a *multiplicative scale* and RD is a measure on an *additive scale*. Although this issue may appear superfluous here, we revisit this concept when discussing effect measure modification in Chapter 6.

Measuring incidence rate/density requires that the investigator keep careful track of the time that each participant contributes to the study and include that time period in the denominator. Figure 4.1 demonstrates calculation of person–time contributed in a study by summing up the total time that individual participants contribute in the study, and how the magnitudes of incidence proportion and incidence density are substantially different. As above, if the investigator describes the ratio of incidence density/rate between exposed and unexposed groups, the measure is then called *incidence rate/incidence density ratio*; if the difference of the ratios is presented, it is then called *incidence rate/incidence density difference*.

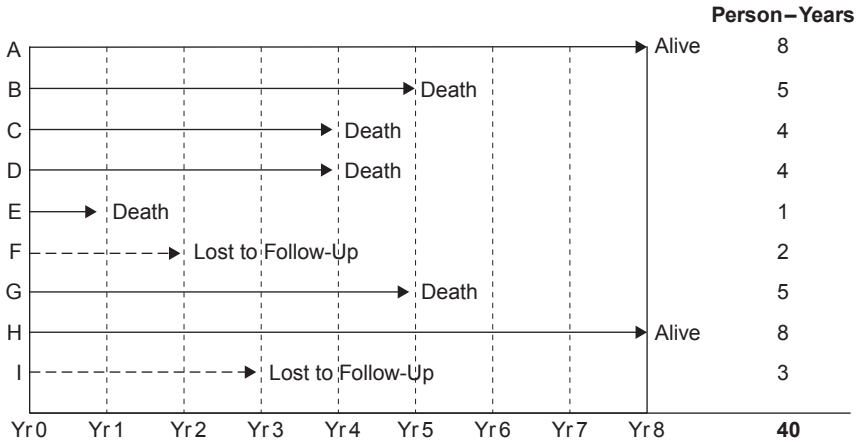


FIGURE 4.1 Calculation of Person–Time This hypothetical study follows up 9 participants (A–I) with oral cancer for 8 years to assess the outcome of chemoradiation therapy. However, different participants contribute different times. Participants A and H are still alive when the study ends after 8 years, so each contributes 8 years in the study. Participants B, C, D, E, and G die after 5, 4, 4, 1, and 5 years from the start of the study and they contribute 5, 4, 4, 1, and 5 years in the study respectively. Participants F and I were lost to follow-up after 2 and 3 years of follow-up respectively, so they contribute those number of years in the study. The total person–time contributed in the study = 8 + 5 + 4 + 4 + 1 + 2 + 5 + 8 + 3 = 40 person–years. Therefore, cumulative incidence of death after chemoradiation therapy in this study = 5/9 = 0.5555 or 55,555.56 per 100,000 persons. Incidence density of death after chemoradiation therapy = 5/40 = 0.125 or 12,500 per 100,000 person–years. The two numbers are very different from each other in magnitude and interpretation.

Correlation

As a concept, correlation indicates how independent (or correlated) two variables are from each other—if the two variables are close to each other then they are less independent, whereas if they are far from each other, they are more independent. Therefore, *correlation measures the departure from independence* between two variables. Measurement of correlation involves two dimensions: *strength* and *direction* of the correlation. Although several types of correlations can be calculated, the two most commonly used in epidemiology are Pearson's product-moment correlation and Spearman's correlation.

Pearson's correlation is a parametric statistic that is derived by dividing the covariance of two variables by the product of their standard deviations. This calculated entity is also known as the "sample correlation coefficient" and is usually denoted as r_{xy} where x and y are the two variables whose correlation is being obtained. It is applicable for interval or ratio scale data and can manifest in two directions: positive and negative. Numerically, r can vary between –1 and +1. If the value of r is between –1 and 0, then the direction of the correlation is negative (indicating that if one variable increases, the other decreases); if the value is between 0 and +1, the direction of r is positive (indicating that if one variable increases, the other also increases). Sometimes r is multiplied by 100 and expressed as a percent (e.g., $r = 0.77$ may be expressed as 77%). The farther the numerical value is from 0, the *stronger* the correlation.

Perhaps the most noteworthy point about a correlation coefficient is that it *only indicates the strength of linear relationship* between two variables. Therefore, it is possible for two variables with $r = 0$ to have a strong curvilinear association. For example, data that are distributed like a sine wave will have $r = 0$ (i.e., no linear relationship), but will have a strong association with two points of inflexion at the crest and trough (see Figure 4.2). This is an important source of confusion in the literature because investigators sometimes assume that low r means no association between two variables; however, it only means *no linear* association. At the same time, stronger r values are often assumed to be evidence for causal association; it implies a linear association between two variables and says *nothing* about a causal relationship between two variables. An astute observer may also point out that if a sine wave is split at the two points of inflexion, then we will be left with three straight lines, each exhibiting very strong r values (i.e., two of those lines would show positive correlation, and the third line would show negative correlation; however, joining them back into a sine wave reduces the correlation coefficient to 0). Therefore, unless there is additional compelling evidence, it not necessarily true that the strong correlation (negative or positive) will also hold for extrapolated values. Figure 4.2 shows hypothetical selected distributions and their correlations.

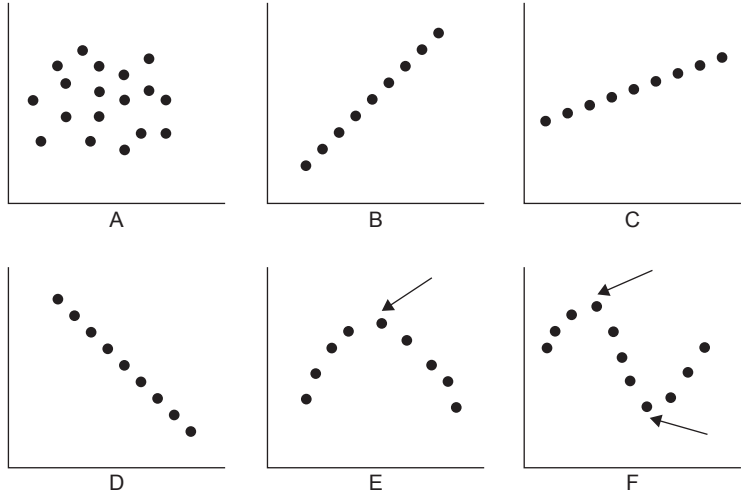


FIGURE 4.2 Correlation A: No linear correlation. B: Strong positive linear correlation. C: Positive linear correlation that is weaker than B. D: Strong negative correlation. E: Curvilinear correlation—no linear correlation. The arrow shows point of inflexion. F: Sine wave—curvilinear correlation. Arrows show points of inflexion.

Spearman's correlation, often called Spearman's rho, is a nonparametric correlation measure that is less "powerful" than the parametric counterpart. It is denoted as r_s or ρ (Greek rho). It assesses arbitrary monotonic function between two variables independent of their distribution assumptions. It is applicable for ordinal scale data and is used to assess the correlation between categorical variables. Its value ranges and interpretations are similar to those of Pearson's product-moment correlation for both strength and direction. It assesses the strength and direction *of linear association* between categorical variables.

Coefficient of determination is the square of the correlation coefficient (r^2) and explains the proportion of variance in one variable that is explained by the other variable. This value is often used in context of regression analysis in assessing how much variation in the data is explained by the regression equation.

Hypothesis testing using correlation coefficient is sometimes a source of confusion. Just because a correlation has been found in a sample does not mean that such a correlation is also present in the population. The null hypothesis for tests of correlation coefficient is that the linear correlation between the two variables *in the population* is zero. Rejecting the null hypothesis means that there is *some nonzero linear* correlation and not necessarily a "strong" correlation (of "cause–effect" type). This is accomplished

by using a t-test or an F-test (the square of the "t" statistic on "d" degrees of freedom is equal to the "F" statistic on "1, d" degrees of freedom). Alternately, a test can be designed specifying a certain threshold of correlation coefficient (e.g., null hypothesis that $r = 0.7$), or testing the difference between two nonindependent correlation coefficients. However, a statistically significant test result does not mean that an observed correlation coefficient in the study is the true correlation coefficient because sample correlation coefficients are *not* unbiased estimates of population correlation coefficient, especially when the sample size is small. An *adjusted correlation coefficient* can be calculated to obtain a relatively unbiased estimate of the population correlation coefficient. Furthermore, depending upon how much the two variables are allowed to vary in the sample data leads to "range restriction" of correlation coefficients, which influences the correlation coefficient compared to what it would be had there been no range restriction. If there are heterogeneous subsamples within the study sample in which the correlation between the study variables differ, then correlation coefficient of the total sample may under- or overestimate the sample correlation coefficient. Therefore, when assessing the correlation coefficient, it is very important to assure oneself that the statistic is being correctly interpreted.

There are no clear structured thresholds about what correlation coefficient values indicate good correlations. However, a general guide is that values between 0.1 and 0.3 (or –0.1 and –0.3) indicate small correlation, 0.3–0.5 suggest medium, and more than 0.5 suggest large correlation. More conservative estimates are also used. In most health science literature, correlation coefficients below 0.6 are generally considered unacceptable. However, investigators must assess their results in view of the problems of making inferences in the absence of adjusted correlation coefficient and without complete assessment of the limitations of their data.

Odds Ratio (OR)

Probability is a proportion that is usually represented as a percentage. Probability measures the proportion of number of events out of the number of times the event can occur. For example, if a person has 32 teeth, the probability of getting dental caries on one tooth is 1/32 = 0.03125 or 3.125%. If an event must happen all the time, then its occurrence is a certainty, and its probability is considered to be 1 or 100%. However, if an event cannot occur, then its nonoccurrence is a certainty, and the probability of the event occurrence is 0 or 0%. Probability values fall between 0 and 1. Probability shows only one side of an event; that is, the occurrence of an event (in the above example, one carious tooth). However, the nonevents exhibit another side of reality (i.e., 31 teeth not having caries). This latter picture is missed by probability estimates. Odds, on the other hand, incorporate information from both sides. If P is the probability of an event happening, then 1 – P is the

probability of the event not happening. Odds incorporate the information and are calculated as Odds = P/(1 – P). Therefore, if the probability of an event is 0.4 (or 40%), the probability of an event not happening is 1 – 0.4 = 0.6 (or 60%); the odds of the event are 0.4/0.6 = 0.67. If the probability of an event is 0.5, then the odds are 1.0, and so forth. Box 4.1 describes an example for calculating odds ratio (OR). *If 2X2 tables are set up in a certain specific way*, then as a mathematical convenience, the OR can be calculated as the ratio of products of cells diagonally across from each other; and is sometimes called the cross-product ratio. Although this *mathematical convenience* may be used, conceptually, OR is not a cross product, but a ratio of odds. Unfortunately, it is common in epidemiologic parlance to define OR as ($a \times d/b \times c$) with the alphabets representing certain cells in a 2X2 table set up in a specific way (see Box 4.1). Therefore, while interpreting the meaning of OR, it is prudent to recall its true mathematical form.

Diseases may occur due to an exposure, and may sometimes occur for reasons other than the particular exposure. However, the focus of a study is to find a *modifiable* factor that can be used as a leverage to either prevent disease, detect it early, or treat it with the best possible results. In this context, studies try to determine the odds of disease among those who have been ex-

BOX 4.1 Probability and Odds Ratio

Probability

Probability (Pr): The relative likelihood that a certain event will or will not occur, relative to some other event.

Mutually Exclusive: Two events are mutually exclusive if the occurrence of one precludes the occurrence of the other.

Additive Rule: For mutually exclusive events: Pr (X or Y) = Pr (X) + Pr (Y)
Independent Events: Additive rule as above. Multiplicative rule: Pr (X and Y) = Pr (X) × Pr (Y)

Conditionally Dependent: Two events are conditionally dependent if the outcome of one depends on the other. Multiplicative Law: For conditional events: Pr (X and Y) = Pr (X) × Pr ($Y \mid X$)

At Least One Event: Pr (at least one) = [1 – Pr (None)]

Calculating Odds Ratio

Odds (O): [Pr (event happening)]/[1 – Pr (event happening)]

Odds ratio (OR): Odds (exposed)/Odds (unexposed)

(Continues)

BOX 4.1 Probability Odds Ratio *(Continued)*

	Disease Present (Oral Cancer)	*Disease Absent (No Oral Cancer)*
Risk factor present (exposed) (smoker)	60 (*a*)	40 (*b*)
Risk factor absent (unexposed) (nonsmoker)	20 (*c*)	80 (*d*)

Pr of disease (exp) = 60/100; Pr of disease (unexp) = 20/100

OR = [(0.6/ 0.4)/(0.2/0.8) = 1.5/0.25 = 6.0

Mathematical Convenience: OR (cross-product ratio) = $[(a \times d)/(b \times c)]$ = $[(60 \times 80)/(40 \times 20)]$ = 4800/800 = 6.0

Interpretation: Those with oral cancer are six times as likely to be smokers as those without oral cancer.

If OR = 0.6, then those with oral cancer are 0.6 times as likely to be smokers as those without oral cancer.

If OR = 1, then those with oral cancer are equally as likely to be smokers as those without oral cancer.

95% confidence interval (CI) around OR = exp[ln(OR) ± 1.96 × SE(lnOR)]

Note: exp = exponential function; ln = natural log function; SE= standard error.

posed to the factor(s) under study and compare them to the odds of disease among those who were not similarly exposed. In this situation, disease occurrence among the unexposed group may be viewed as "background risk" for disease that is not due to the exposure under study. This factor must be adjusted to show the actual odds of occurrence of disease among the exposed. OR, as the ratio of two odds, is calculated as odds among the exposed/odds among the unexposed, and can take any nonnegative rational number value (i.e., 0 to + ∞).

An OR of zero means that the numerator odds (odds of disease among the exposed) is zero; that is, the probability of disease among the exposed is zero. OR = 1 means that the numerator and denominator odds are the same (i.e., the odds of disease occurrence among the exposed and the unexposed are the same). This is interpreted as evidence for "no effect" of the exposure. If the OR is greater than one (i.e., the numerator odds are greater than the denominator odds), then it is interpreted as evidence for exposure to have a

positive effect on disease occurrence, whereas if the OR is less than one (i.e., the numerator odds are less than the denominator odds), it is interpreted as a negative effect of exposure on disease occurrence (sometimes stated as "protective" effect of exposure on disease). Although we have discussed OR in terms of odds of disease occurrence, ORs are calculated in case-control studies where cases are selected and conditioned upon disease status and then exposure is ascertained, which implies that one actually estimates the *odds for having been exposed among cases and controls*. For example, if in a case-control study of oral cancer, the OR for smoking was found to be 8.0, the correct interpretation would be that *cases were eight times more likely than noncases to have smoked* (i.e., the odds of exposure by case status).

Ideally, when examining the association between an exposure and disease, the investigator wishes to explore a causal association by demonstrating the occurrence of new disease after exposure has occurred; that is, calculate the RR of a disease (i.e., the risk of getting a disease after exposure compared to getting it in the unexposed group). However, case-control studies do not provide the opportunity for measuring the RR. Therefore, ORs are calculated in case-control studies as *estimates* of the true RR. If the disease is rare, then ORs are good estimators of RR, but if the disease is not rare, then ORs are poor estimates of RR.

Risk Ratio

As mentioned earlier, $RR = R_1/R_0$, and units of time are not associated with RR. However, when measuring disease incidence rate (or *incidence density*) in a study, different participants may contribute different amounts of time in the study (see Figure 4.1), and the incidence density among the exposed group may be different than the incidence density in the unexposed group. If I_1 is the number of new cases in the exposed group followed up for a total of time T_1, and I_0 is the number of new cases in the unexposed group followed up for a total of time T_0, then incidence rate/density = $(I_1 \times T_1) / (I_0 \times T_0)$. A ratio of the two incidence densities may eliminate the units of time, but the differences in the magnitude of time remain in the measurement. Therefore, incidence density ratios produce different estimates than RR in magnitude and meaning.

Hazard Ratio

In survival analysis, at any time point *t*, the mean survival (time) = (time to event at *t*)/(number of participants with outcome till time point *t*). The survival rate = (number of participants surviving at time point *t*)/(total number of participants). Person–time can be incorporated into the denominator of these equations to provide for person–time survival rate. Hazard = probability of event in time *t* = (number of persons with the event until time

point *t*)/(number of persons at risk at the start of the study). Therefore, hazard is the probability of the event for those participants who survived until time point *t*: that is, the instantaneous probability of the event having survived until that point in time (*t*).

Hazard rate is defined as the instantaneous event rate also called the failure rate at any instant of time *t* given that the event has not yet occurred until that point in time, and will occur in an infinitesimally small instant after the time point *t*. Therefore, the hazard ratio incorporates the probability of having survived until a time point without the event. The hazard ratio is determined in survival analysis as the effect of an explanatory variable on the hazard or risk of an event. It is the ratio of two hazard rates: hazard rate of the exposed and that of the unexposed. If time is viewed as a discrete factor rather than its usual interpretation as a continuous factor, then this instantaneous rate indicates a hazard rate.

Measures of Impact on Population

Risk Difference

Difference measures can be calculated as RD or as incidence rate difference and provide a measure of the absolute effect in an exposed group, as well as provide an assessment of the impact of an exposure in a population. For example, in a study, if R_1 for oral cancer among smokers = 50% and R_0 for oral cancer among nonsmokers = 45%, then RD = 50 – 45 = 5%. Therefore, one can say that only 5% of the oral cancer cases are attributable to smoking. In other words, only 5% excess risk for oral cancer exists due to smoking. In a population of 100,000, if smoking is eliminated, 5000 cases of oral cancer can be prevented. Therefore, it is possible to calculate the number of people who will be impacted due to an exposure or by removal of an exposure. This is the reason that RD is also called *attributable risk*. This measure can be presented as a proportion of the R_0 (attributable risk percent) or may be measured for the whole population instead of the study (see Table 4.1). Attributable risk is defined as the risk that can be reduced when the exposure is eliminated. In the above example, RR = 50/45 = 1.11. Suppose another study found R_1 =5% and R_0 = 4.5%. In this new study, RR = 5/4.5 = 1.11; the same as the earlier study. However, in the new study, RD = 5 – 4.5 = 0.5%, which according to the new study would mean that eliminating smoking would prevent 500 cases compared to 5000 cases estimated by the earlier study. Although RRs reported by different studies may be the same, RDs have a very important application in terms of the actual impact in the population. RD is usually never reported or used in oral health research.

Studies report rates in different ways. For example, the RR from the above studies may be reported as an 11% increase in risk due to smoking; or, those who smoked were 1.11 times more likely to get oral cancer. If smoking

were not associated with oral cancer, then according to the first study, one would expect $R_1 = R_0 = 45\%$; that is, RR = 45/45 = 1, or RD = 1 – 1 = 0. However, the observed RR was 1.11. The 11% increase, therefore, is calculated as RR (observed) – RR (if there was no association) = 1.11 – 1.0 = 0.11 or 11%. If a study were to report an observed RR of 5, it sometimes may state that there was a fivefold increase in risk. This does not translate to a 500% increase. The percentage increase calculated from the above formula would be = 5 – 1 = 4 or 400%. This calculated increase is sometimes called the *relative effect* or RR increase (if RR < 1, then it would be an RR decrease).

Diagnostic Tests

New diagnostic tests come up frequently, and they are assessed against a set "gold standard," which is usually an already established test. Box 4.2 describes the fundamental statistics that are used in characterizing a diagnostic test. The most important properties of a test are (1) its *validity* (i.e., the ability to distinguish between those who have the disease and those who do not—ideally it should also be able to correctly classify people with preclinical disease as positive and people without preclinical disease as negative), and (2) its *reliability* (i.e., it should give the same results when the test is repeated). *Sensitivity* of a test is its ability to identify correctly those who have the disease, whereas *specificity* is its ability to identify correctly those who do not have the disease. Tests are not ideal and often over- or underdetect diseases leading to errors in diagnosis: false positive and false negative outcomes. It is generally assumed that sensitivity and specificity of a test are fixed properties of the test. However, sensitivity and specificity of a diagnostic test may vary depending on population characteristics such as age, disease severity, disease subtypes, and changes in disease characteristics or profile. The need for diagnostic accuracy and certainty depends upon the penalty for being wrong about the patient's true disease status because of the associated burden. *False positive* (FP) results carry the burden of monetary and personal-psychological costs carried by the person if finally found out to be disease-free (all people detected positive need further tests that tend to be more expensive and more invasive). *False negative* (FN) results carry the burden of potential complications, poor disease-related outcomes, and related monetary and psychological costs if found to have the disease at a later stage.

The likelihood ratio (LR) of a positive test result (LR+) is the ratio of the probability of a positive test result if the outcome is positive (true positive) to the probability of a positive test result if the outcome is negative (false positive); likelihood ratio negative is interpreted similarly for negative results (see Box 4.2). High LR+ or low LR– for a test indicates that the test is useful. Greater disease prevalence raises the probability of a positive outcome (see next section for a demonstration of the relation among positive predictive value, sensitivity, specificity, and disease prevalence).

BOX 4.2 Diagnostic Test Statistics with Worked Example

Diagnostic Test Performance

	D+	*D–*	*Total*
T+	a	b	$a + b$
T–	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

	D+	*D–*	*Total*
T+	40	20	60
T–	10	30	40
Total	50	50	100

a = "True positives" (TP)
b = "False positives" (FP)
c = "False negatives" (FN)
d = "True negatives" (TN)
Sensitivity = $a/(a + c)$
Specificity = $d/(b + d)$
Positive Predictive Value = $a/(a + b)$
Negative Predictive Value = $d/(c + d)$
Likelihood Ratio + = Sensitivity/$[b/(b + d)]$
Likelihood Ratio – = $[c/(a + c)]$/Specificity
Likelihood Ratio (LR) = LR+/LR–

True positives = (TP = 40/100 = 40%)
False positives = (FP = 20/100 = 20%)
False negatives = (FN = 10/100 = 10%)
True negatives = (TN = 30/100 = 30%)
Sensitivity (Sn) = 40/(40+ 10) = 80%
Specificity (Sp) = 30/(20 + 30) = 60%
Pos. Pred. Value (PPV) = 40/(40 + 20) = 66.7%
Neg. Pred. Value (NPV) = 30/(10 + 30) = 75%
LR+ = 0.8/[20/(20 + 30)] = 2
LR– = [10/(40 + 10)]/0.6 = 0.33
Likelihood Ratio (LR) = 2/0.33 = 6.06

Diagnostic test performance is measured against a designated "gold standard." It is assumed that the gold standard is correct all the time. Positive (T+) and negative (T–) results from the new test are compared with positive (D+) and negative (D–) true conditions diagnosed by the gold standard, and various statistics are derived as seen in the example above.

Diagnostic tests usually perform well when the disease has clear thresholds that can be used for disease detection. However, most diseases have a "fuzzy" transition phase when latent disease (preclinical phase?) may exist, but is detectable or not detectable by the test at this stage depending on the test's "criterion of positivity" for making a diagnostic call. Criterion of positivity is the threshold test value at which the test outcome is considered positive. Criterion of positivity of tests affects their sensitivity and specificity. In the fuzzy phase, there is a trade-off between specificity and sensitivity of tests. If the threshold is low, then sensitivity increases but specificity suffers, whereas if the threshold is high, then specificity increases but sensitivity suffers.

Decisions about the threshold for criterion of positivity involve weighing the *cost* of false positives (FP) against the *cost* of false negatives (FN). These costs could be economic, social (e.g., burden on health care), or psychological (e.g., anxiety and mental stress, labeling, stigma, long-lasting fear, risk of more invasive tests), all of which cannot be necessarily assessed in monetary terms. A misclassification cost "term" (MCT) can be defined as

$$\text{MCT} = [(1 - \text{P}) \times (1 - \text{Sp}) + \text{R} \times \text{P} \times (1 - \text{Sn})]$$

where R = [costs (FN)/costs (FP)] and specifies an estimated ratio of the costs associated with FP or FN results, respectively, Sp = specificity, Sn = sensitivity, and P = prevalence of disease. The MCT can easily be incorporated in most calculations and appropriate sensitivity analyses can be performed to assess impacts on predictive values of tests. These aspects of diagnostic tests have rarely been discussed in detail in oral health research literature. In general, cost issues are "rarely discussed in the literature and if discussed at all, they are usually incorporated as a fixed summary misclassification cost entity. Utilizing MCTs does not necessarily include specific characteristics of a given entity, but bases itself on the average costs for the entire population, even though MCT may vary substantially at individual level. A cost-neutral approach to classification assumes equal misclassification costs between classes" (Chattopadhyay & Ray, 2008). Risk-benefit analysis of diagnostic tests is key to optimizing beneficial outcomes and reducing costs due to errors. Newer tests with similar performance to the gold standard, but which are perhaps cheaper or have better risk-benefit status might be preferred over the gold standard. The memorandum for the evaluation of diagnostic measures (Kobberling, Trampisch, & Windeler, 1990) emphasizes acceptance of diagnostic tests on the strength of their evidence base. The memorandum suggests a clinical trial type of protocol for evaluating diagnostic tests with four clearly marked phases:

Phase 1: Preliminary technical and methodological investigations
Phase 2: Estimation of test parameters in selected patients
Phase 3: Controlled diagnostic study

Phase 4: Investigation of the effectiveness of continuing risk–benefit analysis

These phases do not necessarily need to occur in strict sequence, but they should be followed at least generally to establish diagnostic superiority of one test over another, although this rarely occurs.

Comparative assessment of different tests can be conducted using their respective sensitivity and specificity statistics. A simple tool that can be used to compare utility of diagnostic tests is the receiver operating characteristics (ROC) curve. Figure 4.3 demonstrates how ROC curves can be used to quickly to assess different tests in terms of their sensitivity and specificity. It is usually plotted using the sensitivity and 1–specificity values of tests; alternatively the same curve can be plotted by using true positive and true negative rates. Because the plot compares two characteristics, it is sometimes called the relative operating characteristics curve. The general idea under ROC analysis is to assess the total area covered by the plot (area under the curve: AUC). AUC is the fraction of the total area that falls under the ROC curve; that is, the higher the AUC, the greater the signal: noise ratio of the test which implies better usefulness of the test to make correct diagnostic decisions. Therefore, a diagonal line such as curve A (see Figure 4.3) would be a worthless test, because the AUC would be only 50% (i.e., 50% of the information lies on either side of the curve), and the probability of a correct decision using such a test would be 50%, which is as good as tossing a coin to diagnose disease. A general guide to interpreting AUC values may look like AUC = 0.5: Noninformative; $0.5 \leq$ AUC 0.7: Less accurate; $0.7 \leq$ AUC 0.9: Moderately accurate; $0.9 \leq$ AUC < 1: Highly accurate; AUC = 1: Perfect test. The top left-hand corner indicates 100% sensitivity and 0% value for 1–specificity, which means 100% specificity. Therefore, it follows that to be useful, the curve has to approach the top left-hand corner so that AUC can increase and the test has better diagnostic power. This feature of ROC curve analysis provides a simple way to assess different diagnostic tests that can be plotted together and visually assessed.

In some situations, more than one test may be needed to confirm diagnosis, the most common example being HIV testing with enzyme-linked immunosorbent assay (ELISA) followed by a Western blot confirmatory test. Such situations mostly arise when screening is conducted and in situations where the confirmatory test is expensive. Such multiple tests may be conducted simultaneously (if either test is positive, result is positive), which leads to a gain in sensitivity at the cost of reduced specificity. Alternately these tests can be conducted sequentially (if both tests are positive, result is positive), which leads to a gain in specificity and loss in sensitivity. The decision to use multiple tests is usually based on objectives of the test (whether for screening or diagnosis), and practical considerations related to the tests (e.g., the setting, length, cost, invasiveness, and insurance coverage).

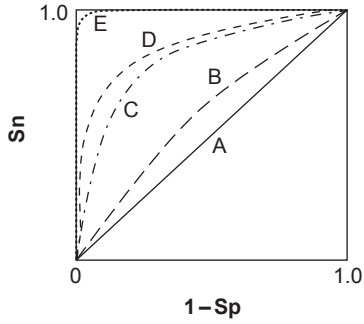


FIGURE 4.3 **ROC Curve** The closer the curve is to the upper left hand corner of the graph, the better the test is. Diagonal A indicates a test equivalent to toss of a coin; B is better than A, C is better than B, D is better than C, and E is nearly an ideal test (near 100% sensitivity and specificity).

Screening Tests

Diagnostic and screening tests need to have different characteristics to be successful in their application. A key reason for this distinction is their roles in disease detection. Whereas diagnostic tests are used in clinical settings with more sick persons trying to find out whether they have a disease, screening is usually carried out in population settings, and the emphasis is usually at early detection. During screening activities, the general goal is to pick up positive cases in the population (i.e., emphasis is on sensitivity), whereas diagnostic tests have the goal of ascertaining whether a person has disease or not (i.e., emphasis is on both sensitivity and specificity). However, both sensitivity and specificity are linked to each other and a trade-off between the two values is generally needed to balance outcomes of a test. Disease prevalence in a diagnostic set-up is usually greater than the population because sicker patients tend to visit or be referred to get diagnosed; whereas screening is usually conducted on apparently nondiseased persons. Therefore, the relationship of test performance to disease prevalence in the population is an extremely important criterion for a screening test.

Positive and negative predictive values (PPV and NPV, respectively) assess performance of tests from a different perspective. These statistics answer the question: Given that the test is positive (or negative), what is the probability that the tested person has (or does not have) disease? This same question is also important to the person undergoing the test. Table 4.2 demonstrates the relationship of diagnostic/screening test results with disease prevalence rates and with each other, if disease prevalence is constant. In either case, PPV is related directly with disease prevalence, sensitivity of

TABLE 4.2 Results from Hypothetical Studies are Presented Below to Demonstrate Relationship Between Sensitivity (Sn), Specificity (Sp), Positive Predictive Value (PPV), and Disease Prevalence

Relationship Assessed	*Disease Prevalence*	*Test Result*	*Disease Present*	*Disease Absent*	*Totals*	*PPV*	*Direction of Relationship*
Disease Prevalence and PPV (Sn and Sp are constant)							*Prevalence* ↓ *PPV* ↓
Sn = 99%							
Sp = 95%	**1%**	T+	99	495	594	**17%**	
		T–	1	9405	9406		
		Totals	**100**	**9900**	**10,000**		
	5%	T+	495	475	970	**51%**	
		T –	5	9025	9030		
		Totals	**500**	**9500**	**10,000**		
	25%	T+	2475	375	2850	**87%**	
		T–	25	7125	7150		
		Totals	**2500**	**7500**	**10,000**		
Sensitivity and PPV (Disease prevalence and Sp are constant)							*Sensitivity* ↓ *PPV* ↓
Sn = 50%							
Sp = 90%	10%	T+	500	900	1400	**36%**	
		T–	500	8100	8600		
		Totals	**1000**	**9000**	**10,000**		
Sn = 70%							
Sp = 90%	10%	T+	700	900	1600	**44%**	
		T–	300	8100	8400		
		Totals	**1000**	**9000**	**10,000**		

Sn = 90%							
Sp = 90%	10%	T+	900	900	1800	**50%**	
		T–	100	8100	8200		
		Totals	**1000**	**9000**	**10,000**		
Sn = 100%							
Sp = 70%	**10%**	T+	1000	2700	3700	**27%**	
		T–	0	6300	6300		
		Totals	**1000**	**9000**	**10,000**		
Sn = 100%							
Sp = 80%	**10%**	T+	1000	1800	2800	**36%**	
		T–	0	7200	7200		
		Totals	**1000**	**9000**	**10,000**		
Sn = 100%							
Sp = 95%	**10%**	T+	1000	450	1450	**69%**	
		T-	0	8550	8550		
		Totals	**1000**	**9000**	**10,000**		

tests, and specificity of tests, if other factors are held constant. Poor PPV values mean greater FP rates. Therefore, a test with poor PPV will be more expensive for the whole program as it will increase the false positive rates, and this situation will be more troublesome when disease prevalence is lower. The same test may provide better results in the diagnostic centers (higher disease prevalence compared to the source population) compared to a screening situation (lower prevalence compared to the diagnostic center). Therefore, the decision to use a test for a screening purpose needs more careful attention to PPV and NPV values apart from sensitivity and specificity of a diagnostic test.

A legitimate question is: Are all diseases suitable for screening? That is, should we set up screening programs for all known diseases? The clear answer to that question is negative. Diseases suitable for screening should have certain properties such as the disease has severe health consequences, the disease is progressive, effective treatment for the disease is available and treatment is more effective at an earlier stage, the disease has a detectable preclinical phase (DPCP), and the DPCP is fairly long and of high prevalence in the target population. A good screening test should be relatively inexpensive; noninvasive, or as minimally invasive as possible; cause minimal discomfort; be easily available; be ethical; and be valid: It must detect what it is supposed to detect.

To be successful and effective, screening programs aim at reducing morbidity and should contain the insight that PPV will increase when sensitivity, specificity, and disease prevalence increase; and that given fixed test characteristics, the programs are usually most productive when directed to a high-risk population. For chronic diseases, one can assess the effectiveness of screening tests by examining the severity of disease at diagnosis, cause-specific mortality rate among persons picked up by screening versus persons picked up by routine care, and cost-benefit and risk-benefit analyses. In oral health research and surveys, screening for several diseases is not a technology-based test, but a clinical examination by trained professionals; for example, oral cancer screening and NHANES (National Center for Health Statistics [NCHS], 1996). Therefore, the reliability of the screening program and proper standardization and calibration of examiners are extremely important factors in the success of the program as well as the validity of the data being generated.

Agreement

Observations about the same thing by the same person at two different occasions can vary, as can the observation of the same thing by two or more different persons at the same time. To ascertain the quality of data in situations where more than one observation has to be made, or where more than

one observer examines different persons, as is common in public health surveys, it is important that examiner observations are valid and reliable. One way of measuring the consistency of observations is to assess the agreement between different observations made by the same observers over different time periods (within-observer agreement), and/or between different observers (between-observer agreement). For example, in a study to assess fluorosis development in seven age cohorts after an 11-month break in water fluoridation, two examiners examined 761 children in grades 3, 4, and 5 in 22 participating elementary schools in Durham, North Carolina. The study was carried out over two phases and in the second phase, each examiner independently reexamined the same group of 23 children with 79% agreement (Burt, Keels, & Heller, 2003).

However, the mere fact of good/strong/complete agreement between two or more observations can occur by chance and thus raises the question: To what extent does agreement between two observers (or two tests) exceed the agreement that would result from chance alone? The *kappa statistic* (K) measures the extent to which observed agreement exceeds agreement expected by chance. It can be calculated as:

> K = [(% observed agreement) – (% expected agreement)]/[100% – (% expected agreement)].

The values of kappa range from 0 to 1, where 0 indicates agreement totally by chance and 1 indicates agreement totally not due to chance. Obviously, in real-world situations, "black and white" kappa values, such as 0 and 1, are not obtained. Just as for correlation coefficients, there is no explicit threshold of kappa values that clearly demarcate the role of chance in agreement. However, as a general guide, kappa: < 0.4 (i.e., < 40%) is considered poor agreement, kappa: 0.40 – 0.75 (40% – 75%) is considered acceptable/good agreement, and kappa: > 0.75 (0.75%) is considered very good agreement. However, investigators should assess kappa carefully, with circumspection and be very clear before accepting the middle category as good evidence for real agreement between examiners. Let us consider kappa for a study of 0.6, which suggests that there is a 60% probability that agreement was real occurrence (or a 40% probability that the agreement was by chance)—confidence in the kappa statistics in such situations must stem from investigators' understanding about the phenomenon under study.

In the fluorosis study mentioned above, the investigators found kappa to be 0.55 (i.e., 55%). The investigators did not find any change in fluorosis prevalence despite a break in public water fluoridation in Durham. Trying to explain the observations, investigators considered the possible role of examiner error as examiners were using more stringent criteria for two different birth cohorts (phase I and phase II of the study). There was a 6-year gap between the two sets of examinations; "examiner drift with respect to criteria is always a possibility in studies running for that length of time. Interex-

aminer consistency scores for fluorosis (kappa) were 0.67 in phase I and 0.55 in phase II, scores that demonstrate a moderate level of interexaminer agreement" (Burt et al., 2003). The investigators then compared the performance of each of the two examiners across both phases and found that one examiner had scored consistently higher than the other throughout the study, but the patterns and the intercohort relationships were much the same for both examiners. Furthermore, the higher-scoring examiner examined 57% of the children in phase I (compared to 43% for the other examiner); the combined data are weighted in favor of the higher-scoring examiner. The investigators acknowledged that "it is therefore possible that the extent of the differences between the cohorts was exaggerated, even though the basic pattern was the same for both examiners" (Burt et al., 2003). However, agreement does not necessarily translate to a correct decision! Despite their perfect real agreement (i.e., very high kappa scores), it is possible that the examiners may both (or all) be wrong. Therefore, as a preparation for surveys, a calibration exercise for examiners must include not only agreement assessment, but also a standardization component.

Precision, P-Value, and Confidence Intervals

Measures of association observed in a study are estimates of the true population parameter and are presented as a single number (point estimate). Therefore, depending upon the characteristics of the study, this point estimate will be different from the true population parameter (in magnitude, and possibly direction). One way to "capture" the true population parameter from a study is to define a range of p-values of possible parameter values so that for the interval of the range of these parameter values, the test p-value exceeds the alpha level (level of significance) that is customarily set at 0.05 for most studies (Rothman, Greenland, & Lash, 2008). This range of defined parameter values is called the confidence interval (CI), where the parameter edges are the confidence limits. The width of this CI is defined by the random variation in the study and by the alpha level (usually 0.05). Therefore, CI = 1 – 0.05 = 0.95 or 95% CI is used most commonly. CIs can be set for any range of confidence levels: 90%, 80%, 86%, and so on. However, with alpha set at 0.05, 95% CIs are normally used, although under more conservative situations, stricter CIs (99%) are defined, as in some genetic epidemiology studies.

Under perfect conditions, studies would be carried out with perfect representative samples, perfect measurement, very high power (100%), and no biases at all. Even under such conditions, studies are subject to random error. However, theory suggests that under the perfect conditions described above, *if and only if* the statistical model used is correct (i.e., *correct model,*

constructed and analyzed correctly), then the 95% CI derived from unlimited number replication of such tests will contain the true population parameter 95% of the time. This would imply that the remaining uncertainty is attributed totally to random error (because by definition, no other errors were allowed) (Rothman et al., 2008). Of course, if a census of the population is conducted (i.e., every individual at risk is measured per study design), the investigator will not need CIs because then the true parameter will have been directly calculated. Real-world studies calculate CIs because studies are not conducted under perfect (or even near-perfect) conditions and are subjected to several errors and biases. Furthermore, such studies are also not "replicated an unlimited number of times," if at all!

Statistical interpretations that are based on the need for unlimited replicated studies are called *frequentist* interpretations, and the followers of this paradigm (most of the investigators of the world) are called frequentists who count the frequency of events and assign probabilities based on such counts. On the contrary, *Bayesian* statisticians are those who follow Bayes's theorem (after Thomas Bayes, 1702–1761) on conditional probability, interpreting statistics on the "degree-of-belief and uncertainty" paradigm of probability. Bayesians state that probability uncertainty and degrees of belief can be measured as probabilities, and these may change based on events. The basic concept is that for frequentist probability, a large (unlimited) number of trials need to be done. What if someone does not conduct a large number of trials, but conducts an event only once? How does someone assign a probability to an event? The usual answer is to look for another person who may have conducted a large number of trials, or conduct theoretical trials and so on, calculate the probability from those trials, and apply that probability to this single trial. This process requires someone to have conducted a large number of trials of that type *earlier*. Bayes's theorem suggests that although probability *requires* multiple trials, we do not know the probability of an event at the time the event occurs, but we gradually become aware of the probability as we experience more such events occurring (i.e., earlier trials determine our *interpretation* of the probability of current events, and if a first-time new event was to occur now, we would not be able to assign a probability for the occurrence of such an event at this time). Therefore, Bayesian methods assign *prior probability* and *posterior probability* to events: Posterior probabilities are *conditional* upon prior probabilities.

When an investigator calculates 95% CI from a single study, the natural question one asks is: "Does the 95% CI calculated from *this* study contain the true population parameter?" Although almost all studies *assume* that the true population parameter is contained in their reported CIs, the correct answer to the question of the investigator is: "We do not know," because unlimited numbers of similar studies have not been conducted. However, as more similar studies are conducted, probabilities can be placed on the "capture" of the true population parameter in the said CI. Bayesian methods of

calculating CIs can answer the question asked by the investigator above. These ideas are also related to the p-value functions across the CI, and to the misinterpretation of p-values in certain circumstances (Rothman et al., 2008). For these and other reasons, epidemiologists on the forefront of methods research recommend that p-values be not reported in epidemiologic study reports, and focus be placed on proper interpretation of CIs instead (Altman, Machin, Bryant, & Gardner, 2000; Poole, 2001; Rothman et al., 2008).

The CI is bound by an upper (UCL) and lower confidence limit (LCL), and provides an interpretation of precision of the point estimate of the study. For ratio measures, the ratio of UCI to LCL (UCI/LCL) provides the confidence limit ratio (CLR), which is a measure of precision of the parameter estimate. For example, let us consider two studies reporting the association of smoking and oral cancer. Study A: RR = 5.2; 95% CI = 3.6, 7.1; and study B: RR = 5.2; 95% CI = 4.0, 6.1. The CLRs for the two estimates are study A: 7.1/3.6 = 1.97, and study B: 6.1/4.0 = 1.53. Comparing the two CLRs (1.53 < 1.97), it can be concluded that study B provides a more precise estimate for the RR than study A. The differences between the CLRs in these studies are relatively minor. However, if 95% CI was 2.0, 15.0, then the CLR would be 7.5, which is substantially greater and also suggests imprecise and unstable estimates. Therefore, CLR, as a measure of precision for risk estimates, can provide a guide to how to interpret the provided evidence. For example, a study that assessed explanatory models of risk indicators for oral hairy leukoplakia among HIV-1 positive individuals reported and compared CLRs to estimate the precision between models as a part of the modeling procedure. This study reported that compared with their full model, all ORs in the final model gained in precision demonstrated by the lesser CLR values (Chattopadhyay et al., 2005). Because RR is a multiplicative scale measure, the ratio of UCI and LCI is used. However, for RD, which is a difference measure, the precision estimate is calculated as UCI – LCI (confidence limit difference: CLD).

Systematic Review and Meta-Analysis

While individual studies contribute important information, by themselves they do not provide complete evidence to accept or refute the solution to a question. Therefore, as more studies are conducted, their results need to be compiled together to arrive at more robust answers. Systematic reviews and meta-analyses serve the purpose of providing these robust answers.

Systematic Reviews

Systematic reviews are literature reviews that focus on a single topical question, identify all high-quality research articles on that topic, and then select

from the list only those that fulfill explicitly defined strict selection criteria (for validity and reliability). Thereafter, these reviews synthesize the information from the selected articles to provide the evidence base to answer the topical question. Such reviews are considered to be good resources for obtaining evidence for a question.

Systematic reviews must incorporate certain characteristics to provide good evidence. They should:

1. Follow an explicit, objective, and transparent methodology
2. Select good quality studies
3. Minimize bias
4. Try to include all quantitative and qualitative data from studies

Systematic reviews should be viewed as ongoing exercises that incorporate new studies as they get published; as conclusions of the reviews are revised in light of new evidence. Systematic reviews generally involve the following seven steps:

1. Identify the problem/study question.
2. Develop inventory of studies through extensive searching and outlining selection criteria.
3. Review each study critically to assess merits for inclusion.
4. Collect data across different selected studies.
5. Analyze the data (including meta-analysis if warranted).
6. Interpret and publish the results.
7. Monitor published literature and update previously published reviews as appropriate.

Whereas systematic reviews are considered strong evidence for the concerned topic, not all systematic reviews are equal (Moher, Tetzlaff, Tricco, Sampson, & Altman, 2007). Systematic reviews should be viewed similarly to other studies and may also be subject to the errors and biases that influence other studies. Use of standardized criteria may bring parity to methods, but if the criteria on which methods are based are themselves inappropriate, then quality may not be properly addressed. However, inclusion criteria and review guidelines may themselves have to be updated as newer research provides new insights.

Founded in 1993, the Cochrane Collaboration is the most well-known resource for healthcare-related systematic reviews. It is an independent, international, nonprofit organization that produces and disseminates systematic reviews of healthcare interventions and promotes the search for evidence in the form of clinical trials and other intervention studies (Cochrane Collaboration, 2002). Reviews of the Cochrane Collaboration are accessible through their website at http://www.cochrane.org/. Oral health-related

systematic reviews of the Cochrane Collaboration may be directly accessed at http://www.ohg.cochrane.org. At the time of writing this chapter (November 2008), the number of oral health topic systematic reviews were 92, including 28 updated reviews and one review that was withdrawn.

Most systematic reviews tend to incorporate only randomized clinical trials as worthwhile evidence. This practice is based on the paradigm that placebo-controlled, double-blind, randomized clinical trials are the "gold standard" for clinical evidence. Clinical trial-related issues are discussed in Chapter 11. However, the decision to use only such trials as sources of appropriate evidence has led to a series of systematic reviews based on a small number of studies. For example, the review "Fluoridated milk for preventing dental caries"(Yeung et al., 2005) included only two studies, and the reviews "Antibiotic use for irreversible pulpitis" (Keennan, Farman, Fedorowicz, & Newton, 2005) and "Slow-release fluoride devices for the control of dental decay" (Bonner, Clarkson, Dobbyn, & Khanna, 2006) included only one study each. Although several reviews cover the topic comprehensively, others are not able to provide firm conclusions or any guidance on the review topic. Although such reports can serve the purpose of identifying areas where an evidence base needs to be developed, whether such inconclusive reviews serve any other useful purpose is an open question. The use of randomized clinical trials as the only source of credible evidence and exclusion of all observational studies is itself perhaps a biased view because most such trials do not report the involved errors and biases. Almost all clinical trials are not population-based studies and do not use stratified sampling methods to obtain a representative sample of the source or target population; neither are real-world exposures randomized. Furthermore, the entire paradigm is based on production of p-value-based evidence, which is essentially a function of effect size and sample size.

Meta-Analysis

Meta data are definitional data that provide information about other data; that is, it is *data about data*. Essentially, the meta-analysis of meta data is a mathematical approach that follows similar steps as in systematic reviews and focuses on obtaining an overall common (average) point estimate from the selected studies that test a hypothesis associated with the topical question. Meta-analyses define a common effect size and adjust for study characteristics. The key goal of meta-analyses is to obtain the *best estimate of the true population parameter* with greater precision from across all available high-quality studies. The central idea is based on replication of studies. As noted above, probabilistic interpretations are based on multiple replicated trials. Whereas the same study may not have been replicated, meta-analysis uses similar studies to approach the replication paradigm and then combines their parameter estimates after appropriate adjustments. This new

"meta-estimate" is considered to be a more precise estimate of the true population parameter. As in Bayesian methods, as more studies are conducted, the meta-estimate may be sequentially revised. These meta-estimates may be incorporated as credible evidence to determine practice recommendations and create guidelines. They also include moderators that help explain study variations. However, they cannot correct for biases that are included in individual studies, publication bias, unpublished results, and poor study design problems. All systematic reviews may not include meta-analysis.

Meta-analysis involves the following sequential steps:

1. Identify the problem/study question, design the study, define the outcome, and specify the determinant–outcome relationship.
2. Develop inventory of studies through extensive searching and outlining selection criteria. For this purpose, electronic databases are usually searched, including Biomedical Research and Informatics, CINAHIL, Clinical Alerts, Clinical Queries, ClinicalTrials.gov, Consumer Health, EBSCO, EMBASE, Environmental Health and Toxicology, Health Information Technology Resources, Health Services Research and Public Health, Human Genome Resources, LILACS, NLM Catalog, NLM Gateway, NLM Mobile, PubMed Central, SCOPUS, and TOXNET, among others. Special search filters may be developed to search through electronic indexes and databases. All resources are not necessarily included in electronic databases. Such resources may be searched through lateral cross-references; accessing conference proceedings; and searching for dissertations, books, published special reports, and unpublished reports, including government internal reports.
3. Review each study critically to assess its merits for inclusion (decision about inclusion/noninclusion of unpublished results can alter the results substantially) and the data extraction from the selected studies. Some meta-analyses incorporate a "fail-safe N" statistic that calculates the number of studies with null results that would need to be added to the meta-analysis in order for an effect estimate to be no longer reliable. This allows investigators to interpret results and keep in perspective the publication's bias toward reporting only non-null results.
4. Make decisions about the dependent and independent measures summary statistics to be used (e.g., means, differences, risk estimates, ORs, relative risk, hazard ratios, and decisions about treatment of statistic variances), and types of analyses to be included depending on the topic of interest (i.e., most clinical outcome-related meta-analyses including clinical trials prefer to include hazard ratios instead of odds or risk-based measures).
5. Collect the data across studies, and analyze the data and the model selection. Decisions about weighting and handling heterogeneity are required before progressing with modeling. Generally regression models

are used in meta-analysis; the common ones are simple regression, fixed-effects meta-regression, and random-effects meta-regression.

6. Interpret and publish the results, monitor published literature, and update previously published reviews as appropriate. Reporting usually involves usage of flowcharts describing study sampling and selection plan with a complete description of the number of items at each stage. *Funnel plots* (see Figure 4.4) are often used to explore the presence of publication and retrieval bias. Funnel plots are scatter plots of the treatment effect against a measure of study size derived from different studies and are used as visual aids to detecting bias or systematic heterogeneity. For example, the theoretical example in Figure 4.4 suggests that the smaller and less precise studies scatter farther than larger and more precise studies. Asymmetric "funnels" indicate a relationship between treatment effect and study size, may indicate publication bias or other problems, and may question the use or appropriateness of a simple meta-analysis. *Forrest plots* (see Figure 4.5) show information and effect estimates from individual studies and provide a view of variation between studies. Forrest plots show the strength of the evidence in

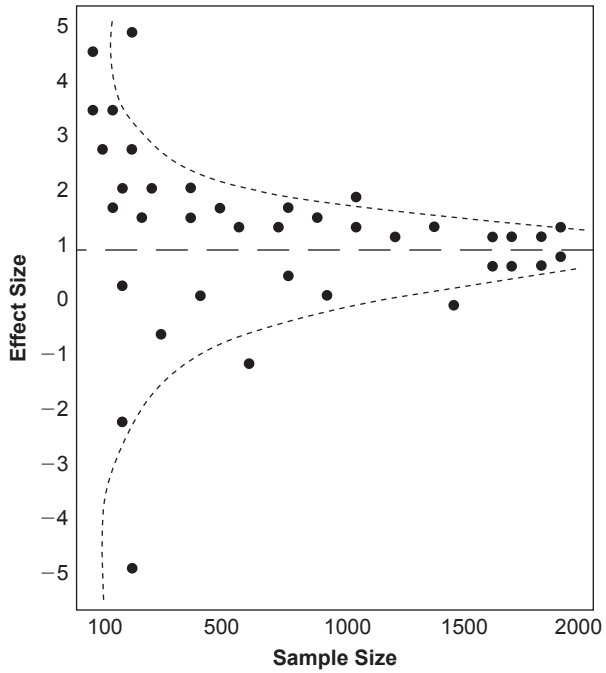


FIGURE 4.4 **Generic Funnel Plot** This generic funnel plot was created from a hypothetical study.

quantitative scientific studies used in the meta-analysis. The effect estimate (and confidence interval) of each study is shown graphically across a vertical line that depicts the "no-effect" zone in the graph. Usually the meta-estimate is also shown on the funnel plot so that readers can make judgments about how estimates from different studies relate to each other and the meta-estimate. Individual studies are usually represented using the same symbol (usually a solid square), and the meta-estimate is shown with a different symbol (usually a solid diamond shape). Meta-analyses generally result in one of the following types of conclusions about strength of evidence: Strong evidence, moderate evidence, weak evidence, inconsistent evidence, or little or no evidence.

Use of meta-analysis enables investigators to appreciate the impact of multiple studies on scientific inferences by shifting the focus from individual studies to collective evidence. It also takes the emphasis away from individual p-value-based reporting to the impact of effect sizes and emphasizes the importance of obtaining the best estimate for the true population parameter as a study goal. The combined estimates are calculated using some kind of weighting techniques. The reason for weighting is to ad-

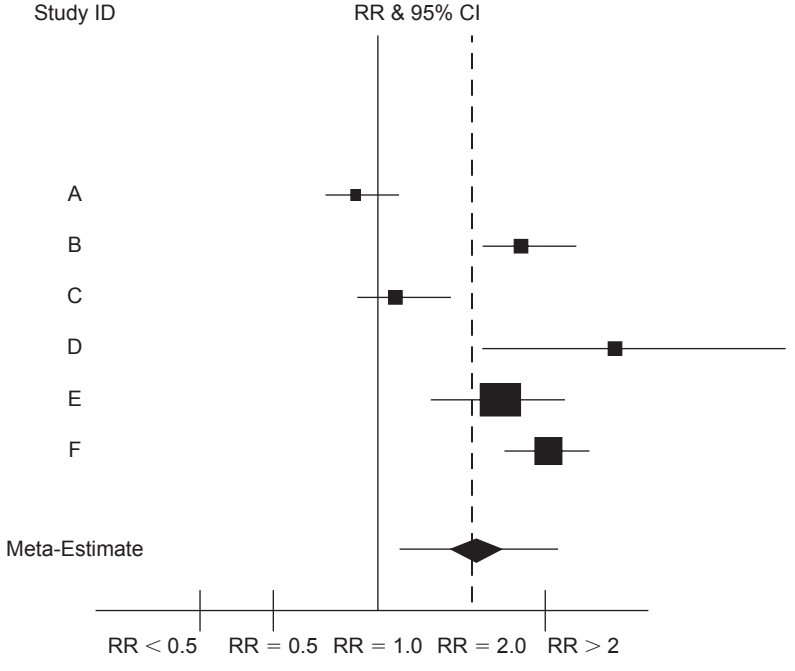


FIGURE 4.5 Generic Forrest Plot This generic forrest plot was created from a hypothetical study.

just for variances and err on the side of conservative estimates (larger variances). Unweighted methods do not always produce larger estimates of observed variance, credibility intervals, and confidence intervals than the sample-weighted method when large sample outliers are present. Commonly used weighting techniques listed by the Cochrane Collaboration include the following:

Dichotomous data fixed effects: Mantel–Haenszel methods (for RR, OR, RD); Peto method (OR)

Dichotomous data random effects: DerSimonain and Laird method (RR, OR, RD)

Continuous data fixed-effect inverse variance model: Weighted mean difference, standardized mean difference

Continuous data random-effects assumption: DerSimonian and Laird methods for weighted mean difference, standardized mean difference

Generic data: Fixed-effect inverse variance; random-effects inverse variance (inverse variance weighting methods weight studies based on the inverse of their variances)

To select a method appropriate to the study being conducted, the Cochrane Collaboration suggests that a decision about which within-trial statistic to use (i.e., OR, RR, or RD) is the primary event and only then should the method of combining the trial data, or meta-analysis, be chosen. Overall, the *Mantel–Haenszel methods* have been shown to be more reliable when there are not many data; the *Peto method* performs well with sparse data and is then the best choice, whereas a *random-effects model* may be better when there is statistical heterogeneity between the studies being reviewed (Cochrane Collaboration, 2002).

Systematic reviews and meta-analyses are important pillars for evidence-based dentistry (EBD) that supplies guidelines to help the clinician make intelligent decisions. By itself, EBD may not give definitive answers. "It does not exchange the tyranny of the expert for the tyranny of the literature" (Goldstein, 2002). It relies first on clinical expertise so critical in dentistry, where the numbers of randomized, controlled clinical trials and prospective cohort studies are limited. Several software products are available for conducting systematic reviews and meta-analysis (both commercial and freeware). The Meta-Analysis Unit of the Faculty of Psychology at the University of Murcia, Spain (http://www.um.es/facpsi/metaanalysis/software.php#3), maintains a website listing commercial and freeware software programs for meta-analysis and reviews of these programs. The Cochrane Collaboration also offers its program RevMan as a freeware for systematic reviews and meta-analysis (Cochrane Collaboration, 2002).

5 Error and Bias

The two central goals in epidemiological studies are obtaining a precise estimate of the true value of the population parameter and establishing causal association between exposures and outcomes, both of which may be threatened by different errors that may arise in almost any part or stage of epidemiological studies (see Figure 5.1). These errors may arise from random variations, biological variations, errors in measurement, errors in study design and planning, and our lack of perfect knowledge about the phenomenon under study that could contribute to the various types of errors. These errors may occur roughly equally between exposed and unexposed (or case/control) groups (*non-differential errors*), or they may occur differently between comparison groups (*differential error*). Information about exposure or disease status is often obtained from participants, who may not be able to relate the information accurately; when measured by instruments, the measurement may be inaccurate leading to erroneous ascertainment. For example, most diagnostic tests have some false–positive and false–negative outcomes that can lead to misclassification of cases as noncases or vice versa. Again, some instruments may consistently give an erroneous value that can be predicted; for example, a sphygmomanometer that consistently reports blood pressure to be 5 mmHg higher than the correct value. Such predictable errors may be amenable to some adjustment in analysis if their correct magnitude is known. Alternately, the same instrument may sometimes record higher blood pressure, and at other times record blood pressure lower than the correct value. Such inconsistencies may not be predictable and therefore it will be difficult to find a method to adjust for such errors. Errors in epidemiological studies are usually classified as unpredictable (random errors) and predictable or systematic (bias).

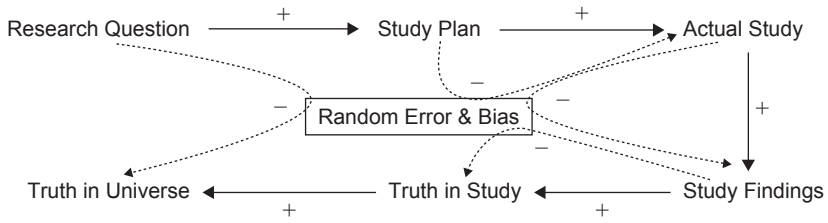


FIGURE 5.1 Random Error and Bias Errors and biases may impact any part of a study. In every step of study design, analysis, and interpretation (solid arrows), there is scope for errors and biases (dotted arrows) resulting from investigator bias, random error, and various study biases. A research question may be set up in such a way that it does not directly answer the intended question and as a result of various errors and biases it obfuscates the truth. Maximizing strengths (+) and minimizing threats (–) at different stages impact making correct inferences. In this trade-off, we try to minimize errors so that they are not large enough to change the conclusions in important ways. Only if the study is able to arrive at a correct inference can the truth about the issue be unearthed.

Random Error

Essentially, errors may be inaccuracies that occur as random variation or due to chance (*random error*). Randomness is defined as an absence of order, purpose, or predictability. Although random processes can be described with probabilistic distributions, outcomes of a repeated random process do not follow any describable deterministic pattern. In common parlance, the term *chance* is often used to describe such a process. The essential quality of randomness is our inability to predict the outcomes of such processes. Random errors can be attributed, at least partly, to lack of knowledge about a phenomenon and the resultant inability to explain the phenomenon completely. In epidemiology, most random error is generated from sample selection (*sampling error*); the overall goal of studies is to minimize random errors by a sampling method that would tend to equalize (at least in theory) the distribution of unknown factors in cases and controls or among the exposed and unexposed groups. At its core, the practice of epidemiology involves a *stochastic process*. "Stochastic" means randomness. In a stochastic process, the behavior of the process or phenomenon is basically nondeterministic because every stage of the phenomenon is determined partly by predictable events and partly by random events. Random differences between samples and the populations it represents, or immeasurable fluxes within populations and/or samples (non-random differences), may compromise complete predictability of future events and states.

Random error can also be perceived as the inverse of precision. Random error in a study may be corrected by improving sampling and study design,

increasing sample size, reducing measurement variability in instruments, or by using strict measurement criteria (e.g., use of regularly calibrated accurate instruments, or using averages of multiple measurement). During study implementation, it may be minimized using stringent quality control measures. Use of appropriate and more efficient analytical methods can also help minimize random errors.

Bias

In contrast to random errors, *biases* are errors that are systematic deviations from the truth. Therefore, essentially, biases are predictable. Although relatively easy to intellectualize, estimation and quantification of biases can be challenging. Bias can be described as inaccuracies that occur in one group and not the other. For example, unpredictable error in sampling would be random error, but selectively choosing a sample of a certain type may lead to a biased sample. Biases threaten the validity of the study. There are three important, somewhat related bias "effects" that may impact epidemiological studies. Careful attention must be paid to avoid these sources of errors.

1. **The Pygmalion Effect:** This "self-fulfilling prophecy," also called the Rosenthal effect, occurs when some participants perform much better than others in certain situations just because they were *expected* to perform better in those situations (Rosenthal & Jacobson, 1992). The effect is named after George Bernard Shaw's play *Pygmalion* in which Eliza Doolittle, Professor Higgin's trainee, makes the comment that the difference between a flower girl and a lady was not their behaviors, but the way they were treated; that is, the manner of treatment of the object (expectation of certain behavior) led to the observed behavior pattern.
2. **The Placebo Effect:** A placebo is a noneffective drug. Some patients may respond to a placebo even if no active drug is given to them, perhaps because the patient believes that the given drug works. The placebo effect is discussed in greater detail in Chapter 11.
3. **The Hawthorne Effect:** This effect implies that individuals will change their behavior if they are aware they are being observed. Conceptually, this is similar to Heisenberg's uncertainty principle in quantum physics, which states that values of certain pairs of conjugate variables (e.g., position and momentum of quantum particles) cannot both be known with arbitrary precision. This situation arises because the particle (e.g., an electron) being measured is smaller than the wavelength of light used to locate it. On contact with a light photon, the momentum of the particle changes instantaneously, so even if it can be "seen," its momentum cannot be gauged due to the change induced by the act of observation with a light photon. This phenomenon is called "observer effect" in physics and it refers to changes that the mere act of observation induces on the phenomenon being observed. However, the Hawthorne effect is

product of human behavior rather than probabilistic uncertainty or randomness.

In general, biases may be minimized through improved design, better accuracy, and strict quality control procedures. Bias cannot be corrected by sample size adjustment. An epidemiologic study should try to balance the threats and strengths within feasible limits while recognizing that "perfect" studies are utopias. In this trade-off, one tries to minimize errors so that they are not large enough to change the conclusions in important or meaningful ways. Box 5.1 lists types of biases and provides their working definitions. In this chapter we discuss some of the important types of biases and how they affect study results, and we discuss confounding bias in Chapter 6.

BOX 5.1 Commonly Reported Biases Occurring in Epidemiological Studies

Bias Category	*Bias Type*	*Description*
Selection bias	Berkson's bias	Occurs due to differential rates of hospital admission for cases and controls. Patients with risk factors and disease are admitted to hospitals more frequently than patients without risk factors or controls with risk factors. Similarly, studies based in ambulatory settings may miss more severe cases who are usually admitted to hospitals. Therefore, hospital-based studies may be biased due to absence of a control group representing the population. It can be controlled by using careful control selection criteria in hospital-based studies.
	Loss to follow-up bias	Occurs especially in cohort studies. Those missing follow-up appointments may be sicker, or have greater exposure to one or more risk factors compared to those who attend regularly. For example, in examining oral and systemic disease linkages in ambulatory settings, those who develop serious conditions may be admitted to the hospital and may not report for their appointments.
	Nonresponse bias	Occurs when study participants with certain characteristics selectively do not participate in the study. The bias may be directly proportional to response rate and can be reduced by increasing the response rate.

BOX 5.1 Commonly Reported Biases Occurring in Epidemiological Studies *(Continued)*

Bias Category	*Bias Type*	*Description*
	Membership bias	Membership bias involves those people who choose to be members of a group with shared attributes (e.g., gym users) and they might differ from others in important ways. For example, such people are more health conscious, have better oral hygiene habits, and eat a less sugary diet.
	Procedure selection bias	Occurs when treatment assignments are made in a way that result in dissimilar treatment groups. For example, certain drugs may be given only to healthier patients or certain types of surgeries may be conducted on healthier patients. This can be addressed by paying careful attention to inclusion and exclusion criteria between comparison groups to ascertain their equivalence and randomization.
Information bias	Recall bias	Occurs due to inaccurate recall of past exposures or disease status details. When people try to recall remote exposure histories, cases may search their memories more deeply and intently compared to controls leads to better recall among cases.
	Interviewer bias	Occurs when interviewers are not blinded to case status of the participant, and try to seek replies more deeply among cases by "clarifying" questions or seeking clearer explanations from cases as compared to controls. Blinding of interviewers to case status (to the best extent possible) can minimize interviewer bias.
	Observer bias	Similar to interviewer bias, occurs especially in cohort studies when the observers may seek outcomes more deeply among the exposed compared to the unexposed group.
	Respondent bias	Occurs when the outcome information is dependent upon participants' assessment. For example, ascertaining diagnoses of pain, or occurrences of periodic information, ulcers,

(Continues)

BOX 5.1 Commonly Reported Biases Occurring in Epidemiological Studies *(Continued)*

Bias Category	*Bias Type*	*Description*
		and so on may be dependent on participants' observed report and may be subject to inaccuracies compared to an examiner-defined clinical outcome.
	Family history bias	Those affected with a disease in a family may have more information about disease information in a family compared to those not affected; medical information flows differentially in the family depending upon medical status of the person.
Combined selection/information bias	Surveillance/detection bias	Occurs when a presumably related exposure leads to closer surveillance and increased case detection, or when an unrelated exposure causes symptoms leading to the unmasking of an otherwise asymptomatic disease that may or may not be subclinical.
	Incidence-prevalence/Neyman's bias	Occurs when prevalent cases are included in the study as incident cases. This occurs mostly in cross-sectional studies without clear case-definition and inclusion criteria. Inclusion of prevalent cases will always increase the reported incidence proportion/rate. Disease risk estimates will be inaccurate.
	Duration ratio bias/survival bias	If the duration of disease (or prognosis) is associated with exposure; that is, it differs between exposed and unexposed groups, then the prevalence rate ratio (measure of association used) will vary between the two groups independent of the exposure per se.
	Point prevalence complement ratio bias	In general, point prevalence rate ratio underestimates the strength of association between exposure and outcome. Prevalence rate ratio varies depending upon prevalence rates. This bias is more common in high-prevalence diseases, although it may also be serious for low-prevalence diseases.

BOX 5.1 Commonly Reported Biases Occurring in Epidemiological Studies *(Continued)*

Bias Category	*Bias Type*	*Description*
	Temporal bias	Occurs when conclusions are based on erroneous temporal sequences of cause and effect. This occurs mostly in cross-sectional and case-control studies when it is difficult to ascertain the temporal sequence of cause and effect.
	Length bias	Occurs in diseases that have a long detectable preclinical phase. Active screening may detect more cases in earlier preclinical phases compared to nonscreened diagnosed cases (which might be at later phases). Effectiveness of screening programs should consider the potential effect of length bias.
	Lead time bias	Lead time is the time by which diagnosis can be advanced by early detection and screening compared to the usual time when diagnosis is made. Early detection and treatment may impose a false sense of increased survival rate, when in fact the survival time increase would be mainly due to the extra time accrued through early diagnosis and not a better posttreatment success.
	Compliance bias	Occurs due to differential compliance to treatment; for example, taking a once a day dose vs multiple dosages a day.
	Insensitive measure bias	This is a measurement error occurring due to a miscalibrated instrument that gives biased readings.
	Procedure bias	Occurs due to nonstandardized procedural practices in the study when participants in one group receive more or less attention compared to another group.

Several other types of biases have been mentioned in the literature, such as one-sided reference bias, wrong sample-size bias, hot stuff bias, data-dredging bias, unacceptable disease bias, referral bias, volunteer bias, withdrawal bias, attention bias, therapeutic personality bias, gold standard review bias, index test review bias, verification bias, author bias, conflict of interest bias, and publication bias.

Selection Bias Example

Biases can impact study results substantially and alter inferences in important ways. Tables 5.1–5.4 provide examples of how selection bias may alter study results. In this population (see Table 5.1), the true disease-exposure association measured by the exposure odds ratio is 2.0 (i.e., cases were twice as likely as controls to have been exposed). However, even if sample size is reduced substantially and differentially between cases and controls (i.e., reduced by 99% in cases and 99.97% in controls; see Table 5.2), the exposure odds ratio does not change (OR: 2.0; 95% CI: 1.6, 2.5). Thus there is no bias introduced in the study by reducing the sample size as long as the proportions of exposed and unexposed do not change. Note that when we describe the true population parameter value, we do not need confidence intervals. However, when we take a sample from the population, we need to place confidence intervals around the parameter estimate because it is just an estimate of the true parameter value in the population, which we do not know in real-world situations.

Compared to the situation above, if a sample exhibits selection bias differentially by exposure status between cases and controls, then the scenario would be different. For example, a sample selection bias, unknown or unexamined by the investigator, could lead to selection of fewer unexposed cases compared to exposed cases (80% of exposed cases and 50% of unexposed cases selected; see Table 5.3). This biased sample results in an exposure odds

TABLE 5.1 True Status of a Hypothetical Population

	Cases	*Controls*
Exposure present	30,000	4,500,000
Exposure absent	15,000	4,500,000
Total	45,000	9,000,000
Odds of exposure	30,000/15,000 = 2.0	4,500,000/4,500,000 = 1
Exposure odds ratio	[(30,000/15,000)/(4,500,000/4,500,000)] = 2/1 = 2.0	

TABLE 5.2 Hypothetical Case-Control Study: Sample From Population in Table 5.1

	Cases	*Controls*
Exposure present	$(0.01 \times 30{,}000) = 300$	$(0.33 \times 4{,}500{,}000) = 1500$
Exposure absent	$(0.01 \times 15{,}000) = 150$	$(0.33 \times 4{,}500{,}000) = 1500$
Total	450	3000
Odds of exposure	300/150 = 2	1500/1500 = 1
Exposure odds ratio	[(300/150)/(1500/1500)] = 2.0; 95% CI: 1.6, 2.5	

The sample size is small compared to the population (cases: 1%; controls: 0.033% of population) but proportional distribution of disease and exposure is same as the population.

TABLE 5.3 Hypothetical Case-Control Study: Biased Sample From Population in Table 5.1

	Cases	*Controls*
Exposure present	$(0.8 \times 300) = 240$	1500
Exposure absent	$(0.5 \times 150) = 75$	1500
Total	315	3000
Odds of exposure	240/75 = 3.2	1500/1500 = 1
Exposure odds ratio	[(240/75)/(1500/1500)] = 3.2; 95% CI: 2.4, 4.2	

Eighty percent of exposed cases and 50% of unexposed cases selected from Table 5.1; controls same as in Table 5.2.

ratio of 3.2 (2.4, 4.2); a much stronger relationship than the true association resulting from selection bias in the study. Note that the exposure odds among the cases (3.2) are inaccurate (C/F true odds of exposure in cases: 2.0; see Table 5.1). Therefore, even if the exposure status among controls represents the true population value, a biased sample among cases alone will lead to a biased risk estimate.

Until now we have selected a biased sample by changing the exposure status among cases only. However, the selection bias may also affect the controls. Table 5.4 shows another possible scenario where the exposure status of the selected sample may be biased not only among cases, but also among controls. This example demonstrates that the exposure status among cases and controls differs from the true parameters equally (i.e., 80% of exposed cases and controls, and 50% of unexposed cases and controls selected). Note that although the exposure odds among cases and controls are inaccurate due to bias (3.2, 1.6, respectively), the exposure odds ratio remains the same as the true value and previous estimate (OR: 2.0; 1.5, 2.6). The exposure odds ratio remains the same because the magnitude of bias in cases and controls cancel each other completely. Therefore, if the bias is present equally in cases and controls, the overall risk estimate will still be a

TABLE 5.4 Hypothetical Case-Control Study: Biased Sample From Population in Table 5.1

	Cases	*Controls*
Exposure present	$(0.8 \times 300) = 240$	$(0.8 \times 1500) = 1200$
Exposure absent	$(0.5 \times 150) = 75$	$(0.5 \times 1500) = 750$
Total	315	1950
Odds of exposure	240/75 = 3.2	1200/750 = 1.6
Exposure odds ratio	[(240/75)/(1200/750)] = 2.0; 95% CI: 1.5, 2.6	

Eighty percent of exposed cases and controls, and 50% of unexposed cases and controls selected from Table 5.1.

correct estimate of the true population parameter, but individual exposure odds within cases and controls will be inaccurate, thus leading to erroneous inferences related to exposure odds among cases and controls.

Exposure Misclassification Example

Tables 5.5–5.8 show examples of the effect of *misclassification of exposure* status on the risk estimates. In these examples, we will keep the sample sizes fixed (i.e., overall sample size as well as the numbers of cases and controls in the study). This hypothetical study has 280 subjects (120 cases and 160 controls). The calculated exposure odds ratio is 2.0 (1.2, 3.3) as shown in Table 5.5. However, if we consider that 20% of the exposed cases and controls were misclassified (*non-differential* or *random misclassification*), then the new contingency table would show different numbers (see Table 5.6). The exposure odds ratio now turns out to be 1.7 (1.1, 2.8). Therefore, nondifferential misclassification leads to a biased risk estimate that is attenuated and is directed toward the null (1.7 vs 2.0). Similarly, we can change the misclassification differentially in cases and controls (*differential* or *non-random misclassification*). Differential misclassification can change the risk estimate either toward the null or away from the null depending upon the specific situation. Tables 5.7 and 5.8 show two examples of differential misclassification and their different effects on the risk estimate directing them away from the null (2.3 vs 2.0; see Table 5.7) and toward the null (1.4 vs 2.0; see Table 5.8).

TABLE 5.5 Hypothetical Case-Control Study

	Cases	*Controls*
Exposure present	80	80
Exposure absent	40	80
Total	120	160
Odds of exposure	80/40 = 2	80/80 = 1
Exposure odds ratio	[(80/40)/(80/80)] = 2.0; 95% CI: 1.2, 3.3	

TABLE 5.6 Hypothetical Case-Control Study Demonstrating Nondifferential Exposure Misclassification

	Cases	*Controls*
Exposure present	80 – 16 = 64	80 – 16 = 64
Exposure absent	40 + 16 = 56	80 + 16 = 96
Total	120	160
Odds of exposure	64/56 = 1.14	64/96 = 0.67
Exposure odds ratio	[(64/56)/(64/96)] = 1.7; 95% CI: 1.1, 2.8	

Study uses same sample as in Table 5.5 (assuming that there is a 20% exposure misclassification in exposed cases and controls).

TABLE 5.7 Hypothetical Case-Control Study Demonstrating Differential Exposure Misclassification

	Cases	*Controls*
Exposure present	80 – 8 =72	80 – 16 = 64
Exposure absent	40 + 8 = 48	80 + 16 = 96
Total	120	160
Odds of exposure	72/48 = 1.5	64/96 = 0.7
Exposure odds ratio	[(72/48)/(64/96)] = 2.3; 95% CI: 1.4, 3.6	

Study uses sample as in Table 5.5 and assumes a 10% exposure misclassification in cases and 20% in controls. The odds ratio is directed away from the null (C/F 2.0).

TABLE 5.8 Hypothetical Case-Control Study Demonstrating Differential Exposure Misclassification

	Cases	*Controls*
Exposure present	80 – 16 = 64	80 – 8 = 72
Exposure absent	40 + 16 = 56	80 + 8 = 88
Total	120	160
Odds of exposure	64/56 = 1.14	72/88 = 0.82
Exposure odds ratio	[(64/56)/(72/48)] = 1.4; 95% CI: 0.9, 2.2	

Study uses sample as in Table 5.5 and assumes that there is a 20% exposure misclassification in cases and 10% in controls. The odds ratio is directed toward the null (C/F 2.0).

Misclassification of Disease Status Example

Tables 5.9–5.14 show examples of the effect of *misclassification of disease* status on the risk estimates. In these examples, a hypothetical cohort study of 1000 exposed and 1000 unexposed persons are followed for 5 years. The true outcome is that in the population, 100 exposed and 50 unexposed develop the disease of interest, and the true risk ratio is 2.1 (1.5, 3.0; see Table 5.9). But, we do not know the real truth because we cannot use a test that is 100% sensitive and 100% specific. However, we have a very good diagnostic test (sensitivity = 90%; specificity = 90%). Therefore, we will be able to identify 90 of the 100 disease cases among the exposed group and 45 of the 50 disease cases among the unexposed group. Similarly, we will be able to correctly identify 810 of the 900 disease negative cases among the exposed and 855 of the 950 among the unexposed cases (see Table 5.10). Combining these observed results, we create a contingency table (see Table 5.11) and calculate the risk ratio as 1.3 (1.1, 1.7). In this example, in reality (see Table 5.9), only 150 persons had disease, but the observed number of "cases" in the study was 320 (see Table 5.12), which is more than double the true number of diseased persons. Such an outcome is a function of the properties

TABLE 5.9 Hypothetical Cohort Study of 2000 Persons Demonstrating Disease Status Misclassification

	Disease Present (D+)	*No Disease (D–)*
Exposure present	100	900
Exposure absent	50	950
Total	150	1850
Disease risk ratio	[(100/900)/(50/950)] = 2.1; 95% CI: 1.5, 3.0	

True association in the study is shown above.

TABLE 5.10 Hypothetical Cohort Study Demonstrating Disease Status Misclassification

	Exposed			*Unexposed*		
	D+	*D–*	*Totals*	*D+*	*D–*	*Totals*
Test +	Sn0.9 × 100 = 90	90	180	Sn0.9 × 50 = 45	95	140
Test –	10	Sp0.9 × 900 = 810	820	5	Sp0.9 × 950 = 855	860
Total	100	900	1000	50	950	1000

Study uses sample from Table 5.9. The diagnostic test used has sensitivity = 90% and specificity = 90%.

TABLE 5.11 Hypothetical Cohort Study Demonstrating Disease Status Misclassification.

	D+	*D–*
Exposure present	90 + 90 = 180	10 + 810 = 820
Exposure absent	45 + 95 = 140	5 + 855 = 860
Total	320	1680
Disease risk ratio	[(180/820)/(140/860)] = 1.3; 95% CI: 1.1, 1.7	

Observed association detected in the study association from Table 5.10. The risk ratio is directed toward the null (C/F 2.1).

of the diagnostic test (i.e., number of false positives) and relatively low prevalence of the disease (disease prevalence was 150/2000 = 7.5%).

The investigators also noted that the observed risk ratio in the study (1.3) is an underestimate of the true risk ratio (2.1)—the risk estimate in the study is biased toward the null. Another important point to note is that the test used had very high sensitivity (90%) and specificity (90%), which is not easy to find in a real-world scene for oral diseases, and yet the bias is substantial. However, what will happen if we use a somewhat less sensitive

TABLE 5.12 Hypothetical Cohort Study Demonstrating Disease Status Misclassification

	Exposed			*Unexposed*		
	D+	*D–*	*Totals*	*D+*	*D–*	*Totals*
Test +	Sn0.8 × 100 = 80	180	260	Sn0.8 × 50 = 40	190	230
Test –	20	Sp0.8 × 900 = 720	740	10	Sp0.8 × 950 = 760	770
Total	100	900	1000	50	950	1000

Study uses sample from Table 5.9. The diagnostic test used has sensitivity = 80% and specificity = 80%.

TABLE 5.13 Hypothetical Cohort Study Demonstrating Disease Status Misclassification

	D+	*D–*
Exposure present	80 + 180 = 260	20 + 720 = 740
Exposure absent	40 + 190 = 230	10 + 760 = 770
Total	490	1610
Disease risk ratio	[(260/740)/(230/770)] = 1.18; 95% CI: 0.9, 1.4	

Observed association detected in the study association from Table 5.12. The risk ratio is directed further toward the null (C/F 2.1 and 1.3).

and specific diagnostic test (e.g., 80% sensitivity and 80% specificity)? Then the scenario changes drastically. Table 5.12 and Table 5.13 show the results using such a test. The risk ratio drops to 1.18 (0.9, 1.4)!

Risk estimates for dichotomously classified diseases mask the true association, and slight changes in sensitivity and specificity of diagnostic tests can make substantial differences leading to seriously erroneous inferences. Even if diagnostic tests with high sensitivity and specificity are used for case ascertainment, bias due to misclassification of disease is substantial. Even if the sensitivity is high, the risk ratio will be affected by specificity, especially if the disease is rare. Because the probability of getting false–positive cases is higher, tests with higher specificity will perform better. If disease misclassification is differential in the exposed and the unexposed group, then the effect on the risk estimates is more complex (see Table 5.14). In this example, the test is assumed to have a 90% sensitivity and specificity among exposed and 80% sensitivity and specificity among the unexposed, and the risk ratio is 0.73; 95% CI: 0.6, 0.9, thereby also changing the direction of the exposure–disease association, potentially leading to a protective effect inference!

TABLE 5.14 Hypothetical Cohort Study Demonstrating Differential Disease Status Misclassification Using Sample From Table 5.9

	Exposed			*Unexposed*		
	D+	*D–*	*Totals*	*D+*	*D–*	*Totals*
Test +	Sn0.9 × 100 = 90	90	180	Sn0.8 × 50 = 40	190	230
Test –	10	Sp0.9 × 900 = 810	820	10	Sp0.8 × 950 = 760	770
Total	100	900	1000	50	950	1000
Disease risk ratio	[(180/820)/(230/770)] = 0.73; 95% CI: 0.6, 0.9					

The diagnostic test has sensitivity = 90% and specificity = 90% among the exposed, and sensitivity = 80% and specificity = 80% among the unexposed. The risk ratio crosses the null (C/F 2.1, 1.3, and 1.18).

Control of Bias

Bias is mostly controlled at the design stage before the study starts and/or at the implementation stage of the study. Bias in some hospital-based case-control studies can be minimized by selecting controls from among hospital patients who may minimize recall bias, Berkson's bias, and loss to follow-up. Using incident cases only prevents incident-prevalence and related biases. Cases and controls can be randomly selected from sampling frame. Loss to follow-up may be minimized by multiple attempts to reach the participant and tracing persons using their recorded identification.

Measurement bias can be minimized by using various techniques such as valid data collection, disease definition, and information collection tools; blinding of data-analyst, interviewer, and clinical investigator to the case/exposure status of the participant; validation of case definition, diagnosis, and exposure data; minimizing time between exposure/event and data collection so that participants do not have to recall events far off in time; using ancillary questions to check responses to important questions that may be subject to recall bias; verifying answers to questions with objective measures or clinical records; measuring exposure and outcomes at the same level to avoid ecological fallacy; using standardized calibration methods to improve within- and inter-examiner agreement; setting up periodic procedure reviews; and monitoring procedures and study protocols.

Bias analysis is a complex topic and is explained in detail in more specialized texts (Rothman, Greenland, & Lash, 2008) to which one may refer. The goal here is to emphasize importance of recognition of the fact that biases may be quantified and should be at least addressed in discussing any study result.

Analytical Control for Selection Bias

Assessment of bias is a necessary and important part of any epidemiologic study. Several recent reports have discussed bias assessment in oral health-related studies (Beck, Caplan, Preisser, & Moss, 2006; Fenwick, Needleman, & Moles, 2008; Kingman, Susin, & Albandar, 2008; Lee, Rozier, Norton, & Vann Jr., 2005; Shelton, Gilbert, Lu, Bradshaw, Chavers, et al., 2003). Accounting for selection bias makes a major difference in the substantive conclusions about the outcomes of a study, and all studies should address biases and report steps taken to recognize, control for, and minimize biases in them.

The best control for bias is to anticipate potential for biases and minimize the sources of bias through a strict and tight study design. Once the sample is selected and information is recorded, the existing biases become a part of the data. Randomized trials have been generally viewed as the "gold standard" of evidence presumably because of their experimental design and randomization that ensures that the distribution of exposure is not biased so that discrepancies could be attributed to random error. Randomization minimizes selection bias but it does not ensure that the randomized groups are equal in all respects. When randomization is not possible, then nonrandom assignment must be adjusted for in statistical analyses.

A recent meta-analysis examined the potential effect of bias from improper methods of allocation concealment and examiner-masking affect on the magnitude of clinical outcomes in periodontal trials (Fenwick et al., 2008). The investigators concluded that there is insufficient evidence to support or refute the theory that the bias from improper methods of allocation concealment and examiner masking impact the magnitude of clinical outcomes in periodontal trials. This study was based on 35 randomized control trials (RCTs) and did not have enough power to detect the differences sought. The authors reported a retrospective power calculation that indicated that they would have needed 265 RCTs to demonstrate a statistically significant effect for the impact of bias on their outcome measure. Apparently, there seems to be a perception that RCTs are somehow free of selection bias, a view presumably emanating from the generally held view (perhaps misplaced and practically unsustainable) that RCTs are the "gold standard" of evidence, which are supposedly correct every time. Clinical trials, however, suffer from selection bias resulting from inappropriate control selection, stage migration, inappropriate inclusion/exclusion criteria, use of multiple subset analyses, investigator bias, and other biases (Longford, 1999; Miller, Rahman, & Sledge Jr., 2001). Randomization and control of exposure by the investigator are the chief advantages in RCTs. However, they are increasingly difficult and expensive to conduct.

Observational studies may be viewed as viable alternatives to randomized trials although they are subject to greater selection bias. Therefore, attempts have been made to demonstrate, assess, and find methods to

minimize and adjust for selection bias in observational studies. Application of model-based selection bias adjustment has been limited in cross-sectional observational data and to continuous outcomes.

Generally, selection biases cannot be overcome using statistical analysis of existing biased data alone. However, the degree of selection bias can be assessed measuring correlation between covariates. One way to make analytical adjustment is to use propensity scores (Leslie & Thiebaud, 2007). Propensity score is the probability that a subject will be assigned to a group under a set of conditions (which could be a set of covariates in the study that may be considered to link to selection bias). To reduce selection bias, groups can be made equal for these selected covariates. However, propensity score can control only for selection on observable characteristics because these characteristics are used for equating the groups. Therefore, adjustment for unmeasured or unobserved factors cannot be achieved using propensity scores which are discussed in greater detail in Chapter 18.

One of the major problems in adjusting for selection bias in statistical analysis is *endogeneity*. Parameters or variables are said to be endogenous if they can be predicted by other variables in the model and may sometimes also result in multicollinearity. In a regression model, endogeneity may also occur when the independent variable is correlated with the error term. Endogeneity results in biased regression coefficients. Regression methods have been proposed to adjust for selection bias. Such models use observed variable values as independent variables (Heckman, 1979). For example, in most recurrent diseases, past disease is considered to be a strong predictor for future disease. Selecting more people with past disease as cases into a study will lead to a selection bias. If the other covariates of both past and future disease are the same or similar, then including past disease as a covariate to predict future disease will lead to endogeneity, because other covariates will be strong predictors of past disease, which is included in the model as a covariate.

A two-stage method uses other available variable(s) called *instrumental variables* that may be used to estimate relationships. An instrumental variable is itself not a variable in the model equation but is correlated with the endogenous variables, conditioned upon other covariates, and cannot be correlated with the error term in the explanatory equation. Just like any other statistical procedure, the credibility of the estimates depends on the selection of an appropriate instrumental variable. Overall, two-stage methods involve a first stage of predicting the main variable using covariates and instrumental variables, and the predicted values are preserved. In the second stage, the actual regression of interest is estimated for the desired outcome variable using desired covariates and the predicted values from the first stage. A working example of the use of a two-stage method published recently has demonstrated its usefulness in controlling for selection bias in

cross-sectional studies assessing dental health services utilization in Women, Infants, and Children (WIC) program participants (Lee et al., 2005).

Shelton et al. (2003) applied these models to include longitudinal studies and binary outcomes. They applied a two-stage probit model using Generalize Estimating Equations (GEE) to account for correlated longitudinal binary chewing difficulty outcomes. They compared their results to results from standard GEE models that ignored the potential selection bias introduced by unobserved confounders. They reported that differences emanating from accounting for selection bias were substantial and were attributable in part to an "adverse selection phenomenon in which those most in need of treatment (and consequently most likely to benefit from it) are actually the ones least likely to seek treatment" (Shelton et al., 2003). Further extension of methods to adjust for selection bias can lead to improving validity of using observational studies as strong evidence equivalent to randomized trials.

Analytical Control for Measurement (Information) Bias

Applicable methods for controlling for information bias varies according to the type, discipline, detail, and nature of information sought. For example, in genetic association studies, population stratification may bias the effect estimates and inflate test statistics. A recent report examined the usefulness of an unlinked genetic single null marker in studies involving one candidate gene (Wang, Localio, & Rebbeck, 2005). The investigators reported that when the distribution of this marker "varied greatly across ethnicities, controlling" for the marker in "a logistic regression model substantially reduced biases on odds ratio estimates." Furthermore, when the marker had the same distributions as the gene across ethnic groups, "biases were further reduced or eliminated by subtracting the regression coefficient" of the marker from the coefficient of the gene in the model. Because the correction of population stratification related bias depended on the distribution of genes and markers, Wang et al. (2005) suggested that "marker choice and the specific treatment of that marker in analysis greatly influenced bias correction."

Bias for mean probing pocket depth or mean clinical attachment–loss estimates varies by site type, number of sites per tooth, and number of quadrants included in the partial recording protocols (Kingman et al., 2008). Estimation of probing depth and attachment level is a common measure of periodontal disease. However, obtaining an overall person-level estimate measuring periodontal sites of all teeth is a time-consuming and expensive method. Alternative and more resource-conservative methods use some form of sampling of teeth so that fewer sites are measured to increase efficiency of resources used. Broadly, these methods can be classified into: (1)

fixed site selection methods, and (2) random site selection methods. Using fixed sites is open to bias because the same sites may not be equally involved in all or most cases, and other sites not included in the fixed site methods may be involved in periodontal disease. Recently, Beck et al. (2006) reported assessment of bias in probing depth and attachment-level estimates using fixed site selection methods and compared those with estimates from random site selection methods using 84, 42, 36, 28, 20, 15, 10, and 6 randomly selected sites. They found that due to bias in fixed methods, probing depth and attachment loss were consistently underestimated. However, the random method selecting 36 sites was the least biased method. They concluded that although both fixed and random methods "underestimated prevalence, especially prevalence of less frequently occurring conditions," most random methods "were less likely to underestimate prevalence than" the fixed methods (Beck et al., 2006).

6 Confounding and Effect Measure Modification

Confounding

Confounding is an essential concept in epidemiology that is central to epidemiological analysis, interpretation, and inference making. For those who are uncompromisingly statistically-oriented and prefer to be guided by p-values alone, it might be a useful reminder that there exists no statistical test for confounding, just as most of life is not lived in black and white, but in the intervening shades of gray. Confounding is a kind of error, different from random errors, that has been described as a "mixing of effects." On the contrary, random errors can be described as inaccuracies that occur equally in either cases or controls, or in exposed or unexposed groups. Bias can be described as inaccuracies that occur in one comparison group and not the other, in a systematic fashion. *Confounding* is a distortion in an observed relationship between exposure and outcome that is brought about by a third factor (the confounding factor) that is associated with both the outcome of interest and the exposure. In other words, confounding is an alternative explanation for an observed relationship. For example, in a hypothetical study to explore the association between the occurrence of myocardial infarction among persons with severe periodontal disease and those without severe periodontal disease, rates of smoking are higher among those with severe periodontal disease (see Figure 6.1). Rates of myocardial infarction are also higher among persons who smoke. In this situation, smoking is a confounder of the relationship between severe periodontal disease and myocardial infarction. Confounding occurs because of the complex and multifactorial relationships between exposures and outcomes.

For a factor to be considered a confounder, it must be associated with both the exposure and the outcome under study, although the associations do not need to be causal in nature. The association of the confounder with

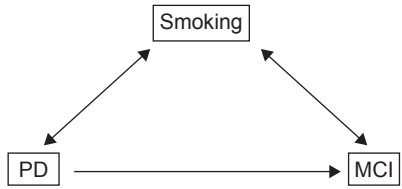


FIGURE 6.1 Confounding Smoking is associated with severe periodontal disease (PD) as well as myocardial infarction (MCI). If there is any observed relationship between PD and MCI, it would not be apparent if the association was being seen because both PD and MCI are associated with smoking. Therefore, smoking confounds the association between PD and MCI.

the disease outcome must be independent of that of the exposure under study, and it cannot be an intermediary factor in a causal chain between the exposure and outcome. For most factors, at the start of the study, it is difficult to know whether the factors under study are confounders. Therefore, for practical purposes, all known or suspected risk factors should be included for measurement and assessment of their confounder status.

Confounder selection is an important issue in modeling. Factors that are in the causal pathway (intermediate variables) between exposure and outcome should not be selected as confounders, and should not be matched. (Rothman, Greenland & Lash, 2008). Adjusting for any factor that is correlated with the outcome and is caused (even partly) by the exposure can bias the association between the exposure and the outcome (Weinberg, 1993). Therefore, some independent risk factors which appear to be confounders should be assessed carefully for their place in causal pathways before they are used as confounders. This point is illustrated with an example in the section under causal models in the modeling section of Chapter 8.

Detection of confounding is conducted through examination of the effect estimates. If the crude effect estimate of the association of an exposure with an outcome is modified substantially by inclusion of the purported confounding factor, then the factor should be considered as a confounder. In the presence of confounders, the real strength of association between exposure and outcome is masked and may lead to incorrect inferences. For example, in a study of oral candidiasis among HIV-1 positive persons, the crude odds ratio for association between low CD4 cell count (below 200 cells/μl) and oral candidiasis was reported to be 9.1 (CI: 3.6, 23.1). However, when adjusted for gender, antifungal drug use, and recreational drug use, the odds ratio (OR) was 6.8 (CI: 2.6, 17.9); that is, a reduction of 15% from the crude rate (Chattopadhyay, Caplan, Slade, Shugars, Tien, et al., 2005). It is not necessary for effect estimates to be attenuated as a result of confounding. For example, in the same study, the crude OR for oral candidiasis

among current recreational drug users was 2.2 (CI: 1.1, 4.1), which increased to 2.5 (1.3, 4.9), a 14% increase after the above-stated adjustments. Increase in the strength of association after adjustment is called positive confounding, whereas reduction in the strength of association after adjustment is called negative confounding.

As with all analysis, it must be recalled that whereas crude estimates are the real, naturally occurring observed estimates of the strength of association, adjusted estimates are artificial and functions of the types of model used, types of variables selected, and construction of those variable categories. The strength of confounding assessment lies in accurate inference-making to develop correct intervention strategies. Confounding can not only change the strength of association, but can also impact the direction of association in some cases. In multilevel variables, confounding may impact different levels differently in terms of strength as well as direction. For example, in the above study, adjustment resulted in a 14% increase of the OR for candidiasis among recreational drug users, but the OR among former users changed in the opposite direction and decreased slightly from 1.3 (CI: 0.8, 2.1) to 1.2 (CI: 0.7, 2.1).

Confounding can be controlled in the study design phase as well as in the analysis phase. In the study design phase, confounding can be controlled by restriction, matching, and randomization; whereas in the analysis phase, confounding can be controlled by restriction, stratification, and multivariate methods.

Causal Diagrams

Drawing line and graph diagrams to depict associations between factors is a commonly used tool to explore and describe these relationships. Description of interactions between factors is also often depicted through such visual medium. Visual depiction has been found to be especially useful in understanding the interplay of more complex interacting factors. Structural equation modeling (SEM) can determine path coefficients of different arms of the causal mechanisms, but require a linearity assumption as in linear regression analyses. SEM is a family of statistical analyses that includes path analysis and factor analysis. It assumes that the relationships between variables are linear. Other graphical techniques used for causal analysis do not require such assumptions. Among the various techniques, a formal graphical depiction for causal analysis is provided by drawing Directed Acyclic Graphs (DAG) (Pearl, 2000).

DAGs are logical visual tools where different variables are interconnected by directed arrows that are not allowed to have a cyclical path (i.e., acyclic) within the total graphical system and require us to set down clearly our assumptions about causal relationships. These graphs have algebraic

equations associated with them, although it is not always necessary to use those equations for understanding relationships. Figure 6.2 (A–F) illustrates the basic principles of working with DAGs which should include all known variables that may be associated with the exposure and outcome. These associations are represented by single-headed arrows. Between any three (or more) variables, the combination of arrows cannot be such that the arrowheads lead from one to another creating a complete uninterrupted cycle. Figure 6.2A shows the sequence Exposure–V1–V3 completes a full cycle between these three factors, violating principles of DAG construction. Figure 6.2B conforms to principles of DAG.

An important use of DAGs is in confounder selection; that is, determining the "net effect" of exposure. Merchant and Pitiphat (2002) illustrated the potential for use of DAGs in assessing confounding in dental research. Basic principles of using DAGs for confounding assessment in causal research were outlined by Greenland, Pearl, & Robins (1999). They suggested that a

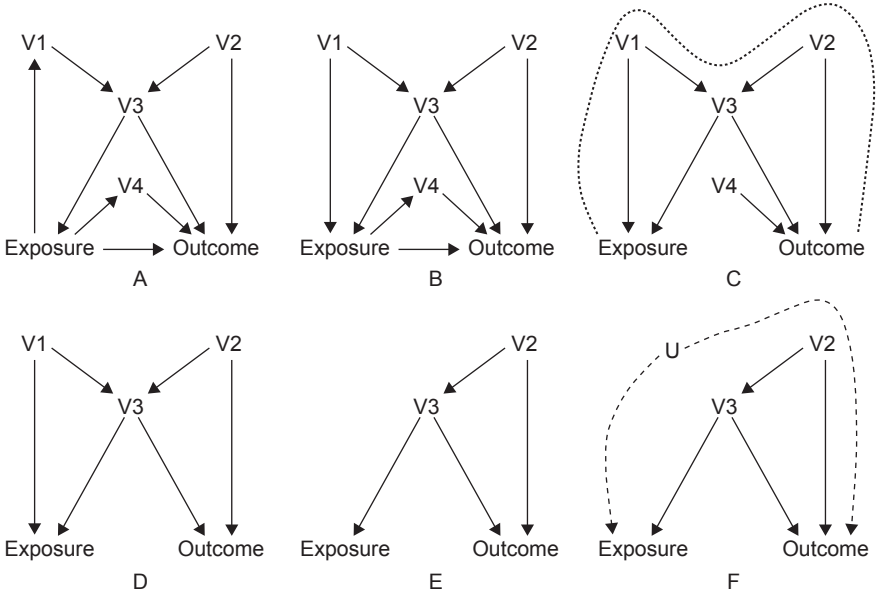


FIGURE 6.2 Directed Acyclic Graphs (DAGs) DAGs must not contain a closed look of directed arrows. DAG-A: Exposure–V1–V3 loop is a closed directed look; DAG-B Exposure–V1–V3 loop is not cyclic. DAG-C shows the left-over figure after removing all arrows originating from the exposure, and unblocked backdoor paths. DAG-D includes relevant variables for consideration. DAG-E shows one option for selecting variables for confounder-adjustment. DAG-F shows the potential effect of unmeasured confounders (U).

variable from which an arrow comes out is a *parent*, and the variable to which the arrow goes is a *child*. For example, in Figure 6.2B, V1 and V2 are parents, whereas V3 is a child of V1 and V2. In the sequence V1–V3–Exposure, V1 is an *ancestor* of exposure, and exposure is a *descendent* of V1 (in the direct sequence V1–Exposure, V1 is the parent and exposure is the child). A sequence of directed arrows with involved variables is called a *directed path* (for example, V2–V3–Outcome). After removing all the outgoing arrows from the exposure, any set of connected paths between the exposure and the outcome is a *backdoor path* (for example, Exposure–V3–Outcome). Any backdoor path that passes from a parent to child and then to another parent is a *blocked path*, and the common factor child between two parents is a *collider* or *blocker* (for example, Exposure–V1–V3–V2–Outcome, shown as a dotted line in Figure 6.2C, is a blocked path because V3 is a child of V1 and V2, and V3 serves as a collider; that is, the backdoor path is blocked at V3).

Confounders can be assessed by drawing an appropriate DAG (see Figure 6.2B), and then removing all single-headed arrows that originate from the exposure under study. This is equivalent to removing all effects of the exposure. In Figure 6.2C, one can check for existence of any unblocked path that leads from the exposure to the outcome (i.e., if exposure and outcome have a common ancestral relationship). This is equivalent to assessing the association between exposure and outcome if all exposure effects are removed. To find the backdoor path, arrow heads are ignored, and a connector line between exposure and outcome is explored. If no backdoor unblocked paths exist, then the net exposure effect is not confounded and there is no need to adjust for the variable. However, if unblocked backdoor paths exist, as in Figure 6.2D, then confounding is present and variable adjustment is required. Greenland, Pearl and Robins (1999) give the following rules for identifying a confounder using DAGs.

1. Define a set "S" consisting of variables in the DAG such that none of these variables are descendents of the exposure or the outcome.
2. Delete all arrows coming out of the exposure.
3. Link all pairs of variables that share a child or a descendent within the set "S" (i.e., descendents and child of variables other than the exposure and outcome are allowed).
4. Check for any unblocked backdoor path that does not pass through the set "S"—there should not be any.
5. Variable(s) in the set "S" should be sufficient for control of confounding.

In Figure 6.2D, the question arises that while V3 should be adjusted for, is it necessary to adjust for V1 and V2 (which are ancestors of V3)? Because V1 would be associated with the outcome through some strata of V3; and V2 would be associated with the exposure in some strata of V2, they would be confounders. Therefore, either one or the other should be adjusted for

(Greenland & Brumback, 2002). Depending on the nature of these variables and the mechanisms of action to produce effects on V2, exposure, and outcome, the variable for adjustment can be decided. For example, in Figure 6.2E, it is assumed that all the effect of V1 on the outcome in some strata of V3 acting through V2 is minimal, and the only important pathway is V1–Exposure–Outcome, making V1 a distal cause to exposure in the causal chain for the outcome.

Measurement error and improper adjustment for confounders can result in *residual confounding*. Knowledge is neither finite nor comprehensively fixed at any time-point because a completely deterministic future does not exist, and complete preservation of the past does not occur. Therefore, unmeasured confounders (known and unknown) may also complicate causal inference (see Figure 6.2F). However, improper adjustment of variables can induce extraneous confounding in an analysis. Bias can result even through adjustment for factors that are partially in the causal pathway between the exposure and outcome (Weinberg, 1993). In causal analysis, sound background knowledge of the subject matter is as important as the study design and quality data analysis procedures (Robins, 2001), especially when selecting confounders that may be adjusted in multivariable analysis. The case of a single variable may be relatively straightforward, but where multiple confounders are to be considered simultaneously, graphical methods used with the knowledge of background subject matter are very useful (Greenland, Pearl, & Robins, 1999). For example, Hernán, Hernández-Díaz, Werler, and Mitchell (2002) reported a case-control study on association of folic acid supplementation and neural tube defects (spins bifida) and demonstrated a selection bias if cases were restricted to live-births (because the neural tube closes about the 28th day of fertilization [Sadler, 2006]), and did not include still-births; the incidence and also the association of folic acid with prevention of neural tube defects is underestimated. Greenland & Brumback (2002) have demonstrated that causal graphs "can illustrate qualitative population assumptions and sources of bias not easily seen with other approaches; sufficient-component cause models can illustrate specific hypotheses about mechanisms of action; and potential outcome and SEM provide a basis for quantitative analysis of effects." They further suggested that these different approaches may be used in a complementary manner to improve causal interpretation of conventional statistical results.

Effect Measure Modification

Statistical interaction has been usually interpreted as evidence for differential biological effect in different levels of factors under study. However, the interactions observed in statistical models do not imply biological activity, but demonstrate differential effect estimates across different levels of factors

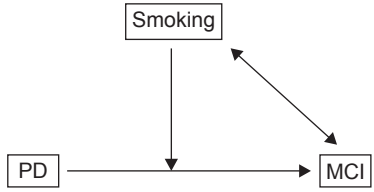


FIGURE 6.3 Effect Measure Modification Smoking modifies the association between PD and MCI.

involved. To distinguish the potential biological mechanism of action implications directly from reported statistical interactions, Rothman, Greenland, & Lash (2008) have suggested use of the term *effect measure modification* (EMM) to denote the impacts of statistical analyses, because the measures of effect begin to have different values across different levels of the concerned variables under study. Therefore, EMM implies a differential association (i.e., strength and/or direction) between exposure and outcome across the levels of a third variable (see Figure 6.3). EMM is also known as effect modification, interaction, statistical interaction, and heterogeneity of effects.

EMM can be demonstrated through stratified analysis. Tables 6.1 through 6.4 show the results from a hypothetical case-control study of oral cancer, assessing the role of tobacco consumption. The overall OR measuring the association of tobacco with oral cancer was 2.2. It was hypothesized that differential association effects between tobacco and oral cancer would be found for xerostomia, education level, and alcohol consumption status of the participants. Results showed no association between tobacco and oral cancer across the categories of xerostomia (OR 1.0; Table 6.2). There was a slight association between tobacco and oral cancer across education levels (OR: 1.2; Table 6.3). However, the OR was the same across both categories of participants' education level, showing no differential effects across categories. Thus, it was concluded that no EMM was demonstrable for xerosto-

TABLE 6.1 Overall Results From a Hypothetical Study to Assess Risk Factors of Oral Cancer

		Oral Cancer		
		Cases	*Controls*	*Total*
Tobacco	High	700	800	1500
	Low	2000	5000	7000
Total		2700	5800	8500

Assessing modification of effects by different factors of the relation between oral cancer and tobacco consumptions. Overall OR = 2.2.

TABLE 6.2 Hypothetical Oral Cancer: Tobacco Consumption Study, Stratified for Third Factor (Xerostomia)

		Oral Cancer		
		Yes	*No*	*Total*
Tobacco	High	800	500	1300
	Low	1500	950	2450
Total		2300	1450	3750

Third factor: xerostomia present. OR: 1.0.

		Oral Cancer		
		Yes	*No*	*Total*
Tobacco	High	100	100	200
	Low	3000	3050	6050
Total		3100	3150	6250

Third factor: xerostomia absent. OR: 1.0.

TABLE 6.3 Hypothetical Oral Cancer: Tobacco Consumption Study, Stratified for Third Factor (Education Level)

		Oral Cancer		
		Yes	*No*	*Total*
Tobacco	High	850	450	1300
	Low	1500	950	2450
Total		2350	1400	3750

Third factor: no education. OR: 1.2.

		Oral Cancer		
		Yes	*No*	*Total*
Tobacco	High	18	182	200
	Low	332	4218	4550
Total		350	4400	4750

Third factor: some education. OR: 1.2.

mia and participants' education level. However, tobacco consumption was associated with increased oral cancer risk among those who consumed alcohol (OR: 3.3) and those who did not (OR: 1.5) (see Table 6.4). The association between tobacco and oral cancer varied across the two levels of alcohol consumption factor, and it was concluded that EMM was present. The association between tobacco and oral cancer was modified by alcohol consumption

TABLE 6.4 Hypothetical Oral Cancer: Tobacco Consumption Study, Stratified for Third Factor (Alcohol Consumption)

		Oral Cancer		
		Yes	*No*	*Total*
Tobacco	High	400	300	700
	Low	1000	2500	3500
Total		1400	2800	4200

Third factor: alcohol consumption positive. OR: 3.3.

		Oral Cancer		
		Yes	*No*	*Total*
Tobacco	High	300	500	800
	Low	1000	2500	3500
Total		1300	3000	4300

Third factor: no alcohol. OR: 1.5.

in a way that the OR increased among alcohol consumers (overall from 2.2 to 3.3) and reduced for alcohol nonconsumers (overall from 2.2 to 1.5).

The term *effect measure modification* clearly suggests that the modification may be apparent in one *measurement scale* and not in another. EMM on ratio scale implies absence of EMM in additive scale and vice versa. Table 6.5 demonstrates this effect. If the rate ratio (multiplicative scale) is kept fixed, then the rate difference (additive scale) varies across levels of age groups; whereas if the rate difference is kept fixed, the rate ratio changes across the levels.

TABLE 6.5 Effect Measure Modifications and Scales of Measurement

	Rates of Disease (100,000 person–years)			
Age group (years)	Exposed	Unexposed	Rate ratio	Rate difference
Fixed Rate Ratio				
20–29	125	50	2.5	75
30–39	300	120	2.5	180
40–49	450	180	2.5	270
Fixed Rate Difference				
20–29	210	70	3	140
30–39	310	170	1.8	140
40–49	420	280	1.5	140

EMM is assessed through "interaction contrasts" demonstrated in a hypothetical example in Figure 6.4. The interaction table shows that compared to the doubly unexposed (RR_{00}—no tobacco, no alcohol exposure), risk increases for all other exposure groups (single exposures [RR_{01} : 5.0; R_{10} : 10.0] and the doubly exposed group [RR_{11} : 20.0]). Antagonism between exposure factors would occur if the risk for the doubly exposed group (RR_{11} : 20.0) was lower than any of the singly exposed groups or the sum of the risk of the two singly exposed groups. In this example, the risk for doubly exposed groups exceeds the sum of the risks of singly exposed groups, so synergy can be assumed. Koopman's interaction contrast ratio (ICR) is calculated as $ICR = RR_{11} - RR_{10} - RR_{01} + 1$. If the value of ICR is a product (or a higher or smaller fraction) of the risk of the two singly exposed groups, the EMM would be multiplicative (i.e., in this example, EMM would be multiplicative in a multiplicative scale). However, in this example, the ICR value 6 suggests that the EMM is additive (i.e., the overall sum is greater) in a multiplicative scale (RR is multiplicative scale statistics).

EMM can be tested in statistical models in several ways such as assessing whether the difference in risk estimates is "substantial"—using a comparison with R_{00}, or using ICR. In stratified analysis, testing for homogeneity of odds ratio may employ the Breslow Day statistic or the

Odds Ratio for Outcome: Oral Cancer

		Tobacco	
		Yes	No
Alcohol	Yes	20.0	5.0
	No	10.0	1.0

Risk:
R_{00} Doubly unexposed
R_{01} Singly exposed, factor 1
R_{10} Single exposed, factor 2
R_{11} Doubly exposed

Synergy vs. Antagonism
$R_{01} + R_{10}$ =; ; >; < R_{11}

Additive vs. Multiplicative EMM
$R_{01} + R_{10}$ vs. R_{11}; $R_{01} \times R_{10}$ vs. R_{11}

Koopman's Interaction Contrast Ratio (ICR):

$ICR = RR_{11} - RR_{10} - RR_{01} + 1$

$ICR = 20 - 10 - 5 + 1 = 6$

(Additive "interaction" on a multiplicative scale)

- Departure from additively: low EMM
- Departure from multiplicity: strong EMM

FIGURE 6.4 Interaction Contrasts A working example of interaction contrasts from a hypothetical study.

Woolf Chi-squared test statistic. Monte Carlo methods for sparse data situation have also been used. Model-based strategies for testing "interaction" include conducting a Chunk test (H_0: at least one pair has interaction) and testing for important factors declared a-priori.

If EMM is present, then the overall effect estimate is not interpretable, and stratum-specific effect estimates should be presented. Confounding assessment within stratum should then be conducted. *EMM and confounding do not preclude each other*. These are separate issues and both may be present in substantial degrees in a study. When EMM is present, different strata may exhibit different degrees of confounding. EMM and confounding are different and mutually independent concepts. There are four possibilities: (1) EMM present, confounding absent; (2) confounding present, EMM absent; (3) neither confounding nor EMM present; and (4) both confounding and EMM exist. (Note that confounding may be different in the different levels of the factor exhibiting EMM depending upon causal and other associations.)

EMM should not be regarded as a mere statistical measure or convenience. On the contrary, it should be considered as a causal framework for developing deeper understanding and gaining insights into causality and mechanisms of the phenomenon under study. The differences in associations depicted by EMM reflect biological differences and provide important clues to biological mechanisms. EMM can be used in hypothesis testing as well as hypothesis-generating situations.

7 Analytical Approaches: The Data

Once a study has been designed and conducted, and data has been collected, the investigator starts to analyze the data to produce results for making inferences from the study. Generally, it is a good idea to have a fair sense of how the data will be collected, what it will look like once collected, and how it will be analyzed, *when the study is designed, if not before*. However, most often, investigators start worrying about the data structure after data have already been collected. Having a good perception of variable construction and data structure often improves a study's design, the collection of information, and the ability to make valid inferences. For example, if the goal of a study is to assess the role of smoking on, say, periodontal disease, the investigators may wish to determine how they will collect the data for smoking. If collected as a dichotomous variable: current smoker yes/no, then historical impact of smoking would be missed, and the potential for spurious association may be increased because periodontal disease onset may have occurred earlier than the smoking start-date, and/or perhaps a dose–response association existed with smoking and periodontal disease, which would not be possible to evaluate. Therefore, if the investigator wishes to address dose–response, he or she would want to collect information about smoking start-date, number of cigarettes, duration of smoking, continuation or interruption of smoking, and so forth. On the other hand, if the investigator wishes to investigate a biological mechanism that suggests that smoking must act in an internal biological compartment, then perhaps cotinine may be measured in serum, saliva, or urine, and their associations with self-reported smoking can be assessed. All data have their limitations; for example, the half-life of cotinine in vivo is about 20 hours, and making cotinine levels a good measure of recent smoking, but not of lifetime exposure to smoking. However, if DNA adducts or other biological alterations can be shown to be good surrogate markers for cumulative smoking burden, then perhaps such markers may serve as better variables to "capture" smoking exposure.

Boxes 7.1 and 7.2, and Figures 7.1, 7.2, and 7.3 describe main features and general properties of data, and main properties of common types of distributions for a quick review of these topics. The most important aspect of working with data is to understand in which form the data would provide the most usable variables for analysis; that is, how to convert the collected data into useful information for insightful inferences. A key to appropriate and efficient data usage follows the paradigm:

Data ≠ Information; and
Quantity of data/information ≠ Quality of insight.

BOX 7.1 Terms Commonly Used in Describing Data

Data

Discrete Data: Can assume only whole numbers.

Continuous Data: Can take any value within a defined range.

Nominal Variable: Consists of named categories with no implied order among the categories.

Ordinal Variable: Consists of ordered categories where the differences between categories cannot be considered to be equal.

Interval Variable: Has equal distances between values, but the zero point is arbitrary.

Ratio Variable: Has equal intervals between values and a meaningful zero point.

Data Visualization

Bar Charts: Bars have spaces between them.

Histograms: Bars are adjacent and joined. Bars represent areas (width has value and X height that has value).

Dot Plots: Bar charts with line of dots.

Scatter Plots: Displays value of two variables for a set of data using Cartesian coordinates. It shows the linear and nonlinear relationships between two variables.

Stem-and-Leaf Chart: Stem: Shows the tens and hundreds place (multiple if the category includes multiple levels for the unit or tens place). Leaf: Shows the unit number as in the data. If turned counterclockwise, it looks like a histogram.

(Continues)

BOX 7.1 Terms Commonly Used in Describing Data *(Continued)*

Frequency Polygon: Used for interval/ratio data (cumulative frequency polygon is a variant). It joins the midpoints of histogram bars.

Data Description

Central Tendency (Variance): Nominal data – mode; Ordinal – median (inter-quartile range: IQR), mode; Interval data – mean (SD), median (IQR), mode; Ratio data – mean (SD), median (IQR), mode.

Mean: Arithmetic > Geometric > Harmonic. Harmonic mean used for sample size estimation for obtaining conservative estimate to maximize power. Mean is affected by extreme values but median is not.

Median: Value such that half of the data points fall above it and half fall below it. It is less affected by extreme values. Insensitiveness to data values makes it not very conducive to testing for differences. Median based tests – nonparametric tests (distribution-free tests) used when normal distribution assumption is violated.

Mode: Most frequently occurring category.

Range: Difference between highest and lowest values.

IQR: The difference between lower (Q_L) and upper (Q_U) quartiles, comprising the middle 50% of the data.

Range Approximation of SD: For small sample size = range/4; for large size = range/6. A "quick" way to estimate standard deviation (SD).

Index of Dispersion: Similar to the coefficient of variation—essentially it is the ratio of variance to the mean. It describes dependence between successive arrivals of an arrival process. An indicator of how well or how rapidly a factor will be dispersed.

Mean Deviation: $[E(|X - \bar{X}|)]/N$

Standard Deviation: Square root of variance $\text{Sqrt}[E(X - \underline{X})^2]/N$. If we add a constant to every number, the Variance/SD does not change.

Skewness: Refers to symmetry of curve. Named according to the direction of the tail of the curve (i.e., left (–ve)-, or right (+ve)-skewed data. Median is not affected by skew but mean gets pulled toward the skew; that is, for right-skewed data, mean > median > mode; for left-skewed data, mean < median < mode.

Kurtosis: Measures flatness/peaking of data. Mesokurtic – normal distribution (has kurtosis = 3, but reported as 0 for ease of comparison as 3 is subtracted from all kurtosis values). Leptokurtic: higher rising data with greater peak. Platykurtic: flatter data, lower peak.

BOX 7.1 Terms Commonly Used in Describing Data *(Continued)*

Box Plot: The box contains middle 50% of the data. Median may be at any point in the box (not necessarily at the center). Step = 1.5 × IQR. Whiskers join the box to the inner fence ~ 1.5 × IQR. Outer fence ~ 3 × IQR. Ninety-five percent of data fall within the inner fences; 99% of data fall within the outer fences (similar to 2 and 3 SD in a normal distribution). Values between inner and outer fences are outliers; whereas values beyond the outer fences are far outliers.

BOX 7.2 Properties of Common Distributions

Distributions

Normal Distribution: Symmetric bell curve. Mean = median = mode; Kurtosis = 0; Mean = 0; SD = 1; 68.2% data in mean ± 1 SD; 95% in mean ± 1.96 SD; 99.2% data in mean ± 2.9 SD. Curve approaches the *x*-axis at the tails but never reaches the *x*-axis (asymptotically approaching *x*-axis).

Central Limit Theorem: If we draw a large number of equal size samples from a nonnormal distribution, the distribution of the means of these samples will still be normal, as long as the samples are large enough.

Standard Z Score: $z = (X - \overline{X})/SD$. Shows how far away the individual score stands from the mean in the distribution. Results from two interval scoring methods of the same phenomenon can be assessed by comparing the Z score from each of the methods—this will show how far the individual is from the mean of the respective scoring methods, thereby allowing comparison across the two scoring methods. Thus if the Z scores are similar, then the two scores from the two systems are also equivalent. The raw mean score has $z = 0$ (any observation that is equal to the mean score will have a $z = 0$). Sum of all Z scores = 0 for normally distributed data. SDs of all the Z scores = 1.

Binomial Distribution: Shows the probabilities of different outcomes for a series of random events, each of which can have only one of two values. Mean = $n \times p$; Variance = $n \times p \times q$; SD = Sqrt ($n \times p \times q$). (n = no. of successes; p = no. of failures; q = no. of trials)

Poisson Distribution: Variance = mean. Poisson random variable can take any nonnegative integer value. Count data follows Poisson distribution. Poisson distribution is used to model occurrence of rare events.

t-Distribution: Student's t-distribution is a symmetric distribution around 0 like the normal distribution developed to describe the behavior of a random variable descriptor of two population means.

(Continues)

BOX 7.2 Properties of Common Distributions *(Continued)*

F-Distribution: Skewed to the right and often appropriate for modeling the probability distribution of the ratio of independent estimators of two population variances.

Chi-Square Distribution: Nonsymmetric, hypergeometric distribution, (right skewed), describing behavior of nonnegative random variables. Used widely in analyses of categorical data.

Hypergeometric Distribution: A discrete probability distribution that describes the number of successes in a sequence of n draws from a finite population without replacement. The p-value of a two-sided Fischer's exact test can also be calculated as the sum of two hypergeometric tests.

Information is viewed as a materialized message having measurable entropy (degree of disorderliness; *information entropy*: amount of information that is missing before reception). Information may also be viewed as a pattern that involves a separation between the object and its representation. In epidemiological studies (or other scientific research studies), information occurs as a sensory input to a device or organism (including animals, humans, and also including the investigator[s]) that may be perceived as very

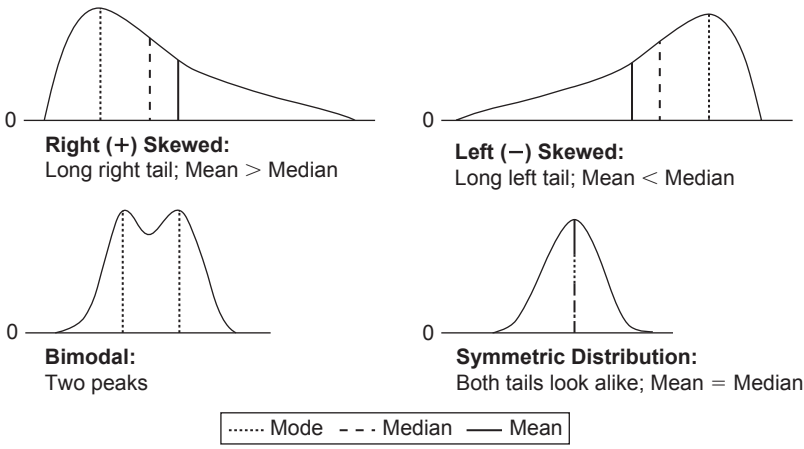


FIGURE 7.1 Shapes and Properties of Certain Distributions Skewness measures the length of the tail of the distribution.

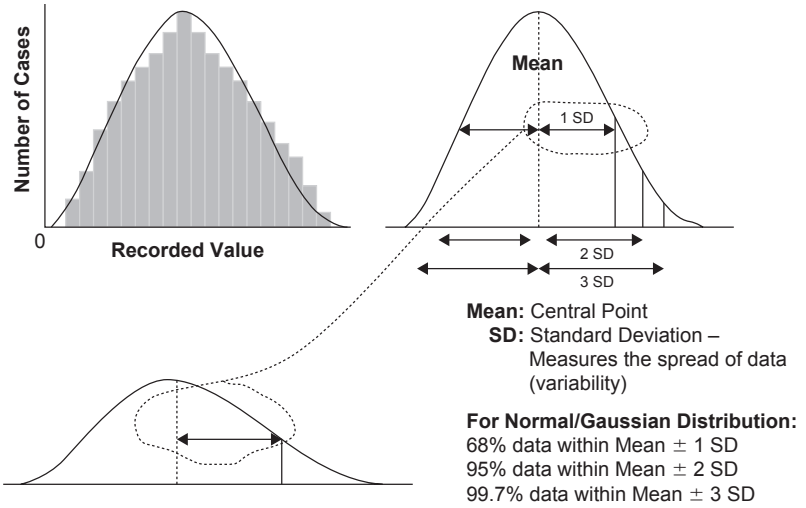
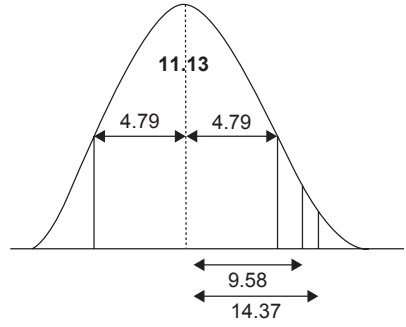


FIGURE 7.2 Gaussian or Normal Distribution and Its Main Properties A wider distribution (lower left) has greater variation measured by variance or standard deviation. Standard deviation is the square root of variance.

Standard Deviation
- Measures spread around the Mean
- The larger the SD, the more the variability
- SD = 0 implies no spread, i.e., all obs same
- Units of SD are same as that of the data

For Normal Distribution:
68% data within Mean ± 1 SD
11.13 ± 4.79 = 11.13 − 4.79;
11.13 + 4.79 = 6.34; 15.92



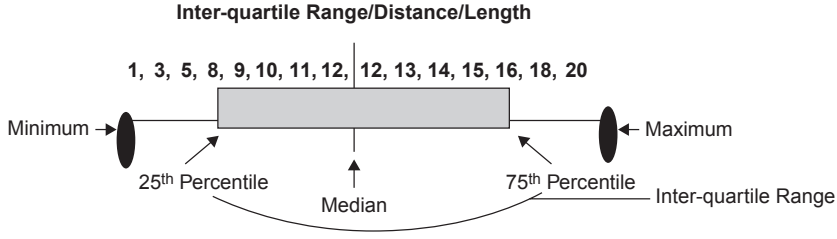


FIGURE 7.3 Example of a Normal Distribution with Interpretation of Measure of Spread (Standard Deviation) Lower figure provides an example of obtaining median and its spread (inter-quartile length). Whenever a measure of central tendency is stated (i.e., mean or median) the measure of spread should be mentioned with it (SD or IQR respectively) to provide better description of the data. Mean and SD are calculated from the data displayed.

important, partly important, or unimportant, and it may or may not be recorded. Information influences transformations of systems leading to expected or unexpected, desirable or undesirable, or useful or useless outcomes. Converting data into information requires the intervention of the transforming agency, (in this case the investigator[s]), and is subject to the perspective of that transforming agency. Qualitative and quantitative research methods differ in the latitude they allow to the transforming agency in applying their individual personal perspective to data, its analysis, and final interpretation.

Types of Variables

Variables may be continuous, or categorical (i.e., ordinal, nominal, dichotomous, or multilevel categories). The total amount of useful information that can be extracted from data depends on the way data have been collected. In general, continuous data provide more information than categorical data—overall an information-content gradient exists for data types: dichotomous < nominal < ordinal < continuous data. Continuous data allow the most information and can usually be converted into other data types; for example, age, and $CD4^+$ cell count in the blood can be categorized into whatever categories suit the investigators' analytical needs. If possible, and if study logistics permit, it is always better to collect data in continuous form rather than in categorical form. Sometimes, several pieces of data are combined to make a continuous variable suitable for data analysis depending on need. For example, to measure smoking in pack-years, the investigator would need to collect data about how many cigarettes the participant smoked and the number of years the participant smoked. There are many such variables; for example, DMFT/S and periodontal attachment loss are extremely important but need to be constructed from collected data.

Whereas continuous variables lend themselves easier to mathematical operations, such options are relatively limited for categorical variables. Categorical variables are generally used to represent a proportion (i.e., what proportion of the total belongs to a category). For example, if dental caries were to be recorded as a dichotomous categorical variable, then one can record the percent of participants who had caries vs percent not having caries. The total information from such a variable is limited. On the contrary, if dental caries were recorded as a continuous variable at person level, then DMFT/S scores of individual participants would be calculated, and an overall mean and standard deviation could be derived, which would provide much more information compared to the dichotomous variable. If need be, different categories for describing caries can always be created from DMFT/S data. If a choice exists, it is always better to collect more detail in data rather than less.

Distributions, Parametric, and Nonparametric Tests

Statistical tests are developed based on certain underlying assumptions. If tests make some assumptions about the underlying probability distribution of data, then the tests are called *parametric tests*, whereas if tests do not make any assumptions about underlying data distributions, they are called *nonparametric tests*. At first glance, it might appear that nonparametric tests would be the best choice because they are "distribution-free" tests. It is also true that these tests would be easy to use when the distribution of data is unknown, when data does not have a numerical interpretation, and when data is ranked. Because of these advantages, some investigators prefer to use nonparametric tests as they are less prone to misuse (being robust to violation of distribution assumptions). However, if both parametric and nonparametric tests could be conducted on the same data, then parametric tests are substantially more powerful than their nonparametric counterparts. To get the most out of data analysis, a simple strategy would be to first assess the data distribution and optimize the data so that it fits the distribution requirements of the parametric analyses. It is also true that most common parametric tests are fairly robust and withstand *minor* violations of distribution assumptions.

DMFT/S is generally treated as continuous variable at a person level. Once DMFT/S is calculated and averaged over a set of observations, it is transformed into continuous data. However, at an individual level, DMFT/S is a count of number of carious involved teeth, just as the number of accidents in a given day, or number of procedures performed by a surgeon. At tooth level, DMFT/S is count data that can take limited values (DMFT: 0 or 1 only; DMFS 0–5 only). Such data are "count" data because these are a set of counts and are inherently different in nature from continuous data such as CD4+ cell count and RBC, serum C-reactive protein (CRP), HbA1c level, and salivary IgA concentration, which are free to vary over a much larger range of values. The difference is that count data can take only nonnegative integer values (i.e., 0, 1, 2, 3…) and cannot have decimal values that truly continuous data should be able to assume. Most continuous variables are treated as belonging to Gaussian (i.e., normal) distribution, whereas the underlying distribution for count data can be Poisson, binomial, or negative binomial. Therefore, statistical treatment of count data requires different handling compared to usual normally distributed continuous data. Most biological secretions, such as cytokines, other inflammatory mediators, CRP, and so on are not normally distributed, but exhibit a skewed distribution (see Figure 7.4). Distribution data for such variables must be examined carefully and appropriate modifications must be done to transform the variable so that its distribution conforms to normal distribution prior to analysis.

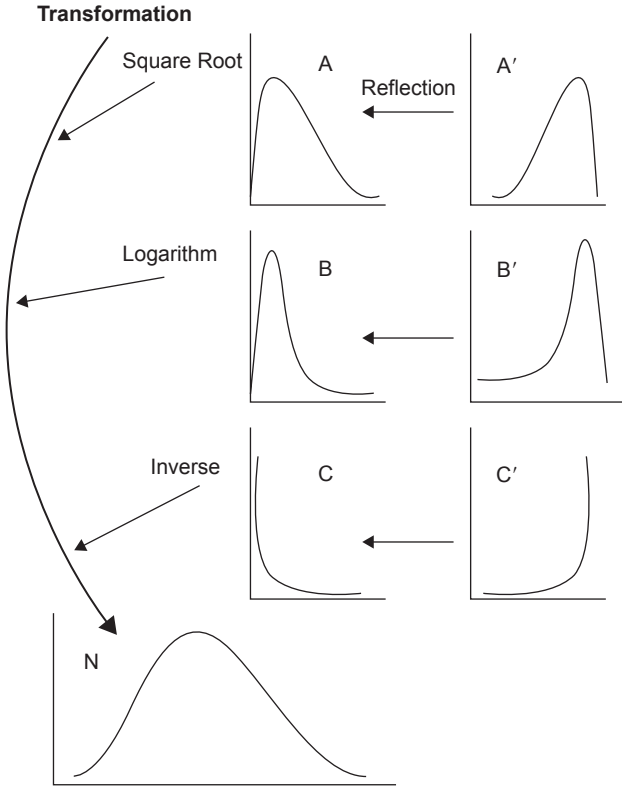


FIGURE 7.4 **Distribution** Distributions A–A′, B–B′, and C–C′ are mirror images. Reflection is the process of converting distributions (i.e., A′, B′, and C′) to their positive counter parts (A, B, and C respectively). Specific transformation (square root transformation for A, log-transformation for B, and inverse transformation for C) convert these distributions to normal or near normal (N) so that parametric analysis can be conducted.

Data Integration

Data analysis is preceded by data integration: *Data visualization* and *data optimization* start with a good look at the nature of data and individually assess each variable by conducting a thorough univariate assessment. Each variable data are optimized before proceeding with a full description of the variables or a comparison between variables.

Transformation of variables is sometimes viewed skeptically as if a black-box phenomenon is under operation. This need not be so. The commonest transformation used in oral health research is log-transformation. By using the natural log or base-10 log value of a variable, the investigator con-

verts the regularly collected data having a skewed distribution into a transformed variable that has a normal distribution so that parametric statistical analyses can be done using the transformed variable (see Figure 7.4). Transformations are a remedy for adjusting for outliers, and for deviation from normality, linearity, and homogeneity of variance. However, real-life interpretation of the importance of the variable values should be done on "untransformed" variables rather than transformed variables. Sometimes, interpretation of transformation may not be intuitive, but a careful selection of the type of transformation performed can help in conducting analyses correctly, leading to valid and interpretable inferences. Some transformations used commonly in science include square-root transformation, reciprocal transformation, logarithmic transformation, and arc–sine transformation.

Figure 7.4 demonstrates some common distributions and common types of transformations that can normalize those distributions to allow use of parametric analysis. If data is severely positively skewed, logarithmic transformation works well to "normalize" the data. It does not matter what base is used to log-transform data (i.e., $\log_{10}$ or natural log, ln or another base). Because logs are defined only for positive numbers, if the original values have negative numbers, they may be converted to positive numbers by mathematical operations (e.g., adding a constant to all values of the variables) before log-transforming the variables. Square-root transformation of count data (i.e., when variance is proportional to the mean) is sometimes helpful as it compresses the upper tail of the distribution. If the range of values is not large, careful exploration of square-root transformation needs to be conducted before deciding to use the square-root transformed variable. If a substantial number of extreme values form the tail of distribution, then reciprocal transformation can be useful as a way to collapse the range of values. Although arc–sine transformation also works well for count data, it works best for proportions. It stretches the tails of the distribution.

Sometimes data distribution appears substantially different than what we usually expect (see distributions A′, B′, and C′ in Figure 7.4). Careful observation reveals that many such distributions are "mirror images" (or reverse) of distributions we are more familiar with (see distributions A, B, and C in Figure 7.4). In such situations, the "mirror image" data can be "reflected" to convert to the more familiar pattern, and then transformed as required. The process of "reflection" means that all values of the concerned variable for the entire data are multiplied by –1 to provide a "reflected image" data distribution. For example, (data A′) × (–1) = (data A). Now the "reflected data" can be transformed as usual.

Hartley, Ho, McConkey, & Geh (2005) reported a meta-analysis to assess the usefulness of different chemoradiotherapy regimen/schedules in rectal cancer management using pathological complete response as a marker for the efficacy of preoperative chemoradiotherapy. They normalized pathological complete response rate by using arc–sine transformation prior to

weighted linear modeling and demonstrated that use of continuous infusion 5-Fluorouracil (5FU), the use of a second drug, and radiation dose were independently associated with higher rates of pathological complete response following chemoradiotherapy for rectal cancer.

Another way for optimizing data is to "trim" the heavy-tailed data or drop extreme outliers. However, some investigators do not like the idea of dropping observations, even if they recognize the distortions caused due to extreme values. In such situations, extreme data can be recoded to an acceptable threshold value; for example, extreme outliers can be reassigned a value that is equal to mean ±3 times the standard deviation that would keep the observation in and yet not distort the analysis. If data are markedly skewed or variances are heterogeneous, then transformations may be helpful. Data transformation is a useful tool, but should be used only if necessary because nonjudicious use of transformation can bias analyses. The key to using a transformation is to realize that it is a tool for optimizing data, and not a tool to derive a good-looking p-value!

Outcomes of data analyses depend on well-conducted data integration procedures to optimize the data. The processes of data optimization should be viewed as procedures to make certain that data analysis can be planned and conducted appropriately in ways that results confirm to the data, and analytical procedures used are valid in arriving at the results, which can be relied upon.

Missing Data

Missing data threatens the reliability and validity of studies and is common in almost any study. Recognition of this fact allows us to design studies to minimize the potential of missing data; devise a plan of action about what to do if data are missing; and address the missing data by implementing those plans if missing data occurs in the study. Although there is increasing recognition of the role of missing data in analyses, there exists a general lack of understanding about how missing data may impact study results and how such problems may be overcome. Overall, most oral health-related published studies do not mention if any data were missing; and if so, how much, how the issue was handled, and if the missing data made any impact or were expected to make an impact in interpreting the study outcomes. Even if missing data are reported, most reports limit the reporting to mentioning how much data were found to be missing. The apparent paradigm for working with missing data seems to be, at best, to report the volume of missing data, and proceed with the analyses assuming that the missing data have no impact at all. Not only does the amount of missing data impact study results, but the *reason* for which data may be missing may bias the study. It is, therefore, imperative for the investigator to pay

TABLE 7.1 Types of Missing Data

Type of Missing Data	*Description*
Unit missing data	Entire data for the unit of observation is missing.
Missing values	Only parts of data for the unit of observation are missing.
Missing wave	Data for one (or more) follow-up visit is missing (in longitudinal study).
Missing completely at random (MCAR)	A random phenomenon. Missing data occurs randomly and is not associated with values of the missing variable or any other variable in the study directly or indirectly. MCAR has minimal potential impact on bias and statistical analyses.
Missing at random (MAR)	Missing data is not associated with values of the missing variable but may be associated with other variables in some way. Therefore, if the involved variable is assessed by itself, data may *appear* to be missing randomly.
Missing not at random (MNAR)	Missing data occurs in a systematic way and may be correlated directly or indirectly with values of the missing variable and other variables in study. MNAR has maximum potential impact on bias and statistical analyses.

close attention to missing data, and not treat it as just another number to be reported.

Missing data may arise due to the design of the study (e.g., too detailed and long questions and responses required, questions dealing with difficult recall, time-consuming questionnaire); due to the participant (e.g., unclear questions, questions perceived as threatening or potentially privacy-violating-type questions); or due to the interaction of participant with the study design (e.g., no suitable response found by the participant, or inconvenient question-delivery mechanism). Table 7.1 enumerates some of the types of missing data commonly found in epidemiologic studies. Such classifications are a generic attempt at describing missing data because for the same observation/group, data for different variables may be missing for different reasons and be of different types. Furthermore, missing data for an individual or a group of individuals may or may not be correlated with each other, study variables, or unmeasured variables; or they may occur in different ways across different variables and times in longitudinal studies. Missing data in multilevel studies can occur at the individual level or at the group level. Data may be missing due to chance, study design errors, characteristics

of participants, measurement errors, data collection conditions, and data-entry or data management errors.

Missing data may impact studies in several ways: They may reduce the total amount of usable information, increase error variance, reduce reliability, incorporate spurious associations, dispel randomization, and induce selection bias. Power of a study may be impacted directly by the amount of missing data. However, if data are systematically missing, then the results may be biased. For example, if those who have poorer outcomes drop out of studies, then analyses not accounting for the dropouts would be biased.

Data missing completely at random (MCAR) may reduce the total number of observations available for analyses, but due to the random nature of "missingness" it does not alter the relationships between variables. Although two variables may have missing values, as long as they are not correlated in a way that "missingness" of one also determines the value of the other variables, there should not be much of a problem in the analyses. For example, people who do not report income in a study and also do not report their insurance status are considered MCAR unless their nonreporting of insurance status was correlated to their individual income levels. *Missing at random* (MAR) data may show a correlation between two variables. For example, those in certain employment groups may not report their incomes. For example, employment and income information, when viewed individually, will be missing randomly, but when assessed together, the correlation may appear. However, the "missingness" of income would be correlated to certain employment types, but not to the income values itself; that is, the probability of nonreporting is not related to the income level. In *missing not at random* (MNAR) data, the "missingness" of data is a function of data values. In the above example, if high-earners (or low-earners) do not report their income, then the "missingness" of income becomes a function of the income values itself, and such data should be viewed as *missing systematically*.

Management of Missing Data

MCAR data, if not too large, can be used in analyses because the resultant estimates are not biased in absence of data. MAR data will need special handling to produce relatively meaningful and unbiased estimates. However, MNAR data is problematic because of the bias it introduces. The only way to obtain unbiased estimates from MNAR data is by modeling the "missingness" of data itself. Such models may produce some possible correction factors that may be applied to estimates, or produce models that may be then be incorporated into other models to correct for missing data.

While reporting studies, it is important to mention the amount of missing data and their relationship to important variables in the study,

how missing data were handled, and what the impact of missing data in the study was. Such clear reporting will allow the investigators and readers to put the issue of missing data in perspective and infer study results accordingly.

Overall, a general process to address the issue of missing data may involve several steps:

1. Anticipate situations that might lead to missing data and prevent or minimize the sources of missing data.
2. Anticipate potential impacts of missing data in the study and develop strategies for addressing those issues as they arise.
3. Assess the amount, pattern, type, and reasons of missing data once the study data are collected and analyze the impact of missing data compared with observations without missing data.
4. Carry out sensitivity analyses to assess the actual impact of missing data in the study.
5. Consider the possibility and impact of compensatory methods to address missing data, and carry out suitable analyses if needed.

One approach to address missing data is to discard data. *Complete case analysis* is a commonly used method where only cases that have complete data are analyzed. The default in several major statistical analytical programs running regression analyses is to use complete case analyses. This issue has some important implications in modeling and is discussed in Chapter 8. If there are many variables in the model to be used, then the number of complete cases may be few and data analyses may be incorrect. Furthermore, if data are not MCAR, then biased model results are likely. *Available case analysis* examines different aspects of the study with different data subsets. For example, if age is available for all persons, then age description will represent 100% of the data. If income is missing for 30% of persons, then income description will be based on 70% of the data and so forth. The major problem with this approach is that results from different data subsets may not be directly comparable. It also assumes that data are MCAR because all available information is *assumed* to be fully representative of the full data set. *Nonresponse weighting* is a method where the sample is reweighted, taking into consideration the missing value, and a new complete case analysis data set is made with the new weighting. This process gets complicated when several missing data points occur or when several variables have missing values.

Data Imputation

Another approach to handling missing data is to "fill in" the missing information instead of discarding data. However, before proceeding with impu-

tation, the investigator must make certain that biases do not exist in the "missingness." Most imputation methods require data to be MCAR or at least MAR. Data imputation is the substitution of some value for a missing data point. Obviously, the major advantage of imputation is that it produces a "complete data set" for analyses. Once all missing data has been imputed, the data set can then be analyzed as usual. However, imputation is a double-edged sword, and one must understand the process, requirements, and limitations of imputation before applying it. Data imputation has been used very successfully in astrophysics, solar physics, and several other sciences including genome-wide linkage analysis (Baier & Wernecke, 2003; Wang, Zhu, & Keen, 2003). Imputation is often perceived as a "black box" phenomenon—one often does not realize that use of imputation in large public use data sets is very common. For example, some income, smoking and self-reported health data among others in NHANES is derived using model-based imputation techniques (Centers for Disease Control, 1993). Several imputation methods have been developed.

- **Substitution:** Information about a subject is extracted from an alternative database and imputed into the data set with missing information.
- **Estimator:** Answers to other questions are used as a guide to develop the most plausible answer to the missing question and the data derived from mathematical operations is imputed.
- **Mean imputation:** A common method is to use the mean value of the data set/strata/group to which that individual belongs to impute the missing value.
- **Last value carried forward:** In longitudinal studies, if data are missing for certain intervals or follow-up visits, often the pretreatment value is imputed for the missing value under the assumption that the error will be toward a more conservative value (i.e., *before* the treatment had an effect).
- **Indicator variable for missingness:** In categorical data, a separate "missing" category is added and used as a level for the variable in analysis.
- **Logical rule imputation:** Some logical rules are developed using "informed" assessment of the involved variable, and a logical value is imputed for the missing value.
- **Cold deck:** A "perfect" response set with fixed values for all data points is prepared as the source from which missing data are imputed.
- **Hot deck:** Other respondents are selected from the data set based on characteristics similar to the respondent with missing values,

and the corresponding information is imputed in place of the missing value.

- **Simple random imputation:** A random value from the variable for which an observation has a missing value from the data set is used to replace the missing value.
- **Regression-based imputation for a single variable:** A regression model is developed to predict the variable for which values are missing, and then this model is used to predict the missing values.
- **Maximum likelihood (ML) methods:** ML methods first estimate the parameters on the basis of available data that then forms the basis of estimating missing data. Thereafter, the parameters are reestimated based on the imputation iteration, and the sequence is continued until models converge on a solution that provides the final imputed values.
- **Multiple imputation (MI):** MI is a Monte Carlo technique in which the missing values are replaced by several simulated data sets. The predicted values generated are modified with the addition of an error component to adjust for uncertainty in obtaining correct variances. MI is a relatively easily adoptable method that is now available in most major statistical software programs. As statistical software usage becomes more common, imputation is also gaining in popularity, although such procedures must be applied with the utmost care. MI makes it possible to assess the impact of missing-data uncertainty on the variances of estimators and revise variance estimates to reflect this uncertainty. In another step, the parameter estimates used in imputing data are drawn randomly from a posterior probability distribution of the parameters. Overall, MI methods help in reducing nonresponse bias and sampling variance. In general, whenever possible, use of a larger number of auxiliary variables rather than fewer are advised when using an MI procedure.

Imputation should not create extra variances and lead to biases or distributional changes in the data. Ideally, the process should rely on data from the sample rather than making assumptions about the nature of missing data. Furthermore, estimates should not be developed that are heavily dependent on imputed values—*imputation should help fill in missing data, and not create most of the data*. Although imputation can be done manually and/or be automated, the best way to conduct imputation is to have an objective and automated reliable procedure. All imputed data must carry a flag to indicate imputation so that imputed values can be assessed separately.

Data Sources

WHO maintains a health-related database including oral health data that can be accessed through their Data and Statistics website at http://www.who.int/research/en/. Oral health specific data can be directly accessed from the Data and Statistics oral health website at http://www.who.int/infobase/report.aspx?rid=112&ind=ORA. Different countries, especially developed countries, maintain statistical databases that can be accessed through their governmental websites.

In the United States, the CDC maintains a comprehensive searchable website of Mortality and Morbidity Weekly Reports, and its special series reports (SS-), which include several reports related to oral health. Several other important reports can be downloaded from websites of related professional organizations. For example, the Future of Dentistry Report can be downloaded from the American Dental Association's website, and the Surgeon General's Report on Oral Health is available at the Department of Health and Human Services webpage. Whereas these reports are freely available, some of the key problem elements in dental public health research is to obtain, assess, and analyze oral health data of populations, to make inferences about the data, and develop suitable strategies to improve oral health of the populations.

Several national surveys are conducted that include some basic oral health-related information, such as the Behavioral Risk Factor Surveillance System, NHANES (currently the fourth survey is ongoing), National Health Interview Survey, Medical Expenditure Panel Survey, Surveillance, Epidemiology, and End Results, and a series of surveys conducted by the National Institute of Dental and Craniofacial Research (NIDCR). Data for these surveys are public-use data and available from their respective websites. The NIDCR also maintains a directory of usable dental data that can be obtained on CD-ROM from their office. However, there is a shortage of local-level data (i.e., state level, and especially county-level data). Several states run their surveillance programs periodically to collect oral health-related data. These states feed the data to a central point to disseminate basic information about the status of oral health in the state, which forms the National Oral Health Surveillance System (NOHSS).

Data tables for oral health indicators: The NIDCR and CDC Dental, Oral, and Craniofacial Data Resource Center maintains the Catalog of Surveys and Archive of Procedures Related to Oral Health that provides selected data tables from different years (national and state). Recent updates are provided for each of the oral health indicators in both HTML and PDF formats that can be viewed, saved, and printed from the website. This catalog includes examined indicators, such as dental caries, oral and pharyngeal cancers, periodontal assessment and disease, sealants, smokeless tobacco lesions, tooth loss, and self-reported indicators such as dental visits, self-assessed oral health status, usual source of dental care, orofacial pain, tobacco use, and dental insurance. (Chattopadhyay, Arevalo, & Sohn, 2008)

Other data sources that can be utilized to obtain oral health-related data include the websites of the American Dental Association (ADA), ASTDD, CDC, Center for Medicaid and Medicare Services, European Global Oral Health Indicators Development (EGOHID), National Immigrant Survey, NCHS, NIDCR Oral Health Data Repository, Pan American Health Organization (PAHO), state-specific Departments of Health, and the U.S. Census Bureau.

Electronic Oral Health Records

Electronic oral health record (EOHR) is an electronic database of patients' oral health and related information. Patients' complete medical and oral health history and charts are included in the EOHR. This allows easy access to patients' general and oral health information such as patient demographics; practitioner characterization; immunizations; health history; health conditions and problems; examinations and findings; treatment plans and clinical orders; diagnostic observations; radiographs, laboratory data and other investigation reports; prescribed medications; all therapeutic interventions; hospital admissions and attendances; scheduled events; patient encounters; and possibly payment records. WHO has defined EOHRs explicitly as:

> All personal health information belonging to an individual is entered and accessed electronically by healthcare providers over the person's lifetime. EOHR extends beyond acute inpatient situations including all ambulatory care settings at which the patient receives care. Ideally it should reflect the entire health history of an individual across his or her lifetime including data from multiple providers from a variety of healthcare settings, primarily to support continuing, efficient, and quality health care. (WHO, 2006)

EOHRs are required to meet legal, confidentiality, and retention requirements of the patient, the attending health professional, and the healthcare institution and country where it originates, passes through, and resides. Currently, the use of EOHRs and their outcomes are still considered as scientifically reportable events globally and clear quality assurance guidelines are derived from the Health Insurance Portability and Accountability Act (HIPAA) in the United States (Chattopadhyay, Souza, & Arevalo, 2009).

Although EOHRs have several benefits such as easy data storage, retrieval, and utilization of data; efficiency; and economy among others, from a data analysis standpoint, EOHRs improve the accuracy, precision, and quality of data recorded in a health record; reduce errors in data recording; and increase the ability for data sharing and linking to other databases, enabling more complex analysis (Chattopadhyay, Souza, & Arevalo, 2009). Despite these benefits, EOHRs are not universally adopted—rather, early

adopters have appeared and are incorporating EOHRs in their dental clinical environment (Chattopadhyay et al., 2009). Lack of standard data entry procedures, uniform coding system, standard terminology, data coding issues, different data formats and reporting systems, and lack of skill in using disease classification systems are some of the major barriers and pitfalls in the use of EOHRs. For the epidemiology investigator, these issues must be kept in mind when accessing EOHRs from disparate sources to link together in research studies. Although "chart abstraction" becomes much easier with the use of EOHRs, they are not designed from a research standpoint, but are essentially mechanisms to allow smooth clinical functioning that may place limits on the usefulness of clinical data from EOHRs for research studies.

As more practice-based research networks are established and expand, more data inflow using EOHRs will occur (Chattopadhyay, Souza, & Arevalo, 2009). The need to integrate clinical practitioners and increase their awareness toward data creates needs for research to also increase. As clinical studies shift from the traditional academic setting to practice-based networks, dental practitioners (both generalists and specialists) will play an increasingly more significant role in research. Increasing use of EOHRs, the need for skillful data management, and data analysis will be required to draw scientifically valid inferences after accounting for all sources of variations in the study, including the possibility for ascertaining the validity of data arising from a variety of sources.

8 Analytical Approaches: Analyses

Overall, data analysis may be viewed as containing three steps: *descriptive statistics*, *exploratory analysis*, and *confirmatory analysis*. Outcomes of data analyses depend on well-conducted data integration procedures to optimize the data. Although analytical approaches may sometimes be viewed as being independent of the study design, in essence, an analytical plan is an integrated part of the study design. For example, definitive causal inferences require a design that allows the interpretation of exposure having preceded the effect, and must be analyzed in a specific way, outputting statistics that may be used to interpret casualty (e.g., calculating risk ratios in a prospective cohort compared to odds ratios from case-control studies will require advance planning in study design to permit analyses for appropriate inferences).

Hypothesis testing involves clear understanding and formulation of a research plan and proceeds according to a sequence of steps (see Box 8.1), facilitating proper interpretation of study results. Inferential statistics play a major role in helping to distinguish between risk factors and their role in determining outcomes. The general goal of these analyses is to determine whether a difference exists between exposed and unexposed (or cases and controls) for a certain outcome. These groups exhibit a distribution for factors under contention, and statistical analysis tries to determine whether these distributions are different from each other. Figure 8.1 depicts this phenomenon. If there is a substantial overlap between the two distributions, it is generally concluded that the two distributions are not distinct (i.e., the two groups do not differ from each other for the factor under study). Alternately, if the distributions do not overlap, then two distinct distributions are seen, and the conclusion of a difference between the groups is made. However, in a real-world situation, clear distinctions are not common, and a certain degree of overlap between two distributions is always seen. It is the role

BOX 8.1 Important Definitions and Steps of Hypothesis Testing

Null Hypothesis: Hypothesis states that there are no differences between the groups being compared.

Alternative Hypothesis: Hypothesis states that groups being compared are different from each other, and usually states the relationship that the investigator is trying to establish.

Type-I Error (Alpha Error): Probability of concluding that there is a difference when there is no difference (i.e., rejecting null when null is true).

Type-II Error (Beta Error): Probability of concluding that there is no difference when there exists a difference (i.e., not rejecting the null, or accepting it, when it is false).

Power: 1-β. Probability of concluding that there is a difference when there really is a difference (i.e., rejecting the null when it is false). Increasing sample size increases power up to a certain limit beyond which the gain in power with increasing sample size is minimal. The relationship between sample size and power is positive and exponential.

Significance: Statistical significance is usually considered to be a necessary precondition for assessing of clinical importance but says nothing about the actual magnitude of the effect. The "alpha level" is an arbitrary probability cut-point (i.e., 0.05 or 5% by convention), indicating the probability of obtaining by chance, a value as extreme or more extreme as is observed in a study.

Hypothesis Testing Steps

1. State the overall (global) research question.
2. State the research question as a statistical hypothesis and develop the null hypothesis.
3. Design the study to answer the question.
4. Decide which test to use to answer the question.
5. Select a significance level a-priori (i.e., usually 0.05, double sided, but may be set lower for some situations).
6. Estimate sample size for the study (i.e., sample size is dependent on the effect size and power).
7. Collect data with accurate explicit measurement protocol.
8. Calculate statistic.
9. Interpret data within the framework of its study.
10. Draw conclusion.

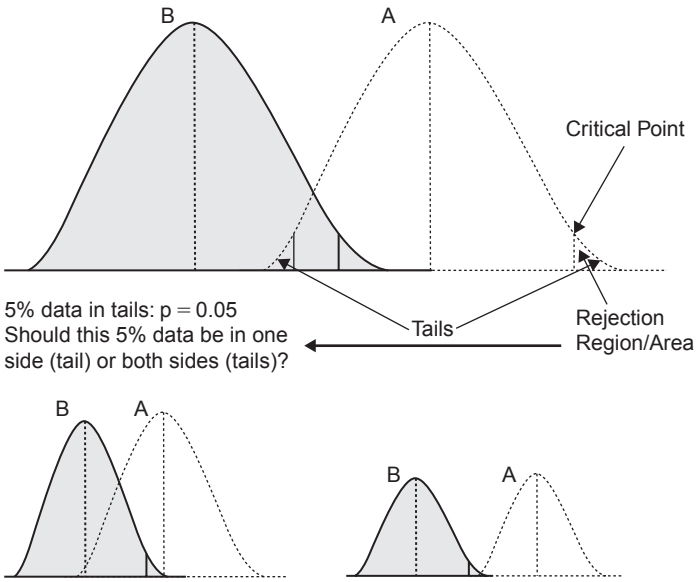


FIGURE 8.1 Comparing Distributions The amount of overlap between distributions (A and B) is tested using statistical tests. The further apart the distributions are, the more different the groups are considered to be.

of biostatistics to determine whether the two distributions are different enough within an acceptable range as per the assumptions made for the involved tests. The conventional critical regions are defined by the significance level (i.e., usually at 0.05 or 5% level) that identifies the critical region for rejecting or failing to reject the null hypothesis. However, the alpha level is a mere convenience that is conventionally agreed upon and does not have ingrained scientific truth to it. This level can be changed depending upon the chances the investigator is willing to take. For example, if multiple tests are going to be conducted, it is common to adjust the p-value to more conservative levels (e.g., 1% or lower), as is commonly done in genetic epidemiologic studies.

The P-Value Conundrum

P-value is the most used statistic in scientific literature and has become the most controversial statistic in epidemiologic research. By definition, p-value is the probability of obtaining a result at least as extreme as the one that is actually observed, *provided* that the null hypothesis is true (commonly stated: "under the null/null hypothesis"). The p-value in a study is a ran-

dom variable that is defined over the conditions of the study such that under the null hypothesis, its *distribution is uniform* over the interval 0 to 1. For the same study, several p-values can be defined. Correct interpretation of the p-value must emphasize the last part of the definition—the *assumption* for p-value; that is, *that the null hypothesis is true*. Therefore, the p-value does not provide evidence that the alternate hypothesis under study is true or false. The logical fallacy of mistaking the null p-value for the probability that the null hypothesis is true is common. For example, a p-value of 0.03 implies that *if the null hypothesis* were true, an association *at least* as strong as the one observed in the study would occur with a probability of 3%. This *does not* mean that if this association is observed, the null hypothesis has a probability of 3% (the latter is a common misinterpretation of p-values in the literature).

Whereas in general, p-values around 0.05 provide almost no evidence against the null hypothesis, p-values below 0.05 provide very little evidence against the null hypothesis other than what they appear to provide at face value (Poole, 2001; Selke, Bayarri, & Berger, 2001; Goodman, 1999a, 1999b). Many investigators believe that "the lower the p-value, the less the influence of chance. Unfortunately, this extremely common use of the p-value is a misuse and an abuse of that statistic" (Poole, 2001). The chosen significance level (most commonly 0.05 or 5%) is an arbitrary value, and the use of this level as a rigid finish line is questionable. For example, what if a study finds the p-value to be 6%? The all or none mechanism of use of p-value has been questioned because the important factor in study results is *strength of evidence* it provides, and not whether it reaches an arbitrarily defined "holy grail" that would miraculously answer all questions and solve all problems. One of the important issues in studies is repeatability of results across different samples and populations. Keeping this in mind, an alternative to the classic p-value has been proposed: p-rep or prep, which is meant to represent the probability of replication of an effect (Killeen, 2005). However, the statistic has been criticized as just a modification of the classic p-value that does not solve any of the problems with p-value, especially addressing the strength of evidence in favor of the goals of the study.

Routine statistical testing does not answer questions such as: How often is the null hypothesis true when we fail to reject it? When we do reject the null hypothesis, how often is the alternative hypothesis true? These are the probabilities of ultimate concern in significance testing—the predictive values of significant and non-significant statistical tests. It has been suggested that we should avoid exact interpretation of p-values in observational research where they may lack theoretical basis—that is, stop interpreting p-values as if they measure probability of alternative hypotheses (which should be done using Bayesian methods)—and we should get serious about the precision of effect estimates and look for narrow confidence intervals instead of low p-values to identify results that are least influenced by random

error (Poole, 2001). However, for most statistical tests that are conducted under assumptions, a "healthy-looking" p-value continues to be the most sought after statistic.

Time in Epidemiological Research

Another knotty issue in epidemiology is the handling of time as a variable in analyses. Time is usually interpreted as a quantity that only has one dimensional flow (making it a vector quantity), although it is treated in different, but inadequate ways. In epidemiological analyses, either (1) time is of no consequence (ignore time in measurement and analysis); (2) it has value, but is not adjusted for (timely events measured, but not incorporated in analyses); (3) it is considered extremely important, but weakly adjusted (inadequately incorporating time in analyses); or (4) it is considered very important and accounted for (proper adjustment for time in measurement and analyses). However, in all these methods of handling time in analysis, only its magnitude is assessed or utilized. Philosophically, Bayesian methods, by placing prior and posterior probabilities to statistical interpretation, do some justice to the directionality of time.

Statistical Tests

Descriptive statistics help in expounding the properties of the variables under study. In the literature, most of this description is expressed as defining the measure of central tendency only. However, it must be understood that although statistical tests assess the difference between means and proportions, this assessment is done in context of the variability in the data. All descriptive statistics must be accompanied by a measure of dispersion (variability) of the data. Table 8.1 outlines some of the commonly used measures of central tendency and variability. For example, the means should be qualified with the standard deviation (SD). Some studies present mean values and the standard error (SE) instead of the SD. The SE looks numerically smaller than the SD. If the sample size is properly described, it is possible to derive the SD, but most readers do not bother to do the calculation, and merely interpret SEs as SDs. SE represents that standard deviation of the sampling distribution, and is used to derive the population mean from the sample mean, but it does not reflect the variability in the sample data from which the sample mean is calculated.

Test statistics generally take the form: (Observed value – Expected value)/(Variability). This has been equated to an assessment of signal: noise ratio. The expected value is what one would expect under null hypothesis, and the numerator is a measure of excess value in the data for the factor under study (signal). Similarly, if this excess is consumed by the inherent variability of the data itself (noise) then the distributions of the two factors

TABLE 8.1 Measures of Central Tendency and Data Dispersion of Variables

Type of Data	*Measure for Central Tendency*	*Measure for Dispersion*
Nominal	Mode*	—
Ordinal	Median	Range
		IQ range
Interval (arbitrary "zero")	Mean	SD
	Median	Range
		IQ range
Ratio (Meaningful "zero")	Mean	SD
	Median	Range
		IQ range

*Mode is a common measure of central tendency for all data types.

cannot be distinguished from each other, and a conclusion of no-difference (i.e., failure to reject a null hypothesis) is made.

Most statistical tests involve some assumptions about the factors being examined and about the data. Common assumptions for usual statistical tests include:

- Data come from a random sample (this is often violated, but most tests are robust and not too sensitive to violation of this assumption).
- Independence of observations—the observations are independent of each other. If this assumption is violated, special statistical handling is required, as in using paired or repeat measure analysis.
- Data come from a normal distribution (violations of this assumption are rescued by the central limit theorem if the sample size is large enough).
- Homogeneity of variance between groups being compared. Data are heteroscedastic if the random variables have different variances. Violation of this assumption requires special statistical management.
- The sample size and cell sizes are large. If this assumption is violated, special tests, called exact tests, must be conducted. Exact tests do not rely on the assumption of large samples; they compute "exact" probability of observing the data in the given study if no association was present and are computationally intensive (particularly for large datasets). Because of intensive computation, exact tests have been historically used mostly as a back-up for the chi-square test ("Fishers exact test") when samples are small. However, with increasing computing power and speed in

computers, these tests are now available even for more complex analyses including regression analyses.

- For linear regression, other assumptions are that the relationship between Y and X is linear, and that the dependent variables are not correlated. It is always possible to draw a straight line through a data using linear regression. The central question is whether that line truly represents the relationships under study. Sometimes curved lines better explain the relationships than do straight lines.
- Logistic regression also assumes the relationship between (logit of Y) and X is linear and that all important variables are included.

Statistical tests further assume that the data is optimized properly for missing values, out-of-range data values, and other errors. These are critical steps at the data entry level. Although spreadsheets and some statistical packages may allow direct data entry in a spreadsheet format, such methods are suitable only if the data is very small and every entry can be manually checked for correctness. However, preparing data entry forms in Microsoft Access, Epi-Info, or other program allows programming to control data entry with steps such as programmed range checks for out-of-range values. Double data entry and comparison of the two data sets is another good data practice to minimize data entry errors. Unresolvable errors should be compared against the original data instrument such as a questionnaire, recording, and so on for correctness to eliminate early transcription errors.

Setting Up the Analysis

All analyses should start with careful univariate assessment of each variable and decisions about unusual values. Extreme values may impact results and may be logically handled using a logical decision system. Some approaches may delete few unusually high or low values: recode the extreme values to an a-priori threshold so that useful information is not thrown out, devise special categories for such variables within the routine analyses, or carry out a separate analysis for those observations. In looking for associations or testing for differences, the overall goal should be to assess these associations or differences after accounting for the potential effects of several other factors in the study. Bivariate analysis assesses just two factors, whereas multivariate analysis assesses the differences between two (or more) factors of interest after accounting or adjusting for other important factors that may impact study results. Often, bivariate analysis is used as a starting point to select variables that show strong associations that may be used later in multivariable analysis. In some situations, the primary interest is only in assessing the difference between two or more groups, and bivari-

ate analysis would suffice—such situations are relatively uncommon in epidemiological studies, although they are often invoked for exploratory analyses or in assessing the importance of including a variable in a multivariable analytical plan. Box 8.2 enumerates tests that are commonly employed in epidemiological studies, whereas Box 8.3 describes salient features of some of the more common tests. Table 8.2 provides a rough guide for selecting common types of tests encountered most often depending upon the nature of the variables under study.

BOX 8.2 Commonly Used Tests

- **t-Test:** Compare means (first check for equality of variance). Single sample/two sample; paired (correlated measure)/unpaired.
- **ANOVA (Analysis of Variance):** Compare means for more than two groups. Multiple comparisons between multiple groups can be done after adjusting for multiple comparison (e.g., Scheffe's pairwise test).
- **Repeat measure ANOVA:** ANOVA for correlated measures.
- **ANCOVA (Analysis of covariance):** Compare means of two factors that covary with each other.
- **Mann–Whitney or Wilcoxon Summed Ranks Test:** Nonparametric alternatives to t-test.
- **Chi-Square Test:** Compare proportions; (Observed count – Expected count)^2/Expected count.
- **Mantel–Haenszel Method:** Analyzes the relationship between two dichotomous factors. The Cochran–Mantel–Haenszel (CMH) test compares two dichotomous groups, adjusting for control variables. Helps in assessment of confounding factors.
- **Pearson's Correlation (symbol = "r"):** A measure of strength of the linear relationship between two continuous variables (values vary between –1 and +1).
- **Spearman Correlation:** Correlation analysis for categorized variables.
- **Linear Regression:** Uses equation for a straight line: $Y = a + bx + e$. Used for predicting the value of a response (dependent) variable from one or more explanatory (independent) variables. Models change in the dependent variable given a change in the independent variable.
- **Logistic Regression:** Regression analysis used for outcome variables that are categorical in nature. Most common is binary (dichotomous) outcome (e.g., yes or no).

BOX 8.3 Properties of Commonly Used Parametric and Nonparametric Tests

t-Test

Overall points: Difference between means—single sample/ two-sample; equal sample/ unequal sample; equal variance/ unequal variance (Satterwaite test for equality of variance); Unpaired/ paired.

General form: $t = (\overline{X1} - \overline{X2}) / \text{Sqrt}\,[(s1^2 + S2^2)/n]$. Pooled variance used for unequal sample size tests.

ANOVA

Overall points: Difference between means: more than two groups; Overall H0: $\mu 1 = \mu 2 = \mu 3 = \mu 4 \ldots$ tested by the model "F" test as ratio of between and within group variances—if statistically different, then multiple comparison adjustment using Tukey's/ Scheffe's/ student–Newman–Keul's/ Dunnett's pairwise tests for checking which group is different. Overall $df = n \times (k - 1)$ (k=number of groups).

Comparisons: Planned comparisons are hypothesis specified before the analysis commences; post-hoc comparisons are further exploration of the data after a significant effect has been found.

Bonferroni correction: Reducing alpha by dividing it by the number of comparisons—too conservative.

Factorial ANOVA: ANOVA with multiple factors—main effects and interactions can be tested.

Statistical Interaction: The interaction between two variables is the extent to which cell means depart from an expected value based on addition of the marginals.

Random factor: Contains only a sample of the possible levels of the factor, and the intent is to generalize to all other levels.

Fixed factor: Contains all levels of the factor of interest in design.

Crossed factor: Two factors are crossed if each level of one factor occurs at all levels of the other factor.

Nested factor: Two variables are nested if each variable occurs at only one level of the other variable.

Repeated Measures: Repeat measure ANOVA.

MANOVA: Multivariate ANOVA. Total sum of squares are partitioned—between the groups & error.

(Continues)

BOX 8.3 Properties of Commonly Used Parametric and Nonparametric Tests *(Continued)*

ANCOVA

ANCOVA involves both nominal and continuous independent variables. It can be used for adjusting baseline differences when randomization is not possible and differences between groups exist. ANCOVA can improve the sensitivity of the statistical test by removing variance attributable to baseline variables.

Correlation & Regression

Correlation coefficient: A number between +1 and –1 whose sign is the same as the slope of the line and whose magnitude is related to the degree of *linear association* between the two variables. Most commonly used is the Pearson's product-moment correlation.

Multiple Correlation Coefficient (R): It is derived from a multiple regression equation and its square (R^2) indicates the proportion of the variance in the dependent variable explained by all the specified independent variables.

Least squares estimation: It computes a line so that the squared deviations of the observed points from that line are minimized.

R^2: R-square, the coefficient of determination, expresses the proportion of variance in the dependent variable explained by the independent variable. R^2 = 1 – (ratio of variance of X & Y). Varies between 0 – 1.

Predicted and Residual Scores: The regression line expresses the best prediction of the dependent variable (Y), given the independent variable (X). However, nature is rarely (if ever) perfectly predictable, and usually there is substantial variation of the observed points around the fitted regression line. The deviation of a particular point from the regression line (its predicted value) is called the residual value. The smaller the variability of the residual values around the regression line relative to the overall variability, the better is the prediction from the equation.

Covariance: Covariance of X and Y is the product of the deviations of X and Y from their respective means.

Multiple regression: Multiple regression involves the linear relationship between one dependent variable and multiple independent variables.

Partial F-test: The partial F-test is the test of the significance of an individual variable's contribution after all other variables are in the equation.

Hierarchical Stepwise regression: It introduces variables, either singly or in clusters, in an order assigned in advance by the researcher.

BOX 8.3 Properties of Commonly Used Parametric and Nonparametric Tests *(Continued)*

Goodness of fit (Logistic regression): It is a chi-square test = –2Log Likelihood (–2LL) test. Similar in concept to R^2 in linear regression. The –2LL test can test differences between two hierarchically well-formulated models to assess contribution of variables in the models.

CHI-SQUARE AND OTHER NON-PARAMETRIC TESTS

Chi-square Test

Nominal, categorical data based non-parametric test for differences between proportions. $\chi^2 = \Sigma [(\text{obs} - \text{Expec})^2/\text{Expec}]$ @ df = (row – 1) × (col – 1). Test statistic has a hypergeometric (Chi-squared) distribution. At p = 0.05, for 1df, $\chi^2 = 3.84$. So, $\chi^2 > 3.84$ for statistical significance.

Rule of Thumb: Value of χ^2 needed for significance at 0.05 level = number of cells. For *multiple testing* of sub-group differences, we need to decompose the χ^2 table into smaller sub-tables.

Small cell size: When expected frequency < 5 in any one cell. Yates correction; Fisher's exact test.

Paired data: McNemar's test.

Two factors: Mantel–Haenszel test.

Multiple factors: Log-linear analysis. CMH – H_0: X and Y are conditionally dependent upon third factor Z. The M–H test: tests strength of association by estimating the common odds ratio. Breslow–Day statistic: tests null hypothesis of homogenous odds ratio. Low p-value of B–D test means that stratum specific ORs are not homogenous and common OR cannot be used.

Goodness of fit: Goodness of fit of a statistical model describes how well it fits a set of observations e.g., Pearsons's Chi-square and Kolmogorov–Smirnov tests. It is a one-tail (upper-tail test).

Measures of Association: Phi coefficient, Cramer's V, Yule's Q, Cohen's Kappa & weighted Kappa. Of these, Kappa can be used for larger than 2 × 2 tables.

Ranked Data Tests (Significance)

Two independent groups: Mann–Whitney U- / Wilcoxon Rank Sum tests.

More than 2 groups: Kruskal–Wallis one way ANOVA.

(Continues)

BOX 8.3 Properties of Commonly Used Parametric and Nonparametric Tests *(Continued)*

Repeated measures: Wilcoxon Signed Rank Test / Friedman two-way ANOVA.

Ranked Data Tests (Association)

Spearman's coefficient

Others: Kendall's Tau, Kendall's W, Point–Biserial correlation.

Survival Analysis

Survival analysis models time to event data. Event is any outcome in which the investigator is interested such as occurrence of a disease, side effect, or a symptom. These events are called "failures" in survival analysis. Therefore, survival analysis analyzes failure rates in one or multiple groups and has the ability to adjust for covariates to derive adjusted failure rates. Usually, the failure rates of two groups (such as exposed and unexposed groups) are compared. The general model used in survival analysis is a proportional hazards model which is a "distribution-free" regression model. Although it makes no assumption about distribution of underlying data, the survival time, or the nature and shape of the hazard function, the model assumes that the hazard rate has a multiplicative relationship between the underlying hazard function, the log-linear function of the covariates used in the model (*the proportionality assumption*), and the baseline hazard (and that there is a log-linear relationship between the independent variables and the underlying hazard function).

- Cox proportional-hazards—The proportional hazard model is a generic term for certain models (especially survival analysis) that determines the hazard rate as a function of a set of covariates. The effect of an independent variable on the hazard rate is assumed to be multiplicative. The Cox principle estimates all proportional hazard models without knowing the hazard function or base hazard rate, but estimates the effects of the covariates (though it cannot estimate the effect of time/duration). These models treat time as 'nuisance' factor.
- Kaplan–Meier method—The Kaplan–Meier estimator, also known as the product limit estimator, estimates the survival function from life–time data. The analysis produces a Kaplan–Meier survival curve. The difference in survival distribution between two samples is tested by the Log Rank test (non-parametric test).

TABLE 8.2 Guide Matrix for Selecting Common Tests Under Common Situations

Variable 1/ Dependent Variable	*Variable 2/ Independent Variable*	*Test(s)*
Continuous	Dichotomous, unpaired	Student's t-test
Continuous	Dichotomous, paired	Paired t-test
Continuous	Categorical	ANOVA
Ordinal	Dichotomous, unpaired	Mann–Whitney U-test
Continuous	Continuous	Pearson's correlation, ANCOVA, Linear regression
Categorical	Categorical, unpaired	Chi-square (exact for small sample size)
Categorical	Categorical, paired	McNemar's test
Continuous	Continuous, categorical	Linear regression
Categorical	Continuous, categorical	Logistic regression (dichotomous); Proportional odds (polytomous)

Multiple Comparison

Erroneous inferences arising out of multiple comparisons are a problem in dental literature that is encountered often. With the alpha (α) level set at 0.05 there is a 5% chance of making a Type-I error. For independent comparisons, the pair-wise comparison error rate is given by the formula: Multiple comparison error rate = $1 - (1 - \alpha_{\text{per comparison}})^{\text{number of comparison}}$. If the comparisons are not independent, then multiple comparison error rate is $\leq \alpha_{\text{per comparison}} \times$ number of comparisons. With alpha set at 0.05, if one conducts 10 independent tests, then the multiple comparison error rate = $1 - (1 - 0.05)^{10} = 0.4$ i.e., 40%. Therefore the probability of making at least one Type-I error is 40% and not the alpha level of 5% that an investigator might assume. Consideration of the number of comparison-planned a-priori is critical to a study because sample size can be selected according to the needs of the study to reduce Type-I error rates. In the above example, the sample size needed for a study requiring 10 comparisons would be larger than the situation where only one comparison is being planned.

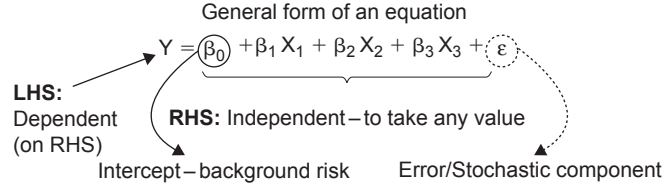
It may be argued that once the data is collected, all information contained in the data is already contained in the properties of the resultant dataset, so post hoc multiple testing should be valid. In post hoc situations, because the sample size cannot be increased to increase the power of the study to accommodate multiple testing, allowances should be made to statistically adjust the alpha level downwards to "back-calculate" adjustments to the Type-I error rate.

A simple way to adjust the alpha level for multiple comparisons is to divide it by the number of comparisons (Bonferroni correction), which is considered to be too conservative. Other approaches include using omnibus testing such as Tukey's test in ANOVA, Student–Newman–Keul's test, Duncan's test and Scheffé's method, among others. With increasing computer speed, efficiency and capacity, multiple testing adjustment based on bootstrapping and Monte Carlo simulations are increasingly being used, especially in large-scale comparison data such as microarrays.

Multivariable Analysis

Multivariable analysis has become the mainstay of epidemiological studies. Although the Mantel–Haenszel method conducts multivariable analyses, modeling has become the ubiquitous method of choice across most studies, perhaps because of explosive growth in computational power, and constant development of software able to handle more complex analyses that includes several other analyses as subroutines, which are outputted in these complex analyses by default anyway. However, predictability of outcomes, the wide applicability of regression analyses, and its ability to handle covariates in an efficient manner has also contributed to the phenomenal growth of modeling as a main method of multivariable analysis. Often, modeling is perceived as "cook-book" panacea of all ills, which of course, is a misplaced conception, unless modeling is applied in an appropriate manner.

Figure 8.2 describes the basic principle of a regression model. An outcome (Y) on the left-hand side (LHS) is explained by a set of variables (X_i) on the right-hand side (RHS) of the equation. LHS is the dependent variable that is not free to take any value, whose value is determined by the RHS.



LHS:
Continuous variable, e.g., white blood count; viral load etc.
Dichotomous variable, e.g., disease yes/no; treatment success/failure etc.

Polytomous variable, e.g., # co-occurring disease; multiple stages of disease

Multivariate vs Multivariable
LHS RHS

FIGURE 8.2 Handling Multiple Variables in Analysis

Variables on the RHS are free to take on any value, and are therefore called independent variables. Each independent variable is qualified by an associated coefficient (β_i) that describes how the variable X_i influences the outcome Y. However, another coefficient, (β_0), exists on the RHS that is not associated with any independent variable. If all independent variables take on the value "0," then effectively, the equation reduces to $Y = \beta_0 + \varepsilon$. Assuming ε to be zero, all of Y is explained by β_0 in such an equation. This is the reason why β_0, which is the intercept of the equation, is defined as the background risk (i.e., the risk of an outcome that exists when all the putative risk factors are zero). A model of the type $Y = \beta_0$ is called an "intercept-only" model, and includes no risk factors. The other element on the RHS, "ε," is the stochastic or random error component. The regression model form, therefore, describes three "groups" of sources of variability that may impact the outcome Y: background risk, risk factors, and random error. In most analyses, we ignore the random error component, and the regression equation used in epidemiological study reduces to its practical form: $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 \ldots + \beta_kX_k$. Although we ignore the random error component conceptually in the models, comprehensive interpretation and explanation of the world phenomenon must place this issue in context to arrive at appropriate answers to questions being asked, especially in etiological modeling. Often, the terms *multivariate* and *multivariable* analyses are used interchangeably. However, there are situations where more than one dependent variable is modeled at the same time, using the same set of independent variables: $Y_i = \beta_0 + \beta_iX_i + \varepsilon$. Therefore, several individual models may be derived from such an equation, each for a specific type of condition in which the dependent variable may exist. Such analyses are called multivariate analyses. Therefore, it is prudent to use the term *multivariable* models for a single outcome explained by multiple variables and the term *multivariate* models for situations where more than one dependent variable is modeled.

Linear regression is generally used for continuous dependent variables, whereas if the dependent variable is categorical, then logistic regression is used. In logistic regression, the logit of the outcome is modeled in terms of a set of independent variables; that is, logit $(Y_i) = \ln(P_i/(1 - P_i)) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 \ldots + \beta_kX_k$. The property of logistic regression that has made it such a useful tool in epidemiology is that by exponentiation of the value of β, one obtains the "odds ratio" for the associated factor. If only one factor exists in the model, then the obtained OR is the "crude OR," whereas if there are multiple variables, then the obtained OR is an "adjusted OR" that is interpreted as the OR after accounting for or adjusting for other variables in the model. Conceptually, this procedure "takes away" the impacts of other variables, and leaves the variable under study with its magnitude of effect only. Often, variables with substantial or statistically significant adjusted ORs are called "independent" risk factors (i.e., the impact continues to remain, independent of the impacts of other factors in the model). Mostly,

the dependent variable in logistic regression is a binary outcome (i.e., yes or no; disease present or absent). However, dependent variables with multiple categories can be modeled using the logistic regression paradigm. *Proportional odds models* and *generalized logit models* are methods that are applicable to multilevel categorical dependent variables. The proportional odds model requires that the dependent variable be ordinal, whereas the generalized logits model can analyze nominal categorical variables.

Oral health data often presents as counts (e.g., bacterial colony counts, number of road accidents, number of insurance claims, number of births and deaths, number of new disease cases of a disease, number of persons with oral cancer vs total population, and number of persons with caries reversals vs person–time for the trial). Such data are well described by Poisson distribution—the Poisson process is a stochastic process that is defined in terms of the occurrences of events. Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate, and are independent of the time since the last event. The probability that there are exactly X occurrences (i.e., X being a nonnegative integer: $X = 0, 1, 2, \ldots$) while the expected number of occurrences is λ, which is sometimes taken to be the rate; that is, the average number of occurrences per unit time. The mean and variance of a Poisson distribution are the same. Person–time in denominator data with count data in numerator can define incidence density function. Incidence density of an event can be modeled using Poisson regression and may be very useful in etiological modeling.

It is possible to use the same dataset without incorporating the observed time period and model the logit of an outcome, or incorporate time in a Poisson regression to model incidence density, or model a time-to event in survival analysis. Alternatively, analyses may even be restricted to basic bivariate assessment or even just univariate descriptive statistics for the same data set. An analytical approach to data, therefore, is predicated upon the type of research question that is being asked and the type of answer being sought. Sometimes, the same data may reveal deeper insights by employing suitable multivariable or multivariate techniques.

Modeling

Analysis of variance (ANOVA) essentially detects differences between groups, but regression analysis also includes the ability to detect changes in the dependent variables based on changes in the independent variables. Therefore, regression analysis can be used for causal assessment and prediction purposes. Whereas ANOVA assesses categorical variables on the RHS, analysis of covariance (ANCOVA) handles both categorical as well as continuous variables in the RHS, as does regression analysis. The basic

purpose of modeling includes control for confounding and finding the best-fitting, most parsimonious, plausible biologically/socially reasonable model, to describe the relationship between the dependent and independent variables.

Depending upon the goal of modeling, three main types of models can be conceptualized.

1. **Prediction models:** These models are developed to predict a dependent variable based on a set of independent variables. Because the future value of the dependent variable is being sought, such models must select variables and categorize them very carefully. For example, if dental insurance status is to be predicted, and the independent variable "employment" includes a category such as "unknown employment status," then the utility of such a model will be minimal, even though it might explain a data set from which it was generated. Prediction models need not necessarily include only causally involved variables because these models are primarily concerned with finding markers for a certain outcome, whose prediction is the goal. However, if effect estimates for each independent variable are being sought to explain their role in the predictability of the outcome, then appropriate confounding control is necessary. In general, past disease has often been shown to be the best predictor of current disease. However, interpretation from such models should clearly state that the prediction sought was for a subsequent disease event, and not the first disease event. A model predicting a subsequent disease event may be substantially different from one predicting the first event. This distinction is often missed in most models that incorporate past disease experience as an independent predictor variable.
2. **Causal models:** Causal or etiological models aim to unearth causal associations between independent and dependent variables. The nature of the goal is such that these models must explicitly seek debate on inclusion of variables in the RHS of the equation. While modeling to ascertain potential etiological mechanisms, Rothman, Greenland and Lash (2008) have suggested that variables falling in a direct causal pathway should not be used as covariates in the same model. Possible reasons for this decision include potential collinearity, possibility of effect-measure attenuation, and induction of spurious causal associations that may result. For example, in a study of oral candidiasis among HIV-positive persons, seeking a causal model, the investigators decided not to include plasma HIV-1 RNA as a variable in models that included blood $CD4^+$ cell count because in an etiological pathway, infection of CD4 cells by HIV-1 and subsequent destruction of the CD4 cells leads to their depleting numbers resulting in low blood CD4 cell count. Because CD4 cell was the main exposure variable, use of plasma viral load was

precluded as a covariate in the multivariable models (Chattopadhyay, Caplan, Slade, Shugars, Tien, & Patton, 2005).

3. **Explanatory models:** These models are data-driven models that are not intended for prediction or for eliciting etiological mechanisms. These models just "explain" an outcome based on whatever variables are available in the data set, and may be used for exploratory purposes. However, just because these models are "exploratory," scientific responsibility to assess associations for the purpose of understanding disease processes and health outcomes is not precluded. Therefore, indiscriminate, thoughtless "throwing" of variables in a model serves no useful purpose. A property of regression analysis is that the more variables are added to the model, the better is the "explanatory" power of the model, assessed through R^2 (linear regression or pseudo-R^2 [logistic regression]). Explanatory models are prone to fall victim to an increasing "explanatory" power by including more variables even though it may be a function of the equation mathematics, rather than a function of the variable(s) under consideration.

The Hierarchically Well-Formulated Model (HWF) Principle

This principle states that given any variable in the model, all lower-order components of the variable must also be contained in the model. For example, let us consider the set of equations below.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 (X_4 \times X_5) \quad (8.1)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad (8.2)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 (X_4 \times X_5) \quad (8.3)$$

Deriving Equation 8.2 from Equation 8.1 is a legitimate modeling application because all variables in Equation 8.2 are also contained in the "parent" Equation 8.1. However, although all variables in Equation 8.3 are also contained in Equation 8.1, the HWF principle is not satisfied in deriving Equation 8.3 from Equation 8.1 because Equation 8.3 includes the "higher-order" interaction variable $\beta_4 (X_4 \times X_5)$, but does not includes its lower-order variables $\beta_4 X_4 + \beta_5 X_5$. Therefore, a hierarchy of equations is not established, and the logic of selecting models by sequentially removing or adding variables would not be satisfied. Furthermore, the interaction variable $\beta_4 (X_4 \times X_5)$ is not interpretable in an equation that does not contain the main variables for which the interaction is being tested (i.e., $\beta_4 X_4 + \beta_5 X_5$).

The process of selecting a final model may use a forward or backward principle. The backward principle is generally used more often. One starts with a "full model" that contains all the variables under study and sequentially removes variables that do not contribute to the model. Statistical software packages present automated solutions that base variable selection solely on p-values. However, there are times when important variables may

not necessarily be significant, or would have turned significant had they been retained till later stages in model building. It is for this reason that automated procedures are not much preferred. Generally, the contribution of each variable in the model and its statistical significance, its importance in the phenomenon, and its potential coaction with other variables are considerations that help in deciding to keep or remove a variable in a model. Guiding rules to select between hierarchical models could include the use of Type-II or Type-III tests, a change in R^2, a change in effect estimates of important or main variables by a certain predecided threshold, or the use of –2 log-likelihood tests (in logistic regression analysis).

Selecting a final model is as much an art as it is science. The "artistic" component in model selection comes from insight into the problem and the ability to understand the potential impact of different variables on each other. For example, if a certain factor is established as an important predictor or etiologic component, but in a specific study that variable turns out to be statistically not significant, then one is faced with the dilemma of either throwing out the variable on a strict basis of including only statistically significant variables, or continuing to include the variable to let it play its role because such has been established already in other studies. Such decisions need to be grounded on the understanding of the phenomenon being modeled, and a strong logical explanation of the choices.

Often, authors prefer to present a "full model" that includes all of the variables, crude effect estimates, and the final selected model in their reports. Such a practice allows readers to make their own assessments about the role of different variables and contributes to transparency in interpreting results. Models that include independent variables must be better than "intercept only" models; that is, the included independent variables must contribute to the model in a significant way. These models should also explain substantial variation in the data and make sense toward the goals of the analysis. Sometimes, models may include statistically significant variables but explain very little variation in the data—interpretations from such models should exhibit caution because important sources of variations exist outside the model structure.

Repeat Measure Analysis

The general principles of model building are applicable to all types of statistical models. An important assumption for modeling is the assumption that observations are independent. The assumption of independence of observations is violated when analyzing data based on multiple observations from the same unit of analysis; for example, factors measured from the patient before and after intervention, multiple visit measurement, and measuring codependent factors such as carotid intimal thickness and saturated fat levels. Such data tend to be correlated in some way. For example, patients'

follow-up information is usually a function of their baseline pathophysiological state or their individual physiology. One way to address this situation is to take the difference of magnitude between the baseline and follow-up measurements, and assess these difference scores as independent scores. Therefore, for a variable measured on a continuous scale, the difference in baseline and follow-up can be assessed for one group by using a one-sample t-test on the difference score, or for two groups by using a t-test as in a two-sample independent test (or ANOVA for multiple groups). However, such an analytical paradigm disregards the inherent correlation between the baseline and follow-up information. Using paired t-test and repeat measure ANOVA (and McNemar's test for assessing differences between proportions) would be appropriate strategies in these situations. Difference scores, however, have been criticized for being generally unreliable under some conditions and not being totally independent of their component scores, which may sometimes lead to biases. They should be used carefully, mindful of the impact of regression to mean in the analysis and the need for corrected difference scores in some analyses. Similar concerns have been raised with relative change scores. Repeat measure ANOVA, assesses baseline–follow-up correlation by time interaction and minimizes regression to mean (Bonate, 2000).

Modeling of correlated data may also follow similar principles. For example, one way of analyzing correlated measures is to model the difference between the observations from the two time periods. For example, modeling a new variable such as "baseline extent of attachment loss–follow-up extent of attachment loss" in a periodontal outcome study may be undertaken. Although this approach may be somewhat useful in some situations, it is not suitable for multiple time-period-based observations. Correlated observations need special handling in multivariate model-based analysis due to the need for repeat measure adjustment in regression methods.

Generalized Estimating Equations (GEE) are being increasingly used in health research to analyze correlated data, especially for categorical data with repeated measurements. This method has a broad application and is suitable methods even for measurements including time-dependent variables, continuous explanatory variables, overdispersed Poisson counts, and the partial proportional odds model (Stokes, Davis, & Koch, 2000).

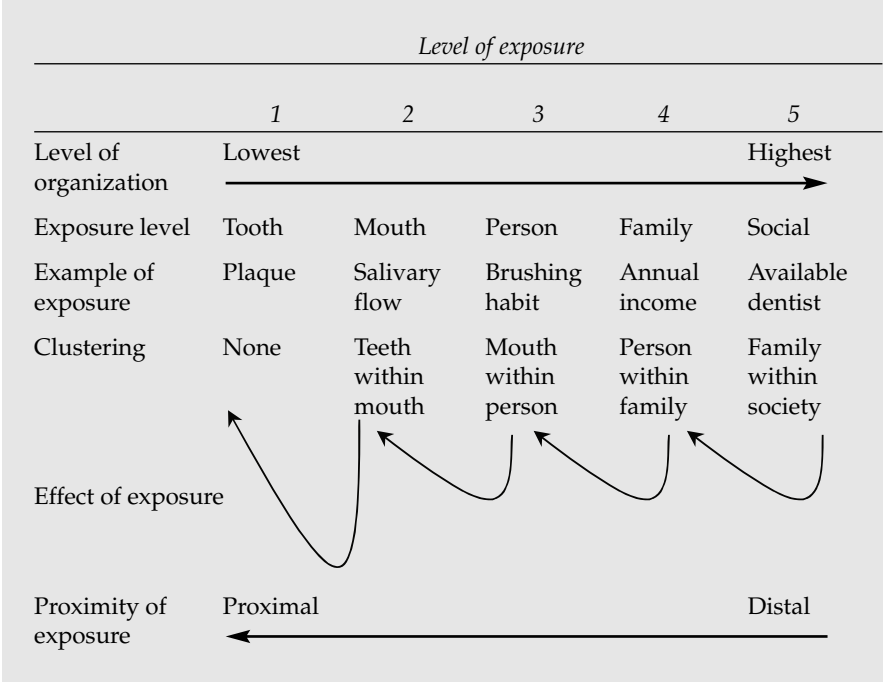
Multilevel Analysis

In linear regression analysis, all effects are modeled as if they occur at a single level. However, different exposures may occur at different levels of organization or clustering. Box 8.4 describes the levels of exposures that occur in the oral cavity. The unit of disease (e.g., the tooth in dental caries) is clustered in the mouth of a person and is directly affected by changes in the local environment such as the dental plaque. However, although saliva is

secreted directly in the mouth, its interaction with tooth surfaces varies according to the position of the tooth in the dental arch and the anatomy of the tooth. Furthermore, although a person from a lower socioeconomic position has a greater burden of dental caries, the effect of monetary deprivation does not directly affect the tooth, but acts through several other potential factors such as low education, inability to take preventive care, and so on, which in turn contribute to the effects on the teeth through complex interaction. The average caries outcome of any tooth type is different from another tooth, and that average is different from the person-level average. Similarly, the individual-level averages between persons are different. Therefore, the level of exposure of risk factors and outcome measurement of the teeth varies and this differential hierarchical level of exposure should be taken into consideration for correct etiological modeling.

Hierarchical linear modeling (HLM), also known as multilevel analysis and random coefficient analysis, addresses these hierarchical exposures.

BOX 8.4 Levels of Exposure for Dental Caries and Clustering in Oral Cavity as the Basis for Multilevel Analysis



Different exposures are shared at different levels of clustering in a hierarchical manner.

HLM may be viewed as a more advanced form of linear regression that allows variance in outcome variables to be analyzed at multiple hierarchical levels. Multilevel analysis has been extended to include multilevel structural equation modeling, multilevel latent class modeling, and other more general models. Although the different intercepts of multilevel factors are not estimated, the variance of the intercepts is estimated. Assumptions of multilevel analysis are similar to those for linear regression: Data should be normally distributed, and the residuals are uncorrelated, and random intercepts and random slopes are normally distributed. The main advantage in using multilevel analyses is that it improves estimations of SEs. Binomial and multinomial logistic methods for categorical outcomes, Poisson, and survival multilevel analyses are available that can also be employed in longitudinal studies and multivariate situations. Multilevel analyses are available in major statistical software such as SAS, SPSS, STATA, R, and in other packages such as MIXOR, MIXREG, HLM, SYSTAT, and EGRET.

Resampling

Resampling is a computer-based process of multiple sampling from the original data, and is used frequently to answer questions that may have taken several large studies (some of which are certainly prohibitive or impossible; e.g., in rare outcome situations). The bootstrap is a general-purpose empirical, seemingly heuristic Monte Carlo simulation approach that can be used for assessing the accuracy of the estimate of quantities such as the mean, the median, the correlation between two variables, or the slope of a regression line for predicting one variable from another that may have been established from experimental data. Bootstrapping means using a special process to perform a task that one would be unable to do in general. Therefore, if situations where multiple sampling or multiple studies may not be possible, resampling from a smaller available data set can be used to draw a distribution of a population parameter. It is permissible to take a random sample from measured data if they are independent and equally probable values through the measurement process.

The fundamental idea of the method is that each measurement taken is considered as an equally likely and valid representation of those which could occur with an infinitely large sample from the process being studied—the values actually measured (e.g., "N"), which are assumed to be derived from a random sample of the whole population. Then, independent random samples are taken from the observed data, until "N" such draws have been made, resulting in a new data set. This new data set is called a bootstrap replicate of the original data. A large number (i.e., 1000–100,000 or more) of such bootstrap replicates are then obtained by the resampling method using statistical software. Once a large number of resamples are drawn, 95% CIs can be calculated from bootstrap replicates. Figure 8.3

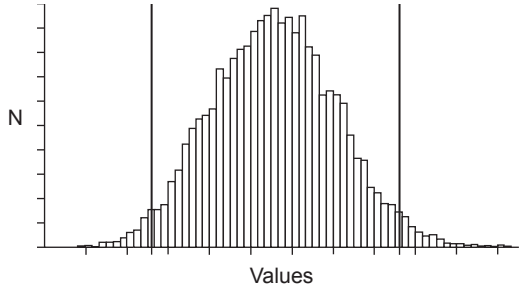


FIGURE 8.3 **Example of Bootstrap Resampled Distribution of Means Across 50,000 Replications** The vertical lines indicate the 95% CI of the resampled estimate.

shows the distribution of mean data from a hypothetical bootstrap resample method replicated 50,000 times.

Markov Process

In probability theory, a discrete-time stochastic process has the Markov property if the conditional probability distribution of future states of the process, given the present state and all past states, depends only upon the present state and not on any past states; that is, it is conditionally independent of the past states (the path of the process) given the present state. The key characteristics of Markov processes are that the current state is conditional upon the previous state, memorylessness of states, the process address time issues, and address random process. Although not much used in oral epidemiology, Markov-based analyses can address important questions about recurring diseases such as oral infections, ulcers, and other events. One of the main points in Markov modeling is that the event modeled should be a function of its previous state only. For example, oral candidiasis in an immune-compromised state presents as a recurring infection. The traditional modeling methods model prevalent disease, or incident disease, as a function of baseline and follow-up variables. However, this strategy can only analyze one event. Past events are usually incorporated in vaguely measured variables such as "history of past disease." Such variables do not take into account the number of past events, timing between events, or multiple recurring events. The Markov process assumes that the current event is a function of the immediate past state and is better able to model recurring processes.

9 Qualitative Research

Qualitative research, common in social sciences, as opposed to quantitative research, is being increasingly used in epidemiology to address certain study aims. Qualitative research allows investigators to understand people's attitudes, behaviors, value systems, concerns, motivations, aspirations, culture, or lifestyles, and is useful in describing a process or phenomenon, or mapping the features of a phenomenon; explaining social phenomena; understanding perspectives, motivations, and frames of references; and generating new ways of perceiving and understanding a social phenomenon. The central premise of qualitative research has been described by Merriam (2002) as follows:

> The key to understanding qualitative research lies with the understanding that meaning is socially constructed by individuals in interactions with their world. The world, or reality, is not the fixed single, agreed upon, or measurable phenomenon that it is assumed to be in positivist, quantitative research. Instead there are multiple constructions and interpretations of reality that are in flux and that change over time. Qualitative researchers are interested in understanding what those interpretations are at a particular point in time and in a particular context. Learning how individuals experience and interact with their social world, the meaning it has for them, is considered as interpretative qualitative approach. If you were interested in studying the placement of a child in foster care, for example, you might focus on understanding the experience from the perspective of the child, the foster family, the agency involved, or all three.

Key characteristics of qualitative research include:

1. Emphasis on understanding the meaning that people uniquely assign to their world and experiences individually.

2. The investigator is the primary instrument for data collection and analysis through verbal and nonverbal means. The subjective interpretations of the investigator are not removed, but explored further to understand how it effects the interpretation of data.
3. It is an inductive process: The investigator searches for possible constructs and pieces of information that can be later tied into a theory, which can then be tested using hypothesis testing analyses.
4. Qualitative research reports are highly descriptive (i.e., a prolix report rather than a brief or numerical one would be a natural outcome of qualitative research).

The premise of qualitative research described by Merriam (2002) above suggests that "fixed single, agreed upon, or measurable phenomenon" do not exist, and therefore, individual subjective interpretations through vivid descriptions, modified by monitored investigator biases, are the key to understanding the world. Whereas such a description may fit into certain limited schemes, this cannot be the complete description of the world. For example, individual perception of risk or benefit, or health belief, does not change the pathophysiology of a disease process unless the disease mechanism involves a substantial psychosomatic component. Occurrence of the same pathophysiological phenomenon in different individuals with the same disease is the basis of etiological pattern recognition, which is verified by recovery from the disease on removal of the causal agency or agencies. Grouping of common patterns together and accounting for the sources of variation allows for effective therapeutic and preventive interventions. This does not negate idiosyncratic responses or low-frequency different side effects, but it provides enough rational basis for efficient use of scarce resources in designing health interventions.

The debate is not whether realities are multiple or fixed single, but whether enough numbers of homogeneous groups can be found so that they would respond uniformly within acceptable limits to efficiently carry out mass-based health programs. Again, the debate is not whether quantitative or qualitative research serves health care best, but is about finding those situations in which qualitative research works, and where it does not. If one has to following the multiple-reality paradigm, then it would be impossible to develop any treatment for any disease because the paradigm structure would mandate individualized testing for every intervention before institution of therapy. Furthermore, under such a paradigm, the realities keep changing at every point in time, and because we do not control the flow of time, by definition, the philosophy professes an impossible situation even for qualitative research because by the time the research study is over, all realities of every individual and investigator would have completely changed, and the study would have become invalid based on its own premise.

Qualitative as well as quantitative research submits to the central dogma of science that the universe is interpretable, and differ from each other in their applicability to specific situations, but not on the philosophical objectivity or subjectivity of reality. They emphasize different aspects of reality and therefore complement each other. Qualitative research differs from quantitative research in that its samples are convenience samples; focus is on certain characteristics of the samples about which more information is sought; the investigator's personal bias directs the studies; data analyses are done differently, and the aim is to discern common themes across the sample (*despite differing individual realities in flux*). Qualitative research aims to understand the process of behavior and social changes—how, why, what, where, and when decisions are made, in contrast to quantitative research that, in an equivalent situation, may assess how much decisions impact groups or differ between groups. Therefore, qualitative research provides wide latitude in asking questions, is helpful in developing insight into issues, and is not limited by sample size or power requirement.

Qualitative research is better viewed as hypothesis-generating rather than hypothesis-testing. For example, if one wishes to understand how much importance people give to their dentition, the investigator can use an initial open qualitative format and can simply record all answers given to a generic question: "What, in your opinion, is the value of dentition?" and can ask other probing questions. Once all answers are collected, the investigator may look for phrases and key words used by the participants and then develop domains and/or question items based on those responses. Alternatively, in assessing cultural competence of a work setting, in-depth open-ended interviews designed as a qualitative investigation may record the full responses of participants and then extract a list of items referred to by different participants that might help in arriving at a general understanding of the needs about the cultural competence intervention. Qualitative data is in the form of words rather than numerical statistics. The investigators' impression takes an important position in such research. Qualitative research may be used to provide additional insight into individual variations, which might then be assessed as a within-group variation.

In sociological and social work research, where individualized parameters are the main focus of research, qualitative research serves very well. For example, if individual low-income families in a certain area have transport problems that prevent them from utilizing sealant-application programs, then this "experience" can be captured by qualitative assessment. However, reduction in the burden of caries of the same community following the prevention intervention cannot be demonstrated through qualitative studies, although a change in their coping strategies may be demonstrated using qualitative research. Although different authors have classified qualitative

studies in different ways (i.e., almost 45 "types" have been described), commonly used qualitative studies may be classified as interpretive, phenomenological, grounded theory, case study, ethnography, narrative analysis, critical research, and postmodern research studies.

Interpretation in Qualitative Research

These studies try to understand a phenomenon from the point of view of participants. Recurring patterns and common themes are identified in data that are then described vividly. For example, Carvalho, Costa, and Marcelo (2008) studied 20 dental students (5 men and 15 women) in a Brazilian dental school to understand their perspectives about the importance of basic life support (BLS) medical emergencies in dentistry. They conducted in-depth one-to-one interviews (20–40 minutes each) with these students. Questions asked in the study included "If I was your patient and I suddenly became unconscious during dental treatment, what would you do?" "Have you ever heard of BLS?" "Have you ever faced a life-threatening situation with anyone?" "How do you consider dentistry within the health sciences?" "Do you think that dentists should be able to perform BLS?" The investigators identified two themes describing the dental students' perspectives. The students perceived that dentistry should focus on the whole patient and not be limited as a sectorial oral-cavity-based discipline, and they felt insecure about handling medical emergencies and were not able to perform proper BLS techniques.

Phenomenology

Phenomenology is a philosophical method that analyzes objective phenomenon as a function of experiential–conscious interpretation. A phenomenological study focuses on the nature, essence, and structure of the experience. To an extent, phenomenological studies use a reductionist approach because they try to show that complex meanings are "built" of fundamental simple units of experiences. This approach assumes that there is a core of the shared experience. Philosophically, this assumption implies that an objective "core" reality of the experience exists, which is at variance to the premise that no fixed single, agreed upon, or measurable phenomenon exist (Merriam, 2002), and only multiple realities are the essence of qualitative research. In phenomenological studies, experiences of different people are aggregated and common themes are extracted. For example, a recent study reported a qualitative and quantitative assessment of an ePortfolio assignment in the operative dentistry clinical simulation module where the qualitative part reported on student self-reflections on the ePortfolio experiences (Gardner & Aleksejuniene, 2008). The qualitative research component of the

study demonstrated that students valued ePortfolio learning as a positive experience.

Grounded Theory

Grounded theory is an investigative method that emphasizes the generation of theory from data in the process of conducting research. This type of research first describes the outcomes, generating the theory that is "grounded" in the collected data from which it arises, and then also verifies this resultant theory. The grounded theory contains all components of a theory (i.e., description of categories and their properties), and hypotheses about the relationships between categories and their properties. These hypotheses, however, are not tested.

For example, Freeman and Stevens (2008) used grounded theory procedures and techniques in a study by obtaining qualitative data from mothers to answer the question: "Why do mothers persist in prolonged bottle feeding?" They conducted in-depth interviews with 34 mothers of children with nursing caries. Their study suggested that mothers used the feeding bottle to "purchase" time by silencing crying children. This conclusion contributed to developing the theory that prolonged bottle feeding time helped mothers purchase time that is used to increase "babyhood closeness" between mother and child.

Ethnography

Ethnography assumes that properties of a system cannot be comprehensively and accurately understood independently of each other (i.e., the system has to be studied in a holistic manner). This approach has been commonly used in cultural anthropology studies. Sociocultural interpretation of data is a mandatory requirement for an ethnographic study. Ethnographic studies are completely defined by the way data is interpreted. For example, Muglali, Koyuturk, and Sari (2008) recently reported a study conducted in Samsun City and neighboring villages in Turkey in which they assessed folkloric beliefs and superstitions held by parents to understand the variety of tooth extraction methods for children's teeth. They found that folk beliefs were concepts handed down through tradition, were strong and existed in rural and urban areas, and appeared across education groups among participants.

Narrative Analysis

In narrative analysis, a first-person detailed narrative personal account of events is recorded as told by the participant and is then analyzed. This may

be viewed as a personal story of the participant. Narrative analysis is also known as autobiography, autoethnography, biography, life history, life narrative, and oral history. In this type of research, the focus is on the details of the personal account, experiences, and context. In many ways, narrative analysis follows the idea of recording the chief complaint of patients when recording their medical history because by definition, a chief complaint is recorded verbatim, as reported by patients in their own words. However, narrative analysis involves more detailed responses and analysis of the statement(s), unlike a chief complaint that is only the initial recording of patients' problems.

Analysis of the narrative stories is conducted either as biographical (e.g., relation of person to the society, gender, socioeconomic position), psychological (e.g., internal thoughts and motivations), discourse analysis (e.g., tone, pitch, pauses while speaking, writing style), or a combination of these approaches. For example, at Karolinska Institute in Sweden, Nordenram, Norberg, and Bischofberger (1994) assessed ethical problems of dentists caring for demented patients who forcefully refuse dental treatment, giving rise to feelings of frustration and anxiety in the treating dentist. The authors recorded a semistructured interview with four hospital dentists and studied the recorded narratives to focus on each interview as a whole to analyze the ethical conflicts and dilemmas of the treating dentists. The study found that hospital dentists' concepts in conflict situations were nebulous, which suggests that education and training programs focused on handling of ethical conflicts in dental management of severely demented patients are professionally needed for practicing dentists.

Case Study

Case studies, which are detailed descriptions of clinical cases or procedures, are very common in oral health research, and are perhaps the most useful forms of qualitative research for new or rare phenomena for which little prior information is available. One of the highly celebrated case studies is the report of clinical findings in eight young homosexual men in New York with Kaposi's sarcoma showing some unusual features (Hymes, Cheung, Greene, Prose, Marcus et al., 1981). The unusual occurrence of Kaposi's sarcoma in a population much exposed to sexually transmissible diseases led the investigators to suggest that such exposure may play a role in the pathogenesis. These cases and subsequent similar case reports added to the literature a new disease form that is now described as HIV/AIDS. Unless there is something very unusual to report among HIV/AIDS patients, such case studies are generally not considered worth reporting today. As the above example suggests, case studies can often contribute greatly in understanding unknown phenomena, including natural history of diseases.

Critical Research

Critical qualitative research is based in the theory of knowledge. It examines our thinking that is rooted in social, cultural, and psychological assumptions, and critiques those assumptions to expand the horizons of the limiting assumptions or remove those limits to expand understanding and develop more informed thinking. This form of research is common in the field of education; for example, traditional clinical dental teaching that addresses patients' clinical condition as an isolated phenomena led to a more holistic, "comprehensive care" approach by viewing the disease condition as a part of the patient (Holmes, Boston, Budenz, & Licari, 2003). The current development of a "community-based dental education" approach is another example of the outcomes of critical research. Dunning, Durham, Aksu and Lange (2008) explored "the little-understood process of evaluating the performance of assistant and associate deans at dental colleges in the United States and Canada." To identify the methods, processes, and outcomes related to the performance appraisals of assistant and associate deans in dental schools, the investigators surveyed deans and associate deans using closed and open-ended questions. They critiqued several important issues such as differences in perspectives on performance reviews, the importance of informal feedback and job descriptions, the influence of an assistant or associate deans' lack of tenure, and the length of service of deans.

Postmodern Research

Postmodernism is the concept of rejecting the idea of the self as a processor of true characteristics and accepting a plurality of voices and deconstruction of what one believes to be true to make way for multiple realities (Hertlein, Lambert-Shute, & Benson, 2004). As the name suggests, postmodern research, also known as poststructural research, does not follow any specific defined research structure, format, or rhythm of plan. Perhaps it is a natural outcome of the central premise of qualitative research that views the world to have no fixed single, agreed upon, or measurable phenomenon and relegates events to individually perceived multiple realities. By definition, such a paradigm should decline to classify or group people in any way because each component individual member of the group has his or her own separate perceived reality that is different from every other member. The very use of any type of classification or grouping in postmodern qualitative research, therefore, becomes an antithesis to itself.

> In contrast to the 'modern' world, where reality is predictable, research is scientific, and there are assumed to be universal norms, for truth and morality, the postmodern world is one of uncertainty, fragmentation, diversity, and plurality. There are many truths, and all generalizations, hierar-

> chies, typologies, and binaries (good/bad, right/wrong, male/female) are 'contested,' 'troubled,' or challenged. (Merriam, 2002)

The foundations of postmodern research as a scientific method of inquiry have been questioned. Postmodern research appears as a "crisis of representation provoked by postmodernism and challenges some of our most venerable notions about scientific knowledge and truth" (Ellis, 1997). In education, even a novel has been accepted as postmodern research methodology in some institutions (Eisner, 1996).

> The core of postmodernism is the doubt that any method or theory, discourse or genre, tradition or novelty, has a universal and general claim as the 'right' or privileged form of authoritative knowledge. Postmodernism suspects all truth claims of masking and serving particular interests in local, cultural and political struggles. . . . No method has a privileged status. The superiority of [social] science over literature—or from another vantage point literature over [social] science [research] is challenged. (Richardson, 1994)

Perhaps such "research" methods work well in the arts, creative writing, or may even be the trigger for creative thought experiments, but whether such research truly represents the scientifically verifiable method of inquiry is questionable. Hertlein et al. (2004) investigated doctoral students' understanding of postmodern family therapy research. The study reported that the "students indicated that postmodern research is characterized by its flexibility in methods, translates into a 'new way' of conducting research, and creates a natural bridge between family therapy researchers and clinicians." The concern in these results is that neither the investigator nor the students clarified how they defined scientific research. There appears to exist some blurring of conceptual boundaries because all "research" is somehow assumed to be scientific research.

All inquiry may not necessarily follow scientific methodology, but practitioners of alternative methods of inquiry should make it clear in their reporting. For example, commentaries that are often published in scientific journals are clearly marked to indicate that the contents are essentially personal views, and not necessarily scientifically verified results. The same goes for perspectives, views, and letters to the editor or similar sections in different journals. Interpreting any inquiry as a scientific inquiry should be questioned, and readers of reports should be made aware about the role such methods play in science.

Perhaps the time has come to clearly demarcate the use of the term *research* from *scientific research*. In the basic sciences, this may not be much of an issue, but in population sciences such as epidemiology, health services, health behavior, and applied biostatistics the distinction needs to be clear. One might suggest that the use of thought experiments is legitimate scien-

tific inquiry, and the investigator of the experiment is free to vary everything, including his or her interpretation of the outcomes, an example of postmodern qualitative research. Lest it be construed differently, the most venerated scientist, Albert Einstein, used thought experiments actively (the most famous being its role in his special theory of relativity), but he did not present those thought experiments as scientific results or truths. Instead, he developed rigorous scientific proofs for the phenomenon under study to which his thought experiments may have provided exciting insights. A good example of how thought experiments may be used to advance scientific inquiry is demonstrated in the Einstein–Podolsky–Rosen paradox paper (Einstein, Podolsky, & Rosen, 1935), where the authors challenged traditional ideas about the relationship between the observed values of physical quantities and the values that can be accounted for by a physical theory, but demonstrated a mathematical proof of the results of their thought experiments. Such work is inherently scientific and should not be confused with postmodern qualitative research.

Data Collection and Analyses in Qualitative Research

Data Collection

Qualitative research employs a variety of methods including in-depth interviews, focus groups, observation, and documentary analysis. The choice of method depends on a number of issues such as the research question, practical issues such as ease of access, relative importance of social context, depth of individual perspective required, and sensitivity of the subject matter (Bower & Scambler, 2007). In qualitative research, the use of two or more methods in conjunction with theory to look at the same issue from different perspectives is called *triangulation*.

Three main data sources in qualitative research are *interviews* (structured/semistructured/unstructured; open ended/close ended), *documents* (written, oral, visual sources of information; cultural artifacts, and so on that may be available in the public domain; secured archives, or personal documents and communications; food inventory record logs; activity logs; life history narratives), and *observations* (usually invisible observer undisclosed to the participant, but active participatory observers are also used in some situations), use of which are determined by the study question. Several sources of data may contribute to one study. In such situations, one primary and one or more secondary data sources are identified. For example, in a study of oral health awareness and hygiene habits of nursing home residents, investigators might interview the residents about their awareness,

and stated importance of oral hygiene and habits, and then ascertain their actual oral hygiene practices through nursing home attendants and any accessible health records. Qualitative research encourages researchers to use more than one method to unearth information about the participants in their study. Data may be collected using traditional methods of one-on-one interviews, mailed interviews, e-mailed/online data interviews, observations in chat rooms, Web forums, and discussion boards.

Data Analysis

Qualitative research data analysis is not usually concerned with the biases and confounding that play a major role in quantitative research. Furthermore, issues such as sample size, power, multiple comparisons, and the like are not impediments to qualitative study design and data analysis. One of the characteristics of qualitative research is that data analysis can start as soon as the data for the first observation comes in (i.e., data analysis proceeds parallel to data collection, simultaneously). This also means that results are being constantly generated as the data collection goes on and the availability of real-time results become a function of the speed of data collection and analysis. Data collection uses an inductive strategy (i.e., inductive reasoning makes generalizations based on individual instances). For certain types of data, there exist software that can data-mine and analyze—so such mechanisms, if installed in an automated environment (such as batch processing), can produce automated, ongoing real-time results. If certain results indicate that a modification of data collection is needed, then real-time information can be used in feedback and feed-forward loops to alter the data collection and/or analysis process to efficiently redirect the study. Some commonly employed data analysis methods in qualitative research are as follows:

Psychological strategy: Psychological survival strategies such as placing "protective shields," making jokes, distancing from enemies, "blocking out" problematic experiences, and so on are used in helping the participant to respond to investigators' queries.

Sociolinguistic strategy: Recorded and linguistic data are used together with data about the social setting where the data was recorded so that each can be interpreted within context of the other.

Literary strategy: Involves the use of self-questioning while reading and assimilating literary works.

Constant comparative method: Combines an inductive process with a simultaneous comparison of all observed events—events are recorded and classified and are simultaneously compared across other data and categories.

Epoch: The state where all belief in the existence of the real world is suspended by the participant. It requires the participant to set aside *all* preconceived notions and beliefs before responding. Epoch assumes people can separate their personal knowledge from their life experiences.

Bracketing: Setting aside preconceived notions and beliefs *about a particular assumption* by the participant, and responding as undisturbed as possible by personal knowledge and experiences. The participants are therefore required to "bracket" their previous thoughts and experiences about the phenomenon and suspend those. Bracketing also assumes people can separate their personal knowledge from their life experiences.

Imaginative variation: Participant is required to imaginatively manipulate different features about the phenomenon from different vantage points, such as opposite meanings and various roles, and distinguish between important and unimportant features.

Deconstruction: The investigator searches deeply into the recordings and data of the participant and seeks to expose deep-seated contradictions in a work by delving below its surface meaning. The investigator takes the stand that words relate only to other words, and the true meaning of statements cannot be directly deciphered from the actual text other than by searching for meaning.

Rhizoanalysis: Reality is viewed through a tree metaphor with interconnected roots and branches that are constantly changing according to each others' responses and are constantly evolving—there are no "fixed" events or outcomes.

Genealogy: The study or investigation of ancestry and family history and involves looking for small inconsistencies, discontinuities, and recurrences.

Archaeology: Historical study into past cultures and their practices.

Schizoanalysis: The reverse of reductionism (i.e., a simple situation is made complex). Four circular components are involved: generative, transformational, diagrammatic, and machinic components.

Data Analysis Software

Because qualitative research uses unstructured information, data analysis is a tedious process. However, as computer technology has advanced, several software solutions have been developed to help qualitative research by managing, shaping, and making sense of the collected information quickly and easily. Computer-assisted qualitative data analysis software (CAQDAS) tools help in classifying, sorting, and arranging information; discover patterns; identify themes; and glean insights from the data. CAQDAS helps to automate and speed up the coding process, permits assessing relationships in complex ways, and provides a structured method for writing and storing

memos to develop analyses. Furthermore, various theoretical concepts and structures can be tested by varying analyses in different ways. However, software is only as good as the way it is used. Automated processes tend to generate too much dependence on "fishing" exercises without understanding the data. Several software products are available for conducting different types of qualitative analysis. Information about these products can be obtained on the Internet through the American Evaluation Association's website at http://www.eval.org/Resources/QDA.htm.

10 Survey Sampling and Surveillance

The core functions of public health include assessment, policy development, and assurance. Carrying out these functions requires regular collection and dissemination of data on health status, community health needs, disease prevalence and incidence, and risk factors of disease states. Public health assessment goals aim to understand disease distribution in populations and explain the causes or determinants of the disease so that a policy can be developed and implemented to control and prevent the disease. In general, surveillance implies close observation of person(s) or group(s). Surveillance of people's health states take on an important meaning in the practice of public health because it allows monitoring of diseases and prompt response to change in rates at which diseases occur in populations. Public health surveillance may be defined as ongoing systematic collection, analysis, and interpretation of outcome-specific data for use in the planning, implementation, and evaluation of public health practice (Thacker & Berkelman, 1988). Surveillance should not be viewed as an end unto itself, but as a tool to help policy development and program monitoring. Due to its ongoing nature, it also needs to be refined and modified to adapt to the goals of public health programs. The need for surveillance is closely integrated with the timely dissemination of these data to those who need to know for effective, evidence-based decision making and rational establishment of priorities. Timely and accurate data also facilitate earlier epidemic detection and control.

Clear delineation of oral health surveillance occurred with the development of oral disease surveillance systems in the late 1960s by WHO, with a focus on dental caries among children, leading to the first ever global map demonstrating prevalence of dental caries among 12 year olds showing high prevalence of caries in industrialized countries and generally low values in developing countries (Barmes & Infirri, 1977). Thereafter, the utility and

need for oral health surveillance data was felt worldwide. Although almost half a century has passed since then, the development of a robust and comprehensive oral health surveillance system (or systems) across the world is not yet uniform, as it is concentrated only in a few developed countries.

Survey Sampling

Surveys have become an accepted and established method for collecting summarizable information on a population and are widely used in public service, private enterprises, and scientific research. The key to successful information collection from surveys lies in the application of appropriate sampling methodology. Essentially, samples are a subgroup of the target population. A common type of sample is a convenience sample that is selected because it may be easily obtainable for a given study. Such a sample usually is not representative of the target population and is considered a biased sample. In order to draw inferences about a target population, the sample must represent it faithfully. This representativeness may be achieved through a variety of techniques based on statistical principles. Samples may be probability samples or nonprobability samples. In a *probability sample*, the probability of selection of any participant in the sample is known or calculated and this probability is not equal to 0 (known, nonzero probability). In a *non-probability sample*, the probability of selection is not known, therefore, the probability of getting a particular sample cannot be calculated. Nonprobability samples do not meet the stringent scientific criteria and do not represent a target population. Based on non-probability samples, the inferences one can draw about populations are very limited. *Sampling fraction* is a term that describes the proportion of the sample size selected from the total population available for selection (e.g., if we select 150 students for an oral health exam from a school of 1000 children, then the sampling fraction is 15/1000 = 1.5%). *Sampling frame* is the framework from which we choose to do the sampling. For example, if we choose to conduct an oral health-related needs assessment survey in a university, then a list of all students, staff, and faculty of the university will be the sampling frame.

Random Sample

Box 10.1 summarizes the general properties of different types of commonly used samples. The simplest sampling design is the *simple random sample* (SRS) where a certain number of participants are selected randomly from among the population (i.e., derived from a random number table or computer-generated random list). In an SRS, every subject in the sampling frame has an equal probability of selection and contributes the same weight to the sample. It is important to emphasize that the selection probability of a subject into an SRS is not only known, but is equal in sampling without re-

placement. In the scientific literature, one often comes across convenience samples derived from hospital attendees erroneously referred to as SRSs; selection probabilities in this case would be unknown. Correct representation of sampling methodology is a key to valid inference making. *Systematic samples* select an initial number by some mechanism or randomly, say X, and then samples every Xth subject from a sampling frame based on a predefined starting point. Modification of the process may be done by selecting the first subject using a random number generator or another mechanism. Further modification may be done in selecting the Xth subject depending upon numerical or other characteristics of the sample to be selected. Subjects have the same selection probability in a systematic sample as in an SRS given equal sampling requirements and equivalent sampling frames.

For example, let us assume that an investigator needs to conduct an oral health needs assessment survey of a small town with a population of 10,000 using an SRS. The town is fairly homogeneous in its demographic attributes (e.g., age, sex, race/ethnicity, education, and employment of residents). The survey team decides to include 10% of the population as the sample (i.e., sampling fraction = 10%; sample size = 1000). The team can obtain a listing of all persons from the township office after obtaining the necessary clearances, and then transfer those names into a computer and instruct the preferred software package to randomly select 10% of the sample. Most commonly used statistical software packages can do this. The resultant list indicates the sample participants who would be approached for participation in the survey to be conducted. Adjustments for nonresponse and other errors can be made at the sampling stage; for example, by selecting a sample that is larger than the required sampling fraction.

Statistically, an SRS is the easiest sampling method and works very well in homogeneous populations (i.e., demographically similar within the defined geographic areas). An SRS will produce a representative sample provided it is sufficiently large, as it requires little knowledge of the population. An SRS assumes that populations are infinite and homogeneous for all attributes. However, it has several drawbacks. In practice, achieving a correct SRS is difficult because it requires an accurate sampling frame consisting of a comprehensive list of the whole population, and the persons may be scattered over a wide geographic area that may not be easily accessible given logistic and budgetary constraints of the study. If a population is not distributed homogeneously (e.g., clusters of persons with certain characteristics such as race/ethnicity, income levels, and cultural practices aggregate in some places), an SRS may not produce a representative sample. Thus, to get a representative sample from a diverse population spread over a large area can be a difficult proposition. To account for these drawbacks of an SRS, several alternative sampling designs have been developed. A systematic sample is easy to analyze and provides better precision than an SRS although systematic changes in a population may lead to biases.

BOX 10.1 Properties of Sampling Methods

Sampling Probability	*Sample Type*	*Fundamental Properties*
Sampling probability unknown (non-probability sample)	Convenience sample	A sample is selected because it is conveniently available. Also called haphazard/accidental sample.
	Purposive sample	Sample selected based of opinion of expert(s). Also called expert/judgment sample.
	Quota	A quota (or a proportion) for inclusion of a particular group is determined by some criteria, and within this group, anyone is selected.
	Snowball	The first participant refers a friend or another person who refers another person and so on.
	Case study	Participants are limited to particular cases under study.
Sampling probability known and is nonzero (probability sample)	Simple random sample	Selected subjects have equal selection probability. The population is treated as generally homogeneous. Statistically simple and efficient. May create logistic problems.
	Systematic sample	Samples are selected in a fixed systematic manner starting from an initial selection point.
	Stratified sample	Population is divided in strata and sampling is done within each strata. All strata are represented in the total sample.
	Proportionate stratified sample	Strata sample sizes are proportional to strata population sizes (uniform sampling fraction).
	Disproportion-ate stratified sample	Strata sample sizes may be selected disproportionately to strata population sizes—useful when groups have to be oversampled to obtain enough numbers for making a meaningful and valid estimate.

(Continues)

BOX 10.1 Properties of Sampling Methods *(Continued)*

Sampling Probability	*Sample Type*	*Fundamental Properties*
	Network/multiplicity sample	Subjects picked through probability sampling methods are asked to identify friends/relatives/known persons who have certain attributes of interest—done using certain counting rules.
	Dual frame sample	Two sampling frames are used in sampling and then these samples are combined. Each sample can focus on a sample with specific attributes. Careful analytical methods are needed.
	Cluster sample	Clusters are identified and a sample of clusters is selected. Each cluster should be heterogeneous.
	Complex sample	Combination of stratified and cluster design.
	Two-phase sample	Some information is collected from subjects in one sample (phase I), and a subsequent second survey collects other information from a subsample of the first phase sample (phase II). If done for more than two phases, it is called multiphase sampling.
	Replicated sample	Several samples are collected from a population using identical designs, called replicated samples. These samples are then combined into a large sample. This allows measurement of variable nonresponse, and also compares estimates across each replicated sample.
	Panel sample	A longitudinal survey where data are collected from the same subjects over time across several rounds of data collection, each round forming a panel of data. Tracking of each participant is a problem. It also does not account for populations changing over time, and sample selected may not continue to be representative of the changed population.

Stratified Sample

The sampling method can be modified to divide the population into groups by defined shared attributes within which sampling can be done separately within the groups or strata, and then these group-based subsamples are combined to create a total sample that is representative of the population. A s*tratified sample* categorizes the population by certain important criteria (e.g., by race/ethnicity, age, income, area of residency, or other attributes) and then selects samples within each of these strata. These within-strata samples may be drawn as SRSs or in other ways. This method is useful when the population is not homogeneously distributed based on certain characteristics that may be deemed important based on the associations being investigated. For example, it may be useful to stratify populations based on race/ethnicity to see impacts of cultural attributes. To be able to draw a stratified sample, one needs to know the population size of the strata and whether it is possible to select samples of a given size within strata. Also, there must be at least two observations in each stratum for obtaining an estimate and a standard error (Kalton, 1983). A stratified sample allows one to capture information from select groups, enables disproportionate sampling, and almost always results in greater precision if the strata are homogeneous for the characteristics under study. Compared to an SRS, administration logistics of stratified sampling are easier, and it also provides a better representativeness (extended coverage) of the population. However, it also adds complexity to some logistic issues and analytical methods. If inappropriate analytical methods are used, it might lead to an invalid estimate of standard error. Stratified sampling requires a good understanding of the population prior to sampling so that appropriate population-specific strata can be determined.

For example, for the hypothetical town described in the SRS example above, suppose the survey team wants to estimate the average number of school-hours lost per week due to dental-related problems by children attending the only elementary school in the town. The investigators believe that the number of hours lost will vary between classes and decide to conduct a stratified random sample instead of an SRS. They stratify the population of students in the school into five strata: children in grades 1 through 5. Once the strata are decided, the investigators select a random sample of students from each class-grade as participants. This strategy (i.e., stratified random sample) allows for increasing the probability of selection of students from each class-grade compared to an SRS from the whole school, and estimates are expected to be more precise than with an SRS.

Samples may be selected in one stage or by using more than one stage (multistage samples). An SRS is a single-stage sampling technique. A two-stage sampling technique is often used in dentistry to select schools in the first stage, and then subsequently to select children from each school in a sec-

ond stage. In this case, the children are the units of analysis. Multiple stages can be added while making the sample selection. Examples of multistage sampling include National Health and Nutrition Examination Survey (NHANES), Medical Expenditure Panel Survey (MEPS), and Behavioral Risk Factor Surveillance System (BRFSS) (NCHS, 2008; AHRQ, 2008).

Cluster Sample

In a population, people usually exist in groups that share common characteristics such as neighborhoods, cultural attributes, and rural or urban centers; or clusters that share common environment exposures such as fluoridated or non-fluoridated water sources. Such groups (i.e., clusters) challenge the assumption of an SRS that populations are homogeneous. To get a representative sample, it is important to account for clusters; sampling methodology can be modified to obtain greater numbers from different clusters to improve the precision of the estimates. Therefore, clusters can be selected for the characteristics of their subjects that are important for the study. *Cluster sampling* includes a sample of well-defined clusters. Sometimes, the strategy for cluster and stratified sampling can be confusing. In stratified sampling, a separate sample is selected for each group or strata, whereas in cluster sampling, a sample of clusters is selected and some or all subjects of the selected cluster are included in the sample. In the selected clusters, if all subjects are included, then it is a one-stage cluster sampling (i.e., selecting of clusters is the only stage and all subjects are included by default).

For example, in a survey for estimating caries prevalence, if schools are selected based on neighborhood status, and every child in the school is selected, then the sampling scheme would be a one-stage cluster sampling. Continuing with our hypothetical example, the state dental director noticed the success of the investigators and asked them to conduct a statewide survey to investigate the utilization of dental services by the residents, but as usual, funds were very limited and the survey team was asked to provide the best possible estimates for the least amount of money. The investigators immediately realized that the state population is not as homogeneous as the town population; that is, people of different race/ethnicity, income, and education levels live in different proportions in different places. The investigators decide to conduct a single-stage cluster sample survey. They identified "clusters" of groups of counties that looked similar to each other, but looked different from other clusters. They then select a random sample of these clusters and decided to include every person in a cluster as a survey participant. This made the job easier logistically compared to visiting people all over the state (as in an SRS); they would visit people living close by, thus making the most efficient use of resources. At the same time, this strategy also provided better representation of people with the attributes that were identified for clustering.

If, however, after selecting the clusters (selection stage one), only some of the subjects are selected from within the clusters (selection stage two), then a second selection stage is introduced, making it a *two-stage cluster sample*. In the above example, if the investigator chose to select a sample of children from the selected school clusters (e.g., third grade only), then the sampling scheme would become a two-stage cluster sampling. Similarly, clusters may be selected based on different characteristics using multiple selection stages—such sampling schemes are generically referred to as *multistage cluster sampling*.

Stratified and cluster sampling are done to improve the representativeness of the samples. In stratified samples, because strata are all represented in the sample, they should preferably be *homogeneous* to minimize within-strata variation. On the other hand, because cluster sampling samples a set of clusters, it is important that the selected clusters also represent the unselected clusters. Therefore, in contrast to the requirement for strata in stratified sampling, in cluster sampling, the clusters should be as *heterogeneous* as possible so that characteristics of unselected clusters are incorporated. Cluster sampling makes sampling efficient by easing logistics, saving traveling time, and reducing costs. However, if clusters are very homogeneous, then they are less likely to represent the target population. Compared to an SRS of the same size, cluster samples produce larger sampling error and must be correctly analyzed to obtain valid estimates of standard errors. One of the main arguments favoring cluster sampling is the cost efficiency it produces. The trade-off between representativeness of samples, costs, and estimate precision is a constant source of strife for setting up periodic samples and surveillance systems.

Sampling schemes of national surveys such as BRFSS, MEPS, and NHANES are multistage cluster samples (NCHS, 2008; AHRQ, 2008). A good example of cluster sampling in oral epidemiology is the sampling for estimating the oral health attributes of children. Because there is strong evidence that dental caries aggregates disproportionately severely among children from poor families (i.e., clusters), most samples identify school clusters according to participation in free lunch programs and/or location according to neighborhood using lunch as a marker for poverty.

Complex Samples

In practice, stratification is commonly used in different stages of selecting clusters and subjects to be samples in multistage surveys (*stratified multistage cluster samples*) mentioned above. Sampling designs using stratification and cluster techniques are generally called *complex samples*, perhaps to faithfully represent their working mechanisms. The stratification used in the first stage to select clusters results in selected clusters forming *primary sampling units* (PSUs). Further sampling stages and subject selection proceed

based on these PSUs. All further schemas of stratifying and selection mechanisms are generally viewed to be *nested* within the PSUs. Sometimes, cluster sizes are very small, and they may not get selected in adequate numbers to contribute useful information. In such situations, survey designers may make a conscious determination to incorporate more numbers of such clusters or subjects from such clusters or strata (i.e., *oversample*) to provide for the ability to increase precision of the parameter estimate. The efficiency and variances from SRS and cluster samples can be compared. The *design effect* is a measure that is useful in comparing all sampling strategies with the SRS. The design effect is defined as the ratio of the variance of the estimator from a non-SRS design to the variance of the estimator based on an SRS of the same size—the smaller the ratio, the better the design. Design effect can be used in many ways: One use is to estimate the non-SRS sample that will provide the same precision as an SRS of a particular precision, or to estimate the gain or loss in precision by using a non-SRS over an SRS.

Analysis and Interpretation of Surveillance Data

Analysis of surveillance data is usually straightforward. Before analyzing survey data, it is helpful to refer to the survey methodology documents and follow the analytical guidelines. The first thing to keep in mind is the sampling technique and how weighting was done in the survey. The analysis of survey data requires the application of special methods to deal appropriately with the effects of the sample design on the properties of estimators and test statistics. Analysis then proceeds with a decision about which rates to calculate (e.g., incidence, prevalence, case fatality), standardization of rates (i.e., age standardization is common; direct or indirect standardization), use of the reference population for standardization (i.e., in the United States, currently the 2000 census population is mostly used as the reference population), controlling for confounding when making comparisons, and accounting for missing data or unknown values.

An important decision is which date should be used for examining trends: date of report or date of diagnosis? Date of diagnosis provides a better estimate for disease occurrence, but if there is a long delay between the dates of diagnosis and report, then this method will underestimate incidence in more recent intervals. Date of report is easy to work with but usually is involved with errors and irregularities. Adjustment of recent counts for reporting delays incorporates an element of estimation of the rates rather than reporting of the true observed rates. If changes in incidence and prevalence patterns and time trends are noted, then the investigator needs to ascertain whether the changes are real or artifactual. Artifactual changes in reported rates could arise due to changes in use or reporting due to holidays or other events, changes in staffing of the surveillance team, sudden proactive case finding efforts for various reasons including a change in in-

terest in the surveillance objectives, changes in surveillance procedures, and changes in diagnostic criteria.

For example, Beltrán-Aguilar et al. (2005) reported surveillance estimates for dental caries, sealants, tooth retention, edentulism, and enamel fluorosis in the United States using NHANES-III data (1988–1994) and continuous NHANES data (1999–2002) for comparison. NHANES made changes in their sampling design between these surveys. The continuous NHANES oversampled certain population subgroups (adolescents aged 12–19 years, persons aged > 60 years, Mexican Americans, non-Hispanic blacks, and persons of low income) to improve reliability of epidemiologic estimates. In contrast, NHANES-III oversampled children aged < 6 years, persons aged > 60 years, Mexican Americans, and non-Hispanic blacks. Also, NHANES-III recorded the presence of dental root caries and restorations at the tooth level but continuous NHANES assessed dental root conditions at the person level. The surveillance report observed four main trends across these surveys: (1) prevalence of dental caries in primary teeth among children aged 2–11 years did not change, (2) prevalence of caries in permanent teeth reduced among children aged > 6 years, (3) in dental sealants among persons aged 6–19 years increased, and (4) reduction in edentulism among persons aged > 60 years.

Usually surveillance results are reported for geographical areas within political boundaries (e.g., counties, states, nations). Reporting rates from areas encompassing political borders can be tricky. Because hospital-based surveillance systems may incorporate people from across the borders, the report will need to estimate the variability and address an adjusted rate with clarification of the methods used to calculate the estimates. Use of the Geographical Information System (GIS) can be helpful in a surveillance report. Apart from making interactive and intuitive graphical representation, GIS data can provide pointed localization and detailed analysis of the spatial distribution of exposures and diseases.

Comparisons of rates across different groups, times, and surveys using statistical testing may find several statistically significant differences, especially if samples are large. In some situations, data measurements may need statistical adjustments such as standardization so that comparison can be made. Even if statistically significant differences are noted, such evidence will have to be assessed against the question: Is the difference meaningful? For example, one study estimated the amounts charged for dental care during 1996 for the U.S. adult population and evaluated whether dental expenditures had increased since 1987 by using data from the 1996 MEPS that was conducted and named differently in 1987 (National Medical Expenditure Survey; NMES) (Chattopadhyay, Slade, & Shugars, 2003). However, the purchasing power of money changes according to inflation rates and this change in the value of money must be adjusted when making a dollar-to-dollar comparison across different time periods. To compare charges from the 1987 NMES to the 1996 MEPS, the investigators converted 1987 expenditures to

their "constant dollar" value in 1996 by multiplying the 1987 expenditures by the ratio of the average consumer price index (CPI) for 1996 divided by the average CPI for 1987. CPI data used for this purpose were obtained from the U.S. Department of Labor Statistics. They then statistically evaluated the null hypothesis that real expenditures (adjusted for inflation) for patients did not change between the two survey periods by subtracting the mean (inflated) expenditures in 1987 from the mean expenditures in 1996. Investigators then divided the difference by the pooled standard error to produce a test statistic that was evaluated for significance by referring to the critical values of the "Z" distribution. In this study, although the annual per capita expenditures for employed persons was significantly greater than unemployed persons ($185 vs $165, respectively), the difference was considered to be not meaningful because no dental procedure could realistically be conducted for the difference amount (i.e., $20). A similar application of meaningful difference is apparent in another study where the investigators examined the professional charges not paid to dentists using the 1996 MEPS data (Chattopadhyay, Caplan, & Slade, 2009). While defining the categorical outcome variable (i.e., professional charges paid or not paid), they decided to use a cut-point (> $50) for the "unpaid amount" because $50 reasonably represented a typical fee charged for the least costly dental service.

Weighting and Variance Estimation

In the absence of epidemiological and biostatistical support, although not explicitly stated, it is generally assumed that all the surveys and samples are SRSs, and investigators proceed with analyses under that assumption. These issues are usually not a problem if only the parameter estimate is being sought. However, if estimate precision and/or statistical tests are sought, as is common, then difficulties may arise. It is also usually assumed that the primary *unit of sampling* (i.e., usually the subjects selected in a sample) is also the *unit of analysis*. However, in cluster samples, units of sampling are the PSUs, even though the units of analyses are the subjects who are finally selected. If the sampling scheme is not an SRS, then even if the parameter estimate is correctly determined, its precision will not be correctly reflected in the standard error calculated from an SRS-based analysis strategy. It will help us recall that for all statistical tests, even if one assesses differences between parameter estimates, this is done in terms to estimates of variation, and standard error is the key statistic in all parametric tests. Furthermore, CIs are also functions of standard error. Obtaining the correct estimate for standard error is a crucial factor in statistical testing.

The need for accounting for the type of study design used to collect data during data analysis has been emphasized for some time. "We recommend that if the study goal is to estimate the magnitude of either a population

value of interest (e.g., prevalence), or an established exposure–outcome association, adjustment of variances to reflect complex sampling is essential because obtaining appropriate variance estimates is a priority" (Caplan, Slade, & Gansky, 1999).

In an SRS, every subject in the sample contributes equally (i.e., is weighted equally). However, in non-SRS samples, the selection probability of subjects into the sample is not the same because of the stratification, clustering, and multiple stages of selection that are imposed by the sampling technique. Unlike an SRS, subjects in other sampling designs do not contribute an equal amount of information, and have to be weighted differently depending upon the sampling scheme. Once the sample differs from an SRS, analytical schemes must incorporate a strategy for weighting. Weighting schemes generally mimic the pattern of sampling schemes, and variable weights seriously affect the estimation of variances and also impact the point estimates. In simpler designs such as one-stage stratified samples, weighting may be easily accomplished by providing a single numerical weight representing its selection probability. However, in cluster samples and complex samples that involve several stages of selection, across different strata and clusters each with different selection probabilities of subjects, the weighting mechanism becomes more complex than simply using a single numerical weight for every subject. Such designs usually apply weights in stages using the PSUs and then nesting weights depending upon sampling schemes. Therefore, complex samples also have complex weighting schemes. Until recently, SUDAAN was the only statistical package that had the ability to adjust variances for complex sample analyses. It is still a forerunner in complex sample analyses. However, several well-known statistical software packages (e.g., SAS, STATA, and SPSS) have incorporated routines to allow variance adjustment for complex sample analyses. It is not acceptable to analyze survey data without adjusting for the sampling scheme using any of these easily available programs for appropriate analyses.

Before analyzing survey data, it is helpful to refer to the survey methodology documentation and follow the recommended analytical guidelines. Sometimes, surveys are carried out in different phases meant to be analyzed only if the data is complete and combined across all phases. NHANES-III was carried out in two phases and some papers analyzed the first (phase I: 1988–1991) and second phases (phase II: 1991–1994) separately as the data came out. However, several discrepancies were noted in prevalence estimates and measures of association for periodontal disease in studies reporting from the two phases of the NHANES-III. For example, phase II estimates of gingival bleeding and probing depth were as much as 56% lower than estimates from phase I and both these estimates were different from that of the combined-phase data, although attachment–loss statistics were consistent between phases (Slade & Beck, 1999).

> Prevalence differences between phases could be explained in part by examiner variations. Odds ratios for probing pocket depth (PD) differed between phases by as much as one-third, although the direction and precision of associations were not affected, and differences were reduced after controlling for examiner. Combined-phase estimates of gingival bleeding (GB) and PD prevalence and extent differ from previously published estimates derived from phase I, apparently because estimates in at least one phase of the NHANES-III study are biased. However, associations with selected risk indicators were fairly consistent between phases. (Slade & Beck, 1999)

Although Slade and Beck contended that estimates from one phase were biased, the example also suggests that the correct analysis strategy was to use the combined-phase data as suggested in the NHANES-III analytical guidelines and not analyze phases separately. Replying to the above contention, Winn, Johnson, and Kingman (1999) stated that:

> While differences were found among dentists in the prevalence of pocket depth of 4 mm or more, for each group of sample persons assessed by a reference examiner–examining dentist pair, the reference examiner's periodontal measurements closely corresponded to measurements made by the examining dentists. Differences between dental examiners in prevalence of periodontal conditions may be due in part to the fact that examinees were not randomly assigned to examiners. As a result, the sample persons examined by each dentist may not have been alike in characteristics thought to affect periodontal disease status. These findings suggest that the observed declines in periodontal health status between phases is not due to examiner bias. This unexplained decline may be the result of sampling variation. It is recommended that combined six-year survey results be presented whenever possible.

Validity and Reliability of Surveys and Surveillance Quality Control

Surveillance reports for notifiable diseases take the forms of (1) individual case reports, (2) aggregated data on total number of persons with the disease, and (3) aggregated data for diseases if and only if an outbreak is judged. Surveillance reports for other diseases usually occur as periodic aggregated data in a planned manner. These reports must include details about initiation and sources of data used in the reports including collection and analytical methods, routing and timing of reports, and relevant policy issues in reporting the disease (e.g., periodicity of report, case definition, case ascertainment, and data storage). Errors (i.e., magnitude of random and systematic errors) in survey data must be measured. Construction of the questions used in a survey must be carefully done and reassessed over

time with every iteration of the survey. *Content, criterion, and construct validity of the questionnaire must be carefully assessed every time the survey is reissued. Even though a questionnaire may have been validated in a prior survey, it may lose validity due to changes in the nature of the population or other events that might have occurred in the interim period*. Similarly, reliability of the questionnaire should be assessed through evaluation of test–retest reliability, interrater reliability, and internal consistency reliability.

Robust quality control procedures must be established for surveillance programs to be successful. These procedures may include information technology monitoring; periodic quality control check and monitoring of data collecting and reporting system; data entry quality check protocol; checklist for the database manager to ascertain data quality; data sharing and integration mechanisms; system maintenance and data security protocol; periodic documentation and training of the surveillance team; periodic analytical method reassessment; and modification of the reporting system in response to system changes, new diseases, or changes in disease pattern as needed.

Oral Health Surveillance in Practice

The U.S. National Oral Health Surveillance System

The National Oral Health Surveillance System (NOHSS) was established in the United States in 2001. It is a collaborative effort between the CDC's Division of Oral Health and the ASTDD and 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 (CDC, 2008). The NOHSS includes indicators of oral health, information on state dental programs, and links to other important sources of oral health information. A total of eight individual-based oral health indicators that NOHSS covers include dental visit, teeth cleaning, complete tooth loss, loss of six or more teeth, dental sealants, caries experience, untreated tooth decay, and cancer of the oral cavity and pharynx. NOHSS also includes a System level indicator: fluoridation status of the drinking water in the United States. The NOHSS website allows interactive querying of the database for national- and state-level information by years (where data are available), and also informs the user from which survey the statistic was calculated. It also includes basic Web-based national- and state-level maps for the indicators recorded, apart from including synopses of state and territorial dental public health programs (CDC, 2008). Therefore, one can compare state estimates with national estimates or those of other states for one of the indicators covered (reviewed in Chattopadhyay et al., 2008). The U.S. NOHSS can be accessed on the Internet at http://www.cdc.gov/nohss/.

Global Oral Health Surveillance

WHO has established a global oral health surveillance system through its "oral health information systems" program. As of 2000, oral health data were available from 184 countries. The United Nations lists 192 member countries in the world, whereas the United States recognizes 194 independent States in the world. The United Nations lists "Congo" as one country: Republic of the Congo, whereas the United States recognizes "Congo" as two separate States: Republic of the Congo-Kosovo, and the Democratic Republic of the Congo. Additionally, the United States also recognizes Kosovo as an independent country. Although an independent country, Vatican City is not mentioned in either list. Therefore, the most comprehensive list would suggest that there are 195 countries in the world.

> The burden of oral disease and needs of populations are in transition and oral health systems and scientific knowledge are changing rapidly. In order to meet these challenges effectively public health care administrators and decision makers need the tools, capacity and information to assess and monitor health needs, choose intervention strategies, design policy options appropriate to their own circumstances, and to improve the performance of the oral health system.
>
> The WHO/FDI goals for oral health by the year 2000 urged Member States to establish oral health information systems and this remains a challenge for most countries of the world. The WHO Oral Health Program is prepared to assist countries in their efforts to develop oral health information systems which include data additional to epidemiological indicators. (WHO, 2008)

In 1981, WHO and the FDI (Fédération Dentaire Internationale, World Dental Federation) jointly formulated goals for oral health to be achieved by the year 2000 as follows:

1. Fifty percent of 5-to-6-year olds to be free of dental caries.
2. The global average to be no more than three DMFT at 12 years of age.
3. Eighty-five percent of the population should retain all their teeth at the age of 18 years.
4. A 50% reduction in edentulousness among 35-to-44-year olds, compared with the 1982 level.
5. A 25% reduction in edentulousness at the age of 65 years and over compared with the 1982 level.
6. A database system for monitoring changes in oral health to be established.

Although WHO goals for the year 2000 were not achieved, their formulation and programmatic impetus emphasized the need for setting up oral health surveillance in different countries. The WHO has set up an international data bank of oral health-related information to which countries contribute national data.

> At WHO, information systems are being established for surveillance of global trends in oral disease and risk factors. The WHO Global Oral Health Data Bank compiles valuable information for monitoring the global epidemiological picture and trends over time in oral health and the WHO Oral Health Programme has initiated integration of the existing database with other WHO health databases and surveillance systems on risk factors. The main surveillance tool is called STEPS (STEP-wise approach to surveillance), a simple approach which provides countries with core standardized methods but leaves them flexibility to expand tools by adding information relevant to the local situation. (WHO, 2008)

The STEPS program can be accessed through the WHO website at http://www.who.int/chp/steps/en/. WHO suggests that nations establish a system for oral health information that could be collected and categorized under the following headings:

1. Epidemiological surveillance
2. Service coverage of the population
3. Service records and reporting
4. Administration and resource management
5. Quality of care provided
6. Oral health program monitoring and outcome evaluation

The Australian Research Center for Population Oral Health (ARCPOH) at the University of Adelaide maintains the oral health surveillance system and tracks a wide variety of conditions related to distribution and determinants of oral health, burden and impact of oral disease, testing the effectiveness of population oral health interventions, and oral health services and labor force research; and it conducts oral health policy analysis. These projects reflect varying oral health and epidemiological perspectives and applied research on the provision of dental health services. These revolve around the distribution and determinants of oral health and disease burden. These also assess the impact of oral disease, testing the efficacy/effectiveness of population oral health interventions, oral health services and labor force research, and oral health policy analysis. The ARCPOH website can be accessed at http://www.arcpoh.adelaide.edu.au/project/.

The European Global Oral Health Indicators Development Project (EGOHIDP) was developed by the European Commission Health and Consumer Protection Directorate-General as a pathfinder project in 2002. In its first phase, the program aimed at establishing priorities for a specifically European context in coordination with the existing program, to make new recommendations for improving health system performance when necessary, and to recommend a list of essential oral health indicators. It was conceived that these indicators would facilitate further promotion of oral health and non-communicable disease surveillance in Europe to collect information, to monitor changes, to assess the effectiveness of the service, and to plan oral health services within the framework of an intersectorial preven-

tive policy based on health determinants. The first phase terminated in 2004 and a final report was published.

Following that event, the second phase (European Global Oral Health Indicators development—phase II: EGOHIDP-II) aims at developing and coordinating the health information and knowledge system to provide quality, relevant, and timely data, information, and knowledge in order to support public health decision making at European national, subnational, and local levels to help promote oral health and contribute to non-communicable disease surveillance in Europe. Upon completion of these objectives, the next step is envisioned as actual implementation of these instruments in the national health interview survey, and the national health clinical survey, and to evaluate their performance. The project can be accessed at http://www.egohid.eu/. The EGOHIDP-II has four designated subobjectives:

1. To develop recommended common instruments for national health interview surveys.
2. To develop recommended common instruments for national health clinical surveys.
3. To develop a methodology for improved NHIS and NHCS data, routinely collected in 25 European countries at the primary oral health care level.
4. To develop methods to adjust national data to allow cross-national comparisons.

Although robust oral health surveillance systems have only started to develop, as the importance of surveillance and data is appreciated, several countries have set up national oral health policies, and started to organize sentinel oral health surveys. For example, the PAHO reports oral health status of several of its member countries in its report "Health in the Americas 2007" (accessible at http://www.paho.org/English/DD/PUB/HIA_2007.htm).

11 Pharmacoepidemiology

Pharmacoepidemiology can be defined as the application of epidemiologic knowledge, methods, and reasoning to the study of the effects (beneficial and adverse) and uses of drugs in human populations (Hartzema, Porta, & Tilson, 1987; Porta & Hartzema, 1987). Pharmacoepidemiology aims to describe, explain, control, and predict the effects and uses of pharmacologic treatment modalities in a defined time, space, and population (Porta & Hartzema, 1987). Primarily, pharmacoepidemiology has been concerned with clinical trials and the study of adverse drug effects. Genetics, molecular biology, and genetic epidemiology studies are finding that a variety of polymorphisms occur in genetic sites associated with drug metabolism, receptors, transporters, and drug target sites. These varieties of potential interaction sites increase the probability of adverse drug reactions among patients. However, this variety also provides the opportunity of using a drug specifically for a certain genetic make-up. *Pharmacogenetics* is the study of these varying responses to drugs due to genetic variation resulting from underlying genetic mutations. It usually focuses on one or a few genes at a time and assesses the role of genetic variation in drug metabolism and identifying target candidates for adverse drug reactions. *Pharmacogenomics* is broader in scope than pharmacogenetics and views the genetics of drug response in the context of the entire genome. It emphasizes the application of genomic technologies to new drug discovery and further characterization of older drugs with the aim of finding population-specific effective drugs that might lead to personalized medicine by using an individual patient's genetic make-up to tailor a patient-specific drug to maximize desired clinical outcomes and minimize adverse reactions.

Drug Development Process

Before a drug can be marketed, it must go through a series of complex development and testing processes in preclinical and clinical (human testing) phases that must comply with good laboratory practice (GLP) and good manufacturing practice (GMP) norms (Kaitin, 1995). Human clinical testing involves four clearly defined phases (phases I, II, III, and IV) that together are called "clinical trials." Sometimes totally new compounds are developed from scratch, and sometimes existing compounds are modified to elicit a certain response. Developing a totally new drug from scratch to its marketing stage may take 10 to 30 years. An estimate suggests that only 1 in 1000 compounds survive from the preclinical to the clinical stage. Of 100 drugs that enter the clinical phase, some 30% are rejected during phase I; 37% during phase II; 6% in phase III; and 7% during the review by the regulatory body (Kaitin, 1995). According to this estimate, only about 20% of the drugs entering the clinical trial phase finally make it to the market.

The preclinical phase of developing a new drug includes research into etiopathogenesis of a specific disease (target disease). This is followed by development of in vitro and/or animal models. Following this, potential compounds that may be useful in treating the target disease are designed and screened for use. Thereafter, these compounds are tested in animal models to determine safety, toxicity, reproductive toxicity, mutagenicity, carcinogenicity, pharmacokinetics, and pharmacodynamics. Sometimes, the in vitro and/or animal testing phases are called as phase-0 trials. At this stage, the company files an Investigational New Drug (IND) application with the regulatory body. IND applications are needed for new drugs and new indications/dosages/dosage forms of already approved drugs.

In the United States and most developed countries, the approval process for a new drug is a structured and sequential process that involves in vitro and in vivo testing (clinical trial). Testing of devices is called device trial. For testing to be conducted, the drug must have a clearly stated goal to treat a target disease. The clinical trial phases commence at least 30 days after the IND application has been filed unless the FDA provides other directives. Each of the phases may involve a series of several studies. Phase III trials, being the key stage in the approval process of a drug, are keenly followed by the scientific community, the media, the stock exchange, and the public. Due to this intense public gaze, many people mistake phase III trials to be the entire clinical trial, ignoring the existence of phases I and II.

Clinical Trial Process

Clinical trials are experimental studies and follow epidemiological study principals. NIH defines clinical trial as "research study to answer specific

questions about vaccines or new therapies or new ways of using known treatments" (NIH, 2008). NIH classifies clinical trials into five types:

1. **Treatment trials:** These studies test experimental treatments, new combinations of drugs, or new approaches to surgery or radiation therapy.
2. **Prevention trials:** Prevention trials look for better ways to prevent disease in people who have never had the disease or to prevent a disease from returning. These approaches may include medicines, vaccines, vitamins, minerals, or lifestyle changes.
3. **Diagnostic trials:** These trials are conducted to find better tests or procedures for diagnosing a particular disease or condition.
4. **Screening trials:** These studies test the best way to detect certain diseases or health conditions.
5. **Quality of Life trials (or Supportive Care trials):** These trials explore ways to improve the comfort and quality of life for individuals with chronic illnesses.

NIH defines *community-based clinical trials* as those clinical trials that are conducted primarily through primary care physicians rather than academic research facilities. Such trials should not be confused with "community trials" that compare different communities and are not randomized studies. Sometimes, under special circumstances, or in specific intractable and intransigent disease situations, the FDA may permit use of the study drug before a full approval occurs. *Compassionate use trials* provide experimental therapeutic agents *before* "final FDA approval to patients whose options with other remedies have been unsuccessful. Usually, case by case approval must be granted by the FDA for such exceptions" (NIH, 2008).

Clinical trials must follow good clinical practice (GCP) and GLP norms. Some of the important characteristics of good clinical trials are listed in Box 11.1. Historically, clinical trials have been conducted on men unless the drug was aimed at a woman-specific target disease. Women and children were generally excluded for convenience and logistic reasons. The assumption of using the drugs for a target disease among women was that the drug would act in the same way in women as in men. Drug action was also assumed to be the same in children except for their evolving metabolic capacity, for which only the dosage was downsized. Essentially, it meant that in such trials, there was no evidence about the action of the drugs in women and children. It is now understood that if a drug is intended to be used in women and children, trials must include them in the study sample. Unless a drug is intended for use in pregnancy, pregnant women are excluded from trials. If nursing mothers are included in trials, their babies should be monitored for effects of the drug.

Phase I trials represent the first trial on humans and are usually conducted with a few volunteers (20–100) who do not have the target disease

BOX 11.1 General Properties of Clinical Trials

A Well-Designed Clinical Trial Includes:

1. Clearly defined population from which participants are recruited.
2. Clearly outlined question and hypothesis with robust sample size estimation.
3. Robust data safety monitoring and management procedures.
4. Explicitly defined inclusion and exclusion criteria.
5. Principled system to provide information to participant and obtain informed consent.
6. Clearly specified and defined primary outcome measure and its measurement methods.
7. Clearly defined specified secondary outcome measures and their measurement methods.
8. Robust randomization protocol.
9. Effective blinding and blinding quality control.
10. Efficient participant compliance-monitoring strategies including participant safety and adverse event-monitoring system.
11. Established functional ethics committee to oversee the trial.
12. System to follow GCP and GLP protocols.

Properties of Disease/Condition Involved in Clinical Trial

1. It must be possible to monitor disease and demonstrate changes in disease state and response.
2. Valid and reliable diagnostic criteria must be available.

Inclusion Criteria Guidelines

1. Participants must be able to understand and provide informed consent of their own volition.
2. Participants should be able to comply with trial protocol.
3. Inclusion criteria must be developed based on natural history and mechanism of disease to improve patient safety and avoid errors in diagnosis and measurement.

Exclusion Criteria Guidelines

1. Participants not confirming to inclusion criteria.
2. Diseases and outcomes that could confound diagnosis and outcome measurements.
3. Participants with known allergic reactions.
4. Participants with other serious medical problems or complex concomitant diseases.

Data Analysis Guidelines

1. Intention to treat paradigm should be followed.
2. Data analyst should preferably be blinded about intervention groups.

(Kaitin, 1995). A phase I trial aims to assess the pharmacokinetics and pharmacodynamics of the drug, establish a safe dose, and assess and minimize drug toxicity. Three main types of studies are conducted in phase I trials. In *single ascending dose* (SAD) studies, patients are given a single dose of the drug and followed up. If no adverse effects are seen, sequentially higher doses are given, followed by observation, until the maximum tolerated dose is reached. In *multiple ascending dose* (MAD) studies, patients are given multiple low doses and the pharmacokinetics and pharmacodynamics of the drug are assessed using biological samples from the patients. Within a safety margin, the doses are then gradually increased to higher predetermined levels, with the patients being monitored continuously. *Food-effect studies* are short lasting and examine how eating food affects absorption, pharmacokinetics, and pharmacodynamics of the drug.

Phase II trials are conducted on a small number of patients who have the target disease. These trials aims to assess efficacy of the drug on its target disease and to find the daily dosage that the drug would be used for. In this phase, further assessment of pharmacokinetic properties of the drug and common adverse events are also important goals. Phase III trials are conducted on a large number of patients (usually between 500–3000) with the target disease to assess the efficacy and toxicity of the drug. Phase III is the most important phase and if successful, permits the company to submit the reports to drug regulatory authorities (i.e., the FDA in the United States) for their approval for release of the drug in the market. Phase III trials must include a randomized clinical trial or trials (typically placebo controlled). Once enough evidence is collected, the company may file a New Drug Application (NDA) with the regulatory authority. The NDA contains all the scientific data, preclinical and clinical, collected by the company, and samples and proposed labeling with details of the text of the package insert (Kaitin, 1995).

Drug Approval Process

The drug approval process lays emphasis on ethical conduct of trials; compliance with GLP, GMP, and GCP norms; and valid scientific conclusions. The regulatory authority assesses the information submitted by the company in the NDA. This involves several experts from different disciplines—physicians as well as scientists (Kaitin, 1995). Once the regulatory authority has approved the drug for marketing, new trials starting at phase II would be needed for new indications. Drugs for life-threatening conditions may be fast-tracked and bypass the Dispatcher-Assisted Resuscitation Trial (DART) process.

The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) is an international body involving regulatory authorities of the United States, Europe, and Japan and experts from their respective pharmaceutical industries to

discuss various aspects of product registration. The ICH aims to reduce redundant duplicate testing of products in different places during the drug development process to make the process more economical, efficient, and reduces delay in drug development without compromising safety, efficacy, and quality of the drug. The ICH-E1A guideline addresses the question of how much safety information is needed before a new drug is approved. The objective of this guideline is to present an accepted set of principles for the safety evaluation of drugs intended for the long-term treatment (chronic or repeated intermittent use for longer than 6 months) of non-life-threatening diseases (FDA, 1995). The guideline suggests that it is expected that short-term adverse event rates that have a cumulative 3-month incidence of about 1% will be well characterized. However, rarer events or those that take a long time to occur will need a longer time for characterization depending on their severity and importance to the risk-benefit assessment of the drug. The guideline also clearly states that during safety evaluation, characterization of adverse events occurring at less than 0.1% are not expected during clinical drug development (FDA, 1995).

Phase IV Clinical Trial: Postlaunch Safety Surveillance

Phase IV trials are the continued monitoring of the drug by the company after the drug is released in the market (*postmarketing surveillance*). The drug is also monitored by the regulatory authority (*pharmacovigillance*). The goal for a phase IV trial is to collect data for long-term safety, efficacy, effectiveness, adverse events, and interactions with other drugs or food items. Postmarketing surveillance must be conducted for at least 5 years.

After approving authorities are satisfied with the evidence in the drug dossier, drugs are approved for use and their indications are prominently displayed in the labels and package inserts. However, after approval in a real-world situation, the approved drugs, devices, and materials are used for a broad range of activities for which they were never approved (*off-label use*) that is not strictly monitored. Decisions for off-label use of drugs are made in clinics on an individual, case-by-case basis. For example, Methotrexate, approved for the treatment of choriocarcinoma, may be used for the medical treatment of an unruptured ectopic pregnancy (McLaren et al., 2008) or for a severe form of polyarteritis nodosa (Boehm & Bauer, 2000). Although off-label use is legal, advertisement, marketing, and promotion of off-label use of drugs is illegal.

Drug Safety and Adverse Drug Reactions

Adverse drug reactions have traditionally been separated into those which are the result of an exaggerated, but otherwise usual pharmacological effect

of the drug (*Type A reactions*) vs those which are aberrant effects (*Type B reactions*) (Strom, 1994). Other classes of adverse drug reactions are chronic effects (*Type C reactions*), delayed effects (*Type D reactions*), end-of-treatment effects (*Type E reactions*), and failure of therapy (*Type F reactions*). Almost all drugs show adverse events in the oral cavity, although most may be minor and not reportable. Box 11.2 lists the types of adverse drugs reactions seen in the orofacial region.

BOX 11.2 Types of Adverse Drug Reactions in the Orofacial Region Reported from a Variety of Drugs

Salivary Gland and Salivary Disorders

Xerostomia, hypersalivation, salivary gland swelling, salivary gland pain, saliva discoloration

Tongue Disorders

Taste disorders, ulceration, loss of papillae, mucosal disorders mentioned below

Oral Mucosal Ulceration

Drug-related burn, fixed drug eruptions, mucositis, neoplasms and preneoplastic lesions, vesicullo–bullous lesions, pemphigus, erythema multiforme, epidermal necrolysis, systemic lupus erythematosus, lupus-like disorders

Mucosal White Lesions

Lichenoid lesions and lichen planus, lupoid reactions, candidiasis, papillomas, hairy leukoplakia, leukoplakia

Mucosal Color Disorders

Superficial transient discoloration, intrinsic pigmentation

Swellings

Gingival enlargement, lip and mucosal angioedema, cheilitis

Neuropathies

Trigeminal neuropathies, involuntary facial movements, orofacial pain, dysesthesia

Tooth Disorders

Extrinsic discoloration, intrinsic discoloration, enamel and dentine structural defects, tooth erosion, internal tooth bleaching, dental caries, altered tooth sensitivity, root resorption, fluorosis

Adapted from Scully & Bagan, 2004; Tredwin, Scully, & Bagan-Sebastian, 2005.

Drug safety and adverse reactions are monitored by several countries and international bodies. In the United States, the FDA monitors adverse drug reactions. The FDA has established the Adverse Event Reporting System (AERS), which is a computerized information database of all approved drug and therapeutic biologicals. In the United States, reporting of adverse events by healthcare professionals and consumers is voluntary. However, if the drug manufacturer receives any information about an adverse event, it must report that information to the FDA. Healthcare professionals, consumers, concerned citizens, or groups may also directly report adverse events to the FDA. Internationally, the WHO and the European Medicines Agency (EMEA) are two of the agencies responsible for monitoring postmarketing studies. Most countries do not have a separate established agency, and the monitoring is effected through their ministries of health.

Drug Regulatory Bodies Internationally

The WHO has two roles in effective drug regulation: (1) development of internationally recognized norms, standards, and guidelines, and (2) providing guidance, technical assistance, and training in order to enable countries to adapt global guidelines to meet their specific drug regulatory environment and needs. The WHO Department of Technical Cooperation for Essential Drugs and Traditional Medicine (HTP/TCM) and Department of Medicines Policy and Standards (HTP/PSM) are centered in Geneva. The HTP/PSM has three technical units and assesses issues such as antimicrobial resistance, blood products and related biologicals, counterfeit medicines, drug safety, utilization and pharmacovigillance, international scheduling of substances of abuse and access to controlled medicines, newsletter and drug alerts, norms and standards for pharmaceuticals, prequalification, selection of medicines, supply management, and rational drug use. The HTP/TCM also has three technical units, but it focuses on addressing issues related to prices of medicines in different regions of the world. The International Conference of Drug Regulatory Authorities (ICDRAs), held regularly since 1980, provide drug regulatory authorities of WHO Member States with a forum to meet and discuss ways to strengthen collaboration. The ICDRAs have been instrumental in guiding regulatory authorities, WHO, and interested stakeholders; and in determining priorities for action in national and international regulation of medicines, vaccines, biomedicines, and herbals.

Clinical trials have now become globalized with the number of multinational trials increasing rapidly. This has increased the need for establishment of robust control mechanisms in different countries. Some of the important regulatory bodies internationally include United Kingdom: Medicines and Healthcare Products Regulatory Agency; The Netherlands: Staatstoezicht op de volksgezondheid; Germany: Bundesinstitut für Arzneimittel und Mediz-

inprodukte; France: Agence Française de Sécurité Sanitaire des Produits de Santé; Denmark: Lægemiddelstyrelsen; Sweden: Läkemedelsverket; Australia: Therapeutic Goods Administration; Hong Kong: Department of Health: Pharmaceutical Services; India: Central Drugs Standard Control Organization, Ministry of Health; Japan: National Institute of Health Sciences; and New Zealand: Medicines and Medical Devices Safety Authority.

Notes on Design and Analysis of Clinical Trials

Bioequivalence, Noninferiority, and Superiority Trials

A *superiority trial* is designed to detect differences between treatments to prove that the new drug is superior to a comparator that could be an active drug or a placebo. This paradigm views disease outcome as the only important outcome in the trial (i.e., if the new drug treats the disease better, then that would be all that counts). The other paradigm, bioequivalence, views clinical trial outcomes more broadly. *Bioequivalence trials* aim to prove that the new drug is therapeutically equal to an active standard treatment. Another closely related study, a *non-inferiority trial*, intends to demonstrate that the new drug is not worse than that of an active standard treatment. Although bioequivalence and non-inferiority trials may appear to be synonyms, bioequivalence trials conform to a two-tailed test, whereas non-inferiority trials conform to a one-tail testing situation.

Drugs have desired effects and adverse effects. Therefore, it is possible that a certain standard treatment, while being a very good treatment for the target disease, may also cause severe adverse events, nullifying its usefulness. In order to develop a new drug, under the superiority paradigm the new drug must be superior to the standard therapy. However, under the alternative paradigms, the new drug may be only just as good as the standard drug. In the latter case, if it can be demonstrated that the new drug has fewer side effects, improves quality of life of the patients, has better compliance, has better effectiveness, has lesser cost, or that it works better in certain specific populations, and has other better outcomes, then the argument of approval of such a drug becomes stronger. Therefore, bioequivalence trials usually include several different types of outcome measures and assess primary and a variety of secondary end points. Success of a bioequivalence trial relies very heavily on strict adherence to GCP norms.

The reference drug in a bioequivalence trial must have a proven efficacy in the patient population being studied. The "*acceptance range*" is the prespecified tolerable distance between the reference drug and the new drug that a regulatory agency is willing to accept as a claim for equivalence. Acceptable range varies according to the disease under study. For most generic drugs, the acceptance range is +/− 20%.

When possible, the FDA requires bioequivalence trials to include a placebo along with an active control—whereas, if equivalence of two active drugs is shown, it is possible that both drugs, although being equal to each other, may be better or worse than a placebo. In the latter situation, bioequivalence would not translate to clinical effectiveness (Temple, 1982). *Bioequivalence drift* may provide challenges to bioequivalence trials. For example, let us assume that for the case under study, the accepted range is +/– 20%. If drug A (new) is 20% below (–20%) the reference drug (R), and another new drug B is 20% above R, then it may be concluded that both drugs A and B are bioequivalent to R and therefore to each other also. However, the drift of drugs A and B to the opposite ends of the acceptance range makes them about 40% apart from each other, which is double the acceptance range on any one side, and therefore the conclusion would be that drugs A and B are not bioequivalent to each other—a conclusion that is opposite to the earlier conclusion.

Placebo and Placebo Effect

A placebo is a therapeutically inactive (inert) compound or procedure given to a trial participant in the guise of a therapy. If a placebo evokes therapeutic effect in the recipient, it is called a *placebo effect*. Effects similar to a placebo effect may be seen in clinics but they are not the same as the placebo effect seen in clinical trials. Ernst (2007) has distinguished between a "perceived" placebo effect, that is, the change after a placebo intervention (as may be conducted in a regular treatment clinic setting), and the "true" placebo effect, the effect caused by placebo administration (as in clinical trials). Box 11.3 describes some definitions, misconceptions, and other terms associated with a placebo. A "true" placebo effect has been explained by two theories: (1) the conditioning theory suggests that a placebo effect is a conditioned reflex as the participant learns to improve after medical treatment (e.g., he or she may have benefited from treatments, or visits to doctors), and (2) the expectancy theory suggests that a placebo effect is a result of increased expectations of the patient from visiting a doctor and brings about a symptomatic effect (different from a pygmalion effect, see Chapter 5). Biological mechanisms involved in the placebo effect may involve endogenous opioids such as dopamine, endorphins, and other neurotransmitters. Ernst (2007) provides an excellent discussion about the placebo effect and the role of placebos in clinical trials, to which the interested reader is referred.

European guidelines state that "when placebo-controlled trials are considered unethical, the demonstration of comparable efficacy to that of a standard therapy may also be acceptable provided that a suitable standard therapy exists" (Committee for Medicinal Products for Human Use [CHMP] Working Party on Efficacy of Medicinal Products, 1995). The FDA takes a similar view of the use of a placebo when it is contrary to the interest of the patient (Temple, 1982).

BOX 11.3 Placebo and Placebo Effect: Variety of Conceptions

Definitions: *Placebo*

- An inert treatment, given as if it was a real treatment.
- A sham treatment without biological activity, used in pharmacology to control for the activity of a drug.
- An inert substance or procedure that alters a physiological or psychological response.
- An intervention designed to simulate medical therapy, which at the time of use is believed not to be a specific therapy for the condition for which it is offered.
- Any therapeutic procedure that has an effect on a patient, symptom, syndrome or disease, but which is objectively without specific activity for the condition being treated.

Definitions: *Placebo Effect*

- A change after a placebo intervention (termed a "perceived" placebo effect).
- An effect caused by placebo administration (termed a "true" placebo effect).
- An effect of patient–provider interaction.
- Any effect attributable to a pill, potion, or procedure, but not to its pharmacologic or specific properties.

Misconceptions About Placebo and Placebo Effect (*And the Truth*)

- The change of symptoms seen in the placebo group of a clinical trial are owing to the placebo effect (*There can be numerous contributors to this "perceived" placebo effect*).
- The placebo effect is about one-third of the total therapeutic effect (*It can vary from 0–100%*).
- About one-third of the population respond to a placebo (*There is considerable context-dependent variation*).
- Placebo-responders (people who reproducibly respond) are distinct from nonresponders (*There are no such characteristics*).
- Only "imagined" complaints respond to a placebo (*Improvements after placebo have been demonstrated for most symptoms*).
- Placebo effects are invariably short-lived (*Long-term effects have been documented*).

(Continues)

BOX 11.3 Placebo and Placebo Effect: Variety of Conceptions *(Continued)*

Placebo-Related Ambiguous Terminology (*And the Explanation of its Meaning*)

- Doctor–patient relationship; therapeutic relationship; therapeutic intent; context effect; Iatro–therapeutic effect; Iatro–placebogenic effect (*Effects caused by patient–provider interaction*).
- Nonspecific effects (*1. Effects of all factors except the specific effect of an intervention; 2. Effects that are not unique to a given intervention*).
- Incidental effect (*Therapeutic effects that are not characteristic effects*).
- Characteristic effect (*Effects that, according to current theory, are responsible for the therapeutic effect*).
- Hawthorne effect (*Effects caused by the fact that study participants are under observation*).

Adapted from Ernst, 2007.

Blinding

Randomization in clinical trials is done to minimize selection bias. Randomization does not ensure that the randomized groups are equal in all respects but does ensure that the distribution is not biased so that discrepancies could be attributed to random error. Blinding is an important method to reduce bias. In a *single-blind trial*, only the participating subject is not aware of the treatment assignment. If the participant as well as the investigator and associated staff are not aware of the treatment assignment, the trial is a *double-blind trial*. Most clinical trials are double-blind, placebo-controlled, and randomized experimental studies. Most trials implement blinding and assume that it worked well. However, this assumption must be tested. For example, the examining clinician can be asked to fill out a log after examining the subject to assign them in one of the treatment assignment groups. Even if the physician is blinded, if he or she can regularly predict the treatment group of the subject, then blinding may not work! Therefore, it is important to test for unblinding of the trial personnel regularly as a quality control measure. Bias in a clinical trial may also occur during the analysis stage when all the data are being assessed for outcomes. Although analysts are usually driven by the hypotheses and the data, it may be possible to tweak the analyses toward certain desirable ends. Because of this possibility, blinding of the data analyst to treatment assignment may be useful. Such trials are called *triple-blind trials*.

In contrast to blinded trials, some clinical trials do not use blinding, and both the investigators and trial participants are aware of the treatment allocation (e.g., active drug or placebo) of the participant. Such trials are called *open-label trials*. Just because a trial is open label does not mean that randomization is not needed—open-label trials may be randomized.

Intention to Treat

An important analytic paradigm for clinical trials that has now become the standard practice is the intention to treat (ITT) principle. For several reasons, a patient may drop out of the study, may not continue in the trial, may be offered another alternative treatment, or may have switched from a placebo group to the active drug group due to compelling circumstances. The dilemma in such situations is whether to ignore the patient totally, or to consider the patient in the group where he or she started or ended. The ITT principle states that the patient will be assessed as belonging to the randomly assigned group at the start *regardless of* his or her adherence to the entry criteria of the trial, the actual treatment he or she received, any deviation from protocol, or his or her withdrawal from assigned treatment. This principle avoids breaking the randomization of the trial and allows for noncompliance, avoids deviations from treatment policy rather than focusing only on the effects of specific treatment, avoids bias associated with systematic exclusion of participants from randomized groups, and provides for conservative estimates of differences between the study groups. ITT is considered to be good for superiority trials but not for bioequivalence or non-inferiority trials. ITT may give misleading results of similarity in bioequivalence or non-inferiority trials when random non-adherence of study participants to the assigned treatment regimen occurs (Blackwelder, 2004). Therefore, in such trials, maintaining a high degree of adherence to protocol is especially important to maintain the validity of ITT.

In some situations in clinical trials ITT may not be possible. In such situations, a *modified intention to treat* analysis allows exclusion of some participants if valid justification exists. For example, if diagnosis is not immediately available at randomization or at the start of treatment, when the patient is required to start treatment based on certain symptoms, modified ITT may be justified. Modified ITT leads to analysis of a subset of the full data (due to elimination of participants as modified) where analysis proceeds as randomized and not as treated. Another strategy called *per-protocol* that is sometimes used states that only participants who complete the entire clinical trial are analyzed. In per-protocol, dropouts are excluded from analysis, which may lead to bias in the trial that may be difficult to address. *End-point analysis*, also known as *last observation carried forward*, may be a suitable strategy in situations where participants do not adhere to treatment and do not return for follow-up. In situations where non-adherence

and loss to follow-up are substantial, a "best solution" analytical strategy should be drawn and multiple types of analysis may be conducted and reported. Participants with missing information should also be analyzed separately for their characteristics and outcomes. Imputation techniques may also provide an alternative for improving analytical robustness.

Other Data Analysis Issues

Competing events may overtake the primary end point and decrease the number of participants, thereby reducing study power. Alternative and surrogate end-point analysis may be conducted. At this time no alternatives have been suggested for competing events, and the general practice is to report all major outcome groups in detail.

Despite randomization, it is possible to have groups that are disbalanced with respect to one or more variables. Covariate adjustment and covariance analyses might provide reasonable solutions by reducing variance in the test statistic, or covariate surrogates may be tried. If stratification was used at randomization, analysis should always be stratified. Adjustment for baseline covariate disparity is debatable at this time. One school of thought suggests that "if done at all, analysis should probably be limited to covariates for which there is a disparity between the treatment groups and that the unadjusted measure is to be preferred," whereas the second school of thought suggests "to adjust on only a few factors that were known from previous experience to be predictive" (Canner, Huang, & Meinert, 1981a, 1981b). In the case of a multicenter trial, randomization should be stratified by clinic and analysis should incorporate clinic as a stratification variable.

Quality Control

Compliance of patients to the directions of treatment usage and trial protocol is an important factor in the conduct of clinical trials and must be carefully monitored, assessed, and documented for valid results. Data quality in clinical trials is another very important determinant of its results that must be assured constantly. Main types of data-related problems that arise in clinical trials include missing data, incorrect data, and excess variability in data. A well-managed clinical trial should put in place quality control protocols to minimize these threats to data quality by ascertaining adherence to protocol by participants, trial managers, and clinicians; to minimize inter-examiner variability and erroneous data; and to improve the completeness of trial documents and report sheets. This can be achieved through training; pretesting; blinding; repeat assessment; data-entry validation and programmed range checks; double-data entry; ongoing data quality monitoring; ongoing monitoring for forms, procedures, and drug handling; and regular external data audits.

As a quality surveillance (ongoing quality control) measure, clinical trials develop "the quality plan," which is a document that describes all quality control procedures for the trial and is the most important reference document for trial managers to ensure that the trial follows GCP, GLP, and standard operating procedures. Some of the quality control issues and protocols in the quality plan include plans to ensure the following procedures (Valania, 2008).

Operational quality control: Correct sampling plan to be used; data sources to be used for quality control; data metrics; acceptable quality levels; and appropriate methods to report and distribute results.

Data analysis: The data being generated in the trial must be the data that is required; verification of data in case report forms from the data source; assuring that the data being analyzed are recorded data in the forms; data presented in tables, listings, graphs, and reports are the same as the data in the trial database.

Monitoring conduct of study: Obtaining appropriate informed consent; ascertaining participants' eligibility based on inclusion or exclusion criteria; adverse events and concomitant medication; drug accountability and storage; and verification of source documents such as medical records, lab data, progress notes, and diagnostic tests.

Resolution of queries: Ascertaining completed data-clarification forms and compliance with regulations.

Clinical Trial Issues in Dentistry

It may be possible, that like several other studies, different clinical trials may vary in their results or be of poor quality. For example, one systematic review examining the clinical trials testing the effect of professional interproximal flossing on proximal caries incidence reported that there were significant study-to-study differences, poor reporting of results, and a moderate to large potential for bias (Hujoell, Cunha-Cruz, Banting, & Loesche, 2006). Although self-flossing failed to show an effect, this review, however, concluded that professional flossing in children with low fluoride exposures is highly effective in reducing interproximal caries risk. Methodological differences between different trials and nonstandardized protocols were a cause of worry in clinical trials of powered toothbrushes although most clinical trials were robust in themselves (Robinson et al., 2005). Another study assessing trials testing the effect of ozone therapy on pits and fissures as sealants found that the quality of the studies was modest and many important methodological aspects such as blinding, randomization, patient compliance, and so on were not reported (Brazelli et al., 2006).

Hickel et al. (2007) have provided an in-depth review of clinical trial protocols for dental restorative materials and made recommendations for

better conduct of such studies. Clinical trials and controlled clinical studies about dental restorative materials have vastly improved over the last several decades. Clinical evaluation of restorations not only involves the restorative material, but also different operative techniques. Rigorous trials for restorative materials in dentistry should make modifications in clinical testing protocol and clinical evaluation parameters. Randomized controlled trials have been rarely used in orthodontics (Tulloch, Antczak-Bouckoms, & Tuncay, 1989). Among those reported, only a few studies were found on patient satisfaction in the long term, and most of them showed low scientific evidence (Bondemarka et al., 2007). Most dental materials are tested very carefully in the laboratory. However, there are very few studies that test the strength of the repaired teeth. These materials undergo a variety of stresses and come under varied chemical, biological, and physical attacks in the mouth, yet there are few studies that have assessed their effectiveness (Nicholson & Czarnecka, 2006). Other study design issues (e.g., measurement and sample size) in clinical trials related to dental caries are discussed in Chapter 13.

Ethical Issues

The World Medical Organization (WMO) Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects, commonly known as the Helsinki Declaration, stated "in any medical study, every patient—including those of a control group, if any—should be assured of the best proven diagnostic and therapeutic method" (WMO, 1996). This has resulted in ethical dilemmas in assigning subjects to a placebo group in placebo-controlled trials. FDA regulation section 314.126 allows interpretation of the principle in the Helsinki Declaration by stating that "an effective regimen of therapy may be used for comparison, e.g., where the condition treated is such that no treatment exists, or administration of a placebo would be contrary to the interest of the patient" (FDA, 1998). The regulation also categorically states that "It is often possible to design a successful placebo-controlled trial that does not cause investigator discomfort nor raise ethical issues" (FDA, 1998).

Medical Errors

The Institute of Medicine (IOM) report on *medical errors* (Kohn, Corrigan, & Donaldson, 2001) states that medical errors kill between 44,000 and 98,000 people in U.S. hospitals each year with associated estimated total annual costs (including the expense of additional care necessitated by the errors, lost income and household productivity, and disability) of between $17 billion and $29 billion. Medical errors may occur due to diagnostic,

treatment, preventive, or other errors such as communication, equipment, and system failure. The IOM has suggested several strategies to improve the medical error situation. These include establishing a national focus to create leadership, research, tools, and protocols to enhance the knowledge base about safety; identifying and learning from errors by developing a nationwide public mandatory reporting system and by encouraging healthcare organizations and practitioners to develop and participate in voluntary reporting systems; raising performance standards and expectations for improvements in safety through the actions of oversight organizations, professional groups, and group purchasers of health care; and implementing safety systems in healthcare organizations to ensure safe practices at the delivery level (Kohn et al., 2001). *Medication error* is defined as any preventable event that may cause or lead to inappropriate medication use or patient harm while the medication is in control of the healthcare professional, patient, or consumer (American Society of Health-System Pharmacists [ASHP], 1998). IOM (Kohn et al., 2001) also mentions that medication errors alone kill some 7000 persons annually, a number that exceeded deaths from workplace injuries in 2000 when the report was released.

Linked Automated Databases

Efforts at catching medication errors and adverse drug events are increasingly relying on automated databases. Several health maintenance organizations (HMOs), Medicaid programs, universal health systems, and hospitals maintain automated databases that are used for such studies. The main reason for growth in medical encounter databases is the increasing computing power in keeping with Moore's Law (Moore, 1965), which has led to increased analytic power of the computer. These databases usually have an enrollment file, a pharmacy file, and a medical records file or files that allow the investigator to access complete medical, medication, and confounder information for analysis. Most such studies utilize inception cohorts to control for period effects, baseline covariate values, and other confounders. Pattern recognition methods have enabled data mining of large administrative databases and multiple linked databases to study rare outcomes. Use of highly automated systems to evaluate the relations between prespecified factors, or empirical techniques to search out common relations not specified in advance, can be easily conducted for large linked databases. "Using massive data sets requires that quality control corresponds to the nature of the high-level information that we derive from large databases" (Walker, 2001).

The Food and Drug Administration and Dental Products

Toothpastes fall under cosmetics as defined by the FDA. The Food, Drug, and Cosmetic Act (FD&C Act) defines cosmetics by their intended use, as "articles intended to be rubbed, poured, sprinkled, or sprayed on, introduced into, or otherwise applied to the human body . . . for cleansing, beautifying, promoting attractiveness, or altering the appearance" [FD&C Act, sec. 201(i)] (FDA, 2009). Products included in this definition include skin moisturizers, perfumes, lipsticks, fingernail polishes, eye and facial makeup preparations, shampoos, permanent waves, hair colors, toothpastes, deodorants, and any material intended for use as a component of a cosmetic product. Depending upon the intended use, some products may be classified as cosmetics as well as drugs; for example, fluoride-containing toothpastes. Such products must comply with the requirements for both cosmetics and drugs. Dental X-rays, cements, resins, tooth-bonding agents, surgical, and orthodontic equipment and materials, amalgam, implants and other equipment are all categorized as devices by the FDA. Other medicinal products such as all local anesthetics, steroids, and antibiotics are classified as drugs.

12

Molecular and Genetic Epidemiology

Molecular Epidemiology

Molecular epidemiology includes the application of molecular biological approaches to epidemiological problems, and use of tools and perspective of epidemiological approaches to comprehend observations from molecular biology. A functional definition of molecular epidemiology (Schulte, 1993) defines it as "the use of biological markers or biological measurements in epidemiological research." This definition limits molecular epidemiology only to biomarker study. A straightforward classification of a biomarker classifies its use into four types, those that: (1) improve assessment of exposure, (2) identify the underlying mechanisms of disease and disease transmission, (3) identify subgroups of the population that are more susceptible to the effects of pathogens or pathogenic substances, and (4) identify subgroups of cases with more homogeneous disease to better clarify the role of various etiologic agents (Rabkin & Rothman, 2001). Overall, molecular epidemiology studies measure biologic response (such as mutations) to specific exposures (mutagens) and assess the interplay of host characteristics such as genotype and phenotype in gene expression, and development of disease and response to therapy. Molecular epidemiology is also useful in diagnosis, prognosis, and follow-up of therapeutic results. Molecular epidemiology is a technique-based discipline. Box 12.1 outlines the commonly used techniques and principles in molecular epidemiology.

Biological Molecules

The basic molecules of life are polymers: (1) proteins, made up of chains of amino acids lined by peptide bonds; (2) nucleic acids, made up of nucleotides linked by phosphor-diester bonds [Nucleotides: purines—adenine

BOX 12.1 Properties of Commonly Used Molecular Techniques and Principles

Technique	*Properties*
Polymerase Chain Reaction (PCR)	DNA polymerase is used to amplify a piece of DNA in vitro by enzymatic replication. The generated DNA is then used as a template for replication leading to a chain reaction, amplifying DNA production exponentially. Therefore, a large amount of DNA can be quickly produced from a single or only a few copies of DNA. It is used to amplify specific regions of DNA (one gene, part of a gene, or a noncoding sequence). **Steps:** Initialization, denaturation, annealing, extension/elongation, final elongation, final hold. **Production stages:** exponential amplification, leveling-off, plateau. **PCR problems:** PCR may fail due to contamination, induced mutations, primer–dimer formation, cross-contamination and false positives, or variable sensitivity dependent upon specific primers and probes. **Variations:** Allele-specific PCR, polymerase cycling assembly (PCA), asymmetric PCR, helicase-dependent amplification, hot-start PCR, intersequence-specific PCR (ISSR), inverse PCR, ligation-mediated PCR, methylation-specific PCR (MSP), miniprimer PCR, multiplex ligation-dependent probe amplification (MLPA), multiplex PCR, nested PCR, overlap-extension PCR, quantitative PCR (Q-PCR), reverse transcription PCR (RT-PCR), solid phase PCR, TAIL-PCR, touchdown PCR, pan-Aspergillus and pan-Candida (PAN-AC), and universal fast walking (isolation of flanking genomic segments adjacent to a known sequence).
Restriction Fragment Length Polymorphism (RFLP)	A method for DNA sequencing that breaks the DNA into pieces with restriction enzymes that recognize specific short sequences. These sequences are then analyzed to determine the length/size of the fragments. A polymorphism for the restricted length is said to occur if the length of a detected fragment varies between individuals. **Uses:** Used in genome mapping, locating disease genes, genetic fingerprinting, and paternity testing; still used in marker-assisted selection. **Fate:** Slow and cumbersome method, overtaken by several other alternatives—terminal restriction fragment length polymorphism (TRFLP).

(Continues)

BOX 12.1 Properties of Commonly Used Molecular Techniques and Principles *(Continued)*

Technique	*Properties*
Pulsed Field Gel Electrophoresis (PFGE)	Modified from regular gel electrophoresis for DNA separation. In contrast to the regular method, PFGE periodically directs the electrical voltage in three directions in three equally timed "pulses." Larger pieces run slower—allows separation of very large sections of DNA. **Use:** Genotyping.
DNA Sequence Analysis	The chromosomes are broken into shorter pieces (subcloning step); uses each short piece as a template to generate a set of fragments that differ in length from each other by a single base that will be identified in a later step (template preparation and sequencing reaction steps); the fragments in a set are separated by gel electrophoresis (separation step); new fluorescent dyes are used to separate all four fragments in a single lane on the gel; the final base at the end of each fragment is identified (base-calling step; this process recreates the original sequence of As, Ts, Cs, and Gs for each short piece generated in the first step); automated sequencers analyze the resulting electropherograms, outputting a four-color chromatogram showing peaks that represent each of the four DNA bases; after the bases are "read," computers are used to assemble the short sequences (in blocks of about 500 bases each, called the read length) into long continuous stretches that are analyzed for errors, gene-coding regions, and other characteristics (Human Genome Project, 2008).
Single-Cell Gel Electrophoresis Assay (Comet Assay)	Common procedure for detecting DNA damage. Commonly used in evaluation of DNA damage/repair, biomonitoring, and genotoxicity testing. Cells are encapsulated in a low-melting point agarose suspension and lysed in neutral or alkaline conditions. Thereafter, the lysed cell suspension is electrophoresed. Then, the DNA is stained by fluorescent dye and visualized. Extent of DNA damage is assessed using the distance the fragments of DNA travel compared to the starting point or "head." Heavier fragments travel for shorter distances, whereas smaller fragments travel farther. The longer the "tail" of the DNA "comet," the more the

(Continues)

BOX 12.1 Properties of Commonly Used Molecular Techniques and Principles *(Continued)*

Technique	*Properties*
	damage. The assessment may be performed using imaging software manually or automatically.
DNA Pooling	A genetic screening method that combines DNA from many individuals in a single PCR reaction to generate a representation of allele frequencies. Pooling allows efficient measurement of allele frequencies in groups of individuals with fewer PCR reactions and genotyping assays compared to regular genotyping. It reduces the cost of large-scale association studies to identify susceptibility loci for common diseases (Sham, Bader, Craig, O'Donovan, & Owen, 2002).

(A), and guanine (G); and pyrimidines—cytosine (C), thymine (T), and uracil (U). A, T, G, and C form DNA, whereas A, U, G, and C form RNA]; (3) polysaccharides, made up of monosaccharides linked by glycosidic bonds; and (4) phospholipids, made up of a diglyceride, a phosphate group, and a simple molecule such as choline. Phospholipids have a hydrophilic end and a lipophilic end (also called hydrophobic), which makes them line up in certain ways to form a bilayer. Phospoholipds form the biological membranes that have fluid properties and form the membranes of cells and subcellular organelles.

The linear organization of amino acid chains form the *primary structure* of proteins, which is then twisted over itself in complex ways making the helix, beta sheet, and beta turns to make stable spatial arrangements held together by hydrogen bonds (the *secondary structure*). Further twisting of the polypeptide chain gives rise to the tertiary structure that is held together only by hydrophobic interactions of the polypeptide component molecular structure. This folding is called "conformation" of protein in three dimensions. Combinations of secondary and tertiary structures give rise to folds that bring together different parts of the structure that would otherwise be far apart in a linear sequence. The *tertiary structure* of proteins gives rise to structural, functional, and topological domains of proteins that account for protein function. Proteases are enzymes that break down proteins to their smallest functional elements, and this mechanism is used to identify the functional active site of a protein. When two or more polymeric chains combine

into a single protein structure, they further twist over their individual tertiary structure to give rise to a *quaternary structure*. Such proteins are called multimeric proteins.

A DNA sequence is therefore a series of letters arranged in order representing a sequential nucleotide structure of the DNA that carries genetic information. The DNA sequence is the genetic code. A succession of five or more nucleotides is generally considered a sequence. For example, AAAGTCTGAC is a DNA sequence. A DNA sequence is interpreted as the genomic pattern of the organism. DNA sequencing is the process of determining the exact order of the 3 billion chemical building blocks (called bases and abbreviated A, T, C, and G) that make up the DNA of the 24 different human chromosomes. Overall, each human cell has 46 chromosomes comprised of 2 meters of DNA, having 3 billion bases. There are approximately 30,000 genes that code for proteins.

Sequence Analysis

In the laboratory, DNA sequence analyses involve a series of steps: (1) the chromosomes are broken into shorter pieces (subcloning); (2) each short piece is used as a template to generate a set of fragments that differ in length from each other by a single base that will be identified in a later step (template preparation and sequencing reaction); (3) the fragments in a set are then separated by gel electrophoresis (separation); (4) new fluorescent dyes are used to separate all four fragments in a single lane on the gel; (5) the final base at the end of each fragment is identified (base-calling). (This process recreates the original sequence of As, Ts, Cs, and Gs for each short piece generated in the first step.) Automated sequencers analyze the resulting electropherograms, outputting a four-color chromatogram showing peaks that represent each of the four DNA bases; and (6) after the bases are "read," computer programs assemble the short sequences (in blocks of about 500 bases each, called the read length) into long continuous stretches that are analyzed for errors, gene-coding regions, and other characteristics (Human Genome Project [HGP], 2008). Sequence analysis has a major component that is conducted using computers as a part of bioinformatics. Sequence alignment involves comparison of sequences and finding similar and dissimilar sequences using computer algorithms. Computer programs are also used for identification of gene structures; reading frames; distributions of introns, exons, and regulatory elements; single nucleotide polymorphisms (SNPs); and comparative genomic assessments. DNA sequences for a particular phenotype, however, may vary in the population, leading to a variation of prevalent sequence in the population. Such variation may occur in several ways: insertions and deletions, differences in the copy number of repeated sequences, and single base-pair differences.

The Human Genome Project (HGP), sponsored in the United States by the Department of Energy and the National Institutes of Health, has created the field of genomics—understanding genetic material on a large scale. The medical industry is building upon the knowledge, resources, and technologies emanating from the HGP to further understanding of genetic contributions to human health. As a result of this expansion of genomics into human health applications, the field of genomic medicine was born. Genetics play an increasingly important role in the diagnosis, monitoring, and treatment of diseases. The HGP estimates that there are some 20,000–25,000 genes within human DNA as well as their controlling regions, and they are recorded in DNA sequence maps. Sequencing of human genes in the HGP has been succinctly described by Stodolsky (2008) of the U.S. Department of Energy, Office of Biological and Environmental Research, Office of Science.

> The human genome reference sequences do not represent any one person's genome. Rather, they serve as a starting point for broad comparisons across humanity. The knowledge obtained from the sequences applies to everyone because all humans share the same basic set of genes and genomic regulatory regions that control the development and maintenance of their biological structures and processes.
>
> In the international public-sector, Human Genome Project (HGP) researchers collected blood (female) or sperm (male) samples from a large number of donors. Only a few samples were processed as DNA resources. Thus donors' identities were protected so neither they nor scientists could know whose DNA was sequenced. DNA clones from many libraries were used in the overall project.
>
> Technically, it is much easier to prepare DNA cleanly from sperm than from other cell types because of the much higher ratio of DNA to protein in sperm and the much smaller volume in which purifications can be done. Sperm contains all chromosomes necessary for study, including equal numbers of cells with the X (female) or Y (male) sex chromosomes. However, HGP scientists also used white cells from female donors' blood to include samples originating from women.
>
> In the Celera Genomics private-sector project, DNA from a few different genomes was mixed and processed for sequencing. DNA for these studies came from anonymous donors of European, African, American (North, Central, South), and Asian ancestry. The lead scientist of Celera Genomics at that time, Craig Venter, has since acknowledged that his DNA was among those sequenced.
>
> Many polymorphisms—small regions of DNA that vary among individuals—also were identified during the HGP, mostly single nucleotide polymorphisms (SNPs). Most SNPs have no physiological effect, although a minority contribute to the beneficial diversity of humanity. A much smaller minority of polymorphisms affect an individual's susceptibility to disease and response to medical treatments.
>
> Although the HGP has been completed, SNP studies continue in the International HapMap Project, whose goal is to identify patterns of SNP groups (called haplotypes, or "haps"). The DNA samples for the HapMap

Project came from 270 individuals, including Yoruba people in Ibadan, Nigeria; Japanese in Tokyo; Han Chinese in Beijing; and the French Centre d'Etude du Polymorphisme Humain (CEPH) resource. (Stodolsky, 2008)

Mutations

Mutations are the swap of one nucleotide for another: the deletion, insertion, or inversion of one to millions of nucleotides in the DNA of one chromosome, and the translocation of a stretch of DNA from one chromosome to another (Rabkin & Rothman, 2001). Box 12.2 describes the types of mutations. Mutations may occur during cell division because of copying errors or by exposure to ionizing and ultraviolet radiations or to chemical substances (mutagens), induced by viruses or self-induced by the organism. Mutations may occur spontaneously due to molecular decay, loss of a purine base, change of one purine base to another purine base, change of a purine to a pyrimidine, change of a pyrimidine to a purine, or change of the type of base from typical to atypical. Various chemicals, viruses, or radiations may induce mutations. While mutation rates vary across species, mutations do not necessarily occur randomly in DNA. "Hot spots" occur in DNA where mutation frequencies are several times higher than the rest of the DNA.

BOX 12.2 Types and Classes of Mutation

Usefulness

- Advantageous mutation
- Disadvantageous mutation
- Neutral mutation

Cell Type

- Germ cell mutation
- Somatic mutation

Small Scale Structural Change

- Point mutations
  - Silent mutation (mutations that do not cause a change in a protein sequence)
  - Missense mutation (a single nucleotide change causes substitution of a different amino acid)
  - Nonsense mutation (mutation that causes a premature stop or nonsense codon in mRNA that leads to a nonfunctional protein)
- Insertions
- Deletions

(Continues)

BOX 12.2 Types and Classes of Mutation *(Continued)*

Large Scale Structural Change

- Amplifications (gene duplications)
- Deletions
- Juxtapose separate pieces (chromosomal translocations, interstitial deletions)
- Chromosomal translocation (interchange with nonhomologous chromosomes)
- Chromosomal inversions (reversed orientation of homologous chromosomes)
- Loss of heterozygosity (loss of one allele by deletion or recombination)

Functional Change

- Amorphic mutations (loss of function)
- Neomorphic mutations (gain of function)
- Antimorphic mutations (dominant negative mutation)
- Lethal mutation
- Reversion (restores original sequence and phenotype)

Heritability

- Heritable
  - Homozygous (identical mutation in both alleles)
  - Heterozygous (mutation in one allele)
  - Compound heterozygous (two different mutations in the two alleles)
  - Wildtype (typical nonmutated type of naturally occurring organism)
- Nonheritable
- De novo mutation (new mutation not existing in parental generation)

Impact on Protein Sequence

- Frameshift mutation (insertion or deletion of triplet codon that shifts the reading frame leading to a totally different protein translation)
- Silent mutation
- Missense mutation
- Nonsense mutation

Other Mutation Types

- Conditional mutation (has wild type phenotype or mutant phenotype under different conditions)
- Hypermutation (programmed immune adaptation by which the organism adapts the immunoglobin genes to respond to new antigens. Hypermutation affects only individual immune cells and are not heritable.)

Mutated genes that encode altered proteins or that cannot be controlled properly cause numerous inherited diseases (e.g., sickle cell anemia). Mutations in noncoding sections of DNA do not cause immediate effects and may be considered "indifferent mutations" or "neutral mutations." However, such mutations may play a major role in evolution of the organism by creating new genes, or by creating new regulatory sequences for already existing genes or for new genes. Most of the noncoding DNA consists of repeated sequences that may move from one part of the genome to another, and may sometimes integrate into genes; damaging, activating, or suppressing them. Such mobile elements account for some 45% of the human genome. Genomes from viruses may get inserted into a host genome and may replicate as a part of it, leading to evolutionary changes in the host genome. Diploid organisms carry two copies of alleles; haploids carry one copy. Haploid cells during meiosis lead to an independent assortment of alleles carrying mutations. Whereas recessive mutations lead to a loss of function (i.e., both alleles must be mutant for phenotypical expression of a recessive trait), dominant mutations express the mutant trait even when present on one allele. These mutations may represent either gain or loss of function. *Penetrance* is the actual phenotypical expression of mutation. It is the proportion of persons with mutation who actually show the disease (phenotype). For example, if penetrance is 80%, then 80% of persons with the mutation will show its effects.

Mutations create variability in the gene pool, are responsible for intraspecies variation, and are also an important contributor to evolution. Mutations, however, can be good, bad, or indifferent. Unfavorable mutations (deleterious) are reduced in the gene pool by natural selection. Artificial selection can, however, maintain unfavorable mutations in the gene pool. Favorable mutations (beneficial/advantageous mutations) are maintained and accumulated by natural selection because these provide better adaptability and better odds of survival and subsequent procreation for the organism. Neutral mutations do not affect the survival of the organism, and may accumulate to be repaired. Genetic repair mechanisms usually repair most mutations before they become permanent and heritable.

Polymorphism

In general, polymorphism is the occurrence of two or more phenotypes in the population of a species. In molecular biology, the term *polymorphism* describes certain point mutations in genotype. By far the commonest type of polymorphisms described is an SNP. An SNP occurs when the DNA sequence varies by the occurrence of difference in a single nucleotide only (A, T, C, or G). For example, if the sequence AAAGT**C**TGAC is changed to AAAGT**T**TGAC by changing one "C" to "T," the change would be an SNP. Most common SNPs have only two alleles. For a polymorphism to be designated as an SNP, the variant sequence type must have a frequency of at least 1% in the population.

More than 99% of human DNA sequences are the same across the population. SNPs make up about 90% of all human genetic variation and it is estimated that 1.4 to 3.1 million SNPs exist in the human genome. These SNPs are common, evolutionarily stable (i.e., not changing much from generation to generation), and have much lower mutation rates than do repeat sequences, which makes them easier to follow in population studies. Furthermore, SNP detection is amenable to automated analysis. An estimate from the National Human Genome Research Institute (NHGRI) of NIH suggests that there are more sites that are polymorphic in the entire human population, "than the number of sites that differ between any particular pair of chromosomes. Altogether, there may be anywhere from 6 million to 30 million nucleotide positions in the genome at which variation can occur in the human population" (National Human Genome Research Institute [NHGRI], 2008). The same estimate suggests that overall, approximately 1 in every 100 to 500 bases in human DNA may be polymorphic.

By themselves, SNPs do not cause disease, but they can help determine the likelihood that a person with the SNP may develop a particular disease, and are therefore suitable as biomarkers. Information about SNPs may be used in three ways in genetic analysis:

> First, SNPs can be used as genetic markers in mapping studies. SNPs can be used for whole-genome scans in pedigree-based linkage analysis of families. A map of about 2000 SNPs has the same analytical power for this purpose as a map of 800 microsatellite markers, currently the most frequently used type of marker. Second, when the genetics of a disease are studied in individuals in a population, rather than in families, the haplotype distributions and linkage disequilibria can be used to map genes by association methods. For this purpose, it has been estimated that 30,000 to as many as 300,000 mapped SNPs will be needed.
>
> Third, genetic analysis can be used in case-control studies to directly identify functional SNPs contributing to a particular phenotype. Because only 3–5 percent of the human DNA sequence encodes proteins, most SNPs are located outside of coding sequences. But SNPs within protein-coding sequences (which have recently been termed cSNPs) are of particular interest because they are more likely than a random SNP to have functional significance. It is also undoubtedly the case that some of the SNPs in non-coding DNA will also have functional consequences, such as those in sequences that regulate gene expression. Discovery of SNPs that affect biological function will become increasingly important over the next several years, and will be greatly facilitated by the availability of a large collection of SNPs, from which candidates for polymorphisms with functional significance can be identified. (NHGRI, 2008)

Biomarkers

Molecular biological markers, or biomarkers, are natural products that can be traced to a particular biological origin. They are powerful tools that can

be used to trace diseases, drugs, and environmental contaminants in modern systems. Biomarkers indicate the biological state—they are most often used as indicators—for disease occurrence or as prognostic indicators. Genetic and molecular biomarkers are divided into three types (Frank & Hargreaves, 2003): *Type-0* (markers for natural history of disease, including disease occurrence, prognosis, and treatment outcomes); *Type-1* (markers for drug action); and *Type-2* (surrogate markers for diseases and outcomes). If the state of a biomarker is altered by the treatment, then levels of such a biomarker is often used as a surrogate end point in clinical studies. Biomarkers are useful because they improve assessment of exposures; help in identifying underlying mechanisms of disease and disease transmission; identify population subgroups that are more susceptible or are immune to certain pathogens; identify subgroups of cases with different disease profiles to understand the natural history of the disease and the role of the various etiologic agents better; and identify population subgroups that respond to treatment differently (and these outcome profiles can be used to better target treatments and develop personalized intervention schemes). Biomarkers play an important role in the detection, prevention, and treatment of oral cancers.

> Biomarkers can take the form of genetic and molecular indicators, which characterize the function of chemopreventives and cancer processes such as oral carcinogenesis. Biomarkers cannot provide all the required information for risk assessment or possible activity of the chemopreventives. Other methods, such as epidemiological analyses and techniques, must be used to enhance our understanding of the risk for oral cancer in human populations. One common epidemiologic method, the questionnaire, helps to determine the use and carcinogenic potential of tobacco and alcohol during oral carcinogenesis. Genetic and molecular changes in human patient populations may result in a reduction in the number and function of tumor suppressor genes. If these changes are to be assessed, the tissues (e.g., buccal mucosa) must be accessible and harvested in a reliable and consistent manner for the acquisition of DNA, mRNA, and protein. Oral tissues provide sufficient quantities of these molecules and, under stringent conditions, the quality required for the isolation of these molecular constituents. In conjunction with epidemiologic techniques, various genotypic polymorphisms, such as glutathione-S-transferase (GSTM 1) or cytochrome P 450 (CYP450Al), have indicated a loss in carcinogen detoxification or the processing of internal growth control signals. Biomarkers are composed of a large diverse group of genetic and molecular structures. Some of these biomarkers are indicators for programmed cell death (PCD), while others describe malignant tumor growth. Many of these classes of molecules are oxidative-responsive (e.g., tumor suppressor p53, Bcl-2, growth factors, immune-derived proteins, and death-inducing molecules) and induce PCD by triggering a cascade of cysteine proteases and regulators (e.g., caspases, death receptors). This pathway results in cell-cycle alterations and DNA fragmentation. (Schwartz, 2000, 92)

DNA Damage

Damage of DNA is frequent, but the effects of DNA damage are minimized by the repair mechanisms in cells. However, some damages may accumulate and lead to adduct formation or fragmentation of DNA during life course. Figure 12.1 demonstrates fragmented DNA visualized using single-cell gel electrophoresis (SCGE), also known as comet assay. It is a sensitive and rapid technique for quantifying and analyzing DNA damage in individual cells originally developed by Östling and Johansson in 1984. Singh et al. later modified this technique in 1988, when it became popular as the alkaline SCGE (Singh, McCoy, Tice, & Schneider, 1988). The name of the assay comes from the image of the electrophoresis gel, which resembles a "comet" with a distinct head and tail. The head is composed of intact DNA, while the tail consists of damaged (single-strand or double-strand breaks) or broken pieces of DNA. While most of the applications of the SCGE have been to study animal eukaryotes, there have been reports of successful application in the study of plant cells (Fairbairn, Olive, & O'Neill, 1995).

The SCGE is a "new" procedure for evaluating DNA lesions and can be used to detect DNA damage caused by double-strand breaks, single-strand breaks, alkali labile sites, oxidative base damage, and DNA cross-linking with DNA or protein (Collins, 2004). The SCGE is also used to monitor DNA repair by living cells and is applicable to any eukaryotic organism and cell type (Rojas, Lopez, & Valverde, 1999). The extent of DNA liberated from the head of the comet is directly proportional to the amount of DNA damage (Collins, 2004; Rojas et al., 1999; Anderson, Yu, & McGregor, 1998). The comet assay is inexpensive once a laboratory infrastructure has been set in place, gives results within a few hours, and is an appropriate tool for environmental monitoring (Moller, Knudsen, Loft, & Wallin, 2000).

Gene Expression Profiling

Gene expression profiling is the measurement of the expression and activity of a large number of genes (several thousands) at once. Today's computer and chip technology allows for rapid simultaneous visual analysis of a large number of genes in a small physical device. The main advantage of such a system is the easy availability of a huge amount of information about an individual. The nature and variety of information that can be extracted from gene expression profiling is staggering. For example, it may be possible to distinguish between cells and individuals who may react differently to drugs or environmental exposures, between cells that may have different dividing rates, and between cells with differing responses to treatment. A gene is considered to be "on" if it is used to produce mRNA; otherwise it is considered "off." Genes may be in on or off states at different times of the day depending upon the stage of cell-cycle that the cell is in, depending

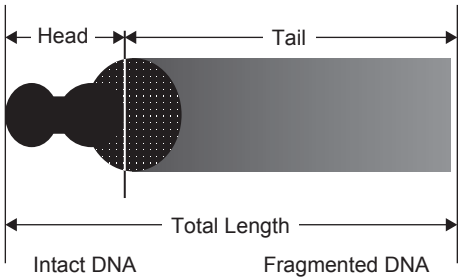


FIGURE 12.1 DNA Comet DNA damage detected in SGE/ Comet Assay Lighter fragments travel further and form the "tail" of the DNA "comet." Intact DNA does not travel at all and forms the "head" of the DNA "comet." The body of the DNA "comet" is formed by heavier fragments of damaged DNA. 100% DNA comet will show the "head," but no discernible "tail." Length of the DNA "comet" is measured from the "head" to the "tail" and is the metric for DNA damage (i.e., extent of DNA damage = length of the DNA comet).

upon the local environment, and depending upon the presence of up- and down-regulating factors. Gene expression profiling allows us to study the different states of the cell, and how genes respond to stimuli. The measurement metric is usually the amount of mRNA under different situations and environments.

The two most common forms of gene expression profiling are the serial analysis of gene expression (SAGE) and microarray analysis. The SAGE technique, a sequence-based sampling technique, produces a snapshot of the messenger RNA population, and is based on the principle that a 10- to 14-bp sequence ("tag") obtained from a unique position within a transcript can uniquely identify a transcript. SAGE permits examination of the "changes in the absolute levels of transcripts in a cell and, because it does not require an a-priori knowledge of the transcriptome (set of all mRNA molecules produced in one cell or a population of cells), can uncover novel genes expressed therein" (Weeraratna, Nagel, De Mello-Coelho, & Taub, 2004). The technique however is labor-intensive, technically challenging, and expensive. In contrast, microarray technology, although older, is easier to conduct. "Since these early studies, microarray profiling has been significantly refined and modified to optimize the sensitivity of the assay as well as the number of genes examined in a given experiment" (Weeraratna et al., 2004). Microarray techniques have been widely used to monitor gene expression for tumor diagnosis and classification, prediction of prognoses and treatment, and understanding of molecular mechanisms, biochemical pathways, and gene networks (Fanland & Ren, 2006).

> In general, microarrays are thousands of spots or probes printed on a solid surface such as glass or silicon that can be hybridized simultaneously to fluorescently labeled experimental samples. And the mRNA from each sample is labeled with fluorescent tags (i.e., Cy3 or Cy5), and then is hybridized to the microarray, which contains either cDNA [~300 bp; either polymerase chain reaction (PCR)–amplified inserts or whole plasmids] or oligonucleotide probes (~40 to 60 bp; custom oligonucleotides) for each gene of interest . . . then these combined targets are hybridized to a glass slide; finally, a confocal laser scanner is used to scan the slide, image analysis software is used to quantify image signals, and the data are converted into a text format showing the relative intensity for expression of each gene. (Wang, 2008)

Microarray data analysis involves two levels: (1) lower-level analysis—microarray experiment design; quality control of the microarray; experiment; microarray image analysis; and preprocessing, filtering, and normalization of raw microarray measurements; and (2) higher-level analysis—advanced data mining to answer problem-specific questions such as tumor classification and marker gene prediction; peak detection in the array comparative genomic hybridization (CGH) experiment; reverse engineering of gene expression networks; and the analyzing time series microarray dataset (Wang, 2008).

Microarray technology has been widely used to identify complex diseases; in drug discovery and toxicology studies; mutation/polymorphism detection (SNPs); and pathogen analysis. Microarray technique compares diseased state to normal tissue and is a very sensitive technique that may have several problematic areas. The "normal" state must be carefully defined because it serves as the main reference against which diseased states are compared. Tissue specimens should also therefore be collected from a truly "normal" state. Similarly, different parts of diseased tissue may also exhibit different gene expression profiles, and a truly representative tissue sample of the diseased state must be used. "Tumors are frequently a heterogenous mixed population displaying varying degrees of anaplasia, necrosis, and vascular proliferation. Thus, even comparing a single tumor cell type derived from different patients can yield quite varied gene expression profiles" (Weeraratna et al., 2004). Another important issue in the use of representative tissue samples is the use of peripheral blood. Whereas blood is easy to obtain, the gene expression in blood may differ from gene expression in the disease site, and exposure–disease association in blood may be substantially different from those in the diseased tissue because they may belong to different biological compartments. For example, in assessing gene expression in oral cancer in response to smoking, the biological effect(s) of smoking on oral tissues would take place through direct interaction between the exposure agent, and/or a subsequent exposure through systemic absorption and release through subepithelial capillaries in

the submucosa, or through secretion into either saliva or gingival crevicular fluid. However, the exposure to white cells in the blood would occur only after systemic absorption, and the metabolic process in a different biologic compartment (blood) would be expected to be different compared to direct exposure (mucosa). Therefore, it is intuitive to consider that gene expression and substantial DNA damage demonstrated in the blood compartment are indirect evidence of potentially greater damage in local oral mucosa.

Another problem in microarray analysis is the use of mixed cell populations from tissues and organs (i.e., a pool of different types of cells and tissues contribute mRNA to the analysis, whether those are affected or not). Gene expression may be phasic and subject to thresholds of exposures and may vary over time—being factors that may complicate sample selection for microarray analysis. Image analysis is used extensively in microarrays. Image analysis and analysis of microarray data are not yet standardized, and different centers use different methods. Multiple methods may be used to analyze microarray data. Reports extensively use "exploratory multivariate analysis" and "cluster analysis."

> In general, classical clustering techniques start by creating a set of bidirectional distance vectors that represent the similarity between genes and between clusters of genes. An iterative process is then undertaken where each gene profile is compared to all the other gene profiles and clusters until eventually all genes are in one cluster (Carr, Bittner, & Trent, 2003). There are numerous hierarchical clustering algorithms that differ in their starting point and the manner in which they calculate and compare the distances between the existing clusters and the remainder of the data set (Quackenbush, 2001). Bottom-up (agglomerative) hierarchical clustering was first applied to microarray analysis by Eisen, Spellman, Brown, and Botstein (1998). Because this technique produces readily visualized patterns of coordinately regulated genes and is supported by software programs such as Clusteruc© and TreeViewc© created by Eisen (http://rana.lbl.gov/), it has become extremely popular for microarray analysis. Other types of cluster analysis include multidimensional cluster analysis, which uses the similarities between two samples to generate a Pearson's pairwise correlation coefficient. This gives an idea of the magnitude of difference between two samples and, when applied to three or more samples, also provides a direction of the difference between them. Once these samples have been mapped into a three-dimensional plot, the similarity between two samples can be assessed by the distance between them. The more tightly two samples cluster together, the more similar they are (Bittner et al., 2000). Once these classes of genes have been identified, statistical analyses can be used to best determine which genes cause the samples to segregate as they do. (Weeraratna et al., 2004)

Microarray analysis also commonly uses ANOVA without adjustment for the large number of multiple comparisons that are inherent to the technique, as is the need for a large sample size that is often not available for

most studies. The generally accepted level of significance is 0.05. In microarray analysis, where thousands of analyses are conducted (typically using 10,000 genes), the chances of false positives is extremely high. For example, assuming that a microarray analysis uses only 1000 genes, at least 50 genes would be detected by chance (i.e., false positive rate = 50/1000 = 5%). Therefore, p-values must be adjusted for such chance findings in microarray analysis. Ideally, pairwise/familywise adjustments should be performed. However, at the least, a Bonferroni correction or false discovery rate adjustments (both are less conservative compared to pairwise/familywise testing correction) should be incorporated in the analysis. More recent developments in microarray analysis employ resampling methods such as permutation or bootstrap, and other methods such as Monte Carlo estimation, for statistical analysis.

Proteomics

The entire set of proteins expressed by a genome, cell, tissue, or an organism is called proteome. Proteomics is the large-scale study of the proteome—especially the structure and function of the proteins. Proteomics is to proteome what genomics is to genome; however, unlike the genome, which is more or less constant, the proteome of an organism changes according to time and environmental changes. Because proteins are the action molecules, Proteomics provide much more information than genomics. The central challenge in the study of proteomics is the need to understand the pre- and posttranslational modifications in the protein and how those events correlate with function.

The salivary proteome is susceptible to a large number of physiologic and biologic processes that stem from the neurological control of salivation, exposure, and interaction of several microorganisms, foods, and nonfood material in the oral environment (Helmerhorst & Oppenheim, 2007). The major salivary protein families together constitute more than 95% of the salivary protein content. Minor salivary glands contribute only about 10% to the total volume of human saliva released into the oral cavity (Dawes & Wood, 1973). The proteome of minor gland secretions may show significantly different characteristics when compared with the proteomes of parotid or submandibular/sublingual secretions (Siqueira, Salih, Wan, Helmerhorst, & Oppenheim, 2008).

There are several intrinsic difficulties in characterization connected to different factors of salivary proteomic variability such as high frequency of genetic polymorphisms, complicated by individual insertions, deletions, and alternative splicing; complex posttranslational maturations comprehending different proteolytic cleavages, glycosylation, phosphorylation, and sulfation processes; and physiological variations (Messana, Inzitari, Fanali, Cabras, & Castagnola, 2008). Overall, protein modification in whole saliva includes (1) degradation of histatins, acidic proline-rich proteins

(PRPs), and cystatins; (2) protein deglycosylation; and (3) protein–protein interactions in whole saliva such as protein complex formation in whole saliva, and covalent cross-linking of salivary proteins (Helmerhorst & Oppenheim, 2007). What distinguishes glandular salivary secretions from most other body fluids is that their constituents are mostly present as protein families of structurally closely related family members. This diversity is the result of allelic variation, gene duplication, alternative splicing events, and posttranslational modifications. These modifications occur through (1) glycosylation of amylase, proline-rich glycoproteins, mucous glycoprotein-1, mucous glycoprotein-2, secretory immunoglobulin A, and agglutinin; (2) phosphorylation of acidic PRPs, cystatin S and SA-III, statherin and histatin-1; and (3) proteolytic processing of basic PRPs, histatins, and statherin.

Salivary proteomics has the potential to develop low-cost, noninvasive, conductible tests from easily collectible samples for various oral and systemic diseases. Preliminary studies aimed at developing such tests are promising. For example, unstimulated "whole saliva from patients with primary Sjogren's syndrome contains molecular signatures that reflect damaged glandular cells and an activated immune response in this autoimmune disease" (Hu et al., 2007) and once validated, these candidate proteomic and genomic biomarkers may improve the clinical detection of primary Sjogren's syndrome.

Genetic Epidemiology

Genetic epidemiology involves the evaluation of the role of heritable causes of disease and deformities in families and in populations. Genetic epidemiology shares common space with molecular epidemiology and the two fields usually work hand-in-glove, especially when assessing gene-environment interactions in disease causation. Genetic epidemiology aims to uncover heritable patterns, determine causal genes or those genes that have major impact in disease occurrence, and find suitable markers of such genes for easily identifying the potential for disease occurrences to help develop and implement therapeutic and preventive mechanisms. The most widely accepted definition of genetic epidemiology describes it as "a science which deals with the etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations" (Morton, 1997).

Steps in Genetic Epidemiology

Genetic epidemiology achieves its goals through three well-defined, sequential steps: (1) establishing that there is a genetic component to the disorder; (2) establishing the relative size of that genetic effect in relation to other sources of variation in disease risk (environmental effects such as intrauterine environment, physical, and chemical effects as well as behavioral and social aspects); and (3) identifying the gene(s) responsible for the genetic

component (Dorak, 2009). Common methods used by genetic epidemiology include risk studies, segregation analysis, linkage studies, and association studies. Box 12.3 outlines the common methods used in genetic epidemiology. These methods can be integrated into a step-by-step sequence of studies to answer questions in genetic epidemiology. Schwartz (2000) has described similar schemes for oral cancer studies that sequentially lead to a set of outcomes. For example, genetic and environmental factor studies lead sequentially to identification of familial susceptibility genes, host genome description, detection of early molecular and genetic events, and finally, detection of early transformation of tissues and oral cancer. Similarly, examination of DNA adducts and antibodies can be linked to epidemiologic studies such as linkage analysis to detect susceptibility genes, which, with use of technologies such as cryogenetics–karyotypic analysis, aneuploidy and allelic alteration analysis, and loss of heterozygosity assessment may lead to microsatellite sequence detection, genomic markers for oncogenes, tumor suppressor activities, and a series of other outcomes.

Genetic epidemiology studies include case-control studies as well as prospective cohort studies. These studies may be associated with several potential problems depending upon the sources from which cases are taken. Ideally, the study cases should be representative of all cases. If some study cases have risks that are modified due to existing comorbidities and competing causes, then choosing study cases from a select population (such as from a certain locality or a clinic or hospital) will modify the risks for outcomes. These might occur differently for etiologic associations in different studies as a function of the source population of the cases (assuming that controls are correctly selected from the population from which the cases arise). Similarly, if the survival from disease is modified by comorbidities and competing causes, the genetic disease outcomes would also be modified in different unknown ways as a function of the interrelations between the comorbidities and the involved competing causes. Because these complex interrelations are not fully examined, it will be difficult to make clear-cut assumptions of the effects of comorbidities and competing causes. Perhaps such studies should carry out sensitivity analyses as a routine, discussing the best- and worst-case scenarios under reasonable assumptions.

Several other problems may arise in genetic epidemiology studies that need close attention at the design stage, such as appropriate control selection; avoiding selection bias resulting from definition of source and target population; genotyping errors; quality control in molecular methods; inappropriate choice of marker/allele/genotype frequency (for comparisons); failure to evaluate the mode of inheritance in a genetic disease; failure to account for the linkage disequilibrium (LD) structure of the gene (only haplotype-tagging markers will show the association, other markers within the same gene may fail to show an association); likelihood that the gene studied accounts for a small proportion of the variability in risk; true variability

BOX 12.3 General Methods Employed in Genetic Epidemiology

Genetic Risk Studies: Risk studies establish the occurrence of greater risk of a trait/disease in persons with a certain genotype. These studies ask the question:

- What is the contribution of genetics as opposed to environment to the trait?

  Genetic risk studies require family based, twin/adoption, or migrant studies. Risk studies assess familial aggregation of traits and use twin, sibling, and adoption studies to establish the genetic basis of traits.

Segregation Analyses: The next step after establishing the genetic basis is to establish the inheritance pattern of the trait in question. This is done using segregation analyses in families. These studies ask the questions:

- What does the genetic component look like? Oligogenic (few genes each with a moderate effect) or polygenic (many genes each with a small effect, etc.)?
- What is the model of transmission of the genetic trait?

  Segregation analyses require multigeneration family data preferably with more than one affected member. Segregation analyses establish recurrence risk ratios in families. Transmission probabilities of genotype, penetrance for each genotype, and population allele frequencies are data that are utilized in developing the genetic model. These analyses reveal Mendelian or non-Mendelian patterns.

Linkage Studies: Linkage implies co-segregation of loci (and not co-segregation of alleles) within families; that is, joint inheritance of genetic loci or alleles. Linkage analyses are generally used for coarse mapping because their genetic resolution is limited. Linkage implies association only at the family level, and not at the population level. Therefore, occurrence of linkage does not imply population-level allelic association. These studies ask the question:

- What is the location of the disease gene(s)?

  Linkage studies screen the whole genome and use parametric or nonparametric methods such as allele-sharing methods (affected sibling-pairs method) with no assumptions on the mode of inheritance, penetrance, or disease allele frequency. The underlying principle of linkage studies is the co-segregation of two genes, one of which is the disease locus. LOD score (Log Odds, in base 10) is a commonly used statistic in linkage analysis. LOD score = log10 (probability of outcome with given linkage/probability of outcome without linkage). For example, an LOD score of 2.0 means the likelihood of observing a given pedigree if the two loci are not linked is 1 in 100 (LOD 3.0 = likelihood is 1 in 1000, and so on). Conventionally, LOD

(Continues)

BOX 12.3 General Methods Employed in Genetic Epidemiology *(Continued)*

scores greater than 3.0 are considered good evidence in favor of linkage; whereas an LOD score below –2.0 is considered good evidence against linkage.

Association Studies: Association, resulting from direct gene involvement or linkage disequilibrium with the trait/disease gene, is studied at the population level and allows fine mapping. Occurrence of association without linkage is possible, especially when the allele for the trait/disease occurs in a minority subgroup of a population or is a poor marker for a disease in descendents. These studies ask the question:

- What is the allele associated with the disease susceptibility?

  The principle is the coexistence of the same marker on the same chromosome in affected individuals (due to linkage disequilibrium). Association studies may be family based (transmission disequilibrium test-TDT; also called transmission distortion test) or population based. Alleles, haplotypes, or evolutionary-based haplotype groups may be used in association studies.

Modified from Dorak, 2009.

among different populations in allele frequencies, information bias, disease, and exposure misclassification; and potential confounding from a variety of variables (measured and unmeasured).

The Hardy–Weinberg Equilibrium

The Hardy–Weinberg principle (HWP) states that both allele and genotype frequencies in a population remain constant—that is, they are in equilibrium (the Hardy–Weinberg equilibrium; HWE)—from generation to generation unless specific disturbing influences are introduced. Those disturbing influences include nonrandom mating, mutations, selection, limited population size, random genetic drift, and gene flow (Hardy–Weinberg Principle, 2009). Genetic equilibrium is an ideal state that may be producible in laboratory conditions, but in nature, HWE is rarely achieved. Static allelic frequencies are used with the following assumptions: the population size is large; allele pairs are independent; mating is random; there occurs no mutation, no migration (no exchange of alleles between populations), nor natural selection. However, almost all of these assumptions are violated in nature to varying extents. Therefore, it is important to test for deviation from HWE. Testing violation of HWE is performed by employing Pearson's

chi-squared test, using the observed genotype frequencies obtained from the data and the expected genotype frequencies obtained under HWE. If HWE is violated, no allelic association test is used (because an independence assumption is not met). Lack of HWE in controls is usually an indication of problems with typing rather than selection, admixture, nonrandom mating, or other reasons for violation of HWE (HWP, 2009).

Population Stratification

Presence of systematic differences in allele frequencies between subpopulations in a population is called population stratification. Furthermore, pooling of data from two countries might lead to population stratification (Wacholder, Rothman, & Caporaso, 2000; Millikan, 2001). Migration of individuals between populations is the main cause of population stratification. This phenomenon can pose serious problems in association studies using case-control designs because associations could be found essentially due to the structure of the population resulting from population stratification and not due to loci associated with disease, giving rise to spurious associations. This occurs because the assumption of population homogeneity in association studies is violated under these conditions. Similarly, less prevalent disease loci may be missed. Ways to get around the problem include understanding the underlying population structure, compensating for potential bias resulting from population stratification, and using genomic control. *Genomic control*, developed by Devlin and Roeder (1999) and modified by Bacanu, Devlin, and Roeder (2002), involves the use of unlinked markers (i.e., not linked with the trait in question) to control the possible inflation of the number of false positives and false negatives, as it corrects for any inflation of the statistic caused by population stratification.

Gene Flow

Gene flow (also known as gene migration) is the transfer of alleles of genes from one population to another or movement of genes from one population to another regardless of whether (or duration for which) it remains within the same species or is transferred to another species. Gene flow has an important role in the genetic variation within a population, as it changes the allelic frequencies and proportions present in a population. The extent to which gene flow affects a population depends largely on the species' mobility such as migration and mating patterns. If gene flow is regularly maintained between two (or more) populations, then with time, the populations become homogeneous due to the combination of their gene pools and reduced variation. Therefore, gene flow has been considered to act against speciation. Gene flow associated with dental traits has been used in human evolution studies. For example, Stringer, Humphrey, and Compton (1997)

examined the relationships between a range of modern human samples from cladistic analyses of the published population frequencies of tooth crown characters, using new data on the Krapina Neanderthal sample. Their study reconstructing a hypothetical dental ancestor suggested that the similarities between the African and Australasian groups resulted from the retention of symplesiomorphous dental traits (i.e., dental traits shared between two or more taxa). They further suggested that despite expectations from multiregional evolution, recent Europeans are dentally less like the Krapina Neanderthals than are Africans and Australians (Stringer et al., 1997).

Gene flow can occur through hybridization or gene transfer through microbes (bacteria, viruses) that may be affected as horizontal gene transfer, antigenic shift, or reassortment. Uncontrolled hybridization, introgression, and genetic swamping (i.e., gene pollution) in purebred organisms may lead to homogenization or replacement of local genotypes that occurs because of gene flow. Control of gene flow and genetic pollution is an extremely important process in genetically modified organisms and products such as food stuff.

Genetic Drift

Genetic drift, also known as allelic drift, is the accumulation of random changes of gene variants in a population, and is a function of the relative frequency of an allele in a population. Genetic drift is a slow phenomenon, and the differences between two successive generations may not be overtly noticeable. However, cumulative effect of genetic drift over several generations may be large. For example, emergence and spread of drug-resistant tuberculosis has been attributed to genetic drift (Hershberg et al., 2008). Use of molecular epidemiology techniques is based on a basic assumption of molecular epidemiology that lineages of pathogens are, for the most part, genetically stable spatiotemporally (Levin, Lipsitch, & Bonhoeffer, 1999). Using computer-generated simulations of the accumulation of mutations in a human gene pool, McKee (2005) found support for considering genetic drift and suggested that a polygenic model of the probable mutation effect could be a viable hypothesis for an explanation of the dental reduction that has occurred in some human populations over the last 40,000 years. Evidence supporting the role of genetic factors in susceptibility to chronic periodontitis is beginning to accumulate. The role of genetic factors in phenotypic expression can be estimated from the degree of resemblance between relatives, as compared with that of unrelated members of a population. In one study, Dowsett, Archila, Foroud, Eckert, and Kowolik (2002) determined whether there was a familial basis for periodontal disease status in an untreated population in Guatemala using heritability estimates as a measure of familial clustering of disease—heritability may be

used as an estimate of the proportion of total phenotypic variation of a quantitative trait attributable to genetic factors. They did not find much evidence supporting heritability of periodontal disease in their study, perhaps due to an underlying lack of genetic variation within this sample; or it may indicate that, compared with the role of environmental factors, the genetic contribution to periodontal disease phenotypes is relatively minor (Dowsett et al., 2002).

Gene–Environment Interaction

Genotype prevalence may vary tremendously among populations in the world. Some diseases occur just because of the presence of a genotype (i.e., the genotype is a necessary and sufficient cause for the disease; for example, Down's syndrome, trisomy-21). However, for most genes to produce an outcome, some help from other factors derived from the environment are necessary—therefore, even if a certain genotype is present, the absence of an environmental challenge may disallow a disease outcome, the risk for which may become real if the concerned exposure is present. Furthermore, given fixed prevalence of an at-risk null genotype, if the prevalence of exposure increases, the probability of outcomes related to that exposure also increases. Such joint actions between environmental factors and genetic factors to produce a disease are termed as *gene–environment interactions*.

Gene–environment interaction concerns itself with differential risk among people upon exposure depending on their genotypes. The case-only study design assesses gene–environment interaction more efficiently than a case-control design, and also avoids the pitfall of population stratification (Khoury & Flanders, 1996). The background concept for the case-only study is that hereditary factors that control the metabolism of carcinogens or other toxic substances may modulate the risk of disease. Different genotypes may respond differently to environmental risk factors. In this situation, the environmental risk factor would be considered a major risk factor and the genetic factor viewed as a modifier. Unknown genetic susceptibility that predisposes to differential environment sensitivity (i.e., gene–environment interaction), if ignored, could easily conceal the effects of an environmental factor on risk of disease (Andrieu & Goldstein, 1998). Under different models for gene–environment interaction, stratification by underlying genotypes can markedly improve the predictive value of disease risk factors in order to better target prevention efforts (Khoury & Wagener, 1995).

Analytical Issues in Genetic Epidemiology

Several statistical issues that may have profound impact on study outcomes occur in genetic epidemiology. The definition of terms may also lead to counting errors and adversely impact study outcomes.

> Confusion may occasionally arise through wrong usage of the terms allele, gene, or marker in an association study. Some investigators state that they compare allele or haplotype frequencies, but only count each individual once. They, therefore, refer to what used to be phenotype frequencies in serological Human Leukocyte Antigen (HLA) studies, or in the case of genotyping studies, to marker frequencies (MF), which correspond to inferred phenotype frequencies if it is an expressed genotype. Allele (AF) or haplotype frequency (HF) is analogous to gene frequency (GF) in that they are always calculated in terms of the total number of chromosomes not individuals. (Dorak, 2007)

Other common statistical issues that impact genetic epidemiology studies include lack of power, excessive subgroup analyses and posthoc analyses being treated as a-priori instead of exploratory analyses; possible one-tailed tests, ignoring multiple comparison; ignoring correlated data (e.g., using a chi-square test where McNemar's test should be used); and non-consideration of alternative genetic models.

Janes and Pepe (2006) pointed out that although case-control studies are common in genetic epidemiology for estimating the accuracy of a biomarker, the optimum case-control ratio for such studies has not been determined. They have suggested that although equal numbers of cases and controls may suffice for association studies, accuracy of biomarker studies need a different assessment of a case-control ratio. They have provided an expression for the optimal case-control ratio, when the accuracy of the biomarker is quantified by the receiver operating characteristic (ROC) curve using an empirical nonparametric ROC estimator that estimates the constituent survivor functions. The authors stated that there occur uncertainties in estimates of the parameters involved in the optimal case-control ratio and that this uncertainty may be substantial in a small pilot study. "The ROC slope estimator in particular may be sensitive to the choice of kernel or bandwidth. In practice, we recommend that one err on the conservative side and consider a range of plausible estimates consistent with the pilot data" (Janes & Pepe, 2006).

If the main interest of the study is to look for gene–environment interactions only, then such interactions can be accurately and more precisely estimated by case-only design compared to a case-control design (Khoury & Flanders, 1996; Andrieu & Goldstein, 1998; Yang & Khoury, 1997; Rothman et al., 2001; Goldstein, Falk, Korczak, & Lubin, 1997). Gene–environment interactions essentially look for interaction terms in the analyses. If the genotype and the exposures are independent, the odds ratio (OR) obtained from a case-only study (COR) becomes the synergy index on a multiplicative scale derived from a regular case-control study. Under the null hypothesis of no multiplicative effects, the COR is expected to be unity; if there are more than multiplicative effects, the OR will be more than one. ORs and confidence intervals can be obtained in case-only designs by using stan-

dard crude analyses or logistic models after adjusting for other covariates (Khoury & Flanders, 1996). The assumption made for case-only studies is that the exposure and the genotype are independent.

Genetic epidemiology provides the scientific foundation to measure the magnitude of disease risk associated with different alleles both in the presence and absence of certain key nongenetic risk factors for the disease. Like other scientific literature, genetic epidemiology literature also exhibits several biases (such as interplay of selective reporting and language biases) in the postulated epidemiological associations globally (Pan, Trikalinos, Kavvoura, Lau, & Ioannidis, 2005), and there exists a need for a global, transparent, comprehensive outlook in molecular population genetics and epidemiologic studies in general.

SECTION

II

Epidemiology of Oral Diseases and Conditions

13 Dental Caries

Dental caries needs cariogenic bacteria, dental plaque, catchment/stagnation areas, fermentable substrates (sugars), and susceptible tooth surfaces. The cariogenic bacteria share common properties such as they are acidogenic leading to pH <5.0 permitting dissolution of enamel; they can survive this low pH, and yet continue to produce more acids; and they can attach to tooth surfaces and produce glucans to allow production, maintenance, and growth of dental plaque. The cariogenic organisms involved in coronal caries are viridians streptococci (*Streptococcus mutans*, *S. sobrinus*, *S. salivarius*, *S. mitior*, and *S. sanguis*). Within 2–5 minutes of a 10% glucose rinse, the pH falls to close to 5, and takes about 1 hour to recover (Stephan curve). However, in the presence of dental plaque, the low pH is retained over a substantially longer time period. Experiments using sucrose as substrates have shown that compared to a single administration, sucrose is more cariogenic when given in several repeated small doses to maintain plaque activity. Root caries usually is caused by gram-positive pleomorphic rods—*Actinomyces israelii*, *A. gerencseriae*, *A. naeslundii*, *A. odontolyticus*, and *A. georgiae*.

Over the years, epidemiological studies have also provided evidence leading to confirmation of the experimentally observed effects. For example, ecological studies have shown low caries prevalence in populations that have low sucrose consumption; a drop in caries prevalence during a wartime sugar shortage occurred followed by an increase in caries prevalence when sugar became easily available in markets; and low caries prevalence in persons with sugar metabolism disorders such as hereditary fructose intolerance (Schuler, 2001). Various sugar and nonsugar sweeteners have been examined for their cariogenicity as shown in Table 13.1. In general, several studies have demonstrated the risk factors/indicators for dental caries that include increasing age, female sex, poor socioeconomic position, high cariogenic diet, and poor salivary flow.

> Despite progress in reducing dental caries, individuals in families living below the poverty level experience more dental decay than those who are economically better off. Furthermore, the caries seen in these individuals is more likely to be untreated than caries in those living above the poverty level; more than one third (36.8%) of poor children ages 2 to 9 have one or more untreated decayed primary teeth, compared to 17.3% of nonpoor children. In addition to poverty level, the proportion of teeth affected by dental caries also varies by age and race/ethnicity. Poor Mexican American children ages 2 to 9 have the highest number of primary teeth affected by dental caries (a mean of 2.4 decayed or filled teeth) compared to poor non-Hispanic blacks (mean 1.5) and non-Hispanic whites (mean 1.9). Among the nonpoor, Mexican American 2- to 9-year olds have the highest number of affected teeth (mean 1.8), followed by non-Hispanic blacks (1.3) and non-Hispanic whites (1.0). There are also differences by race/ethnicity and poverty level in the proportion of untreated decayed teeth for all age groups. (U.S. Department of Health and Human Services [USDHHS], 2000)

Caries Detection

The primary detection method for dental caries is through visual inspection. Visualization of early carious lesions (precavitated and small cavitated lesions in difficult-to-access areas) is a function of visual acuity. A systematic review of the English-language literature found that point estimates or reasonable range estimates for the diagnostic validity of methods

TABLE 13.1 Cariogenicity: Sugar and Nonsugar Sweeteners

Substrate	*Cariogenicity*	*Notes*
Sucrose	Highest	
Glucose, Fructose	Lesser	
Lactose, Galactose	Still lesser	
Glucose syrups and Maltodextrins	Less than sugars	Hydrolytic products of starch used as bulk sweetners
Hydrogenated glucose syrups and Lycasons	Less than sugars	Hydrolytic products of starch subsequently hydrogenated and used as bulk sweetners
Isomalt	Low	
Xyltiol, Sorbitol, Mannitol, Lactitol, etc.	None	Sugar alcohols
Sacaharin, Aspartame, Thaumatin, Acesulfame K, and Cyclamate	None	Nonsugar intense sweetners

Adapted from Cawson and Odell, 2002.

for the diagnosis of carious lesions could not be established because the number of reports of diagnostic performance involving primary teeth, anterior teeth, and root surfaces were insufficient to draw firm conclusions (Bader, Shugars, & Bonito, 2001a). A year later, another review by the same research group assessed the evidence describing a histologically validated performance of methods for identifying carious lesions comparing visual, visual/tactile, radiographic (film and digital), fiber optic transillumination (FOTI), electrical resistance measurements (ERM)/electronic caries monitor (ECM), and quantitative laser/light induced fluorescence (QLF) (Bader, Shugars, & Bonito, 2002). The investigators concluded that there was insufficient evidence to support generalizable estimates of the sensitivity and specificity of any given application of a diagnostic method. "The literature is problematic with respect to complete reporting of methods, variations in histological validation methods, the small number of in vivo studies, selection of teeth, small numbers of examiners, and other factors threatening both internal and external validity" (Bader et al., 2002). However, published studies from individual studies indicate that in general, the new quantitative methods (FOTI, ERM, and QLF) show high correlation with lesion depth and therefore might be better than visual methods for monitoring small changes in lesions over time.

A review of 29 caries detection criteria systems concluded that the majority of the current systems were ambiguous and did not measure the disease process at its different stages (Ismail, 2004). A recent study (Manton & Messer, 2007) set out to compare, in vitro, the effect of placing opaque and clear fluorescing pit and fissure sealants on the detection of occlusal caries. The investigators confirmed caries presence or absence histologically on serial sections examined under stereomicroscopy. They found that the sensitivity of all occlusal caries detection methods and their correlation with the histological gold standard was low, which led them to suggest that "tactile detection of occlusal caries should be discontinued, and the probe used only to clean the pits and fissures gently for more accurate visual detection, or prior to pit and fissure sealant placement" (Assaf, de Castro Meneghim, Zanin, Tengan, & Pereira, 2006). Because caries detection is a visual-oriented process, intra-examiner variation could be a source of measurement errors while measuring precavitated lesions. This study demonstrated that a high degree of agreement between examiners (Kappa >0.9) could be achieved through intensive standardization and calibration exercises.

The International Caries Detection and Assessment System (ICDAS; 2007, at http://www.icdas.org/) has developed extensive criteria for diagnosing and coding dental caries. A decision tree for coronal primary caries is outlined in Table 13.2. Further, ICDAS developed a two-digit coding method to identify caries associated with restorations/sealants. In this system, the first digit is one of the restoration digits in the second list below (96–99). The second digit represents the caries lesion code from the list above (0–9).

0— Non-restored tooth
1— Sealant placed partially
2— Sealant fully on the tooth
3— Tooth-colored restoration
4— Amalgam restoration
5— Stainless steel crown
6— Porcelain or gold or porcelain fused to metal (PFM) crown or veneer
7— Lost or broken restoration
8— Temporary restoration
9— This number is used as a prefix for special situations related to a missing/part of a tooth
  96—Tooth surface cannot be examined or surface is excluded for some reason
  97—Missing tooth because of caries (tooth surfaces are coded 97)
  98—Missing tooth for reasons other than caries (all tooth surfaces are coded 98)
  99—Unerupted tooth (tooth surfaces are coded 99)

Thus according to the ICDAS, a tooth with early visually detectable caries near a partially retained sealant would be coded 10. Similarly, a sec-

TABLE 13.2 Decision Tree for Coding Coronal Primary Dental Caries

Diagnostic Criteria and Outcome								*Caries Code*
Any carious discoloration when wet?	No	Any discoloration when dry?	No					0
			Yes					1
	Yes	Any cavitation?	No	Any shadowing?	No	Discoloration beyond pits and fissure?	No	1
							Yes	2
					Yes			4
			Yes	Is dentin exposed?	No			3
					Yes	More than 50% of total surface involvement?	No	5
							Yes	6

Adapted from International Caries Detection and Assessment System (ICDAS), 2007.

ondary caries lesion near an amalgam restoration would be coded 42 and an unrestored tooth with a distinct cavity extending into dentin would be scored 06. Although the ICDAS system is considered efficacious in detecting dental caries on coronal tooth surfaces (Ismail et al., 2007), it lacks reliability of caries detection on smooth approximal tooth surfaces. A root caries detection decision tree recommended by the ICDAS is shown in Table 13.3.

Because of a variety of statistical reasons in calculating agreement statistics using ICDAS coding systems that among other reasons include assumptions, number of categories used, and characteristics of the statistic, the simple Kappas may not be valid or comparable across studies. Weighted

TABLE 13.3 Decision Tree for Coding Root Caries

<table>
<tr><th colspan="6">Diagnostic Criteria and Outcome</th><th>Caries Code</th></tr>
<tr><td colspan="7">Primary Root Caries</td></tr>
<tr><td rowspan="3">Can root surface be directly seen?</td><td>No</td><td colspan="5"></td><td>E</td></tr>
<tr><td rowspan="2">Yes</td><td rowspan="2">Any color change after air drying for 5 deconds?[a]</td><td>No</td><td colspan="2"></td><td>0</td></tr>
<tr><td rowspan="2">Yes</td><td rowspan="2">Any cavitation?[b]</td><td>No</td><td>1</td></tr>
<tr><td>Yes</td><td>2</td></tr>
<tr><td colspan="7">Root Caries Associated with Root Restoration</td></tr>
<tr><td rowspan="3">Any color change adjacent to root restoration?</td><td>No</td><td colspan="4"></td><td>0</td></tr>
<tr><td rowspan="2">Yes</td><td rowspan="2">Any cavitation?[a]</td><td>No</td><td colspan="2"></td><td>1</td></tr>
<tr><td>Yes</td><td colspan="2"></td><td>2</td></tr>
<tr><td colspan="7">Root Caries Activity (for codes 1 and 2 from above)</td></tr>
<tr><td rowspan="3">Texture and appearance of base of discolored area</td><td>Smooth and shiny</td><td colspan="4"></td><td>Arrested</td></tr>
<tr><td rowspan="2">Rough and matted</td><td rowspan="2">Sensation on probing</td><td>Leathery</td><td colspan="2" rowspan="2"></td><td>Quiescent</td></tr>
<tr><td>Soft</td><td>Active</td></tr>
</table>

[a]Color change will appear as brownish (light/dark) or black coloration.

[b]Defined as loss of contour greater than 0.5 mm.

Adapted from International Caries Detection and Assessment System (ICDAS), 2007.

Kappa scores should be used when ICDAS codes are used in studies for calibration and standardization of examiners and caries categorization purposes. The ICDAS recommends that investigators provide the following reliability statistics for their study reports:

1. Kappa coefficients for comparisons between the senior examiner and each examiner separately.
2. Kappa coefficients for intra-examiner reliability for each examiner.
3. The Rows × Columns table should be included for all comparisons.
4. Where possible, the Stuart–Maxwell (SM) statistic tests the homogeneity of marginal frequencies should be presented.

Active vs Inactive Lesion

Dental caries is now understood to occur across a continuum from initial difficult–to-visualize subclinical and subsurface changes to the later stage of visually overt lesions manifesting as small cavities, with or without significant dentinal involvement at still later stages. Active carious lesions are in a dynamic state of flux. Precavitated lesions may either progress or regress (remineralization), whereas cavitated lesions may progress or be arrested. Once the lesion is arrested, it transitions to an inactive lesion. The likelihood of transition of an inactive lesion to active lesion is generally considered to be low. Caries activity assessment criteria include visual appearance, response to tactile feeling, and potential for or evidence of plaque accumulation. A systematic review to determine the strength of the evidence for the efficacy of professional caries preventive methods applied to high-risk individuals and the efficacy of professionally applied methods to arrest or reverse noncavitated carious lesions reported the following:

> The results do not indicate that the preventive and management methods reviewed are not efficacious; rather, they demonstrate that not enough is known to determine the efficacy of the methods. Suggestions for strengthening the limited evidence base involve the following: i) increasing the number of studies that examine prevention among high risk individuals and non-surgical management of non-cavitated lesions, ii) including a wider variety of subject ages, iii) targeting aspects of the efficacy questions not yet addressed, iv) strengthening research methods employed in the studies, and v) reporting methods and outcomes more completely. (Bader, Shugars, & Bonito, 2001b)

The ICDAS criteria for distinguishing between active and inactive carious lesions are shown in Table 13.4. The ICDAS codes active lesions between 1 and 3 depend upon the lesion's existence in the plaque stagnation area: pits and fissures, near the gingival, or approximal surface below the contact point; 4 if the lesion is "probably active"; and 5 or 6 if it is in dentine

TABLE 13.4 Characteristics of Active and Inactive Carious Lesions

	Characteristics of Lesion	
ICDAS Code	*Active Lesion*	*Inactive Lesion*
1, 2, or 3	Surface of enamel is whitish/yellowish opaque with loss of luster; feels rough when the tip of the probe is moved gently across the surface. Lesion is in a plaque stagnation area; that is, pits and fissures near the gingival and approximal surface below the contact point.	Surface of enamel is whitish, brownish, or black. Enamel may be shiny and feels hard and smooth when the tip of the probe is moved gently across the surface. For smooth surfaces, the caries lesion is typically located at some distance from the gingival margin.
4	Probably active	
5 or 6	Cavity feels soft or leathery on gently probing the dentin.	Cavity may be shiny and feels hard on gently probing the dentin.

Adapted from International Caries Detection and Assessment System (ICDAS), 2007.

and feels "leathery" or "soft." Inactive cavitated carious lesions that may appear shiny and feel hard on gently probing the dentin are coded 5 and 6, respectively.

However, for successful epidemiological investigations, studies comparable to each other, or valid surveillance programs, firm and universally valid diagnostic thresholds need to be described. A recent study tried to assess the reproducibility of a calibration trial at different diagnostic thresholds of dental caries over a 12-month period (Assaf et al., 2006). This study used WHO criteria and assessed the results of 11 trained, standardized, and calibrated examiners who had previous experience in epidemiological surveys. The examiners were standardized and calibrated five times throughout the study period—an initial training phase (theoretical and clinical) and in five calibration exercises (i.e., at baseline, 3, 6, 9, and 12 months). Although this rather intense program obtained healthy agreement and Kappa values (for DMFT and DMFS), the areas of disagreements occurred mostly in the initial lesion decay component diagnostic call. The question arises that if examiners who are experienced in epidemiological field survey settings continue to have diagnostic call-related disagreement linked to initial caries lesions, then perhaps technological diagnostic aids such as FOTI, QLF, Laser techniques, and electronic caries monitor (Tranæus, Shi, & Angmar-Ma'nsson, 2005) should be tested in field trials, and if found useful, should be used to assist in making a firmer definitive diagnostic call for initial caries lesion—either as an individual test, or using multiple testing strategy

(e.g., only "doubtful" initial lesions could be confirmed using a technological diagnostic aid). Of the technological diagnostic equipment, a laser caries detection aid is already in the market, which now has developed a handier "pen" device. Such a device can perhaps be conveniently used in surveys.

Caries Measurement Issues

The stage at which dental caries is measured significantly affects epidemiologic assessments of disease prevalence and treatment need in a population, as well as dental clinicians' practice decisions.

> If both precavitated and cavitated lesions are counted, a correct prevalence and incidence estimate will be obtained for "total caries burden," and a policy based on this estimate will presumably address the issue better. However, because preventive measures (primary, secondary, and tertiary) for precavitated and cavitated lesions are substantially different in conceptualization, planning, resource allocation, resource use, and measurement techniques, clear diagnostic criteria distinguishing these two formats of disease should be applied, and total caries burden should also be categorized by precavitated and cavitated lesion category. An important issue in caries measurement is the lack of coherent and standardized criteria applicable universally. (Chattopadhyay, Arevalo, & Sohn, 2008)

Whereas diagnostic accuracy of visual examination only to detect cavitated dental caries is not very high (sensitivity of 63% and specificity of 89%), use of a sharp explorer may improve accuracy and increase sensitivity to 92%, while a visual–tactile examination has a sensitivity of 92%. Sensitivity for precavitated lesion detection is similar to that of cavitated although without the additional benefit of using an explorer specificity is further lower, at 69% (Bader et al., 2001a). As noted earlier, since sensitivity and specificity of a diagnostic technique are fixed, false positive and false negative results are functions of diagnostic technique prevalence. Therefore as disease prevalence decreases, the positive predictive value of the test decreases (and negative predictive value increases); that is, fewer of those testing positive actually have the disease. In most developed countries, the general trend over the past few decades has been a reduction of caries prevalence. Therefore, diagnostic accuracy of tests with sensitivity in the range of 60–70% would lead to large numbers of false positives and have a major impact on resources. On the other hand, if prevalence of the disease increases, the positive predictive value will increase, thereby reducing false positives. Of course, a relative decrease of false positives, although heartening, is not necessarily a matter of relief because the amount of reduction may not be large enough to have a major impact in early caries detection. Secondly, it is possible that the increased sensitivity may be fueled by cavitated lesion detection, while precavitated lesions are not affected at all. Such an eventuality will only help progression of precavitated lesions, and not to

their regression. Therefore, the index of measuring dental caries must be able to make a distinction between precavitated and cavitated lesions.

Components of DMFT/S

Dental caries is a complex disease. Over the last several decades, a number of measurement criteria have been developed to identify the presence of dental caries. However, as the understanding of dental caries progressed, the clinical criteria systems remained focused on the assessment of the disease process at only one stage, the so-called "decayed" status (Ismail et al., 2007). The DMFT/S index score (i.e., the sum of the number of Decayed, Missing due to caries, and Filled Teeth/Surfaces of teeth in a person) is accrued over the life-course and is therefore a cumulative caries experience index that indicates a total caries experience of an individual and its sequelae. DMFT/S gives equal weight to decayed, missing because of caries, filled tooth, or tooth surface. Being an irreversible score, a person who gets caries once in life, continues to be counted as "diseased" through life even if no further caries occurs, or the treated lesion does not exhibit secondary caries. Even if the person then loses the tooth due to caries, the index continues to score the person as "diseased"; that is, the person is branded for life—one may be able to get rid of the tooth, but not the branding!

Teeth lost for reasons other than caries do not accrue DMFT/S. However, the "M" component of DMFT/S index has historically been a source of error because it is not ascertained properly, and even if ascertained, is based on a historical recall by the person (subject to recall bias). It has been suggested that recall bias would be more severe in DMFS compared to DMFT because the recall of the number of surfaces would be tougher, but this is a moot point because the number of missing surfaces can be calculated from the number of missing teeth under certain assumptions. Even if examiners ask about the reason a tooth is missing, if the subject replies "lost due to noncarious reasons" (periodontal disease most common), examiners almost never ask whether the tooth ever had caries. These issues make DMFT/S index a very complicated one to interpret, and thus it will always underestimate cumulative caries experience in a population. If DMFS is used, the subjects can never correctly recall the number of surfaces filled in each tooth. Furthermore, mixing of examiner-determined "D" and "F" with a self-recalled element would invalidate the index.

This phenomenon outlined above has important implications for assessment of risk factors and risk indicators of caries in epidemiological studies. For example, if a person had a cavitated or filled tooth scored using DMFT, but subsequently loses that tooth due to periodontal disease, orthodontic reasons, or trauma, and then, subsequent to the tooth-loss, the person's DMFT score woud get reduced contrary to the notion that DMFT/S is an irreversible index (periodontal disease is more prevalent than orthodontic

reasons or trauma in adults, but orthodontic reasons may be more prevalent in adolescents).

Each of the components: the D, M, and F, of DMFT/S, are associated with socioeconomic position (SEP). Low-income persons have greater caries prevalence because of higher incidence rate (D), higher tooth extraction rate (M), and lower treatment rates (F). In this case, reduced F components feeds into increased D and M components depending on the stage of advancement of disease when the person is assessed. This could be a major problem in assessing the poorer sections of society across the world. Because an M component can give rise to potentially greater recall bias, the net effect on the validity of caries estimation using DMFT/S may vary between populations, cultures, and countries.

People may not be able to recall if teeth were lost due to caries or periodontal disease. Therefore, the M component gets to be linked to periodontal disease due to recall bias/error. Both periodontal disease and dental caries are strongly associated with poorer SEP. Therefore, assessment of dental caries burden and caries outcomes would be confounded by periodontal disease status whenever SEP is incorporated in analyses. Figure 13.1 uses a DAG to demonstrate this effect under the social-causation paradigm that SEP is a cause of caries (as of periodontal disease). DAG-1 shows the usual situation that removing the direct path arrow does not leave any backdoor paths. However, because teeth may be lost due to periodontal reasons, a backdoor path opens up (DAG-2) between SEP and caries through periodontal disease. This potential confounding issue has never been considered in dental analyses. However, attempts have been made to estimate and adjust

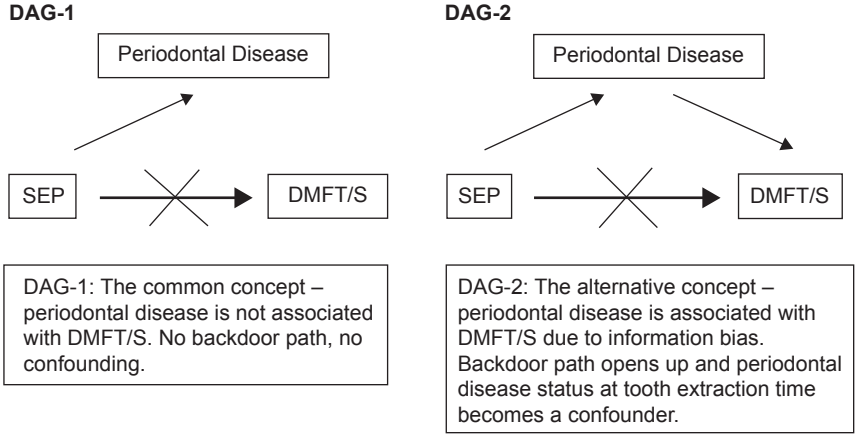


FIGURE 13.1 DAG Showing Relationships Between SEP, Periodontal Disease, and DMFT Dental Caries Index

for the M component to reduce bias related to it in estimating true caries burden in populations.

Adjustment for the M Component

Cross-sectional studies assessing prevalent caries in older adults report a substantial number of missing teeth, which makes accurate caries estimation difficult. To address this issue, several methods for statistical adjustment have been proposed. A "simple" caries increment adjustment at the person level, a "net" caries increment method, and an "adjusted" caries increment method have been proposed to make an adjustment to compensate for the M component in DMFT/S (reviewed by Broadbent & Thomson, 2005). The "simple" method simply takes the difference of DMFT/S of an individual between two time points and assesses the change over the time period. The "net" increment method deducts the "reversals" in caries scoring from the increments to arrive at a net change between follow-up periods (for an example of the problem with reversals, see the next section on cavitated and precavitated lesions). The "adjusted" caries increment method assumes that true reversals (due to demineralization–remineralization) are a rare phenomenon, and most reversals are attributed to examiner errors, and thus use a correction factor involving reversal count and new restoration counts between the follow-up periods. It has been suggested that the "adjusted" method should not be used if the reversal-to-increment ratio is below 1:10. As a better alternative to these strategies, incidence density measure has been used (Caplan et al., 1999) because change of incidence density between follow-up periods is a better indicator of true caries incidence. These methods assume that the follow-up periods are short enough to catch reversals and minimize recall bias about the M component. While these assumptions may be true for short-term planned studies, they do not provide for making stable population level estimates or for surveillance projects where follow-up periods are longer.

Lawrence, Beck, Hunt, and Koch (1996) proposed an adjustment formula to improve the method of estimating caries experience as expressed by the DMFS index in population groups with missing teeth. This formula assumes a follow-up study and in their demonstration example, the investigators used a 5-year follow-up time. They used two steps to arrive at the formula:

Step 1: Estimate the predicted DMFS at the follow-up; and Step 2: Develop a formula using baseline DMFS data to estimate the predicted prevalence. The DMFS-adjustment formula for each participant was calculated as:

$$\text{DMFS adjustment} = C + [N_{miss} \times (C \,/\, N_{press}) \times K]$$

where:

C = coronal or root DFS at follow-up
N_{miss} = number of missing surfaces at follow-up

N_{press} = number of tooth surfaces present at follow-up
K = constant population prevalence ratio for caries in teeth that were lost versus those that remained, expressed by the formula:
$K = (\Sigma A / \Sigma N_{miss}) \times (\Sigma N_{pres} / \Sigma C)$

where:

A = mean number of surfaces affected by dentinal lesions for teeth that were extracted from time baseline to follow-up.

For a new prevalence study, K would not be available, because it has been estimated from data that is external to the new prevalence study (i.e., K was estimated from a follow-up study *prior* to the start of this new prevalence study). This population constant, K, can vary by tooth type, demographic subgroups, and the number of years a particular tooth was missing, although as a risk ratio, K may be reasonably stable and transportable to other study populations. The investigators applied their results to the Piedmont 65+ Dental Study and demonstrated that the adjustment of all M surfaces avoided the biases inherent in the traditional DMFS and DFS indices and that this adjustment formula can be used without obvious bias (Lawrence et al., 1996).

Although this adjustment has not been used in subsequent studies, the potential to validate the adjustment model in different populations allows the development of strong estimates of the constant K as either a universal constant, or as a population-specific constant. If such a constant can be developed (i.e., perhaps revalidated periodically over a long time frame), then it can be applied to all population surveys and provide valid estimates of caries burden, as well as permit us to use the difference between adjusted and unadjusted DMFS as a marker for true burden of caries-related extractions. Such a marker can be used as a public health outcome measure in surveillance and for addressing caries-related extraction disparities.

Cavitated and Precavitated Lesions

DMFT/S is primed for cavitated lesions—once a cavity is detected, the person gets a score of 1, and if no other lesions occur, the score remains the same for the life of the person, even if the person loses the tooth due to caries. Therefore, DMFT/S can be considered to be caries sequela-invariant. However, the trend for early detection and prevention of dental coronal caries in the population of the developed world in an advantageous SEP has changed our view of case-definition for dental caries by turning it into two distinctly different kinds of diseases. In these populations, diagnosing dental caries implies making a diagnostic call on precavitated lesions. The rest of the world, however, still wrestles with cavitated lesions. These two conditions should be viewed as two different diseases for one key reason: precavitated caries may be reversed, but cavitated lesions cannot be reversed

(although an individual's DMFT/S may be reversed as we noted above), and the natural history of caries as a precavitated lesion undergoes a complete change at cavitation.

Measuring a precavitated lesion involves its own set of pitfalls. First, precavitated lesions are sometimes very difficult to distinguish from very mild or mild fluorosis and from other white conditions of teeth, and they are a source of diagnostic dilemma in epidemiological studies. Because both lesions (mild fluorosis and precavitated caries) appear as a structural defect resulting in a white-chalky surface lesion, differential diagnosis often rests upon clinical opinion. Furthermore, both lesions may reverse their physical presentation, opening up the potential for serious diagnostic errors. Examiner disagreements and between-examiner errors have been reported to be high (Sanchez-Figueras, 2003; Stookey, 2005; Stookey, Jackson, Zandona, & Analoui, 1999).

In addition, the reversal of precavitated carious lesions due to the remineralization of the tooth surface creates a diagnostic dilemma in follow-up visits. Because demineralization–remineralization can occur repeatedly over time at different rates, timing the periodicity of follow-up visits and the validity of case-ascertainment in follow-up studies become difficult problems, especially if the rate of change between different participants varies substantially. A direct requirement of this phenomenon is to schedule a series of follow-up visits close enough so that periods of remineralization and demineralization are not missed by the examiner(s). Data coding of this process can give rise to erroneous caries scores. Table 13.5 shows the D component scores (of DMFS) on any one tooth surface of five hypothetical subjects with precavitated lesions over time (baseline plus six follow-up visits). Because such studies assess a group sum of DMF over time (last row), the study shows absolutely no effect because the DMF continues to remain recalcitrant at "2" across all time periods. However, all five subjects have very different caries experience over time and show different susceptibility profiles. If this were a study assessing some kind of an intervention to help remineralization or prevent demineralization, then a summary DMF score would be an invalid measure for caries outcome.

TABLE 13.5 Precavitated Caries Follow-Up of Five Hypothetical Subjects

Subject	*T0*	*T1*	*T2*	*T3*	*T4*	*T5*	*T6*
1	0	1	0	1	0	1	0
2	1	1	0	0	1	1	1
3	1	0	0	0	0	0	1
4	0	0	1	0	0	0	0
5	0	0	1	1	1	0	0
Total DMF	2	2	2	2	2	2	2

Score reversals occur over time due to demineralization–remineralization fluxes (0 = no caries; 1 = caries present).

Another problem that is rarely discussed in the literature is that the risk factors for precavitated and cavitated lesions among adults may be different from those of children. Most research of risk factors for precavitated lesions has focused on children. Because dental caries is progressive from precavitated lesion to cavitation, preventive efforts that fail in the precavitated stages will only increase the recorded DMFT/S score based on caries burden. A tooth having a large filling will have substantially fewer surfaces at risk for primary caries, but a greater risk for secondary caries. Similarly, once extracted, the tooth also drops out of the risk pool for caries although not only does this not impact the denominator data, but the missing tooth continues to be counted as diseased in DMF/S.

Root Caries

The Root Caries Index (RCI) was initially described to estimate the prevalence of root caries and was restricted to subgingival lesions because teeth were not considered to be at risk unless roots were exposed. However, this resulted in the underestimation of root caries. Since then, RCI incorporates both supra- and subgingival root caries lesions, and their scores are recorded separately (Katz, 1996).

$$\text{RCI} = [(\text{Decayed} + \text{filled root surfaces}) \times 100] \,/\, [\text{Decayed} + \text{filled} + \text{sound root surfaces with periodontal attachment loss}]$$

A study found that 55% of the restorations placed on root surfaces were for cervical wear and sensitivity, and not for root caries (Walls, Silver, & Steele, 2000). The investigators developed a correction factor to make allowance for the proportion of restorations placed because of wear and sensitivity.

Apparently coronal and root caries tend to appear together in the same individuals, but fillings attenuate that relationship. The impact of dental treatment on the epidemiology of dental caries appears to be considerable and calls into question whether the F component of the caries index is related to disease as defined by epidemiologic criteria (Beck & Drake, 1997). Therefore, proper handling of restoration-related data is critical to the accurate assessment of caries prevalence.

Caries Risk Assessment

Development of a valid and easy-to-use caries risk assessment tool will go a long way in identifying a high-risk group of children for targeting caries prevention activities and monitoring them for caries status change. Goals of caries risk assessment have been identified by Messer (2000) as (1) to identify high-risk patients before they become caries active, (2) to screen out low-risk patients to allow long recall intervals, and (3) to monitor changes in disease status in caries-active patients. However, dental caries is a multi-

factorial disease with risk factors acting at multiple levels. Several risk assessment tools and models have been suggested—they simply measure the causes of dental caries when disease is present or after disease occurrence or they measure the actual existence of disease. Such methods are self-fulfilling prophecies, or at best, assess the risk for *subsequent* disease. Risk factors for first occurrence of disease may exhibit a somewhat different profile. None of the caries risk-assessment methods have demonstrated a high level of consistent performance across different populations, either at the level of individual patient diagnostics or as a screening test.

Overall, dental caries risk-assessment tools can be divided into several groups as mentioned below.

1. SEP, oral hygiene, and dietary factor-based methods
2. Infant behavioral factor-based methods
3. Past caries experience-based methods
4. Salivary and microbial factor test-based methods
5. Salivary oligosachharide content-based methods

The American Academy of Pediatric Dentistry (AAPD, 2006) has suggested a caries risk-assessment tool in the form of a listing of risk indicators based on history (11 component factors), clinical evaluation (four component factors), and optional supplemental professional assessment (two component factors). The patient is graded under one of the three categories: high, moderate, or low risk for each of the 17 listed factors, and the patient's overall assessed risk for developing caries as the highest level of risk is scored for any of the listed factors. This risk-assessment tool is an irreversible categorization method. For example, one of the components in the history section is that the patient has decay (i.e., yes: high risk; no: low risk). Therefore, once any child gets caries, the child continues to be at high risk forever. Such methods are insensitive to track effectiveness of preventive programs and assume a dichotomous-static disease risk category (i.e., the risk of disease, once acquired, remains static), and they do not deal with the magnitude of the risk and possible changes in the disease risk.

A "Cariogram," which is a past caries experience-based method, has been suggested as a method to describe a new way of illustrating the caries risk profile of an individual (Bratthall & Hänsel Petersson, 2005). The Cariogram is a statistical prediction model "weighted" analysis of the input data, using mainly biological factors. It expresses to what extent different etiological factors of caries affect caries risk. The Cariogram aims to identify the caries risk factors for the individual and provides examples of preventive and treatment strategies to the clinician. It uses the following factors: caries experience, related diseases, diet, contents, frequency, plaque amount, mutans streptococci, fluoride program, saliva secretion, and saliva buffer capacity. The information is then given a score on a scale ranging from 0 to 3 (0–2 for some factors) according to predetermined criteria. A score of 0 is

the most favorable value and a maximum score of 3 (or 2) indicates a high, unfavorable risk value. The output is a color-coded pie chart. The Cariogram does not specify the particular number of cavities that will or will not occur in the future, it just classifies people into risk groups.

A recent study reported that a salivary oligosachharide component-based assay was analyzed using a combination of multiple linear regression and neural net analyses to develop the algorithms that describe the relationship between some salivary mucin (oligosaccharides) patterns and DFT (Denny, Denny, Takashima, Galligan, & Navazesh, 2007). Although this study claimed a comprehensive predictive model, the report was based on a small convenience sample, and some analytical aspects were not fully explained, including the complete sample size (apparently different sizes were used for different analyses), rationale, and methods on which grouping of different teeth was based (an important outcome metric in the study). Although only some mucins are present in saliva (Offner & Troxler, 2000), salivary mucins have been shown to be nutrients for plaque-forming bacteria (Wickström & Svensäter, 2008), and are correlated with persons having gastric diseases harboring *Helicobacter pylori* (Silva et al., 2008), rehardening of root caries-like lesions (Turssi, Lima, Faraoni-Romano, & Serra, 2006), and differential effects on hydroxyapatite crystals (Park, Chung, Kim, Chung, & Kho, 2006). Therefore, at this time, a salivary mucin-based test for caries prediction should be considered to be in the early phases of development.

Current evidence suggests that children who acquire *Streptococcus mutans* early in life are at greater risk for dental caries and that diet and oral hygiene may interact so that if there is a balance of "good" habits by way of maintaining good plaque control and "bad" habits by way of having a cariogenic diet, the development of caries may be controlled (Harris, Nicoll, Adair, & Pine, 2004). However, good oral hygiene habits and avoidance of cariogenic foods can compensate for these risk factors. These results are based on cross-sectional studies; well-designed longitudinal studies are lacking. Some studies have demonstrated that children with caries from deprived and nondeprived backgrounds had a different caries-associated flora—children living in deprivation harbored more *Streptococcus mutans* and Lactobacilli (Beighton et al., 2004). The greater microbial count was, however, associated with a greater amount of cavitated lesions. The argument appears somewhat circular because if the burden of caries is greater in any group, the burden of caries-causing organisms is also expected to be greater in that group. Therefore, such evidence cannot be accepted as evidence for having greater risk of harboring the organisms *before* the caries occurred—such hypotheses can be tested only through well-designed longitudinal studies.

Awareness of the need to utilize more appropriate statistical procedures, such as the assessment of incidence density over traditional logistic

regression-based odds ratio estimation, came through studies that demonstrated that baseline risk factors were significant predictors of caries risk. As discussed in earlier chapters, incidence density assessment helps assess etiological relationships better than the traditional prevalence ratio assessment.

> Recent dental research in the area of risk assessment is focusing on the evaluation of new, technologically advanced methods for the diagnosis of caries, perhaps in the hope that their improved sensitivity and specificity (compared with those of conventional methods) will increase the chance of detecting small treatment effects or discriminate between competing treatment modalities, even in low-risk populations. These methods may also help to identify caries development sooner, long before caries is clinically evident. Ultimately, we want to develop treatments that will delay or suppress caries development, allowing newer diagnostic methods to provide an earlier signal or marker of impending caries development. The use of survival or time-to-event methodologies is clearly relevant in this area, including the work of Hannigan, O'Mullane, Barry, Schafer, and Roberts (2000), exploiting the use of the log–logistic model for clustered survival data, and that of Hujoel and coworkers (1994), applying the Poisson regression model to caries incidence. Cox proportional hazards regression analysis also makes sense to explore caries risk factors, as recently used to determine the relationship between salivary mutans streptococci (MS) counts and caries incidence in Japanese preschoolers (Ansai et al., 2000). Each of these models allows for the inclusion of subject- and surface-specific explanatory variables to identify risk factors that affect caries development. Risk factors identified through the use of survival methods can guide the selection of appropriate high-risk subpopulations for future study—in this case, the subset of tooth surfaces and subjects likely to benefit from preventive therapies. (Johnson, 2004)

Using the Poisson regression model, Powell, Leroux, Persson, and Kiyak (1998) demonstrated that the risk of coronal caries was greater in participants with high-baseline caries risk factors such as male gender, Asian ethnicity, root DMFS, *streptococcus mutans* count, and lactobacillus count.

Caries Prevention

Caries prevention using fluoride systems is discussed in Chapter 17. The mainstay of caries prevention using the clinical method is the use of pit and fissure sealants. "The notion of retention is central because the main function of sealants is to change pit and fissure morphology to form an efficient physical barrier between the enamel surface and oral environment for as long as possible" (Muller-Bolla, Lupi-Pégurier, Tardieu, Velly, & Antomarchi, 2006). In a recent systemic review, these investigators found the following evidence about usefulness of pit and fissure sealants in caries prevention.

1. The retention rate of autopolymerized and light-cured resin-based sealants did not differ significantly.
2. Light-cured resin-based sealants had a significantly higher retention rate than fluoride-containing light-cured resin-based sealants at 48 months or more (RR: 95% CI = 0.80: 0.72–0.89).
3. Using a rubber dam did not affect retention of autopolymerized resin-based sealants.
4. Using a rubber dam improved retention of fluoride-containing light-cured resin-based sealants (2.03: 1.51–2.73).
5. No recommendable best clinical procedures could be offered because not enough studies existed comparing the many combinations. The studies that existed were of poor quality. Potential sources of variations in clinical practice that could impact sealant retention rates, but for which no evidence exists, include tooth-cleaning method, isolation stage, enamel surface preparation and/or acid etching, and adhesive agent application.

Dental Caries Clinical Trial Design

Relatively low prevalence of dental caries in the developed world has thrown up challenges for dental clinical trials. Factors that affect caries clinical trials may be participant characteristic related (i.e., age, gender, SEP, brushing habits, ability to follow instruction, compliance); disease/exposure rate related (i.e., caries prevalence/incidence, fluoride exposure); study design related (i.e., study drop-out rates, exclusion criteria, length of study, stratification, examination techniques used, inter- and intra-examiner variability); or study analysis related (i.e., sophisticated statistical methods that impact study design, data collection, and appropriate interpretation). These challenges arise from four main sources: (1) reduced prevalence and risk of caries and related study power issues, (2) difficulty in addressing the dynamic flux from demineralization–remineralization rate differences and organizing follow-up visits, (3) increasing difficulty of using placebos due to ethical reasons arising out of increasing standards of care, and (4) modifications needed for utilizing more sophisticated statistical analysis to address the challenges mentioned above. To address these challenges, several design-related solutions have been suggested (Johnson, 2004), such as:

1. Use of stratification and use of randomized block design
2. Conducting caries risk assessment
3. Use of properly validated diagnostic methods
4. Baseline adjustment for caries-related risk factors
5. Use of more appropriate statistical data analysis procedures

Several other issues related to clinical trials in general can also effect dental caries trials and have been discussed in Chapters 2 and 11.

14 Periodontal Diseases

Classification of Periodontal Diseases

The three basic questions asked when diagnosing periodontal disease are (Armitage, 2004): (1) What periodontal disease or condition does the patient have? (2) How severe is the problem? and (3) Is the disease or condition localized or generalized? Classifying periodontal disease has been a complex problem for a long time, and it continues to be so. The dilemma faced in classifying and defining periodontal disease stems from newer understanding of underlying disease etiopathogenesis and the fact that like dental caries, periodontal disease may occur differently around different teeth, as well as in different sites of the periodontium surrounding the same tooth! Apart from its location, the different rate of progression of the disease, its different pathophysiological profile, and its range of presentation has added to the difficulty in our ability to define, classify, and measure the disease accurately. The term *periodontal disease* perhaps represents a group of closely related different diseases with similar presentation rather than a single disease entity.

In 1989, the World Workshop of Clinical Periodontics offered a classification of periodontal disease based on the then current understanding of disease pathophysiology emphasizing periodontal disease as an outcome of a host–infective agent interaction (American Academy of Periodontology, 1989; see Box 14.1). The important problems with this classification system include: different classification criteria were used to define disease in different categories (e.g., age-related disease [adult/early onset]; rate of disease progression [slow/rapidly progressing periodontitis]; clinical symptoms based categories [necrotizing ulcerative periodontitis]; and clinical outcome based category [refractory periodontitis]); categories were not

exclusive; and all categories were not homogenous (e.g., "refractory periodontitis" included all types of etiologically based categories not responding to therapy).

Etiology-based classification may become complicated because periodontal microbial flora changes with progression of disease, and its stages may imply that all diseases due to all organisms across all clinical stages need to be grouped into one classification irrespective of their rate of progression, clinical response, and outcomes. For example, the microbial flora at nonresponsive sites have been known to be different than responsive sites; for example, nonresponsive sites may harbor enteric rods, staphylococci, and Candida, whereas those with recurrent disease harbor *Porphyromonas gingivalis*, *Prevotella intermedia*, *Eikenella corrodens*, *Streptococcus intermedius*, or microbial complexes consisting of various combinations of *P. gingivalis*, *S. intermedius*, *Treponema denticola*, *Campylobacter rectus*, *Bacteroides forsythus*, *Peptostreptococcus micros*, and *Fusobacterium nucleatum* (Armitage, 2002). Furthermore, different etiological mechanisms may be involved in different progression rates of periodontal diseases such as innate and acquired host susceptibility, composition and quantity of the subgingival flora, and the nature of genetically determined host response to microbial challenge (Armitage, 2002). Overall, six major problems were identified with this classification system: (1) gingivitis/gingival diseases were not included; (2) age-dependent criteria of periodontitis was not validated; (3) rates of progression crossed over across different periodontitis categories, yet a heterogeneous "Rapidly Progressive Periodontitis" category existed; (4) extensive overlap in the clinical characteristics of the different categories of periodontitis; and (5) "Refractory Periodontitis" and (6) "Prepubertal Periodontitis" were heterogeneous categories (Armitage, 2002). The 1999 International Workshop for a Classification of Periodontal Diseases and Conditions developed another classification with eight categories, each with several subcategories (see Box 14.1).

As genetic and molecular technology becomes easily accessible, and more information repositories are built up, it will become possible to classify periodontal diseases based on sets of organisms, the host's genetic makeup based on response to organisms, the host's immune system, the response to environmental challenges such as smoking, and so on. Whether such systems lead to easily utilizable classification systems, or they overclassify excessively detailed listings devoid of generalizability remains to be seen. The concept of periodontal disease etiopathology has changed as the new paradigm suggests initiation of disease by specific bacteria within a biofilm that stimulates an immunoinflammatory host response, resulting in host tissue destruction in bursts. Disease modifiers, which may be of genetic, environmental, or acquired origin, have been recognized as major determinants of disease severity and progression (Oringer & Williams, 2000).

BOX 14.1 Periodontal Disease Classification by the World Workshop on Clinical Periodontics (American Academy of Periodontology, 1989); and International Workshop for a Classification of Periodontal Diseases and Conditions (1999)

1989 Classification

1.1 Adult Periodontitis
1.2 Early Onset Periodontitis
  1.2.1 Prepubertal Periodontitis
    1.2.1.1 Generalized
    1.2.1.2 Localized
  1.2.2 Juvenile Periodontitis
    1.2.2.1 Generalized
    1.2.1.2 Localized
  1.2.3 Rapidly Progressive Periodontitis
1.3 Periodontitis Associated with Systemic Diseases
1.4 Necrotizing Ulcerative Periodontitis
1.5 Refractory Periodontitis

Six major problems were identified with this classification system: (1) gingivitis/gingival disease not included; (2) age-dependent criteria of periodontitis was not validated; (3) rates of progression crossover across different periodontitis categories, and a heterogeneous "Rapidly Progressive Periodontitis" category; (4) extensive overlap in the clinical characteristics of the different categories of periodontitis; and (5) "Refractory Periodontitis" and (6) "Prepubertal Periodontitis" were heterogeneous categories.

1999 Classification Major Categories (Subcategories)

1. Gingival diseases (dental plaque-induced gingival diseases; non-plaque-induced gingival lesions)
2. Chronic periodontitis (localized; generalized)
3. Aggressive periodontitis (localized; generalized)
4. Periodontitis as a manifestation of systemic diseases (associated with hematological disorders; associated with genetic diseases; not otherwise specified)
5. Necrotizing periodontal disease (necrotizing ulcerative gingivitis; necrotizing ulcerative periodontitis)
6. Abscesses of the periodontium (gingival abscess; periodontal abscess; periocoronal abscess)
7. Periodontitis associated with endodontic lesions (combined periodontal–endodontic lesions)
8. Developmental or acquired deformities and conditions (localized, tooth-related factors that modify or predispose to plaque-induced gingival disease/ periodontitis; mucogingival deformities and conditions around teeth; mucogingival deformities and conditions on edentulous ridges; occlusal trauma)

Adapted from Armitage, 2002.

Measuring Periodontal Disease

Gingival Disease Measurement

Gingivitis is the inflammation of the gingiva in which the junctional epithelium remains attached to the tooth at its original level. Measurement of gingival disease is based on two paradigms: measuring signs of inflammation and measuring bleeding on probing. The Gingival Index (GI) of Löe and Silness (1963) that is still used employs inflammation criteria to grade gingivitis on an ordinal scale of 0 to 3 assessed by probing gingiva on the mesial, distal, lingual, and buccal surfaces of the teeth. The GI scores reflect progressively greater signs of inflammation (0—normal, healthy gingival; 1—mild inflammation [no bleeding on probing]; 2—moderate inflammation [bleeding on probing and other signs]; and 3—severe inflammation [spontaneous bleeding and other signs of inflammation]). All teeth are measured and an average score is assigned for the individual. Generally, a GI of 0.1–1.0 is interpreted as mild inflammation; a GI of 1.1–2.0 is interpreted as moderate inflammation; and a GI of 2.1–3.0 is interpreted as severe inflammation. Higher scores in a single site tend to be compromised over a large number of sites being measured, thereby undermining the sensitivity of the index. Therefore, sometimes the numbers of sites with a certain threshold score are assessed separately for more comprehensive description; for example, *n*% of sites with a score of 2 or more.

The GI is not sensitive to subtle early gingival changes or to changes in gingiva in the mid-ranges of pathophysiology. Two of the main questions that arise for using the GI are: How much inflammation is needed for the gingivitis to exist, and how much probing pressure is required to elicit bleeding? The reliance of the GI on bleeding is a limitation because different examiners may use different pressure (varying between 3 to 130 gm depending on the examiner) to elicit bleeding response, which may invalidate findings and increase between-examiner error rates creating measurement bias. Gingival probing is not recommended for screenings, surveys, or surveillance because of measurement issues and sensitivity issues related to the index.

The GI continues to be the core index for measuring gingivitis—a modified GI (MGI) has been developed based on the original GI. The MGI eliminated gingival probing and redefined the scoring for mild and moderate inflammation. The main reason for removing the probing requirement was to reduce the probability of disturbing plaque, reduce traumatizing the gingiva, and reduce multiple calibration exercises to lessen between-examiner errors. MGI scores gingiva on a 5-point scale: 0—no inflammation; 1—mild inflammation involving any portion of but not the entire marginal or papillary gingival unit; 2—mild inflammation involving the entire marginal or papillary gingival unit; 3—moderate inflammation; and 4—severe inflam-

mation. MGI assesses either full mouth or only partial mouth, and it calculates mean scores at an individual or population level.

Other indices commonly used to measure gingivitis include the Bleeding Index, the Plaque Index, and a Modified Gingival Margin Plaque Index. These modifications usually assess the extent of the gingiva involved in gingivitis and do not add to the basic sensitivity and validity problem of the GI: whether bleeding or a probing-based index is sensitive enough for subtle changes in gingival pathology, and whether bleeding truly reflects the underlying etiopathology of gingivitis.

Periodontal Disease Measurement

The essential feature assessed in measuring periodontal disease involves probing depth, pocket depth, and attachment loss (clinical attachment loss, or CAL) (see Figure 14.1). However, the presence of attachment loss at a given site does not necessarily indicate that there is active periodontal disease at that site! Like the DMFT, attachment loss is a cumulative measure—therefore, attachment loss at a given site indicates only a history of periodontal disease at that site. NHANES measures attachment loss, probing depth, and pocket depth. In NHANES, the periodontal examination was conducted at two sites, mid-buccal and mesiobuccal, for each tooth, in two randomly chosen quadrants, one maxillary and one mandibular, on the assumption that conditions in these two quadrants would represent the disease in the whole mouth. Third molars were excluded because of their frequent extraction in young adulthood, so a maximum of 14 teeth and 28 sites per individual were examined (Borrell, Burt, & Taylor, 2005). Commonly used periodontal indices are outlined in Box 14.2.

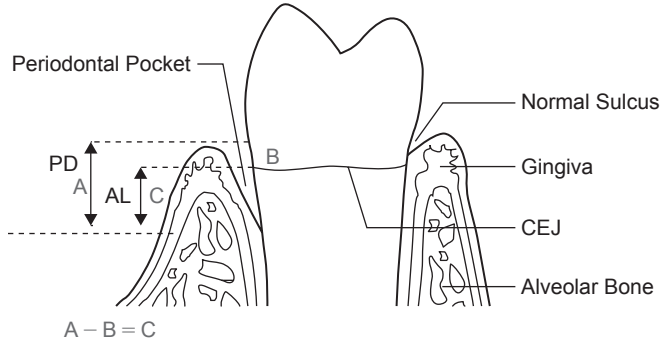


FIGURE 14.1 Clinical attachment loss assessment
PD: Pocket depth; AL: Attachment level

BOX 14.2 Salient Features of Periodontal Indices

Periodontal Disease Index (PDI) (rarely used today)

Procedure

- Assess gingival inflammation at the six teeth
- G0: Absence of inflammation
- G1: Mild to moderate inflammatory gingival changes not extending all around the tooth
- G2: Mild to moderate severe gingivitis extending all around the tooth
- G3: Severe gingivitis characterized by marked redness, tendency to bleed, and ulceration
- Record pockets for the six teeth

The distance from the free gingival margin to the cemento–enamel junction and the distance from the free gingival margin to the bottom of the gingival crevice or pocket should be recorded for the mesial, the facial, the distal, and the lingual aspects of each tooth examined.

PDI Score Calculation

- If attachment loss (AL) is 0 for a tooth, then the PDI score for the tooth is the gingival score
- If AL is present, then
  - <3 mm PDI = 4
  - >3–6 mm PDI = 5
  - >6 mm PDI = 6

The Periodontal Index (PI) of Russel does not include an assessment of clinical attachment loss, and all pockets of 3 mm or larger graded equally, combining both gingivitis and periodontitis on the same 6-point scale (min = 0 max = 8).

Periodontal Disease: NIDCR Protocol

Measurement Coverage

- Periodontal assessment
- Gingival assessment (bleeding)
- Calculus assessment
- Periodontal destruction assessment
- Loss of attachment
- Furcation involvement

(Continues)

BOX 14.2 Salient Features of Periodontal Indices *(Continued)*

Loss of Attachment

- All teeth in facial and mesio-facial of teeth in two quadrants (random upper and contralateral lower)
- Attachment loss assessed by indirect methods of Ramfjord
- NIDCR probe used
- Pocket depth and attachment loss is measured at each site

Furcation Involvement

- Assessed on eight teeth
- Maxillary first and second molars
- Maxillary first premolars
- Mandibular first and second molars
- Assessed at mesial, facial, and distal of maxillary molars; facial and lingual of mandibular molars
- Scored 0 (none), 1 (partial), or 2 (through)

Calculus

0: No calculus
1: Supragingival calculus
2: Supra- and subgingival calculus

Extent and Severity Index (ESI)

- PI and PDI do not provide information on the extent of disease
- ESI developed to provide estimates of extent and severity
- Does not assess gingival inflammation
- Assesses AL using the Ramfjord method
- Same teeth and sites as NIDCR protocol, although often used for full mouth exams
- Must decide on a threshold of disease, such as AL > 3 mm
- Extent score is the percentage of sites in a person who has AL greater than the threshold
- Severity score is the average AL per site among the sites with AL
- ESI = (20, 3.0): 20% of the sites examined had disease, and of the sites with disease, the average AL was 3.0 mm.

The ESI uses estimates of attachment level from probing measurements of 14 sites in a quadrant in the maxillary arch and 14 in the contralateral mandibular arch.

(Continues)

BOX 14.2 Salient Features of Periodontal Indices *(Continued)*

Community Periodontal index of Treatment Needs (CPITN)

Evaluation

- The worst finding in each sextant is coded is recorded and used
- The maximum code for the entire mouth is used for the treatment recommendation

Grading

Grade 0: No signs of periodontal disease
Grade 1: Gingival bleeding after gentle probing
Grade 2: Supragingival or subgingival calculus
Grade 3: Pathologic pockets 4–5 mm deep
Grade 4: Pathologic pockets > = 6 mm deep

Treatment Recommendation

- Maximum score 0: No need for additional treatment
- Maximum score 1: Need to improve personal oral hygiene
- Maximum score 2: Need for professional cleaning of teeth, plus improvement in personal oral hygiene
- Maximum score 3: Need for professional cleaning of teeth, plus improvement in personal oral hygiene
- Maximum score 4: Need for more complex treatment to remove infected tissue

CPITN treats gingivitis and periodontitis in the same scale (i.e., assuming that periodontitis is merely an extension of gingivitis); CPITN is not indicated as a tool to assess prevalence of periodontal disease.

Radiographic Assessment of Bone Loss

- Rarely used for epidemiological studies
- Usually measured from bitewing radiographs
- Distance from cement-enamel junction(CEJ) to alveolar crest
- Expressed as millimeter or percent of root length
- Highly correlated with measurements of AL with periodontal probes
- Several sources of error and difficult to standardize across clinics and research centers for appropriate comparison of resultant data
- Variations in projection geometry
- Variations in film contrast and density
- Obstruction of views

(Continues)

BOX 14.2 Salient Features of Periodontal Indices *(Continued)*

- Computerized programs can detect bone changes as little as 0.5 mm
- Digital subtraction radiography and computer-assisted densitometric image analysis

Ramfjord Teeth

- Tooth #3: Right maxillary first molar
- Tooth #9: Left maxillary medial incisor
- Tooth #12: Left maxillary first premolar (bicuspid)
- Tooth #19: Left mandibular first molar
- Tooth #25: Right mandibular medial incisor
- Tooth #28: Right mandibular first premolar (bicuspid)

Measurement Issues in Periodontal Disease

Like measurement of dental caries, periodontal disease measurement suffers from several conceptual and implementation problems that provide challenges to the validity of the disease measures. Essentially, periodontal disease is measured as an amount of periodontal destruction that has already occurred historically.

Which Teeth to Measure?

The ideal solution would be to measure the entire periodontium—however, such a measurement would be inefficient in terms of time, effort, and involved expenses, but provide useful information about disease burden from regular periodic measurements for comparison, especially if teeth are lost. An extension of the measurement issue also includes a decision about which sites of a tooth should be measured; that is, all six sites (mesiobuccal, buccal, distobuccal, mesiolingual, lingual, and distolingual)? Typically, all sites of the teeth being assessed are measured. One goal of periodontal measurement research is to find the set of teeth that provides completely valid, representative information about the periodontal disease status of a person. "Ramfjord's teeth" includes a partial mouth recording of a set of six teeth presumed to represent the entire mouth during examination (see Box 14.2). Other mechanisms of selecting representative teeth employ methods that use the most affected teeth (overestimate disease if prevalence is low), teeth selection based on stratification of the mouth and quadrants based on some criteria, or random selection of a set of teeth. Beck, Caplan, Preisser,

and Moss (2006) demonstrated that a truer representation of periodontal disease burden in the mouth is obtained through random sampling of measurable sites in the mouth rather than through a fixed set of teeth. They reported that the bias and precision of probing depth and clinical attachment-level estimates of fixed partial examination methods (e.g., Ramfjord teeth) of randomly selected sites in the mouth compared to full-mouth examinations. Reporting their findings from testing six teeth was compared to that of the whole mouth; they suggested that the subset is usually preferred because it is easier and faster to complete (Beck et al., 2006).

It has been anecdotally suggested that among the fixed-site methods, Ramfjord's method is least biased and has the best representation of periodontal disease in the mouth. Beck et al. (2006) compared Ramfjord's method with a method of randomly selecting 36 sites in the mouth and found that Ramfjord's method had greater bias and relative error, thus concluding that randomly selected sites provide a better picture of the true burden of periodontal disease in the mouth compared with any fixed-site-based measurement method. However, as of periodontal disease prevalence estimation, Beck et al. (2006) noted that that although both methods (Ramfjord's method and the random selection method) underestimated the prevalence of periodontal disease; the random-site selection methods were less likely to underestimate prevalence than fixed-site-based methods.

What Is Measured as Periodontal Disease?

The current way of measuring periodontal disease is a historical measure of attachment loss, very similar to the DMF type of measure. However, it does not measure the etiologic process of periodontal disease. For an etiologically representative measure, measurements of host–microbe interaction such as inflammatory cytokines (e.g., PG-E2, TNF-a, Interleukin-1a, Interleukin 1ß) might provide better markers for the disease process. For instance, bursts of periodontal destruction (Cohen, 1993) would be correlated with inflammatory cytokine titers rather than historical attachment loss. Active lesions therefore would be expected to correlate better with cytokine titers, whereas old/inactive lesions would correlate with attachment loss type of measurement. Whereas periodontal disease prevalence may be defined using attachment loss, active lesions may be defined only if lesion activity is detected. To assess incidence of periodontal disease (as opposed to historical burden), it would be of interest to assess the active lesions rather than depend on change of attachment loss alone. Effective screening and disease prevention activities would also require an assessment of the burden of active lesions. If etiological mechanisms are to be assessed, then follow-up visits would need to be planned according to the dynamics of periods of periodontal bursts rather than normally assuming a continuous monotonous linear change in disease occurrence.

Periodontal disease risk factors include inadequate oral hygiene, less frequent dental visits (fewer cleanings, fewer extractions), diabetes, immune deficiencies, AIDS, smoking, lower education, lower income, and race (African American, Hispanics, and Native Americans are more likely to have periodontitis compared to Whites, and are more likely to have more severe forms of periodontitis). It is currently known that sites of advanced bone loss harbor anaerobic, gram-negative species (e.g., *B. gingivalis*, *A. actinomycetemcomitans*), while adjacent healthy sites contain facultative gram-positive aerobes (e.g., *Streptococci*). Of the > 300 types of bacteria in the oral cavity, only a few are implicated as causing periodontal disease.

Epstein–Barr virus (EBV) DNA is detected in 60–80% of aggressive periodontal lesions and in 15–20% of gingivitis lesions or normal periodontal sites. The periodontal presence of EBV is associated with an elevated occurrence of periodontopathic anaerobic bacteria. EBV and cytomegalovirus often coexist in marginal and apical periodontitis. EBV proteins up-regulate cytokines and growth factors, which seem to play a central role in the proliferative response of tongue epithelial cells in oral hairy leukoplakia and in the cell-transformation process of EBV-associated malignancies (Slots, Saygun, Sabeti, & Kubar, 2006). It has been suggested that a herpesvirus active infection in the periodontium impairs local defenses and permits overgrowth and increased aggressiveness of periodontopathic bacteria. In turn, periodontal pathogenic bacteria may augment the virulence of periodontal herpesviruses. Furthermore, interactions among herpesviruses and specific bacterial species may be an important pathogenic feature of periodontitis (Oringer & Williams, 2000; Slots, 2007). A suggested screening test for periodontal disease is the measurement of aspartate aminotransferase (AST) level in gingival crevicular fluid (GCF). As assessed against periodontal disease measured using CAL, AST measurement in GSF has a sensitivity of 93% and specificity of 68% (Oringer & Williams, 2000).

Analysis Issues in Periodontal Disease

Several analytical issues have been raised in assessment of periodontal diseases (outlined in Box 14.3). Across a population, exposures on teeth differ, and teeth should be viewed as clustered inside individuals rather than individual entities that are uniformly exposed to individual or ecology–level exposures. Therefore, one of the main issues is: What kind of modeling strategy should be used for describing periodontal diseases in predictive or etiologic models? Generally, logistic regression models are used. Such models assume that exposures and outcomes are measured in the same affected entity, and that exposures occur uniformly across the exposed entities. However, all exposures do not necessarily follow such a rule. For example, all individual level exposures may not necessarily be equally available across all the teeth, or on all sites in each tooth. For example, salivary exposures to

BOX 14.3 Analytical Issues in Periodontal Disease

Etiologic Model

Does Burst theory correlate well with historic measurement?

Does gradual incremental disease correlate well with historic measure?

Active vs. inactive sites: What are the differences?

Microbial measures?

Predictive models: Does etiology have a role?

Type of model issue?

"Burst" or not: Is this a Markov process?

Predictive Models

What do these predict: past disease/progression/probability of new disease?

Ascertainment and analysis of disease incidence or prevalence?

Progression of disease: Repeat measure?

Any between-site variation?

Does averaging score over several teeth diminish sensitivity?

Other Questions

Number of sites?

Clustering of teeth in mouth?

Episodic nature of disease accounted for?

Both these factors may induce complexities in repeat measure (key to ascertaining disease progression)

Multiple causation (necessary and sufficient cause models?)

Disease definition is usually impacted by selection of cut-point for dichotomizing disease status

What is the validity of cut-points from the viewpoint of the above issues?

Overall We Need to Consider:

Hierarchical/multilevel models?

Markov/Monte Carlo procedures?

Ascertainment and analysis of disease incidence?

Progression of disease: Repeat measure?

Any between-site variation?

Does averaging score over several teeth diminish sensitivity?

teeth in close proximity to salivary duct openings are different than those that are farther from the duct openings. Therefore, cleansing activity, calculus formation, or the salivary protective factor activity, will be differentially affected in different areas of the mouth. Buccal sides of the teeth may differ from lingual, and interproximal areas may differ from other areas with respect to access of cleaning mechanisms and other exposures in the mouth.

For periodontal disease-related exposures and disease process, four levels of exposure are recognized: ecological level, individual level, tooth level, and site level. Therefore, it is only appropriate to conduct hierarchical analysis to address exposure–outcome associations in a correct manner (Axtelius, SoÈderfeldt, & AttstroÈm, 1999). Appropriateness of multilevel analysis in periodontal disease modeling was emphasized by Axtelius et al. (1999) from their studies because the variance components at all the human three levels were significantly larger than zero and the inserted predictors showed 100% sensitivity relating to the subject-level variance. Failure to recognize hierarchical structure in data can also make most results misleading through violation of the assumption of independence of observations, which is basic for classical statistical techniques. Site-specific nature of data is suitably analyzed using multilevel modeling that can be applied to longitudinal analysis as well (Tu et al., 2004a, 2004b). Detailed discussion on multilevel modeling and its role in periodontal disease modeling is discussed in Chapter 8.

Periodontal Disease Surveillance

Although several studies have reported periodontal disease distribution estimates, disease definition used in most of them varied substantially, not permitting easy comparison. Historically, in the United States, national periodontal disease prevalence estimates are usually obtained from NHANES using the Periodontal Index. However, estimates for Hispanics are not available due to sampling issues related to this ethnic group. The Hispanic HANES (1982, 1984) survey provided periodontal disease estimates for Hispanics. The NHANES-III discarded use of the Periodontal Index and instead started to use recording of probing depths, loss of attachment, and presence of calculus and bleeding in partial mouth examinations. Such examinations are resource intensive and demand a large group of examiners whose standardization, calibration, and actual examination process required substantial resources. In 2005 with a funding reduction in place, the clinical periodontal assessment was eliminated in NHANES.

An alternative method of periodontal assessment was attempted by using self-reported measures for periodontal disease. In 2003, the Centers for Disease Control and Prevention (CDC), in conjunction with the American Academy of Periodontology, recommended use of self-reported measures that could be valid to predict the prevalence of periodontal disease

and develop and test questions to be used to survey the U.S. population (Eke, 2005). Self-report is an often-used, efficient, and accepted means of assessing several population characteristics, risk factors, and health-related factors and outcomes. It is expected that the "availability of valid self-reported measures of periodontal disease would facilitate epidemiologic studies on a much larger scale, allow for integration of new studies of periodontal disease within large ongoing studies, and facilitate lower-cost population" periodontal surveillance (Blicher, Joshipura, & Eke, 2005). However, self-reported measures for periodontal disease estimation has several challenges because the person must have visited a dentist and have been assessed for periodontal disease before he or she acquires the knowledge about his or her periodontal status. Therefore, estimation of periodontal status by self-report becomes a function of an annual dental visit. Many persons do not visit dentists periodically, or have never visited a dentist at all, or are otherwise unaware of their periodontal status because periodontal disease may be asymptomatic. Therefore self-report will underestimate periodontal disease.

The CDC Periodontal Disease Surveillance Project in 2003 consisted of a thorough review of the literature of previous studies that measured the validity of self-reported measures to predict the prevalence of periodontal disease (Eke & Genco, 2007). The work group focused on exploring the use of combined self-reported measures (e.g., gum disease, bone loss, history of treatment of gum disease, history of loose teeth, use of mouthwash or dental rinse, cleaning between teeth) and known risk factors (e.g., age, smoking, diabetes) to predict the prevalence of periodontitis among the population. A set of promising self-reported questions that showed evidence of being valid predictors of periodontal disease prevalence within populations were derived from a range of existing datasets analyzed by the panel experts. The validity of six periodontal screening questions were assessed as part of the Australian National Survey of Adult Oral Health, a population-based survey in Australia that has interview and clinical protocols similar to NHANES (Slade, 2007). This report concluded that the questions could be used in large population surveys, attaining useful levels of validity in predicting the prevalence of clinically evaluated periodontal disease. The National Center for Health Statistics is conducting the study using NHANES protocols for interviews and clinical examinations, including a full-mouth examination. If the pilot study proves to be successful, the valid questions will be incorporated into NHANES 2009–2010 (see Box 14.4).

Theoretical basis, item selection, statistical handling, and validity issues related to the self-report assessment tool for population-based surveillance of periodontal disease have been reported in detail (Dietrichet al., 2007; Dye & Thornton-Evans, 2007; Eke & Genco, 2007; Genco, Falkner, Grossi, Dunford, & Trevisan, 2007; Gilbert & Litaker, 2007; LaVange & Koch, 2007; Miller, Eke, & Schoua-Glusberg, 2007; Page & Eke, 2007; Slade, 2007; Taylor

BOX 14.4 Self-Reported Questions to Assess Periodontal Disease Under Trial by CDC

Q1. Do you think you have gum disease?
Q2. Overall, how would you rate the health of your teeth and gums?
Q3. Have you ever had treatment for gum disease, such as scaling and root planning, sometimes called "deep cleaning"?
Q4. Have you ever had teeth become loose on their own without injury?
Q5. Have you ever been told by a dental professional that you lost bone around your teeth?
Q6. During the past 3 months, have you noticed a tooth that doesn't look good?
Q7. Aside from brushing your teeth with a toothbrush, in the last 7 days, how many times did you use dental floss or any other device to clean between your teeth?
Q8. Aside from brushing your teeth with a toothbrush, in the last 7 days, how many times did you use mouthwash or other dental rinse product for treating dental disease or another dental problem?

The responses to these questions use a categorical item response (yes/no/don't know/refused/no comments) along with recording numerical response or one of the several choices in other questions (Adapted from Eke & Genco, 2007).

& Borgnakke, 2007; Tomar, 2007). In a systematic review of self-reported periodontal disease measures, Blicher, Joshipura, and Eke (2005) reported that results varied across populations and types of self-reported measures. They suggested that questions such as "Has any dentist/hygienist told you that you have deep pockets?" was a good self-report measure and had a sensitivity of 55%, a specificity of 90%, a positive predictive value of 77%, and a negative predictive value of 75% compared to clinical pocket depth. They further suggested that higher validity potentially could be obtained by the use of combinations of several self-reported questions and other predictors of periodontal disease.

15 Oral Cancer

In 2006, some 31,000 new oral cancer cases were diagnosed in the United States (Jemal, Siegel, Ward, Murray, Xu, et al., 2006). More than 90% of the oral cancers are oral squamous cell carcinoma. Oral cancer may present as an innocuous white/red patch, a small ulcerated lesion with or without pain, a large fungating ulcer that may include a super-added infection, and as indurated lesions of various sizes. The important epidemiological characteristics of oral cancer that stand out are: (1) oral cancer is mainly a disease of older people and is a sizeable public health problem, (2) there exists a significant racial/ethnic disparity in oral cancer disease burden and outcomes, (3) there occurs a significant delay between its clinical presentation and final diagnosis, and (4) the worldwide 5-year relative survival rate from oral cancer is generally less than 50%, although women tend to have a higher relative survival rate than men. (Arbes, Olshan, Caplan, Schoenbach, Slade et al., 1999; Lavelle & Scully, 2005; Mcleod, Saeed, & Ali, 2005; Mignogna, Fedele, Russo, Ruoppo, & Muzio, 2001; Robinson & Mickelson, 2006; Sciubba, 2001; Silverman, 1998).

Oral Cancer Scenario in the United States

Oral cancer rates have tended to be stable over long periods of times. The Surveillance, Epidemiology and End Results (SEER) program is a premier and authoritative source of information and statistics about cancer incidence and survival in the United States. SEER collects information on incidence, survival, and prevalence from specific geographic areas representing 26% of the U.S. population and compiles reports on all of these, plus cancer mortality, for the entire United States. Its coverage includes 23% of African Americans, 40% of Hispanics, 42% of American Indians and Alaska Natives, 53% of Asians, and 70% of Hawaiian/Pacific Islanders. At present 18

population-based cancer registries form the SEER program. The CDC's National Program for Cancer Registries (NPCR) collects data from cancer registries of all states (http://www.cdc.gov/cancer/npcr).

U.S. Incidence

The age-adjusted incidence rates of oral cancer have been fairly unchanging over the past half a century. From 2000 to 2003, the age-adjusted incidence rate was 10.5 per 100,000 men and women per year. The median age at diagnosis for cancer of the oral cavity and pharynx was 62 years of age. Approximately 0.5% were diagnosed under age 20, 2.4% between 20 and 34, 7.3% between 35 and 44, 20.5% between 45 and 54, 24.3% between 55 and 64, 22.2% between 65 and 74, 17.0% between 75 and 84, and 5.7% 85+ years of age. A similar picture emerges when using a different and large date range. For example, between 2001 and 2005, the median age at diagnosis for cancer of the oral cavity and pharynx was 62 years of age. Approximately 90% were diagnosed in persons above 45 years of age (Ries, Melbert, Krapcho, Stinchcomb, Howlader, Horner, et al., 2008). The age-adjusted incidence rate was 10.4 per 100,000 men and women per year. These rates are based on cases diagnosed in 2001–2005 from 17 SEER geographic areas (Ries et al., 2008).

U.S. Mortality

From 2001 to 2005, the median age at death for cancer of the oral cavity and pharynx was 68 years of age. The age-adjusted death rate was 2.6 per 100,000 men and women per year. These rates are based on the numbers of patients who died in 2001–2005 in the United States (Ries et al., 2008).

Among the U.S. states with SEER data available for 2005, Kentucky has the highest incidence rate for oral cavity and pharyngeal cancer, being substantially higher (11.8 per 100,000; 95% CI: 10.8, 12.8) compared to the national average (10.4; CI: 10.3, 10.5). In a more recent report, the overall incidence of oral cavity and pharyngeal cancer in Kentucky was 12.1/100,000 (Huang, Valentino, Wyatt, & Gal, 2008), which was significantly higher than the overall SEER rate of 11.3/100,000. This study reported that in Kentucky, the differences were more pronounced for males (19.2 vs 16.3/100,000). Most oral cancer cases (62.1%) had a documented smoking history (higher in advanced stage disease: 73%). Incidence rates were lower in Appalachian regions (11.4/100,000) compared to non-Appalachian regions (12.4 100,000). Another study in Kentucky demonstrated that treatment of oral cancer reflects the locally and regionally aggressive nature of these tumors. In advanced disease neither surgery nor radiation therapy can be used as a primary modality. Often, combined treatment is necessary. Chemotherapy has a less well-defined role, although may increase control rates in advanced tumors (Kenady, Arnold, Regine, & Mentzer, 2002). Kentucky consistently

has one of the highest rates of tobacco use in the United States. Thus, there seems a possible ecological basis for considering the high prevalence of tobacco use in the state as an important determinant for oral cancer in Kentucky.

Oral Cancer Survival

U.S. Survival

Oral cancer survival rates in the United States have remained stable at low levels for several decades. In the United States, each year some 31,000 new cases of oral cancer are diagnosed and more than 7500 people die of oral cancer. Overall, the age-adjusted incidence rate (2000–2003) was 10.5 per 100,000 and the age-adjusted mortality rate was 2.7 per 100,000 (Ries et al., 2008). The overall average 5-year survival rate is 58.8%. The 5-year survival rate has not changed substantially over the past 50 years (Ries et al., 2008; Silverman, 1998). In Europe, too, the 5-year survival rates have been stable for decades (Coleman, Gatta, Verdecchia, Estève, Sant et al., 2003).

Kentucky has a higher incidence rate for oral cavity and pharyngeal cancer than the national average and is a good test case for models that may be useful nationally. Currently there are no useful screening methods and most oral cancer cases present for treatment at late stages. Overall, people have also become better educated, insurance coverage has increased, and per capita as well as national median income has been increasing. Yet, the intransigence of oral cancer survival rates despite such advances is baffling. The relative stability of the oral cancer survival rate is not limited to just 5-year survival among those 50+ years at diagnosis (54% in 1976, 54.2% in 1980, 54.7% in 1990, 53.2% in 1993, 54.8% in 2004), but is also noted for any survival period. Relative survival rates (1-year survival to 21-year survival) between 1976 and 1996 are shown in Table 15.1. The Commonwealth of Kentucky, particularly the many rural counties of the state, continues to demonstrate metrics of poor general and oral health, especially oral cancers (Huang et al., 2008). These characteristics are also shared by poorer and more remote regions in the United States.

Oral cavity carcinoma is the sixth most common cancer and is often detected in later stages. Oral squamous cell carcinoma (SCC), the predominant type of cancer found in the oral cavity, is a disfiguring and deadly cancer. About 31,000 new SCC cases occur annually, causing about 7500 annual deaths in the United States. The complications of oral cancer and its therapy have major psychosocial and economic impact on patients with cancer, their families, and society. Although these complications are currently unavoidable, oral healthcare professionals are not aware enough to minimize their negative effects (Elting, Avritscher, Cooksley, Cardenas-Turanzas, Garden et al., 2008).

Global incident rates vary, but are high in almost all parts of the world. Selected rates for oral SCC in the United States are shown in Tables 15.1–15.3. However, the survival rates from oral cancer are low and have remained generally stable at these low levels for more than half a century, which is a matter of concern especially because there have been major advances in cancer diagnosis, management, prognosis, and supportive care during this time. Oral cancers presenting in earlier stages have better survival rates (McLeod, Saeed, & Ali, 2005; Mignogna et al., 2001; Ries et al., 2008; Robinson & Mickelson, 2006).

Table 15.2 summarizes the 5-year survival rate for oral cancer by different factors. The overall 5-year relative survival rate for 1996–2004 from 17 SEER geographic areas was 59.7%. Five-year relative survival rates by race and sex were 61.0% for white men, 62.9% for white women, 36.1% for black men, and 52.1% for black women. For tongue cancer, the overall 5-year relative survival rate between 1996–2004 from 17 SEER geographic areas was 57.7%. Five-year relative survival rates by race and sex were 59.8% for white men, 60.4% for white women, 33.3% for black men, and 39.4% for black women (SEER, 2008). The stage distribution based on historic stage shows that 33% of oral cavity and pharynx cancer cases are diagnosed while the cancer is still confined to the primary site (localized stage), 51% are diagnosed after the cancer has spread to regional lymph nodes or directly beyond the primary site, 10% are diagnosed after the cancer has already metastasized (distant stage), and for the remaining 5% the staging information is unknown. The corresponding 5-year relative survival rates were 82.2% for localized, 52.7% for regional, 28.4% for distant, and 47.7% for unstaged (SEER, 2008). This suggests that screening for oral cancer may have a major impact in improving survival from oral cancer. Apparently, a paradigm shift in the screening and diagnosis policies is needed to improve oral cancer survival, and a systems approach should be able to identify the leverage points to improve oral cancer survival rates which may be addressed by using a systems dynamics modeling approach.

Lifetime Risk

Based on rates from 2003 to 2005, it is estimated that 1.01% of men and women born today will be diagnosed with cancer of the oral cavity and pharynx at some time during their lifetime (see Table 15.3).

The 5-year survival rate for early-diagnosed oral cancer is 75% compared to 20% for late diagnosed oral cancer; the overall average rate being 52% (Ries et al., 2008). However, despite the vast information collected and scientific advancement, the survival rates have not changed substantially. Associated socioeconomic, legal, and political determinants may also be very important factors indicating system-level factors that may impact oral cancer incidence and survival. Data from a survey encompassing oral cancer-related

TABLE 15.1 Relative Oral Cancer Survival Rates (Survival Period: 1976–1996)

	'76	'77	'78	'79	'80	'81	'82	'83	'84	'85	'86	'87	'88	'89	'90	'91	'92	'93	'94	'95	'96
1-year	80.2	80.5	80.0	80.6	79.5	80.4	79.9	79.1	81.2	81.7	80.0	80.7	81.7	79.3	81.1	80.9	79.6	81.3	83.0	82.5	81.5
2-year	67.6	65.6	66.6	65.3	67.4	66.4	66.1	65.4	66.6	68.4	66.4	67.7	66.5	66.7	68.4	67.5	68.0	67.8	70.5	70.4	
3-year	61.4	58.0	60.8	59.1	60.7	59.5	59.0	60.3	59.8	61.1	60.6	61.2	58.7	59.7	61.1	60.3	60.4	61.3	64.7		
4-year	57.4	54.1	57.1	55.2	57.0	56.0	54.4	55.7	56.1	57.3	57.0	57.7	53.7	55.1	57.6	55.5	55.6	56.9			
5-year	54.0	51.5	54.3	52.5	54.2	53.3	51.0	52.3	53.0	54.2	53.9	54.3	50.5	51.6	54.7	52.2	53.2				
6-year	51.7	48.8	52.2	49.9	51.5	51.1	48.2	49.8	50.0	52.3	51.5	52.0	47.5	48.7	52.5	49.4					
7-year	49.5	47.4	49.2	47.7	49.2	48.5	45.9	48.2	47.7	49.7	49.4	49.6	44.5	45.9	50.1						
8-year	48.0	44.9	47.3	44.5	47.9	47.0	44.3	46.4	46.5	48.3	47.5	46.6	42.2	44.3							
9-year	46.6	42.7	45.6	42.6	45.8	45.5	42.6	44.6	44.4	46.2	45.3	44.9	40.4								
10-year	45.4	40.6	42.6	40.7	44.3	43.9	40.9	42.8	42.7	44.9	43.4	43.4									
11-year	43.6	39.3	40.2	39.2	43.1	42.6	39.7	40.9	41.2	43.8	42.1										
12-year	42.5	37.4	39.2	37.8	42.2	41.7	37.8	39.2	39.8	42.2											
13-year	41.5	35.5	38.8	36.7	39.9	39.8	36.9	37.6	38.2												
14-year	40.1	34.2	38.3	35.7	39.3	38.3	36.0	37.1													
15-year	38.3	33.5	36.2	34.5	38.2	35.9	34.9														
16-year	37.1	31.9	35.0	33.3	37.5	34.1															
17-year	36.5	31.2	34.5	32.7	36.5																
18-year	35.3	30.7	33.6	31.9																	
19-year	35.1	29.5	33.0																		
20-year	35.0	29.3																			
21-year	34.5																				

Source: Horner, M.J., Ries, L.A.G., Krapcho, M., Nexman, N., Aminou, R., Howlader, N., et al., (eds). (2009). *SEER cancer statistics review 1975–2006*. Bethesda, MD: National Cancer Institute. Retrieved August 12, 2009, from http://seer.cancer.gov/csr/1975_2006/

TABLE 15.2 Survival Rates per 100,000 Population of Oral SCC in the United States

Description		*Rate*
5-year Relative Survival		
Year at Diagnosis		
1974–1976		53.5
1983–1985		53.2
1989–1996		54.0
1989–1996	*All Stages*	54.0
1989–1996	*Localized*	81.3
1989–1996	*Regional*	43.5
1989–1996	*Distant*	21.4
1989–1996	*Unstaged*	36.4
5-year Relative Survival Rates 1989–1996		
Age at Diagnosis		
< 45 Years		59.9
45–54 Years		55.6
55–64 Years		52.8
65–74 Years		52.3
< 65 Years		55.4
≥ 65 Years		52.1
> 75 Years		51.6

Source: Horner, M.J., Ries, L.A.G., Krapcho, M., Nexman, N., Aminou, R., Howlader, N., et al., (eds). (2009). *SEER cancer statistics review 1975–2006*. Bethesda, MD: National Cancer Institute. Retrieved August 12, 2009, from http://seer.cancer.gov/csr/1975_2006/

questions about prevention, role of tobacco, lesion recognition, diagnostic techniques, and patient management showed a lack of knowledge and skills among graduating dental students (Burzynski, Rankin, Silverman, Scheetz, & Jones, 2002). It is likely that such lack of knowledge and confidence in making diagnoses may translate to a subsequent deficiency in incorporating optimal oral cancer prevention and control procedures in clinical dental practices. Therefore, it is conceivable that the interplay of factors at different levels of the healthcare delivery system, along with tumor biology, clinician knowledge, and environmental factors determine oral cancer incidence (and survival, by extension), demonstrating the need for determining leverage points in the system to increase oral cancer survival. This situation can be viewed as analogous to "policy resistance," which is defined as the tendency for interventions to be defeated by the system's response to the intervention itself (Sterman, 2006).

It is generally known that most cancers are painless growths, at least initially. The pain associated with a cancer occurs due either to its subsequent infection, its ulceration, and/or upon involving a nerve. These events occur after the cancer has existed and spread (at least locally) for some time.

TABLE 15.3 Percent Diagnosed with Oral Cancer in 10, 20, and 30 Years and in Remaining Lifetime, Given Cancer-Free at Current Age by Sex

Current Age (Years)	*Males*				*Females*			
	+ 10 Years	*+ 20 Years*	*+ 30 Years*	*Eventually*	*+ 10 Years*	*+ 20 Years*	*+ 30 Years*	*Eventually*
0	0.00	0.01	0.01	1.47	0.00	0.00	0.02	0.72
10	0.00	0.01	0.05	1.49	0.00	0.02	0.03	0.73
20	0.01	0.05	0.17	1.49	0.01	0.03	0.08	0.72
30	0.04	0.17	0.48	1.51	0.02	0.07	0.18	0.72
40	0.13	0.45	0.89	1.50	0.05	0.17	0.35	0.71
50	0.34	0.79	1.19	1.43	0.12	0.31	0.51	0.67
60	0.49	0.92	1.14	1.19	0.20	0.41	0.54	0.58
	Lifetime risk of being diagnosed = 1.47%				Lifetime risk of being diagnosed = 0.72%			
	Lifetime risk of dying = 0.41%				Lifetime risk of dying = 0.22%			

Source: Horner, M.J., Ries, L.A.G., Krapcho, M., Nexman, N., Aminou, R., Howlader, N., et al., (eds). (2009). *SEER cancer statistics review 1975–2006*. Bethesda, MD: National Cancer Institute. Retrieved August 12, 2009, from http://seer.cancer.gov/csr/1975_2006/

Whereas delay in diagnosing a cancer at an inaccessible part of the body can be attributed to its location, it is dismaying to realize that in the mouth, which is open to inspection at any moment, most cancers present with pain, implying that the cancer had been unnoticed for quite some time. Furthermore, 5-year survival for localized oral cancer is substantially higher than those that spread—rates being 81.3% for localized oral cancer, 51.7% for regional, 26.4% for distant, and 45.0% for unstaged oral cancer (Ries et al., 2008).

Oral cancer prognosis depends on its T, N, M stage: The 5-year survival rate of tongue SCC, whatever the T stage, is 73% in cases with negative nodes cases, 40% in patients with positive nodes without extracapsular spread, and 29% when nodes are metastatic with extracapsular spread. Nodal micrometastases are found in up to 50% of negative node tongue SCC patients operated on the neck (Calabrese, Bruschini, Ansarin, Giugliano, De Cicco et al., 2006). A recent systematic review suggested that the expression of intense psychosocial complaints, higher self-perceived physical ability, and self-reported high physical functioning have been shown to be significantly associated with increased survival in head and neck cancers. Uncertainty about the diagnosis and treatment, being single, poor cognitive function, baseline fatigue, and alcoholism were suggested to be prognostic indicators. Inadequacy of training in oral cancer prevention and screening as self-assessed by physicians, nurse practitioners, and dental health professionals have been noted as important factors for possible delays in diagnosis (Patton, Ashe, Elter, Southerland, & Strauss, 2006).

TABLE 15.4 Summary of Screening Tests for Oral Cancer

Screening Test	*Testing Principle*	*Test Drawbacks*
Routine clinical exam (the "oral cancer exam")	Visual exam and palpation	Poor visualization of high risk areas; high false negative rates.
Toludine blue vital staining	Binds to nuclei having high DNA/RNA content (e.g., malignant cells)	Also detects reactive inflammatory lesions; has high false positive rates.
Tissue autofluorescence	Upon exposure to lasers, autofluorescence differs between normal (green color) and cancer tissue (orange/red color)	This is a new technique under development. Early evidence suggests that autofluorescence may correlate with oral cancer progression. Loss of intensity may fail to distinguish between benign and premalignant lesions.
Oral brush biopsy	Scrapes all three layers of oral epithelial cells; stained and cytology studied for atypical changes	High false positive rate. Cannot distinguish from reactive lesions that may exhibit atypical cells. Not much evidence in support from field screening trials.
Chemiluminescence	Highlights cells with increased nuclear: cytoplasmic ratio imparting those with a whitish color	High false positives due to detection of reactive, inflammatory, and benign lesions.

Epidemiological parameters have not been established for any of the tests in screening trials. Adapted from D'Silva and Ward, 2007.

Screening for Oral Cancer: The Status of Evidence

Screening implies efforts to identify otherwise occult conditions (typically disease or disease precursors) in some appropriate segment of the general population (Katz, 2001). It is the effort to detect disease that is not readily apparent, or risk factors for disease in an at-risk segment of the population. Table 15.4 briefly discusses some screening tests for oral cancer highlighting those that may become important tests in the future. The U.S. Preventive Services Task Force (USPSTF) in 1996 stated that "available screening tests

for oral cancer are limited to the physical examination of the mouth, a test of undetermined sensitivity, specificity, and positive predictive value. Despite the strong association between stage at diagnosis and survival, there are few controlled data to determine whether routine screening in the primary care setting leads to earlier diagnosis or reduced mortality from oral cancer." The USPSTF concluded that there was insufficient evidence to recommend for or against routine screening for oral cancer, but noted that clinicians should remain vigilant for signs and symptoms of oral cancer and precancers in people who use tobacco or regularly use alcohol (USPSTF, 2004a). The Cochrane Collaboration review about oral cancer screening concluded similarly (Kujan, Glenny, Duxbury, Thakker, & Sloan, 2005). It suggested that given the significant morbidity and mortality associated with advanced oral cancer and its treatment, clinicians may wish to include careful examinations for oral cancer in asymptomatic persons at significantly increased risk for the disease. The USPSTF said that prevalence of oral cancer was lower in the United States than other countries. However, oral cancer incidence is high in several parts of the world. Overall, there is insufficient evidence to support or refute the use of visually based examination adjuncts. Given the lack of data on the effectiveness of adjunctive cancer detection techniques in general dental practice settings, clinicians must rely on a thorough oral mucosal examination supported by specialty referral and/or tissue biopsy for oral cancer diagnosis (Patton, Epstein, & Kerr, 2008).

It may be presumed that detecting oral cancer at an early stage would be the most effective means of improving survival and reducing morbidity from this disease. However, from a systematic literature review, Scott, Grunfeld, and McGurk (2006) found that most clinical/tumor factors, sociodemographic variables, and patient health-related behaviors were not related to the duration of patient delay. Patient delay is a problem in oral cancer and yet at present the reasons for such delays are poorly understood and under researched. A recent study suggested that primary care physicians are well suited to providing head and neck examinations, and to screening for the presence of suspicious oral lesions. Referral for biopsy might be indicated, depending on the experience of examining physicians (Epstein, Gorsky, Cabay, Day, & Gonsalves, 2008). However, physicians are not trained in oral etiopathology.

Screening for oral cancer may be able to reduce morbidity and mortality. Started in 2000, the Kerala Trial in India was conducted in a cluster-randomized, controlled setting with 59,894 subjects in the intervention group and 54,707 subjects in the control group. Subjects were 35 years or older. The intervention group received three rounds of screening (oral inspection by trained health workers) at 3-year intervals (Sankaranarayanan, Mathew, Jacob, Thomas, Somanathan et al., 2000). The first intervention demonstrated that 47 cancers (7 resultant deaths; case fatality rate: 14.9%) were diagnosed in the intervention group and 16 cancers (9 resultant deaths; case fatality

rate: 56.3%) were diagnosed in the control group. Although the case fatality rate in the screening intervention group was attractively low compared to the control group (small numbers, low frequency cells), the 2004 USPSTF suggested that the difference in the case fatality rate "between the two groups (14.9% and 56.3%, respectively) could potentially be attributed to lead-time bias" (USPSTF, 2004a). The USPSTF concluded that the evidence was insufficient to recommend for or against routinely screening adults for oral cancer (recommendation rating = I) (USPSTF, 2004b). Its recommendations and rating are noted in Box 15.1.

> The USPSTF found no new good quality evidence that screening for oral cancer leads to improved health outcomes for either high-risk adults (i.e., those over the age of 50 who use tobacco) or for average risk adults in the general population. It is unlikely that controlled trials of screening for oral cancer will ever be conducted in the general population because of the very low incidence of oral cancer in the United States. There is also no new evidence for the harms of screening. As a result, the USPSTF could not determine the balance between benefits and harms of screening for oral cancer.

Clinical Considerations

- Direct inspection and palpation of the oral cavity is the most commonly recommended method of screening for oral cancer, although there are little data on the sensitivity and specificity of this method. Screening techniques other than inspection and palpation are being evaluated but are still experimental.
- Tobacco use in all forms is the biggest risk factor for oral cancer. Alcohol abuse combined with tobacco use increases risk.
- Clinicians should be alert to the possibility of oral cancer when treating patients who use tobacco or alcohol.
- Patients should be encouraged to not use tobacco and to limit alcohol use in order to decrease their risk for oral cancer as well as heart disease, stroke, lung cancer, and cirrhosis. (USPSTF, 2004b)

Risk Factors of Oral Cancer

Several studies have assessed risk factors at different levels related to oral cancer survival, such as sociological, economic, patient behavior (smoking, alcohol consumption, diet, age); clinical presentation (size, site, super-infections, etc); biological (stage at presentation, comorbidity, tumor size and histological properties, tumor ploidy, biomarkers); coinfections, treatment modality; and dental/medical professional readiness (Bagan & Scully 2008; D'Silva & Ward, 2007). Established risk factors for oral cancer include tobacco, alcohol, and human papilloma virus (HPV) infection, whereas possible risk factors include certain mouthwashes and "hot mate" (an infusion

BOX 15.1 U.S. Preventive Services Task Force on Screening for Oral Cancer (2004): Recommendations and Ratings

The Task Force grades its recommendations according to one of five classifications (A, B, C, D, I) reflecting the strength of evidence and magnitude of net benefit (benefits minus harms):

- **A.** The USPSTF strongly recommends that clinicians provide [the service] to eligible patients. *The USPSTF found good evidence that [the service] improves important health outcomes and concludes that benefits substantially outweigh harms.*
- **B.** The USPSTF recommends that clinicians provide [the service] to eligible patients. *The USPSTF found at least fair evidence that [the service] improves important health outcomes and concludes that benefits outweigh harms.*
- **C.** The USPSTF makes no recommendation for or against routine provision of [the service]. *The USPSTF found at least fair evidence that [the service] can improve health outcomes but concludes that the balance of benefits and harms is too close to justify a general recommendation.*
- **D.** The USPSTF recommends against routinely providing [the service] to asymptomatic patients. *The USPSTF found at least fair evidence that [the service] is ineffective or that harms outweigh benefits.*
- **I.** The USPSTF concludes that the evidence is insufficient to recommend for or against routinely providing [the service]. *Evidence that [the service] is effective is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.*

drunk in South America) (International Agency for Research on Cancer [IARC], 1991). There is insufficient evidence at this time to designate Herpes Simplex Virus (HSV) infection and fat intake as risk factors for oral cancer. However, there are early studies suggesting that high butter intake, high saturated fat intake, high proportion of calories from cholesterol, and high fat intake may be associated with a greater risk of oral cancer (Franceschi, Favero, Conti, Talamini, Volpe et al., 1999a, 1999b). Some studies have suggested a role for Fluorides in oral cancer. This issue has been discussed in Chapter 17 along with other issues related to Fluorides in oral health.

Tobacco

Studies have reported that the odds ratio for heavy cigarette smoking among persons with oral cancer is more than four times compared to never-smokers (Silverman, 1998). Tobacco consumption occurs in many forms. Cigarette smoking, snuff, and chewable tobacco are commonly known forms. However, several other forms of tobacco consumption and uncommon smoking styles have been linked to oral cancer such as naswar (tobacco

+ slaked lime), nass (tobacco + slaked lime + ash), paan (betel nut + areca nut + slaked lime + tobacco), reverse smoking, black tobacco—air cured, and blond tobacco—flue cured (Chattopadhyay, 1989; Gupta, Murti, & Bhonsle, 1996; Merchant, Husain, Hosain, Fikree, Pitiphat et al., 2000; Sancho-Garnier & Theobald, 1993). Quitting cigarette smoking brings down the risk of oral cancer over a period of several years. It has been suggested that for those who quit smoking and remained smoke-free for 10 years, the risk of oral cancer was similar to that of nonsmokers (Blot, McLaughlin, Winn, Austin, Greenberg et al., 1988; Franceschi, Talamini, Barra, Barón, Negri et al., 1990).

Alcohol

Risk of oral cancer among alcohol drinkers is higher compared to non-drinkers (Blot et al., 1988; Franceschi et al., 1990). The odds of oral cancer for heavy alcohol consumption is almost nine times (OR = 8.8) compared to non-drinkers (Silverman, 1998). Franco, Kowalski, Oliveira, Curado, Pereira et al. (1989) estimated that drinking over 100 Kg of alcohol over a lifetime increased the risk of oral cancer threefold compared to non-drinkers, whereas consuming more than 400 Kg of alcohol over a lifetime increased the risk more than sevenfold. However, joint exposure to smoking and alcohol is explosive! For those who smoke heavily and also drink heavily the risk is 37.7, demonstrating *multiplicative effect measure modification on a multiplicative scale* for joint exposure to heavy smoking and alcohol consumption; thereby suggesting a serious synergistic effect between smoking and alcohol for oral cancer (Silverman, 1998).

> A recent matched case-control study of 375 participants reported a higher prevalence of current and former smokers among the cases (85.4%) than among the controls (69.9%). After adjustment for alcohol consumption, all measures of tobacco smoking, amount, duration, cessation, and type of tobacco were shown to be strongly associated with oral and oropharyngeal cancer. Additionally, measures of alcohol drinking status, duration, amount, and cessation were also associated with oral and oropharyngeal cancer development. The authors reported a significant supra-additive combined effect between smoking and alcohol consumption (never-smokers who never drank, OR = 1.0 [reference]; never-smokers who ever drank, OR = 1.7 (95% CI = 0.8–3.4); ever-smokers who never drank, OR = 1.6 (95% CI = 0.5–4.7); ever-smokers who ever drank, OR = 12.7 [95% CI = 5.5–29.1; p-value = 0.008]). (Castellsagué, Quintana, Martínez, Nieto, Sánchez et al., 2004, reviewed in Ragin, Modugno, & Gollin, 2007)

Human Papilloma Virus (HPV)

HPV has been found in cervical cancer, tonsillar cancer, and certain types of head and neck cancers (Oh, Kim, Woo, Kim, Jeong et al., 2004). The first

report suggesting the role of HPV in oral SCC assessed the presence of HPV antigens in 40 oral SCC using immunohistochemistry. Forty percent of these lesions were HPV positive with suggestive changes on light microscopy, of which half expressed HPV structural proteins (Syrjänen, Syrjänen, Lamberg, Pyrhönen, & Nuutinen, 1983). The same researchers later examined those biopsies to test for the presence of HPV DNA using in situ hybridization and PCR, and found that 12 of the 40 disclosed the presence of HPV 11, 16, or 18 DNA (Chang, Syrjänen, Nuutinen, Kärjä & Syrjänen, 1990). By now, at least 30 of these types have been detected in the oral cavity (Chang, Syrjänen, Kellokoski, & Syrjänen, 1991; Greenspan, D' Villiers, Greenspan, D'Souza, & Zur, 1988; Hagensee, Cameron, Leigh, & Clark, 2004; Oh et al., 2004). The role of HPV in oral cancer etiology has now been firmly established. A large multicenter international study involving 1670 cases and 1732 controls found that compared to those without oral cancers, exposure odds to HPV (type 16 E6 and E7 proteins and type 18) was about threefold in oral cancer cases (OR = 2.9, 95% CI: 1.7, 4.8) (Herrero, Castellsagué, Pawlita, Lissowska, Kee et al., 2003). The difficulty in providing true causal evidence of HPV's role in oral cancer lies in our lack of understanding of the significance of mechanisms by which HPV leads to oral carcinogenesis and in the limitations in the molecular analysis of HPV (Ha & Califano, 2004). HPV prevalence in cancers is technique sensitive. In research involving HPV, the need for standardized, consistent, and careful sample acquisition and processing methods must be emphasized.

> A review of the literature (Miller & White, 1996) revealed that HPV DNA was detected at a higher frequency by more sensitive assays, such as PCR (37.1%), than by moderately sensitive assays, such as Southern blot hybridization (27.2%), or by low-sensitivity assays, such as in situ hybridization or immunohistochemistry (25.2%) ($p < 0.005$). Although the wide variation in HPV prevalence may be narrowed with the increasing use of PCR as a more sensitive HPV detection method, variability in HPV prevalence still exists among these studies, and may be due to differences in the sensitivity of the various PCR primer sets used. The MY09/11 and GP5+/6+ primers are the most frequently used to amplify HPV DNA in cervical samples. A recent study compared the sensitivity of detecting HPV in oral vs. cervical samples with MY09/11 or GP5+/6+ primers (Remmerbach et al., 2004). The authors reported that, although both primer sets were in agreement when used in cervical DNA samples, the GP5+/6+ primers were more sensitive than MY09/11 for HPV detection in oral DNA samples. To reduce the variation in the literature of HPV DNA prevalence in the oral and oropharyngeal mucosa, one recommendation may be to design more sensitive PCR primers. There is increasing evidence that HPV infection may occur frequently in the normal oral mucosa (Kansky et al., 2003; Lambropoulos et al., 1997; Terai, Hashimoto, Yoda, & Sata, 1999), but this does not mean that the presence of the virus predicts progression to malignancy, since the majority of HPV infections may be transient rather than persistent. (Ragin et al., 2007)

Protective Factors

Limited epidemiological evidence suggests that a diet rich in fruits and vegetables may provide protection against oral cancers. Dark-yellow vegetables, citrus fruits, and carotene-rich foods such as carrots, pumpkins, and fresh tomatoes may be protective against oral cancers (Franco et al., 1989; IARC, 2003). Vitamins C and D, high beta carotene intake (working as an antioxidant), and vitamin E (antioxidant) may be preventive against oral cancers (IARC, 2003). Other micronutrients such as folate, vitamin A, and iron have been associated with prevention of oral cancer though additional epidemiological evidence is needed for definitive conclusions. Firm evidence is not yet available, but initial studies indicate that high fiber intake and olive oil consumption may have a role in oral cancer prevention (IARC, 2003).

Genetic and Molecular Epidemiology of Oral Cancer

Oral cancer, a multistep process, involves several molecular and genetic processes and factors. For example, tumor initiation through exposure to environmental exposures involves DNA repair, missense repair, DNA replication, and chromosomal segregation; tumor promotion through biochemical and genetic alterations involves oncogenes (P53, Rb, Ras, Rab, myc); and tumor progression to cancer involves a host of biochemical and physiologic processes leading to its growth progression, invasion, and metastasis (Schwartz, 2000). Several somatic mutations, such as those in p17, p16, p53, and mutations of epidermal growth factor-1 have been associated with oral and head and neck cancers (Lee, Soung, Kim, Nam, Park et al., 2005).

It is important to develop biomarkers for oral cancer occurrence, treatment, and progression. Biomarkers help in the evaluation of prevention or use of therapies and the detection of the earliest stages of oral mucosal malignant transformation. Biomarkers reveal the genetic and molecular changes related to early, intermediate, and late end points in the process of oral carcinogenesis and refine our ability to enhance the prognosis, diagnosis, and treatment of oral carcinomas (Daly, 1993; Greenwald, Kelloff, Boone, & McDonald, 1995; Page, 1994). Genetic and molecular biomarkers also determine the effectiveness and safety of chemopreventives.

Oral cancer exhibits field cancerization (the potential development of cancer at multiple sites) characterized by the expression of mutations in the exons of tumor suppressor genes, for example the p53 gene—a reduction in tumor suppressor activity by the gene and the development of mutations in p53 have been associated with smoking and an increased risk for oral carcinoma development (Brennan, Boyle, Koch, Goodman, Hruban et al., 1995).

> The mutational change noted in the bases of p53 in oral carcinoma is GC→TA, while the carcinogen benzo(a)pyrene, from tobacco smoke, also causes the identical GC→TA mutations of the bases in the p53 sequence

(Brennan et al., 1995). In contrast, combinations with alcohol and tobacco products produced seven different mutations of the bases for p53 in oral carcinomas. These included GC→CG, GC→TA, GC→AT, AT→TA, AT→GC, AT→*CG, and a frame shift (Brennan et al., 1995). Other endogenous mutations in the gene for p53 include "hot spots" of CpG sites which result from methylation and deamination of cytosine by enzymatic processes that differ from those that produce mutations of p53 bases (Brennan et al., 1995). In oral cancer patients who do not smoke or drink, a predominance of mutations at CpG sites has been observed for p53. (Brennan et al., 1995; Schwartz, 2000)

A CpG site is a region of DNA where C occurs next to G in a linear sequence of bases lengthwise.

Although inherited mutations are rare, inherited susceptibility to oral cancer may be mediated through common genetic alterations increasing cancer risk when present together. Genes coding for enzymes are involved in the metabolism of complex molecules present in tobacco, alcohols, and other risk factors. Penetrance of mutations in such genes may be high or low. Penetrance describes the proportion of individuals who carry a particular form or variant of a gene and also exhibit a phenotype. The Oral Cancer Gene Database at http://www.tumor-gene.org/Oral/oral.html, provides the latest information on the genes involved in oral cancer. Germline mutations in gatekeeper genes (i.e., promote cell removal and death in response to mutations to eliminate mutation heritability threat) and caretaker genes (i.e., tumor suppressor genes that maintain genome integrity by repairing damaged DNA) have been associated with oral cancer. The p53 gene normally acts as a gatekeeper by promoting cell death in response to mutations. Mutations on p53 in the germ line (e.g., in Li–Fraumeni syndrome—inherited p53 defects) substantially increases the risk of several cancers, including oral cancer among affected family members (Prime, Thakker, Pring, Guest, & Paterson, 2001). Similarly, xeroderma pigmentosum patients are at greater risk for early onset of oral cancer (presumably because of their extreme light sensitivity). Xeroderma pigmentosum is caused by defects in nucleotide excision repair genes. However, low penetrance genes contribute to cancer risk through gene–environment interactions. Many of the complex molecule metabolizing enzymes are present in oral tissues.

Glutathione S-Transferase Enzymes

Polymorphisms of genes associated with the metabolism of complex molecules have been associated with oral cancers. Prominent among these are Glutathione S-Transferase (GST) enzymes (GSTM-1, GSTT). The GST family of enzymes plays a significant role in metabolism of polycyclic aromatic hydrocarbons (PAH; e.g., benzo-a-pyrene). GSTM-1 enzyme also detoxifies ethanol, and therefore is an important enzyme for alcohol metabolism. The

null-GST genotype may be inefficient in alcohol and tobacco smoke product metabolism, thereby increasing the risk for oral cancer in such persons because of their inability to detoxify the carcinogens. A pooled analysis and meta-analysis of GST enzyme systems (Hashibe, Brennan, Strange, Bhisey, Cascorbi et al., 2003) found that oral cancer risks were elevated for GSTM-1 (OR = 1.45; CI:1.05–2.00), GSTT-1 (OR: 1.15; CI:0.82–1.63), GSTP-1 (OR: 1.52; CI:1.05–2.20), and all three GST genotypes (OR: 2.06; CI: 1.11, 3.81).

Cytochrome P450 and N-acetyl Transferase Enzymes

The Cytochrome P450 (CYP) enzymes are also important metabolizing enzymes; for example, CYP1A1 metabolizes PAHs, and CYP1E1 catalyzes oxidation of several compounds found in cigarette smoke (N-nitrosamines, benzene). CYP1A1 has been associated with oral cancers (OR: 1.48; CI: 0.77, 2.88) (Hashibe et al., 2003). The N-acetyl transferase (NAT2) metabolizes heterocyclic amines. NAT2 polymorphisms have not yet been associated with oral cancers, although very few studies have been conducted.

Alcohol Dehydrogenase

Alcohol dehydrogenase (ADH) enzyme oxidizes ethanol to acetaldehyde, which is a carcinogen. ADH1B and ADH1C geotyopes involving phase-1 alcohol metabolism enzymes produce acetaldehyde from ethanol. ADH1C*1 and ADH1B*2 alleles encode for a fast metabolism genotype that might lead to quicker and greater availability and accumulation of acetaldehyde. In a pooled analysis, no increased risk of oral cancer was found for ADH1C*1 genotype. However, the ADH1B*1 allele has been associated with esophageal cancer (Brennan, Lewis, Hashibe, Bell, Boffetta et al., 2004).

DNA Repair

DNA repair is a key defensive mechanism that protects the genome and rapidly addresses potential harmful changes at very early stages. Disruption of DNA repair may lead to cancer development. As mentioned earlier, tumor initiation through environmental exposures involves DNA repair and missense repair among other mechanisms and processes, leading to growth progression, invasion, and metastasis (Schwartz, 2000). Several case-control studies of relatively smaller sample sizes have examined the role of DNA repair in cancer. Berwick and Vineis (2000a) recently summarized results from human studies on DNA repair and cancer susceptibility markers, assessing technical validity, such as reproducibility, sample size, and selection of control subjects in these studies.

> Assays of DNA repair capacity used, to date, can be broadly grouped into five categories: (1) tests based on DNA damage induced with chemicals or physical agents, such as the mutagen sensitivity assay, the G(2)-radiation

> assay, induced micronuclei, and the Comet assay; (2) indirect tests of DNA repair, such as unscheduled DNA synthesis; (3) tests based on more direct measures of repair kinetics, such as the host cell reactivation assay; (4) measures of genetic variation associated with DNA repair; and (5) combinations of more than one category of assay. The use of such tests in human populations yielded positive and consistent associations between DNA repair capacity and cancer occurrence (with odds ratios in the range of 1. 4–75.3, with the majority of values between 2 and 10). However, the studies that we have reviewed have limitations, including small sample size, "convenience" controls, the use of cells different from the target organ, and the use of mutagens that do not occur in the natural environment. The evolving ability to study polymorphisms in DNA repair genes may contribute to new understandings about the mechanisms of DNA repair and the way in which DNA repair capacity affects the development of cancer. (Berwick & Vineis, 2000a)

The primary inquisitiveness about the biologic relevance of the assay systems used in assessing DNA repair is a developing research question. For example, lymphocyte-based testing has been used often, but "circulating lymphocytes are dormant cells and any functional tests on them require extensive in vitro manipulation" (Hemminki, Xu, & Le Curieux, 2000). These authors have pointed out that the validity of a DNA repair test can be examined if the test "measures removal of specific DNA damage in the target organ, when it has been demonstrated that DNA repair is the only means of damage removal (i.e., the adduct is chemically stable and no appreciable cell death and replication takes place)" (Hemminki et al., 2000). Other tests that measure the removal of specific UV radiation-induced DNA damage in human skin in situ have been developed (Bykov, Jansen, & Hemminki, 1998). However, a considerable number of studies consistently found an association between biologic tests and cancer at several sites, though their significance is still obscure (Berwick & Vineis, 2000b). It is further pointed out that the biologic relevance of DNA repair studies in lymphocytes "is critical for epidemiologic studies where associations are derived from numerous subjects to define small reproducible alterations that may be important" (Berwick & Vineis, 2000b).

DNA repair genes such as hOGG1 and XRCC1 have been linked with oro-pharyngeal and head and neck cancers (Cheng, Sturgis, Eicher, Spitz, & Wei, 2002; Goode, Ulrich, & Potter, 2002) as some studies have shown positive associations, that is, an important role for the XRCC1 399Gln polymorphism in p53 gene mutation in oral squampous cell carcinomas (Hsieh, Chien, Chen, Liao, Wang et al., 2003). However, others have demonstrated negative associations; for example, Olshan, Watson, Weissler, & Bell (2002) reported "a weak elevation in risk associated with the Arg194Trp polymorphism [odds ratio (OR) = 1.3; 95% confidence interval (CI) = 0.6–2.9] and a decreased risk for the Arg399Gln polymorphism (OR = 0.6; CI = 0.4–1.1). We

found a markedly decreased odds ratio for the Gln/Gln genotype among whites (OR = 0.1; CI = 0.04–0.6) and blacks (OR = 0.01; CI = 0.0004–0.3)." This study reported "a suggestion of an interaction between the Arg194Trp and Arg399Gln polymorphisms and tobacco use."

Despite several scientific advancements in understanding the mechanisms underlying oral cancers and the development of treatment alternatives, the survival rate of oral cancer remains low and largely unchanging over the past few decades. Epidemiological understanding to this effect is vital so that those at greater risk can be targeted for preventive regimens. Current understanding of oral cancer genetics and multidisciplinary intervention mechanisms exist, but the bottleneck in terms of information dissemination needs to be addressed, and public education awareness campaigns are strongly needed to reduce the burden of oral cancer by increasing early detection.

16 Other Oral Diseases and Conditions

Oral Potentially Malignant Disorders

Overall, precancer is a disease condition that has a greater probability of turning into cancer in the future although the magnitude of the probability and time have not been established. In 1978, WHO distinguished between precancerous lesions and precancerous conditions (World Health Organization, 1973). WHO issued the following distinctive definitions: *precancerous lesion*: "morphologically altered tissue in which oral cancer is more likely to occur than in its apparent normal counterpart"; and *precancerous condition*: "generalized state associated with a significantly increased risk of cancer." Accordingly lesions such as leukoplakia and erythroplakia would be categorized as precancerous lesions, whereas conditions such as lichen planus and oral submucous fibrosis would be categorized as precancerous conditions. A workshop comprised of specialists in the fields of epidemiology, oral medicine, pathology, and molecular biology with a special interest in oral cancer and precancer, organized by the WHO Collaborating Center for Oral Cancer and Precancer in the United Kingdom, met in 2005 to discuss current concepts, terminology, classifications, natural history, pathology, and molecular markers for oral precancer (Warnakulasuriya, Johnson, & van der Waal, 2007). This workshop decided to refer all precancerous conditions and lesions together as "potentially malignant disorders." Important oral potentially malignant disorders include leukoplakia, erythroplakia, palatal lesions in reverse smokers (stomatitis nicotina palatii), oral submucous fibrosis, actinic keratosis, lichen planus, discoid lupus erythematosus, and some rare hereditary disorders such as dyskeratosis congenita and epidermolysis bullosa. Box 16.1 outlines the important features of some of these disorders.

BOX 16.1 Characteristic Features of Some Important Potentially Malignant Disorders

Leukoplakia

Definition: Leukoplakia refers to white plaques of questionable risk for cancer having excluded other known diseases or disorders that carry no increased risk for cancer.

Clinical types of leukoplakia: Homogenous and non-homogeneous leukoplakia. Non-homogeneous may be *Speckled leukoplakia* (predominantly white, but having mixed red and white patches); *Nodular leukoplakia* (red or white rounded polypoid growths); and *Verrucus leukoplakia* (white lesions with corrugated or wrinkled-appearing surfaces).

Differential diagnosis: White sponge nevus, frictional keratosis, morsicatio buccarum, chemical injury, acute pseudomembranous candidosis, leukoedema, lichen lanus (plaque type), lichenoid reaction, discoid lupus erythematosus, skin graft, hairy leukoplkaiia, leukokeratosus nicotina palatii.

Erythroplakia

Definition: A fiery red patch that cannot be characterized clinically or pathologically as any other definable disease.

Clinical types: Smooth red patch that may have a granular surface. Erythroleukoplakia may have a mixed red and white patches.

Differential diagnosis: Desquamative gingivitis, erythematous lichen planus, discoid lupus erythematosus, pemphigoids, hypersensitvity reactions, Reiter's disease, erythematous candidiasis, histoplasmosis, haemangioma, Kaposi's sarcoma.

Molecular epidemiology: Studies have assessed DNA content, role of P53 mutation, and role of HPV. No strong evidence of such associations exist at this stage.

Oral Submucous Fibrosis

Definition: Chronic disorder characterized by fibrosis of the oral mucosa. May also involve pharynx and esophagus.

Clinical types: Intact mucosa with fibrotic bands; desquamated/erotic/ulcerated mucosa.

Differential diagnosis: Scleroderma; white and red lesions of the oral cavity

Risk factors: Mostly associated with betel quid chewing and smokeless tobacco use.

(Continues)

BOX 16.1 Characteristic Features of Some Important Potentially Malignant Disorders *(Continued)*

Lichen Planus

Definition: Chronic disorder characterized by mucosal inflammation and immunologic reactions.

Clinical types: Reticular, erosive/ulcerative, plaque type, atrophic, bullous.

Differential diagnosis: White and red lesions of the oral cavity. Lichenoid reactions caused by drugs and other allergens.

Causal/risk factors include: High-stress lifestyle, autoimmunity.

Adapted from Warnakulasuriya, Johnson, and van der Waal, 2007.

Leukoplakia

The definition of leukoplakia continues to be a source of worry. Although the clinical entity "leukoplakia" carries a greater risk compared to normal oral epithelium for cancer, its diagnosis is arrived at by elimination rather than by definitive diagnositc criteria. The previously mentioned workshop recognized that dividing leukoplakia into homogeneous and nonhomogeneous types was not enough to adequately identify the cancer risk of all lesions falling under the definition of leukoplakia. They pointed out that lesions that have mixed white and red patches should be designated as "erythroleukoplakia" because such lesions carry a greater risk for cancer compared to white-alone lesions or normal epithelium.

> A provisional diagnosis of leukoplakia is made when a predominantly white lesion at clinical examination cannot be clearly diagnosed as any other disease or disorder of the oral mucosa. A biopsy is mandatory. A definitive diagnosis is made when any etiological cause other than tobacco, areca nut use has been excluded and histopathology has not confirmed any other specific disorder. (Warnakulasurya et al., 2007)

The problem with this approach is that it almost requires leukoplakia not to have a history of tobacco or other deleterious exposures creating a bias against such associations.

Several follow-up studies have estimated that between <1% and 18% of oral premalignant lesions transform into oral cancer (reviewed by Reibel, 2003). Although epithelial dysplasia is an important feature to assess the potential of a lesion to turn into cancer, not all dysplasia turn into cancer, and not all cancers exhibit prior dysplastic features. Diagnosis of dysplasia is still subjective; there exists a need to improve histological assessment for epithelial dysplasia or develop other methods.

Making a diagnostic call of dysplastic epithelium is subjective and ill-defined. It also leads to disagreement between oral pathologists, especially when distinguishing between mild and moderate varieties (which may result in very different outcomes). A WHO-organized expert group working together as the Collaborating Center for Oral Cancer and Precancer in the United Kingdom met recently to discuss the problems related to making a diagnostic call of epithelial dysplasia. They suggested that classification of dysplasia (currently variably classified as no dysplasia/questionable/mild/moderate/severe dysplasia) into two categories (no dysplasia/questionable/mild dysplasia—all with low risk, and moderate/severe dysplasia implying high risk) would be better because of their view that reducing the number of choices to two may increase the likelihood of agreement between pathologists (Warnakulasuriya, Reibel, Bouquot, & Dabelsteen, 2008). One method that may be applicable (especially in resource-poor settings) to histological sections is the characterization of silver-stainable nucleolar organizer regions (AgNORs) in the dysplastic epithelium. Assessing the potential usefulness of AgNORs, one study suggested that the method could be useful in distinguishing dysplastic lesions that may become cancerous (Chattopadhyay & Ray, 2008). Using cut-points of mean AgNOR counts, the area under the Receiver Operating Characteristics (ROC) curve (AUC) was 74%. Using resampling methods, cut-off points for mean AgNOR counts per cell were suggested as thresholds for dysplastic lesions with greater potential for turning to cancer (nonparametric method mean AgNOR cut-point = 2.42 (CI: 2.43; 2.82) AgNORs/nucleus, parametric method suggested cut-point = 2.57 (CI: 2.31; 2.66) AgNORs/nucleus).

Mithani, Mydlarz, Grumbine, Smith, and Califano (2007) have reviewed the genetic epidemiology studies of leukoplakia that have been conducted, although these involved small samples. However, a pool of similar results is beginning to be accumulated. For example, loss of heterozygosity (LOH) of the chromosome arms 3p and 9p are associated with increased malignant potential in oral leukoplakia, which imparts a 3.8-fold risk for malignant transformation of leukoplakia; and if additional LOH appears at the 4q, 8p, 11q, 13q, and 17p loci, the risk increases to 33-fold. Prevalence of microsatellite instability (MSI) is higher is leukoplakia compared to normal tissue. MSI increases have been associated with increasing histological grades of dysplasia that have greater risk of cancer. Furthermore, telomerase activity has also been demonstrated to increase in leukoplakia. Other changes noted in leukoplakia (compared to normal tissue) include p53 protein expression and increased mitochondrial copy number (reviewed in Mithani et al., 2007).

Molecular studies assessing DNA content (aneuploidy), p53 mutation, LOH, and detection of cell surface carbohydrates and keratins have suggested that markers of epithelial differentiation and genomic markers could potentially be good candidates to assess improving the prognostic evaluation of potentially malignant disorders. However, "as yet, one or a panel of

molecular markers has not been determined that allows for a prognostic prediction of oral pre-cancer which is any more reliable than dysplasia recording. However, these new markers could be considered complementary to conventional prognostic evaluation" (Reibel, 2003).

Another aspect related to early identification and diagnosis of potentially malignant disorders is to respond on a self-report based on self-examination. For the success of such strategies, self-awareness of potentially malignant disorders among the public is extremely important. Shugars, Adesanya, Diehl, Redman, Malley et al. (2007) have assessed data obtained by questionnaire and clinical examination of U.S. veterans at six U.S. Veterans Affairs Medical Centers with clinical outcomes including homogeneous and nonhomogeneous leukoplakia, smokeless tobacco lesion (STL), papilloma, lichen planus, and erythroplakia. Whereas some 40% veterans were unaware of their lesions, their awareness varied with lesion diagnosis. Awareness was more likely with STL and less likely with homogeneous leukoplakia. Furthermore, awareness was predicted by the presence of a lesion on easily visible mucosa (adjusted OR = 11.2) and a history of mouth sores (OR = 11.2) (Shugars et al., 2007). Perhaps, educating the public about developing a habit for self-examination of the mouth every morning and night during daily brushing activity might be useful in catching potentially malignant disorders early in their clinical phase.

HIV-Associated Oral Diseases

Human immunodeficiency virus (HIV)-associated oral disease includes all oral mucosal lesions associated with HIV/AIDS, such as oral candidiasis (OC), oral hairy leukoplakia (OHL), Kaposi's sarcoma, herpes simplex virus infection, apthous ulcer, and necrotizing ulcerative stomatitis; OC and OHL being the two primary diseases. The prevalence of all HIV-associated oral diseases has changed globally following highly active antiretroviral therapy (HAART). At present there are no published studies evaluating risk factors or indicators of post-HAART *incident* HIV-associated oral diseases as a group or for OC/OHL separately. There have been suggestions of some potential risk factors or indicators of HIV-associated oral diseases, though clear independent risk factors or indicators have not been established (post-HAART). Other studies have reported on HIV-associated periodontitis, whereas some have suggested that HIV-associated oral diseases serve as markers for immunosuppression. One study reported incident Kaposi's sarcoma and another reported survival time for OC and OHL after seroconversion, but neither evaluated risk factors or indicators. Some other studies have associated OC/OHL with progression to AIDS.

Oral opportunistic infections are one of the earliest clinical manifestations of HIV infection and may also suggest HIV disease progression. Current

estimates suggest that between 20% and 50% of HIV-infected patients develop lesions in the oral cavity. WHO estimates suggest that some 40–50% of people who are HIV-positive have oral fungal, bacterial, or viral infections, which often occur early in the course of HIV infection. The commonest HIV-associated oral lesions, OHL and OC, are clinical markers of symptomatic HIV infection. Lesions such as oral Kaposi's sarcoma, cytomegalovirus ulcers, and herpes simplex virus ulcers persisting beyond 1 month in duration have also been categorized as AIDS-defining oral conditions, though they are rare. HIV-infected individuals with OC and/or OHL are 1.84 times more likely to have a viral load of 20,000 copies/ml or more (95% CI: 1.02, 3.35) (Patton, McKaig, Eron, Lawrence & Strauss, 1999).

OC is a fungal disease caused mainly by *Candida albicans* and is one of the most common oral manifestations of HIV-positive individuals. One estimate suggested that more than 90% of patients with AIDS will develop OC at some point in the course of the disease. OC suggests immunological decline and may be an initial sign of HIV infection or progression to AIDS. The most common *Candida* species in HIV-infected individuals is *C. albicans*, being found in 63–93% of cases. Other species found are *C. glabreta* (14–21%), *C. krusei* (4–10%), and *C. tropicalis* (2–7%) (Hunter, Gibson, Lockhart, Pithie & Bagg, 1998; Tumbarello, Tacconelli, Caldarola, Morace, Cauda & Ortona, 1997). *C. non-albicans* spp increase in prevalence with immune decline and previous antifungal drug exposure. OC responds to antifungal therapy, but eradication is rarely achieved unless the underlying immunocompromised state is resolved. OC has been found to be associated with low CD4+ T-lymphocyte count (CD4 cell count) ($< = 200/mm^3$).

OHL is an opportunistic viral lesion caused by Epstein–Barr virus (EBV) and is frequently detected among HIV-infected individuals. It is a benign hyperplastic lesion presenting mostly along the lateral borders of the tongue, though it has also been reported on the ventral and dorsal tongue surfaces, and is frequently bilateral in presentation. Rare instances of OHL occurring on palate, buccal mucosa, floor of the mouth and oropharynx have been reported. Hospital-based studies have suggested that clinical presence of OHL may have 100% positive predictive value for HIV-infection/AIDS. OHL has been associated with more rapid progression to AIDS among HIV-infected individuals, and with HIV-viral loads (20,000 copies/ml or more), independent of CD4 cell count. OHL has been found to be associated with low CD4 counts ($< 200/mm^3$), and absence of anti-p24 antibodies in serum and saliva (Katz, Greenspan, Westenhouse, Hessol, Buchbinder et al., 1992; Kolokotronis et al., 1994; Triantos, Porter, Scully, & Teo, 1997).

Associations of HIV-associated oral diseases with known risk factors such as immune-compromised state, smoking, antiretroviral drugs, recreational drug use, and so on have been reported in several studies (reviewed by Patton & Van der Horst, 1997). For example, whereas prevalent OC has

been associated with low CD4+ cell count [<200 cells/μL, adj. OR: 12.7 (CI: 4.9, 32.9)], antiretroviral combination therapy [OR: 0.6 (CI: 0.3, 0.9)], and current smoking [OR: 2.5 (CI: 1.3, 4.8)]; prevalent OHL has been associated with low CD4+ cell count [<200 cells/μL, OR: 7.2 (CI: 2.7, 18.9)], antifungal medication use [OR: 1.8 (CI: 1.1, 2.9)], current recreational drug use [OR: 2.5 (CI: 1.3, 4.9)], and male gender [OR: 2.5 (CI: 1.3, 4.8)] (Chattopadhyay et al., 2005a).

Incidence of oral diseases in HIV-positive population has been reported by two studies (Chattopadhyay et al., 2005b; Greenspan, Gange, Phelan, Navazesh, Alves et al., 2004), of which one was among women only (Greenspan et al., 2004). Incidence rate (per 1000 person–months) was 9.3 for OC, 6.8 for OHL, and 13.5 for HIV-OD. Incidence of OC was associated with low CD4 count [adjusted incidence rate ratio—IRR: 3.0 (CI: 1.7, 5.1)], smoking [IRR: 1.9 (CI: 1.0, 3.8)], and combination antiretroviral therapy [(IRR: 0.3 (CI: 0.1, 0.8)]. Incidence of OHL was associated with antifungal therapy [IRR: 2.8 (CI: 1.5, 5.4)], women [(IRR: 0.5 (CI: 0.2, 1.1)], and low CD4 count that was conditional upon smoking status. For example, compared to those with high CD4 count who were not current smokers, those with low CD4 count who were current smokers were 6.5 times as likely to develop OHL, those with low CD4 count who were not current smokers were 4.9 times as likely to develop OHL, and those with higher CD4 count who were current smokers were 1.3 times as likely to develop OHL (Chattopadhyay et al., 2005b).

Risk estimates for joint occurrence of OC and OHL (OC–OHL) have been examined in one study (Chattopadhyay & Patton, 2007). This cross-sectional study reported a final parsimonious proportional odds model in which occurrence of OC–OHL was independently associated with CD4+ counts <200 cells/μL [OR: 13.4 (CI: 6.6, 27.2)] and CD4+ counts 200–499 cells/μL [OR: 3.9 (CI: 1.9, 8.1)], current smokers [OR: 2.3 (CI: 1.4, 3.8)], and Whites [OR: 1.7 (CI: 1.1, 2.5)]. Combination antiretroviral therapy was protective [OR = 0.5 (CI: 0.3, 0.9)]. However, a model containing all examined variables (fully adjusted model) with estimates is shown in Table 16.1. Examining the risk of jointly occurring oral diseases in HIV-infected persons helps our understanding of comorbidity and their interactions with each other and with the oral defense system. Under this concept, the risk of joint occurrence of disease may indicate a more compromised oral defense capacity, and that persons with OC–OHL are in some ways different from those who get only OC or only OHL. Therefore, for such a concept to hold meaning, the association measure (OR) for OC–OHL must be greater than that for OC or OHL singly. This was indeed demonstrated in the fourfold greater odds of OC–OHL over OC only or OHL only (Chattopadhyay & Patton, 2007), an observation that supports the concept of disease interactions and a potentially limited capacity to mount a successful defense against oral microbial infections or, organism–organism coaction may occur.

TABLE 16.1 Proportional Odds Model for Ordinal HIV-Associated Oral Diseases

	Full Model[a]	
Statistical Test	*Statistic (df)*	*p-value*
Likelihood ratio Chi-square	123.31 (15)	<.0001
Score test (proportional odds assumption)	14.9884 (15)	0.4522
Parameter	**Coefficient (SE)**	**p-value**
Intercept 2 (both OC and OHL)	–5.378 (0.55)	<.0001
Intercept 1 (either OC or OHL)	–2.891 (0.52)	<.0001
Variable	**OR (CI)**	**p-value**
CD4+: <00 cells/ μL vs. 500 + cells/μL	11.7 (5.6, 24.5)	<.0001
CD4+: 200–500 cells/ μL vs. 500 + cells/μL	3.4 (1.6, 7.2)	0.0011
Race[a]: Whites vs Blacks	1.7 (1.1, 2.5)	0.0195
Anti-HIV medication: monotherapy vs. none	0.7 (0.4, 1.1)	0.1141
Anti-HIV medication: combination vs. none	0.6 (0.4, 0.9)	0.0263
Smoking: former vs never smoked	1.1 (0.6, 2.1)	0.7788
Smoking: current vs never smoked	2.0 (1.2, 3.5)	0.0087

[a]Adjusted for age, antifungal drugs, sexual orientation, and recreational drug use.

No HIV-associated oral diseases: 32% Whites; 68% Blacks; Either OC or OHL singly: 39.5% Whites; 60.5% Blacks; Both OC and OHL: 45% Whites; 55% Blacks.

OR reported for: jointly occurring OC and OHL; reference = either OC or OHL (intercept 2), or no OC, no OHL (intercept 1). OC: Oral candidiasis; OHL: Oral Hairy Leukoplakia.

Adapted from Chattopadhyay and Patton, 2007.

Recurrent Aphthous Stomatitis

Recurrent aphthous ulcers (RAU) in the oral cavity are painful ulcers causing substantial morbidity in the United States and elsewhere in the world. From convenience sample studies, prevalence of RAU has been reported to vary between 1 to 66% among adults and 1 to 40% among children. Literature on RAU is published frequently, but a high proportion of those are clinical reviews. Though several studies have investigated the role of various factors in RAU etiology; nevertheless, the epidemiology and etiology of RAU remains unclear. Independent risk factors of RAU have not been clearly established in population-based studies, and most evidence comes from convenience samples and clinic-based studies. Box 16.2 enumerates the commonly discussed risk factors of RAU.

Only three population-based studies have been conducted, two of which are representative of the U.S. national adult population and describe

BOX 16.2 Purported Risk Factors of Recurrent Aphthous Ulcer (RAU)

- Genetic factors
- Negative association with smoking
- Oral contraceptives
- Cyclical association with menstrual cycle stage
- Positive association with T-cell-mediated immune responses
- TNF-α
- Interleukin
- Keratinocyte maturity
- Heat shock proteins
- Hematinic deficiencies such as vitamins B1, B2, and B6, B7 folate deficiency, 21 zinc deficiency
- Defective mucosal epithelial turnover
- Microbes: *Oral streptococci*, *M. tuberculosis*, H. pylori, Herpes viruses, *Varicella zoster* virus, Cytomegalo virus.

the epidemiology of RAU in adults (Axéll & Henricsson, 1985; Chattopadhyay & Chatterjee, 2007; Rivera-Hidalgo, Shulman & Beach, 2004). The reported prevalence of RAU varies widely, most likely due to the differences between samples used for most of the studies that were hospital/clinic-based convenience samples, where more patients with RAU are likely to be found, thereby overestimating the prevalence. Because the NHANES-III data is analyzed with appropriate nesting and weighting statements to adjust the variance for the complex sampling design of the survey, studies using NHANES with proper analytical design provide representative U.S. national estimates. Therefore, there appear to be some 3 million persons of all ages in the United States with RAU, of which about 2 million are ages 17 years or more. "Overall, for all Americans regardless of age, this study found RAU prevalence to be 1030 per 100,000 people (95% CI: 830/100,000; 1220/100,000). The prevalence of RAU among children was 1500 per 100,000 (95% CI: 1090/100,000; 1910/100,000) and was greater than that among adults at 850 per 100,000 (95% CI: 630/ 100,000; 1070/ 100,000)" (Chattopadhyay & Chatterjee, 2007). This study reported the lower vestibule as the commonest site for RAU. Reported risk factors for RAU from multivariable logistic regression model-based analyses included those in the age group 17–29 years (OR: 2.7, CI: 1.4, 5.5), men (OR: 1.7, CI: 0.9, 2.8), those with low serum insulin level (OR: 2.0, CI: 0.9, 4.4), and nonsmokers [compared to those who smoked more than 10 cigarettes per day, RAU risk of never-smokers was greater (OR: 9.2, CI: 2.8, 30.1)] (Chattopadhyay & Chatterjee, 2007).

RAU prevalence rates reported by two studies from the same NHANES data differed slightly (Chattopadhyay & Chatterjee, 2007; Rivera-Hidalgo et al., 2004). This difference was explained because one study excluded all persons belonging to "other" race categories from prevalence estimation (8% of the weighted sample size). Further, the prevalence of RAU was highest in the "other" race group, a fact that could have led to underestimation of the true prevalence. Another source of the slight difference between the studies was the inclusion of recurrent herpes labialis in the RAU-diagnosed group by one study. Herpes labialis, however, is a different disease entity and may confound association of RAU with risk factors.

Endodontic Outcomes

Good endodontic outcomes are important for tooth longevity, and in advancing societies, are being increasingly studied as retention and good maintenance of teeth are becoming hallmarks of rising living standards across the world. Longer retention of teeth with full coverage crowns is becoming common. However, the fate of such therapy is not well documented. Kirakozova and Caplan (2006) examined the predictors of subsequent root canal therapy (RCT) in teeth receiving full coverage restorations using a case-control design with 137 subjects, defining cases as those persons whose crowned teeth received RCT before a predefined cut-off date and controls as those persons whose crowned teeth did not receive RCT on or prior to that date. From the multivariable logistic regression models, they suggested that younger age and greater extent of coronal and root destruction were important predictors of who received RCT subsequent to full coverage crowns (Kirakozova & Caplan, 2006).

Maintaining high clinical management standards of care in private practice is a key ingredient of the prevailing standard of health care in a country. Cost-effectiveness of therapeutic interventions is an important consideration for their being recommended and eventually adopted regularly, especially if the treatment involved is complex and expensive. For example, in a study spanning across treatment types, teeth with crowns were found to have higher effectiveness values (although cost was substantially higher) compared to teeth restored with large amalgams. The cost of an additional year free of catastrophic treatment for crowns was $1088.41 at 5 years and $500.10 at 10 years. Cost-effectiveness ratios favored women and the teeth in the maxillary arch. Perhaps the higher incremental cost-effectiveness ratio for crowns should be considered when making treatment decisions between large amalgam and crown restorations (Kolker, Damiano, Flach, Bentler, Armstrong et al., 2006).

Treatment outcomes also need to understand treatment failures. For example, a large study examining root canal failure found that root canal failure was predicted by teeth with fewer proximal contacts at access (OR: 2.7;

95% Cl: 1.4, 5.1), older age (OR: 1.4; CI: 1.1, 1.9 per 10-years of age increase), facial injury (OR: 3.6; CI: 1.2, 10.5), and more plaque (OR: 1.7; CI: 1.0, 2.6) (Caplan & Weintraub, 1997). Noncompliance with RCT is also an important predictor of overall treatment success. Because patients with greater evidence of past and current oral disease are less likely to have completed RCT, they may require additional counseling about the importance of carrying through with prescribed treatment

An important determinant of RCT outcome is the dental professional who performs the treatment. General dental practitioners with more than 10 years in clinical practice were more likely to recommend (1) referring difficult cases rather than performing endodontic therapy themselves, and (2) extracting perforated or root-fractured teeth prior to obturation rather than continuing treatment (Caplan, Reams, & Weintraub, 1999). However, newer treatment methods and associated technology may play an important role in an experienced practitioner's hand. A recent study assessed the clinical outcomes of RCT in private practice using standardized and nonstandardized protocols. In this study, both protocols were successful in obtaining the desired clinical outcomes (Conner, Caplan, Teixeira, & Trope, 2007). In several situations, however, nonstandardized protocols may not work well. Proper education interventions are required for successful adoption of treatment protocols, which is the key to its final success. Whereas short-term adoption of technology may occur quickly, long-term adoption of technology requires sustained effort. For example, a study assessing the adoption of nickel–titanium rotary instrumentation (NTRI) among general dental practitioners with a short-term as well as a long-term perspective found that lectures in combination with hands-on training resulted in a better short-term acceptance rate compared to lectures alone (rising from 4 to 73%). The long-term adoption rate for acceptance of the new technology required its having a relative advantage over existing technology. "Common reasons for dentists not to adopt NTRI were that they could not get started or that they found no advantage over the old technology" (Reit, Bergenholtz, Caplan, & Molander, 2007).

Although pulpotomy in primary dentition has long been carried out, adult pulpotomy in permanent dentition has been generally considered controversial and suitable only in very young teeth with open apices. This viewpoint has been challenged over the years (Caliskan, 1993; Cvek, Mejare, & Andreasen, 2004). Furthermore, a recent study reported by Menezes, Bramante, Letra, Carvalho, and Garcia (2004) in mongrel dogs suggested that different biocompatible materials used in pulpotomies all resulted in successful pulpal healing from pulpotomy wounds. Apparently, a biological basis for performing pulpotomies in adult permanent teeth exists. Individual studies have reported more than 5 years of symptom-free existence of permanent teeth treated with pulpotomy in adults who were in the early middle ages of their lives (Chattopadhyay & Ray, 2007). In resource-poor

settings, pulpotomy in permanent teeth has slowly gained ground and is becoming an acceptable form of endodontic treatment. Some studies have shown that properly carried out, pulpotomy in adults can be successful, and therefore can be considered as an alternative treatment to the extremes of extraction or RCT (Caliskan, 1993; Cvek et al., 2004; McDougal, Delano, Caplan, Sigurdsson, & Trope, 2004; Santini, 1986). The evidence base for performing pulpotomy in adult permanent teeth is developing and would benefit from larger epidemiological studies assessing the risk factors adequately.

Oral Health and Systemic Health

Oral diseases share common risk factors with the four leading chronic diseases—cardiovascular disease, cancer, chronic respiratory disease, and diabetes—including unhealthy diet, tobacco use, and harmful alcohol use. Poor oral hygiene is also an additional risk factor.

Periodontal Disease and Cardiovascular Diseases

Chronic bacterial and viral infections have received renewed interest as possible risk factors for atherosclerosis and its clinical manifestations. Prospective studies have found that periodontal disease, resulting from infection by anaerobic, gram-negative bacteria, is associated with greater risk of coronary heart disease and stroke events. Two proposed mechanisms for a causal relationship between periodontal infection and atherosclerotic cardiovascular disease are: (1) inflammatory response due to chronic, repeated, systemic exposure to periodontal pathogens, lipopolysaccharide (LPS), and other bacterial products following loss of epithelial integrity in the periodontal sulcus, and (2) systemic exposure to certain species of periodontal bacteria that may promote platelet aggregation and blood coagulation increasing the risk of acute thromboembolic events (Herzberg & Meyer, 1996).

Studies have reported association of hemostatic factors such as fibrinogen, tissue-type plasminogen activator (tPA), and von Willebrand factor (vWF) and others with periodontal disease. A mild to moderate deficiency of factor VIII and vWF often is associated with gingival bleeding. Individuals with severe dental infections also have high levels of vWF antigen, leukocytes, and fibrinogen (FIB). Fibrinolytic activity forms the other side of the coagulation haemostatic mechanism. Whereas tPA converts plasminogen (PLG) to plasmin, plasmin activates the kinin cascade and latent metalloproteinases. It has been shown that the alteration of the plasminogen activator–plasmin system affects the progression of periodontal disease, especially because the periodontal pathogens may bind to metalloproteinases. Gingival fibroblasts may have an affect on the severity of inflammation and degradation of the extra cellular matrix of gingival tissues by producing a

large amount of plasminogen activator in response to LPS. Thrombomodulin (TM) is an endothelial cell protein that binds protein C and thrombin. This reaction leads to activation of protein C by thrombin in gingival crevicular fluid (GCF) at sites with bleeding on probing. Thrombomodulin in gingival epithelium may regulate thrombin activity at sites of coagulation and inflammation with periodontal disease, although inflammation may impair the regulation of thrombin. A paradigm shift in evaluating the periodontal–cardiovascular disease relationship occurred with the suggestion that periodontal disease may act as an exposure/risk factor for cardiovascular disease. Several studies have found higher levels of plasma C-reactive protein (CRP; a general inflammatory marker) in participants with periodontal disease compared to those free of periodontitis (Ebersole et al., 1999; Joshipura, Wand, Merchant, & Rimm, 2004; Mattila et al., 2002; Slade, Offenbacher, Beck, Heiss, & Pankow, 2000).

Hemostatic factors in the blood are consistently associated with both prevalent and incident cardiovascular disease. Most studies have presented main-effects model and have not explored potential interactions between potential effect modification between periodontal disease and important biological factors purportedly involved in associated etiological mechanisms. Firm causal analyses should explore effect modifications in studies that are designed and well-powered to detect real interactions. As has been pointed out earlier, it is possible that attachment loss misclassifies contemporaneous microbial/inflammatory burden. Periodontal microbial titers, levels of periodontal infection, or pro-inflammation disease activity markers in the gingival crevicular fluid may classify the true exposure more accurately. These measures have not yet been used in published studies such as this one.

It has been suggested that periodontal disease may lead to cardiovascular disease outcomes through microbial challenge from the infected periodontal tissues, resulting LPS exposure, and concurrent increase in systemic inflammatory and haemostatic factors in the blood. In one recent study (Mattila et al., 2002), treatment of periodontal disease was shown to lead to reduced plasma CRP levels suggesting the possibility that periodontal disease may act as a trigger for increasing plasma CRP leading to other potential consequences.

Overall, the results from different studies examining the relationship between periodontal disease and cardiovascular outcomes have been mixed. As in several other associations with periodontal disease and systemic health outcomes, residual confounding by smoking and use of clinical measures of periodontal disease rather than measures that represent biological activity representative of the purported inflammatory or infective mechanisms cloud the true nature of the associations. It has been reported from a study of a subset of the Atherosclerosis Risk in Communities (ARIC) Study that clinical periodontal measures were not associated with risk for coronary heart disease (CHD), but IgG antibody levels of several periodontal organisms

were associated with CHD [for example, IgG for *T. denticola* (OR: 1.7; CI: 1.2 to 2.3) in smokers, and IgG for *Actinobacillus actinomycetemcomitans* (OR: 1.7; CI: 1.2 to 2.7)]. "These findings indicate that the quality and quantity of the host response to oral bacteria may be an exposure more relevant to systemic atherothrombotic coronary events than clinical measures" (Beck et al., 2005).

The recently published "American Journal of Cardiology and Journal of Periodontology Editors' Consensus" on the association between periodontitis and atherosclerotic cardiovascular disease clarified the current understanding of the link between atherosclerotic cardiovascular disease (CVD) and periodontitis aiming at providing an approach to reducing the risk for primary and secondary atherosclerotic CVD events in patients with periodontitis (Friedewald, Kornman, Beck, Genco, Goldfine et al., 2009). Variation between different studies related to the association between CVD and periodontitis was attributed to: (1) variations in study populations, including differing age groups, ethnicities, and geographic locations, and (2) differing measures and definitions of periodontitis, with some studies based only on clinical measures (i.e., pocket depth, bleeding with probing, tooth attachment level). Other studies, in which the relation appeared stronger, were based on nonclinical measures such as systemic antibody response or radiographic evidence of alveolar bone loss (Friedewald et al., 2009).

The consensus statement outlined that although the treatment of periodontitis reduces systemic markers of inflammation and endothelial dysfunction, no prospective periodontitis intervention studies have evaluated CVD outcomes. There exists evidence to believe that because untreated or inadequately controlled moderate to severe periodontitis increases the systemic inflammatory burden, periodontitis may independently increase the risk for CVD. The consensus group made a series of recommendations related to clinical practice in situations where periodontal health CVD health issues may be involved that are summarized in Box 16.3.

Periodontal Disease and Pregnancy Outcomes

Periodontal disease is mostly a low grade chronic infection that may serve as a continuously acting generative pump to increase inflammatory cytokine levels in the blood. It may therefore be possible that periodontal disease may impact pregnancy outcomes such as preterm birth and low birth-weight deliveries, placental infection, premature rupture of membranes, preeclampsia and premature labor through systemic maternal infections mediated by inflammatory cytokines, and increased prostaglandin production (Boggess et al., 2003; Dasanayake, 1998; Jeffcoat et al., 2001; Offenbacher et al., 1996, 2001). An attempt to look for a causal association between periodontal disease and pregnancy outcomes should also logically

BOX 16.3 Recommendations of "The American Journal of Cardiology and Journal of Periodontology Editors' Consensus: Periodontitis and Atherosclerotic Cardiovascular Disease"

Clinical Recommendations: Patients with Periodontitis

I. Patient Information

Recommendation A: Patients with moderate to severe periodontitis should be informed that there may be an increased risk for atherosclerotic CVD associated with periodontitis.

Recommendation B: Patients with moderate to severe periodontitis who have one known major atherosclerotic CVD risk factor, such as smoking, immediate family history of CVD, or history of dyslipidemia, should consider a medical evaluation if they have not done so in the past 12 months.

Recommendation C: Patients with periodontitis who have ≥ 2 known atherosclerotic CVD major risk factors should be referred for medical evaluation if they have not done so in the past 12 months.

II. Medical and Dental Evaluations

Recommendation A: Medical evaluation of patients with periodontitis should include assessment of atherosclerotic CVD risk, including past CVD events, family histories of premature atherosclerotic CVD disease or sudden coronary death, diabetes mellitus, systemic hypertension, and dyslipidemia.

Recommendation B: Medical evaluation of patients with periodontitis should include a complete physical examination and annual measurement of blood pressure at rest (seated for 5 minutes with the feet on the floor and attention to appropriate blood pressure cuff size).

Recommendation C: Medical evaluation of patients with periodontitis should include a blood lipid profile (total cholesterol, LDL cholesterol, HDL cholesterol, and fasting triglycerides) and blood glucose measurement. A plasma hsCRP determination is optional but should be considered, because recent studies have suggested that elevated plasma hsCRP may have added value by helping determine how aggressively standard risk factors should be treated, especially lifestyle changes.

III. Risk Factor Treatment: Abnormal Lipids

Recommendation A: Patients with periodontitis and 1 abnormal serum lipid and/or elevated plasma hsCRP are recommended to follow a multifaceted lifestyle approach to reduce atherosclerotic CVD risk according to the National Cholesterol Education Program Adult Treatment Panel III guidelines.

Recommendation B: Drug therapy for elevated LDL cholesterol should be prescribed in patients with periodontitis in whom target LDL cholesterol levels are not achieved with lifestyle changes.

(Continues)

BOX 16.3 Recommendations of "The American Journal of Cardiology and Journal of Periodontology Editors' Consensus: Periodontitis and Atherosclerotic Cardiovascular Disease" *(Continued)*

IV. Risk Factor Treatment: Cigarette Smoking

Recommendation: All patients with periodontitis who smoke tobacco should discontinue this habit because this is a major risk factor for atherosclerotic CVD and periodontitis.

V. Risk Factor Treatment: Hypertension

Recommendation A: All patients with periodontitis and elevated blood pressure should be treated to target levels as defined by the seventh report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure (JNC-7).

Recommendation B: All patients with periodontitis and elevated blood pressure should undertake lifestyle changes.

Recommendation C: All patients with periodontitis and elevated blood pressure not controlled to target levels with lifestyle changes should be treated with pharmacologic therapy.

Recommendation D: Patients with periodontitis prescribed calcium channel blockers for hypertension or any other indication should be monitored for worsening of periodontitis in association with gum hyperplasia.

VI. Risk Factor Treatment: Metabolic Syndrome

Recommendation: Patients with periodontitis meeting criteria for metabolic syndrome should be identified, and all risk factors for atherosclerotic CVD should be treated, beginning with lifestyle changes aimed at weight reduction.

Clinical Recommendations: Patients with Atherosclerotic Cardiovascular Disease with or Without a Previous Diagnosis of Periodontitis.

I. Patients with Atherosclerotic CVD and Previous Diagnosis of Periodontitis

Recommendation: Periodontists and physicians managing patients with CVD should closely collaborate to optimize CVD risk reduction and periodontal care.

II. Patients with Atherosclerotic CVD and No Previous Diagnosis of Periodontitis

Recommendation A: Periodontal evaluation should be considered in patients with atherosclerotic CVD who have signs or symptoms of gingival disease,

(Continues)

BOX 16.3 Recommendations of "The American Journal of Cardiology and Journal of Periodontology Editors' Consensus: Periodontitis and Atherosclerotic Cardiovascular Disease" *(Continued)*

significant tooth loss, and unexplained elevations of hsCRP or other inflammatory biomarkers.

Recommendation B: Periodontal evaluation of patients with atherosclerotic CVD should include a comprehensive examination of periodontal tissues, as assessed by visual signs of inflammation and bleeding on probing, loss of connective tissue attachment detected by periodontal probing measurements, and bone loss assessed radiographically. If patients have untreated or uncontrolled periodontitis, they should be treated with a focus on reducing and controlling the bacterial accumulations and eliminating inflammation.

Recommendation C: When periodontitis is newly diagnosed in patients with atherosclerotic CVD, periodontists and physicians managing patients' CVD should closely collaborate to optimize CVD risk reduction and periodontal care.

Summarized from Friedewald et al., 2009.

imply that treatment of periodontal disease (i.e., removal or substantial reduction of the pro-inflammatory challenge) should also lead to desirable clinical results in these adverse pregnancy outcomes. Although some studies have demonstrated such results, they are generally small-sized and not generalizable (Jeffcoat et al., 2003; Lopez, Smith, & Gutierrez, 2002).

As has been discussed elsewhere, measurement and definition of periodontal disease reflects *historical biological* rather than *current biological* activity. For example, consistently increased inflammatory cytokine levels attributable to periodontal disease at required times during pregnancy have not been clearly established. This offers the possibility of misclassification of exposure that is measured in historical context. An interesting recent study assessed the impact of changed definitions of periodontal disease on the associations between periodontal disease and pregnancy outcomes. This study assessed 14 different ways of defining periodontitis and more than 50 periodontal disease continuous measurements that have been reported in literature examining periodontal disease and pregnancy outcome studies (Manau, Echeverria, Agueda, Guerrero, & Echeverria, 2008). They then applied these definitions to their own cohort of pregnant women and performed logistic regression analysis to assess the association between periodontal disease (as defined by different definitions) and pregnancy outcomes. The study reported that the *significance of the association between periodontal disease and*

pregnancy outcomes varied by the periodontal disease definition or the measurement used. Although this study has opened the possibility of potential spurious associations (at least based on disease definition vs biological plausibility), the study method stated "every case definition and every measurement of periodontal disease described was independently tested for association with each adverse pregnancy outcome. The level of statistical significance was set at 5%" (Manau et al., 2008). Considering the number of post-hoc tests run in the analysis, adjustments for multiple testing should have been made. Nevertheless, the possibility of association between periodontal disease and pregnancy outcomes as a function of disease definition suggests that periodontal disease definitions need more careful attention, especially in terms of relating commonly used measures with biological activity that it is supposed to represent, to explain the observed associations with pregnancy outcomes.

Periodontal Disease and Diabetes

Diabetes mellitus is now recognized to be a group of disorders involving altered glucose, fat, and protein metabolism that manifests as increased blood glucose levels and an altered glucose-tolerance test. Consistently raised levels of glycosylated hemoglobin (Hb1Ac) in the blood indicate poor control of diabetes mellitus. Contrary to an earlier concept, a stage of "prediabetes" is now recognized where overt glucose intolerance is not evident, but blood glucose levels are on the borderline upper levels, and Hb1Ac is raised. Some patients may switch between prediabetes and diabetes depending on their physiological status and lifestyle situations but may eventually shift to a diabetes stage more permanently. Even among children with diabetes, prevalence of periodontal disease is greater than in children without diabetes. It has been suggested that periodontal disease and diabetes mellitus share a bidirectional association because some evidence seems to suggest that while diabetes is a risk factor for severity of gingivitis and periodontitis, periodontitis is also a risk factor for poor glycemic control and complications among diabetics. Periodontal disease may sometimes be the first sign of diabetes mellitus. Long-standing, uncontrolled or poorly controlled diabetes leads to immune compromise, and such patients are at risk of developing oral candidiasis and other opportunistic infections.

Using multivariable modeling to control for other risk factors for periodontitis, Tsai, Hayes, and Taylor (2002) demonstrated "the odds of having periodontitis in adults with poorly controlled diabetes mellitus was 2.9 compared with that in adults without diabetes mellitus. Furthermore, for people who had diabetes mellitus but better glycemic control, the odds ratio was 1.56. This study is important because of the nationally representative population and the consideration of multiple complicating variables" (Lamster, Lalla, Borgnakke, & Taylor, 2008). The risk of death from cardiac or

renal disease for people with severe periodontitis was 3.2 times greater than that of people with no, mild, or moderate periodontitis (Saremi, Nelson, Tulloch-Reid, Hanson, Sievers et al., 2005).

Periodontal Disease and Other Systemic Diseases

Periodontal disease has been associated with several cancers such as oral cancer (multivariable adjusted ORs varying between 1.4 and 5.3); gastric cancer; esophageal cancer, (OR/RR/HRs: 1.3–2.1); lung cancer (HR: 0.58–1.54); pancreatic cancers (OR/RR/HRs: 1.23–2.1); and all cancers (OR: 1.1–1.55) (reviewed by Meyer, Joshipura, Giovannucci, & Michaud, 2008). Smoking is usually identified as an important independent risk factor in these studies. Smoking has been associated with tooth loss, periodontal diseases, and several cancers. This raises the possibility of such as a potential confounding of the association between tooth loss, periodontal disease, and cancer by smoking and other common risk factors (Meyer et al., 2008).

A stroke or cerebrovascular accident is damage to the brain due to a reduction in the blood supply to the brain. Several studies have associated periodontal disease with stroke, implying that periodontal disease could be an exposure factor that leads to stroke through inflammatory mechanisms. In different studies, the multivariable adjusted OR for association between periodontal disease and stroke has varied between 1.27 (95% CI: 1.01–1.61) and 2.90 (CI: 1.49–5.62) (reviewed by Joshipura, 2002). Although a causal relationship between periodontal disease and stroke may exist, especially for recent events and active periodontal lesions, the evidence is not clear as different studies have used varying criteria and disease definitions. It is therefore "unclear whether the associations found between these oral conditions and cardiovascular disease had any causal component" (Joshipura, 2002). It has been suggested that because oral disease and cardiovascular disease share several common risk factors, it is important to rule these out as alternative explanations before interpreting a relationship as causal.

Increasing evidence suggests that clinical signs of periodontal disease are independently associated with renal impairment. Kshirsagar, Offenbacher, Moss, Barros, and Beck (2007) examined the possible linkage of kidney disease with serum antibody to oral pathogens. They reported that high levels of serum IgG to selected periodontal pathogens including *P. gingivalis, T. denticola*, and *Aggregobacter actinomycetemcomitans* were associated with an increased odds for glomerular filtration rate (< 60 ml/min/1.73 m^2). Multivariable adjusted odds ratio for *P. gingivalis* (OR: 1.6 CI: >1.0–2.6), *T. denticola* (OR: 1.8 CI: 1.2–2.8), and *Aggregobacter actinomycetemcomitans* (OR: 1.7 CI: 1.1–2.7) suggests their independent association with impaired renal function after adjusting for traditional risk factors such as race/ethnicity, age, gender, education, diabetes, hypertension, LDL cholesterol, HDL cholesterol, triglycerides, and BMI (Kshirsagar et al., 2007).

The burden of oral diseases and other chronic diseases can be decreased simultaneously by addressing common risk factors such as tobacco use and unhealthy diet, use of protective gears to prevent injuries, and long-term exposure to an optimal level of fluoride. Disease prevention can be implemented through various strategies, especially if common pathways are challenged. WHO suggests that the public health solutions for oral diseases are most effective when they are integrated with other chronic diseases and with national public health programs.

Orofacial Clefts: Cleft Lip and Palate

Orofacial clefts (OFCs) are the most common craniofacial malformation in the newborn and are present in some 171 syndromes (Eppley, van Aalst, Robey, Havlik, & Sadove, 2005). OFCs consist of cleft lip, cleft palate, and cleft lip and palate together. Although atypical presentations may occur, the most common clinical presentation is that of a lateral cleft of the lip through the philtrum with or without extension through the palatal shelves. Some of the OFCs are predictable embryologically while some are not. There exists no universally accepted comprehensive classification for OFCs. Surgical treatment of OFCs leads to a dramatic clinical, functional, and aesthetic improvement, allowing the patient to lead a normal life. Surgical treatment-based classification systems have been developed, with Tessier's classification system being most commonly used (Tessier, 1976). This system numbers OFCs from 0 (midline cleft of the lip and nose) to 30 (mandibular cleft) depending upon their anatomical location (15 locations for clefts, and combinations of several types of clefts and associated malformations) using the orbit as the primary reference point. It has been pointed out that "although this classification system is of value for most craniofacial clefting problems, it is inadequate for the commonly seen cleft lip–cleft palate deformity (the typical cleft lip corresponds in part to Tessier cleft nos. 1, 2, and 3)" (Eppley et al., 2005).

A new classification system has been proposed that scores the surgical complexity of the OFCs for repair. For example, for primary palate clefts, scores range from 0 (normal primary palate) to 12 (complete cleft of the primary palate with contact between the segments). Surgical complexity was incorporated by adding a complexity score (1 unit for every millimeter of separation of the cleft) under the assumption that the greater the separation, the greater the difficulty in surgery. Lip features, including height of the lip; symmetry; sulcus depth; and muscular, skin, and mucosal integrity (among others) were incorporated as scores from 1 (symmetrical lip height) to 9 (presence of cupid arch). To obtain the overall complexity score, each of the subscores from the above categories are added up (Ortiz-Posadas, Vega-Alvarado, & Maya-Behar, 2001). Similarly, separate scores are developed for secondary palate, lip and nose. This system attempts to derive a numerical

score using basic mathematical operations, but the validity of assigning different complexity factors as multiplication units appears to be somewhat arbitrary. For example, the authors state that "from a surgical and aesthetic-functional perspective, the complexity of a bilateral cleft and its repair exceeds the simple summed complexity of the unilateral clefts that form the bilateral cleft. For that reason, bilateral clefts were scored as 1.5 times the sum of the unilateral cleft components" (Ortiz-Posadas, Vega-Alvarado, & Maya-Behar, 2001). It remains an open question as to why the assigned score was not deemed to be another rational number.

Several behavioral and nutritional factors such as alcohol consumption, cigarette smoking, folic acid shortage, steroids, anticonvulsant drugs, an excess of retinoic acid and vitamin A, as well as environmental factors such as altitude, have been associated with occurrence of OFCs. Coffee drinking by pregnant mothers has been suggested to be a contributing factor in OFC of their children. In a recent report based on a large, population-based case-control study in Norway including 573 OFC cases (377 with cleft lip with or without cleft palate and 196 with cleft palate only) and 763 randomly selected controls, Johansen, Wilcox, Lie, Andersen, and Drevon (2009) could not detect evidence of an association between caffeine exposure and OFCs when all sources of caffeine were considered. Adjustment for known confounding factors in general had minor effects on risk estimates. In this report, "compared with that for no coffee consumption, the adjusted odds ratios for cleft lip with or without cleft palate were 1.39 (95% confidence interval: 1.01, 1.92) for less than 3 cups a day and 1.59 (95% confidence interval: 1.05, 2.39) for 3 cups or more. Coffee consumption was not associated with risk of cleft palate only (for > or = 3 cups vs. none, adjusted odds ratio = 0.96, 95% confidence interval: 0.55, 1.67). Tea consumption was associated with a reduced odds ratio of both cleft lip with or without cleft palate and cleft palate only" (Johansen et al., 2009).

The phenotypical effect resultant in OFCs are possibly outcomes of interaction of environmental–behavioral factors with the genotype, involving genes such as: MSX1,TGFb3, TGFA, MTHFR and GABA receptor b3 (Carinci, Pezzetti, Scapoli, Martinelli, Avantaggiato et al., 2003; Krapels, Vermeij-Keers, Müller, de Klein, & Steegers-Theunissen, 2006). However, the results from different studies for some of these gene-effects and gene-environment interactions have been varied, making it difficult to derive firm conclusions. For example, a Human Genome Epidemiology review (HuGE review) reported that:

> Transforming growth factor alpha (TGFA) is a well-characterized mammalian growth factor. Since the first report of an association between DNA sequence variants at the TGFA genetic locus and nonsyndromic oral clefts, 47 studies have been carried out, producing conflicting results . . . Bias, lack of statistical power, and genuine population diversity can explain the diverse results. In the aggregate, TGFA is probably a genetic modifier of cleft-

> ing in humans, which is consistent with the oligogenic model suggested for nonsyndromic oral clefts. (Vieira, 2006)

HuGE reviews identify human genetic variations at one or more loci, describe what is known about the frequency of these variants in different populations, identify diseases with which these variants are associated, summarize the magnitude of risks and associated risk factors, and evaluate associated genetic tests. Reviews point to gaps in existing epidemiologic and clinical knowledge, thus stimulating further research in these areas (Office of Public Health Genomics, CDC, 2007).

Recently, Wyszynski, Sárközi, and Czeizel (2006) discussed methodological factors that account for the wide variation in the reported prevalence rates of anomalies associated with oral clefts and pointed out six reasons that "the published prevalences of associated anomalies vary considerably:"

1. Differences in case definition and inclusion/exclusion criteria
2. Length of time after birth that cases are examined
3. Variability of clinical expression of associated anomalies
4. Knowledge and technology available to produce syndrome delineation
5. Selection of patients, sources of ascertainment, and sample size
6. True population differences and changes in frequency over time

The National Center on Birth Defects and Developmental Disabilities at the Centers for Disease Control and Prevention conducted a workshop in January 2006 entitled "Prioritizing a Research Agenda for Orofacial Clefts." Yazdy, Honein, Rasmussen, and Frias (2007) summarized the state of knowledge and need for further research in oral clefts as:

> The goals of the meeting were to review existing research on OFCs, identify gaps in knowledge that need additional public health research, and develop a prioritized research agenda that can help guide future public health research. Experts in the field of epidemiology, public health, genetics, psychology, speech pathology, dentistry, and health economics participated to create the research agenda. Research gaps identified by the participants for additional public health research included: the roles of maternal nutrition, obesity, and diabetes in the etiology of OFCs; psychosocial outcomes for children with OFCs; the quality of life for families and children with OFCs; and the health care costs of OFCs. To create the research agenda, the participants prioritized the research gaps by public health importance, feasibility, and outcomes of interest. (Yazdy et al., 2007)

Specifically, the following eighteen areas were pointed out for further OFC research (Yazdy et al., 2007).

1. Phenotype characterization to define more etiologically homogeneous categories of OFCs
2. Effects of nutrition and nutritional supplements on the risk of OFCs

3. Early screening measures to identify learning outcomes in children with OFCs
4. Quality of life for children with OFCs
5. Social marketing campaign targeting smoking and OFCs
6. Long-term outcomes for individuals with OFCs
7. Effect of timing of OFC diagnosis
8. Obesity, maternal diabetes, and insulin resistance in the etiology of OFCs
9. Using ''parameters of care'' in treatment of OFCs
10. Ethnicity and population differences in OFCs
11. Effects of maternal medication use in the etiology of OFCs
13. Mental health in adolescents with OFCs
13. Costs associated with OFCs
14. Maternal infection and OFCs
15. Maternal alcohol consumption and OFCs
16. Effect of payor status on outcomes in children with OFCs
17. Dyslexia intervention for children with OFCs
18. Air pollution in the etiology of OFCs

Overall, the goals of these recommendations for research in OFCs are twofold: To increase capacity to prevent OFCs and to improve the quality of life and other long-term outcomes for children and families affected by OFCs.

17 Fluorides

Introduction

For more than five decades, fluoride has been the cornerstone of preventive dentistry in the United States and elsewhere. In the United States, the adoption of fluoridation as a public health measure began in 1962 when the U.S. Public Health Service recommended that communities add 0.7 to 1.2 mg/L of fluoride to the community water supply. In the United States, community water fluoridation reaches more than 160 million people. It has been recognized as one of 10 great achievements in public health of the 20th century because of its causal links to large reductions in tooth decay in many industrialized countries during the latter half of the century. It is ranked along with other great public health achievements in the United States such as vaccination, motor-vehicle safety, safer workplaces, control of infectious diseases, decline in deaths from coronary heart disease and stroke, safer and healthier foods, healthier mothers and babies, family planning, and recognition of tobacco use as a health hazard.

Sources of fluoride include fluoridated drinking water, non-artificially fluoridated municipal water that may have a minute concentration of fluoride, well water, bottled water from a municipal source, spring water, bottled "infant" or "nursery" water, bottled water with added fluoride, and distilled or purified water. Naturally available drinking water in several countries such as India, China, and several African countries may contain high concentrations of naturally occurring fluoride in excess of WHO's recommended Guideline Value of 1.5 mg/L (Lennon, Whelton, O'Mullane, & Ekstrand, 2005). Prevalence of dental and skeletal fluorosis in these areas is also reportedly high. Foods and beverages such as tea, beer, and wine may provide additional fluoride exposure. Table 17.1 shows fluoridated water coverage from the 27 countries that have water fluoridation schemes with

TABLE 17.1 Coverage of Fluoridation in Countries Serving 1 Million or More Population

		Water fluoridation schemes covering populations of 1 million or more	*Drinking water supplies with a natural fluoride concentration of around 1 mg/L covering populations of 1 million or more*
Country	*Population (millions)*	*Percent Population Covered*	*Percent Population Covered*
Argentina	35.9	9	12.5
Australia	19.3	60.6	—
Brazil	172.5	38	—
Canada	31	42.9	—
Chile	15.4	35.1	—
Colombia	42.8	68.7	—
France	59.4	—	3
Gabon	1.3	—	100
Guatemala	11.7	15.4	—
Hong Kong	6.7	100	—
Ireland	3.8	60.5	—
Israel	6.4	67.2	—
Korea	46.1	11.7	—
Libya	5.4	—	18.5
Malaysia	22.6	69.9	
Mexico	100.4	—	3
New Zealand	3.8	60.5	—
Philippines	77.1	6.5	—
Senegal	9.7	—	10.3
Singapore	4.1	100	—
Spain	39.9	10	—
Sri Lanka	19.1	—	14.7
Tanzania	35	—	34.9
United Kingdom	59.5	9.1	—
United States	281.4	60.8	3.6
Vietnam	79.7	5.5	—
Zimbabwe	13	—	20

Adapted from Lennon, Whelton, O'Mullane, and Ekstrand, 2005.

fluoridated water reaching a substantial number of people for which data are available with WHO.

Alternative sources of fluoride have been introduced over the years. While fluoridation was the main source of fluoride in the 1950s in the United States, fluoride is available from many sources today. Fluoride-containing dental products in several forms such as gels and varnish are also applied by healthcare practitioners. The additional sources of fluoride include dietary

fluoride supplements, dentifrices, rinses, and tablets. Fruit juices, carbonated beverages, infant formulas, and certain cereals are also known to contain significant amounts of fluoride and may contribute to the total intake of fluoride (Burt, Eklund, & Lewis, 1992). Different countries employ different vehicles for delivering fluoride to the population for the prevention of dental caries—for example, water fluoridation in the United States and Australia, salt fluoridation in France and Switzerland, and the use of professionally applied fluorides to individuals in Scandinavia are well known methods.

It is generally acknowledged that this increased availability of fluoride has contributed significantly toward the decline in dental caries in both fluoridated and non-fluoridated communities (Ripa, 1993; USDHHS, 1991). Explanations for the caries decline in non-fluoridated areas include the use of fluoride toothpastes and other forms of fluoride, extension of the benefits of fluoridation through consumption of beverages and foods processed in fluoridated areas, improved oral hygiene, and improved restorative care (Brunelle & Carlos, 1990; Ripa, 1993). In the 1940s when caries levels in the United States were high, it was acceptable to think that the more fluoride protection one received the better it was for their caries prevention effort. However, the benefits from exposures to multiple sources of fluoride are not always additive, i.e., they may not provide significant additional benefits. Excessive fluoride exposure may lead to dental fluorosis. The optimum level of fluoride in drinking water is decided on a balance between the maximum caries beneficial effects and the risk of dental fluorosis, an unwanted side effect. Practitioners prescribing fluoride therapy should base their recommendations on an understanding of the potential for fluoride exposure.

Adoption of a fluoride technology varies globally according to environmental, cultural, economic, and political circumstances of the country or community. In communities where fluoride in drinking water exceeds an acceptable level, it is removed to prevent enamel fluorosis and skeletal fluorosis. Exposure to fluoride may occur through air, food, and drinking water; the most common methods of fluoride delivery for prevention of dental caries are water fluoridation and fluoride toothpastes. In the United States, the National Center for Fluoridation website (http://www.fluoridationcenter.org/) keeps fluoridation-related information for public access. At this time, no central and comprehensive global fluoride data repository exists, but some information based on self-reported data is available at the FDI: World Dental Federation website (http://www.fdiworlddental.org).

Mechanism of Action

When the inverse relationship between fluoride in water and dental caries was discovered, it was thought that the caries' inhibitory effect was primarily systemic through ingestion and incorporation of the fluoride ion into developing teeth. Current research indicates that fluoride is more effective

when a low level is maintained in the oral cavity at all times (Burt et al., 1992; Whitford, 1996). Fluoride has been shown to enhance remineralization and inhibit demineralization during the carious process. Fluoride is stored in plaque and released in response to the acidic environment. In the presence of appropriate levels of calcium and phosphate ions, fluoride has been shown to reverse enamel demineralization. In addition, fluoride in plaque is also known to inhibit glycolysis, the process by which fermentable carbohydrate is metabolized by cariogenic bacteria to produce acid. Other mechanisms of fluoride action include bactericidal effect at higher concentration, formation of a temporary layer of calcium fluoride, and making *streptococcus mutans* less acidogenic (Burt et al., 1992).

Studies have not conclusively provided evidence about the pre-eruptive or post-eruptive action of fluoride. However, it appears that the anticaries activity of fluoride may have both pre-eruptive as well as post-eruptive effects. A recent report from a study examining the pre- and post-eruptive effect on permanent first molars in 6- to15-year-old Australian children suggested that pre-eruption exposure was important for a caries-preventive effect in these children since the post-eruption effect alone could not lower caries levels significantly (Singh & Spencer, 2004). The report also demonstrated that a high pre-eruption exposure to fluoride could decrease caries levels significantly in pit and fissure surfaces; although for other surfaces, caries prevention was evident only at high pre- and post-eruption exposure to fluoride. A significant preventive effect was seen in all surfaces by a continuous pre- and post-eruption exposure.

Water fluoridation provides a mechanism for this continuous exposure as a benefit in all surfaces. A predominant part of the anticaries activity of fluoride is a function of its concentration in the fluid environment of the teeth (Ekstrand & Oliveby, 1999) whether the teeth are in development stages in their intra-bony crypts or have erupted into the oral cavity. Before eruption, the main mechanism by which fluoride can get to the immediate environment of the developing tooth is through blood, whereas post-eruption, fluoride can reach the local dental environment through consumed fluids, applied therapy, and constantly secreted saliva. After fluoride exposure, plaque becomes a reservoir for fluoride. In the plaque fluid, fluoride may exist in ionic form and be bound either in plaque, in calcium fluoride, to enamel, and/or to soft tissues. In persons eating a normal diet and living in an area with about 0.2 ppm of fluoride in the water supply, the normal salivary fluoride level is about 0.6 μmol/L (0.01ppm) (Oliveby, Ekstrand, & Lagerlöf, 1987). This concentration is modified by consumption of other fluoride agents such as fluoride toothpaste, other agents, and foods.

Pre-eruptively, fluoride may contribute to anticaries activity by strengthening the enamel structure by formation of fluoroapatite crystal structure. The fluoride ion replaces hydroxyl ions in hydroxyapatite crystal

lattice, leading to formation of fluoroapatite in the enamel. The equation below depicts the formation of fluoroapatite from hydroxyapatite in the presence of fluoride ions, which occurs when fluoride concentration is about 200–500 mg/L (Friedman, Solouki, Gurevitz, Gedalia, & Onisi, 1984; Lazzari, 1976).

$$\underset{\text{(hydroxyapatite)}}{Ca_{10}(PO_4)_6(OH)_2} + 2F^- \rightarrow \underset{\text{(fluoroapatite)}}{Ca_{10}(PO_4)_6F_2} + 2OH^- \quad \text{(Eq 1)}$$

There is an initial rapid absorption of fluoride on the surface of hydroxyapatite with displacement of hydroxyl ions, followed by slow diffusion of fluoride that is ultimately followed by slow diffusion of the hydroxyl ions outside the crystal. Thereafter, recrystallization occurs when already existing Ca_2^+ and PO_4^{3-} ions (which are in equilibrium with hydroxyapatite) precipitate to form fluoroapatite (Lazzari, 1976). Enamel formation occurs from the dentinal side toward the occlusal surface as ameloblasts move in that direction. Therefore, the total time available for these reactions to occur effectively would be more when the surface enamel is formed until the tooth erupts. It has been suggested that the early maturation stage of enamel formation appears to be particularly sensitive to the effects of fluoride on enamel formation (Den Besten, 1999).

Fluoroapatite strengthens the surface enamel to resist organic acid attacks more effectively. The strength of fluoroapatite comes from the fluoride ion being more electronegative than the hydroxyl ion and the total electron density and electron availability around fluoride being more than that of hydroxyl ions. This reduces the total number of attacks on fluoroapatite, making it more stable. Also, the calcium–fluoride bond is stronger than the calcium–hydroxyl ion bond due to the fluoride ion's higher electronegativity. Another factor that may contribute to the strength of fluoroapatite is shorter crystal axes in fluoroapatite compared to hydroxyapatite. Some studies have suggested that fluorides may help alter the tooth morphology by helping produce teeth with shallower grooves and more round cusps, which may reduce the chances of food sticking to the occlusal surfaces of teeth (Reynolds & Riley, 1981). These results have not been consistently reproduced, and it is assumed that even if these effects occur, their effect is probably very limited.

Post-eruptively, fluoride may contribute to anticaries activity in several different ways:

1. The fluoride ion helps maintain the surface enamel fluoroapatite structure through mechanisms similar to those described above (ten Cate, 1999). Furthermore, fluoroapatite crystals take up more lysozyme, which may provide some added benefit in thwarting microbial attacks (Eggen & Rölla., 1983; Twetman, Lindqvist, & Sund, 1986).

2. Caries reduction due to the fluoride concentration in the apatite structure of enamel is further strengthened by a continuous presence of fluoride in the plaque liquid. The surface enamel is in a state of flux due to constant demineralization and remineralization. Thus, the presence of fluoride in the immediate surrounding of the surface enamel pushes the reaction toward forming more fluoroapatite.
3. Fluoride reduces the wetability of enamel surfaces and reduces protein absorption on enamel. These actions may reduce the probability of plaque formation and acid retention in close proximity to the enamel surface, which may reduce the contact time for any effective demineralization to occur as a result of acid attacks.
4. The fluoride ion inhibits growth of *Streptococcus mutans*. In the plaque fluid, fluoride ions may exist in free form during the demineralization–remineralization process and may be augmented when supplied from outside in the form of fluoride rinses, other resources, or from salivary secretion. Although effects of fluoride on microbial organisms have been demonstrated, its anticaries actions probably stem more from its role in demineralization–remineralization (Geddes & Bowen, 1990).

When examining Equation 1 previously, a question can be asked: What happens if more fluoride ions exist in the environment of the hydroxyapatite than the optimal concentration? Equation 2 below shows how this reaction proceeds if more fluoride ions are available for the reaction (two fluoride ions in left-hand side of Equation 1 compared to 20 in Equation 2).

$$\underset{\text{(hydroxyapatite)}}{Ca_{10}(PO_4)_6(OH)_2} + 20F^- \rightarrow \underset{\substack{\text{(calcium fluoride—}\\ \text{powdery substance)}}}{10CaF_2} + 4PO_4^{3-} + 2OH^- \qquad \text{(Eq 2)}$$

The occurrence of dental fluorosis is dependent on the dose, duration, and timing of fluoride exposure (Den Besten, 1999). If a large amount of fluoride is present, then instead of fluoropaptite, the end product is calcium fluoride, which is a powdery substance. In such circumstances, instead of forming robust crystals, the enamel turns into a chalky substance that wears off quickly. A different physical form of enamel may occur in the in-between range of fluoride ion concentration. The net result is physically defective enamel resulting in dental fluorosis.

The decline in dental caries, increase in dental fluorosis, and a change in thinking regarding the mechanism of fluoride action that emphasizes more on its post-eruptive effect have influenced the recent revisions in the recommendations for fluoride use (ADA Council on Access, Prevention, and Interprofessional Relations, 1995). Proponents of water fluoridation stress the long history of safety, effectiveness, and economic advantages, while the opponents raise questions about safety, personal freedom, and the need for

fluoridation now, when a decline in dental caries is being observed in non-fluoridated areas. This decline, however, is seen only in developed countries.

Dental Fluorosis and its Measurement

Dental fluorosis is a hypomineralization of the enamel characterized by greater surface and subsurface porosity than in normal enamel changing the strength and optical properties of the enamel. It includes a range of manifestations from small white opacities on the surface of an intact tooth to yellow–black discoloration and massive destruction of the tooth. Measurement of dental fluorosis continues to be subjective with no clear reliable indices. Table 17.2 provides a comparative assessment of the indices currently used for measuring dental fluorosis (i.e., Dean's Index, the Thylstrup and Fejerskov Index (TFI), Tooth Surface Index of Fluorosis (TSIF), Fluorosis Risk Index (FRI), and the modified Developmental Defects of Dental Enamel (DDE) Index).

Most of the dental fluorosis cases occur in the questionable to mild range and are not easily discernible. This led to the argument that even if there were no major structural defect compromising the masticatory function of teeth in very mild/mild fluorosis, the aesthetic effects of teeth needed to be considered, especially because society has started placing much more emphasis on aesthetic appearances than ever before. This paradigm shift led to several studies that examined aesthetically acceptable/not acceptable dental fluorosis and intra-oral distribution of dental fluorosis (i.e., effect on posterior—not aesthetically important vs anterior—aesthetically important teeth). Studies of the intra-oral distribution of dental fluorosis in low-fluoride areas suggest that teeth that formed later in life were more frequently affected compared with those that formed early. The steady increase of plasma fluoride with age, even under constant fluoride exposure, has been suggested as a possible explanation (Fejerskov, Richards, & Den Besten, 1996; Larsen, Kirkegaard, & Poulsen, 1987; Manji, Baelum, Fejerskov, & Gemert, 1986). The risk period for dental fluorosis in upper central incisors has been estimated to be between ages 15 and 24 months for males and between 21 and 30 months for females (Evans & Darvell, 1995). An analysis from Kingston and Newburgh, New York, however, showed that the aesthetic consequence of exposure to multiple sources of fluoride was less dramatic, as evidenced by the lower frequency in upper anterior teeth compared with posterior teeth. The longer maturation process of the posterior teeth and the thicker enamel appear to be the likely explanation for the higher occurrence of dental fluorosis in posterior teeth (Kumar, Swango, Haley, & Green, 2000). Like most issues related to fluorosis, the issue related to aesthetically acceptable or not acceptable fluorosis and time of impact of fluoride vis-à-vis

TABLE 17.2 Comparison of Indices for Fluorosis Measurement

Criteria	*Dean's Index*	*Tooth Surface Index of Fluorosis (TSIF)*	*Thylstrup–Fejerskov Index (TFI)*	*Fluorosis Risk Index (FRI)*	*Developmental Defects of Dental Enamel (DDE) Index*
Measurement method	Visual	Visual	Visual	Visual	Visual
Measurement scale	Ordinal 6-point	Ordinal 8-point	Ordinal 10-point	Ordinal, 4-point. Also includes 2-point scale for excluded and nonfluoride defects.	Modified index: nominal 10-point for type of defect; ordinal 4-point for extent of defect
Characteristic measured	Severity of effect	Extent of surface involvement	Severity of effect	Extent of involvement	Defects on buccal and lingual surfaces of teeth—white and yellow deformities and hypoplasia
Level of measurement	Tooth level and individual level	Surface level	Surface level	Tooth level	Tooth level
Notes				Multiple zones (34) on tooth structure related to tooth development. Teeth divided into two timed, developmentally distinct zones.	Multiple coding systems and examinations may increase between-examiner and intra-examiner error; nonstandard definition of enamel defects.

tooth formation, and systemic vs local action of fluoride impacting caries prevention have been mired in controversy stemming from measurement problems associated with dental fluorosis.

> This confusion in the period at risk for dental fluorosis is due in part to the differences in the duration, type, and level of fluoride exposure, and also to the indices used to measure it. In addition, some limitations of the studies that attempt to explore associations between fluoride exposure and fluorosis on specific teeth may explain the lack of consistent results. First, studies conducted in fluoridated areas are ecologic, and therefore, all the constraints of ecologic studies are applicable. The fluoride exposure from water is measured at the aggregate level and may not reflect an individual's fluoride consumption. Second, the exact time of formation and mineralization of specific teeth at the individual level is not known. Third, the bioavailability of ingested fluoride is subject to variation at the individual level. Finally, the role of loosely bound fluoride in the skeleton and its carry-over effect will make it impossible for the timing of exposure to low levels of fluoride to be defined. Further, these problems are compounded by the difficulty in incorporating blindness into the fluorosis examinations, lack of methods to confirm the diagnosis, and concerns about the validity of exposure data from sources other than water water fluoridation. (Kumar, Swango, Haley, & Green, 2000)

Another major problem that confronts fluorosis studies is the lack of clear diagnostic criteria for dental fluorosis. Currently, diagnosis of fluorosis is based on a visual inspection of the tooth surface—fluorosis is diagnosed as white, chalky discoloration of the surface in its early forms, and therefore it is difficult to distinguish from other conditions that have a similar appearance, including several developmental disturbances and even early carious lesions. Box 17.1 describes definitions of some of the conditions that may cloud the diagnosis of dental fluorosis and notes the salient differences with mild fluorosis. Almost all enamel defects could be included in a list of differential diagnosis for dental fluorosis. This situation raises the question about potential disease misclassification that is more likely to be differential toward dental fluorosis rather than non-differential because of the greater awareness of fluorosis issues, and because water fluoridation is also a politically sensitive topic.

Dean's Index continues to be the most popular index of fluorosis, presumably because of its ease of use and simplicity, although it does not provide adequate information on the distribution of fluorosis within the dentition. Two key criticism of Dean's Index made more than 20 years ago are still valid: "The use of the term 'questionable' is too vague; and that the Index appears to describe the milder forms of fluorosis accurately but is not sensitive enough to distinguish between degrees of fluorosis in high-fluoride areas" (Clarkson, 1989). Furthermore, and by definition, a category designated as "questionable" disease is a built-in differential misclassification

BOX 17.1 Differential Diagnosis of Dental Fluorosis and Terms Used [(definition): synonyms/ alternative terms] in Describing Enamel Defects that May Be Confused with Dental Fluorosis and Should Be Included in Differential Diagnosis of Dental Fluorosis

Characteristic	*Milder Forms of Fluorosis*	*Nonfluoride Enamel Opacities*
Area affected	Usually seen on or near tips of cusps or incisal edges.	Usually centred in smooth surface; may affect entire crown.
Shape of lesion	Resembles line shading in pencil sketch; lines follow incremental lines in enamel, form irregular caps on cusps.	Often round or oval.
Demarcation	Shades off imperceptibly into surrounding normal enamel.	Clearly differentiated from adjacent normal enamel.
Color	Slightly more opaque than normal enamel; paper-white. Incisal edges, tips of cusps may have frosted appearance. Does not show stain at time of eruption (in these milder degrees, rarely at any time).	Usually pigmented at time of eruption often creamy-yellow to dark reddish–orange.
Teeth affected	Most frequent on teeth that calcify slowly (cuspids, bicuspids, second and third molars). Rare on lower incisors. Usually seen on six or eight homologous teeth. Extremely rare in deciduous teeth.	Any tooth may be affected. Frequent on labial surfaces of lower incisors. May occur singly. Usually one to three teeth affected. Common in deciduous teeth.
Gross hypoplasia	None. Pitting of enamel does not occur in the milder forms. Enamel surface has glazed appearance, is smooth to point of explorer.	Absent to severe. Enamel surface may seem etched, be rough to explorer.

(Continues)

BOX 17.1 *(Continued)*

Characteristic	*Milder Forms of Fluorosis*	*Nonfluoride Enamel Opacities*
Detection	Often invisible under strong light; most easily detected by line of sight tangential to tooth crown.	Seen most easily under strong light on line of sight perpendicular to tooth surface.

Synonyms/Alternative Terms

Dental fluorosis: Enamel fluorosis, mottling, fluorosed opacities

Enamel opacities (qualitative defect in enamel, abnormality in translucency of enamel): Internal enamel hypoplasia, developmental opacities, idiopathic opacities, demarcated, diffuse, confluent opacities

Enamel hypoplasia (quantitative defect in enamel, reduced thickness of enamel): Aplasia, internal and external hypoplasia, hypocalcification, pits, grooves, missing enamel

Discolored enamel (abnormal appearance in enamel): Pigmentation, tetracycline staining

Developmental defects of enamel (disturbances in hard tissue matrices and their mineralization during odontogenesis): Including enamel defects, dental fluorosis, enamel opacities, hypoplasia, and discolored enamel

Adapted from Russell, 1961: Clarkson, 1989.

factor in dental fluorosis diagnosis and estimation. This is not a matter of semantics: If something is questionable, then how is it possible to be classified as a definitive disease entity? In essence, the "questionable" category itself is a candidate that clearly suggests the need for a valid confirmatory (and diagnostic) test for dental fluorosis. Another criticism of Dean's Index is that it overestimates the aesthetic significance of dental fluorosis, because the subject-level classification is based on the lesser of the two worst-affected teeth and anterior and posterior teeth are weighted equally (Horowitz, 1986; Rozier, 1994). The DDE Index is time-consuming, and the analyses of data are complicated. Modifications have now been proposed to make it simpler to use and the data more meaningful. The TFI is related to the histology of fluorosis; however, the initial minute changes observed on dry enamel surfaces are of little aesthetic importance. The TSIF Index overcomes some of the limitations of Dean's Index but remains a classification rather than being a true index that may be usable in assessing disease progress or prognosis.

In essence, all dental fluorosis "indices" are disease classification or categories masquerading as indices because they have been issued an "ordinal" numeric code. These steps perhaps are only the initial few steps in the development of a true index of dental fluorosis. An approach to develop an index involves the following steps, discussed further in Chapter 18: development of a conceptual model, identifying dimensions and potential factors of value that may contribute to the index, testing and finalizing the model and index factors, formulating the index and developing a scoring method for using the index, fine-tuning and finalizing the index, and evaluating the index—testing the index (i.e., validity and reliability). The key factor for creating the index in question is to develop a thorough conceptual model and evaluate each aspect of the model to sequentially fit in the variety of factors into a single working model. Such models must consider causal and confounding issues (Dawid, 2002; Greenland & Brumback, 2002; Hernan, Hernandez-Diaz, Werler, & Mitchell, 2002), key measurement issues involved in quantifying an abstract idea (Falqueto, Lima, Borges, & Barreto, 2004), and tie these pieces into a consistent measurable whole (Slade, 1997).

In the context of dental fluorosis, the key determinant of a valid index would be an unambiguous, valid, and reliable diagnostic test that should be able to demonstrate fluoride in fluoroapatite or calcium fluoride in the enamel. Interestingly, whereas fluoride exposure data requires definitive evidence of fluoride (as in water, toothpaste, etc.) only during diagnosis of fluorosis, demonstration of fluoride is not required despite the potential for misclassification of disease. Although surface discoloration and defect can be used as a primary criterion for diagnosis of fluorosis, a confirmatory diagnosis should be able to demonstrate evidence of excess fluoride in the lesion for it to be attributed as fluorosis. Currently several advanced techniques in physics exist that can be utilized to develop a definitive diagnostic test for fluorosis in a clinical setting—for example, use of fluorescence imaging (Pretty et al., 2006).

Fluoride Exposure Measurement

Most studies measure exposure to fluoride as an ecological factor; that is, they assess people's duration of stay in areas with community water fluoridation or natural fluoride levels in their drinking water sources. Such methods then mathematically calculate a cumulative exposure to fluoride based on water consumption averages in communities by obtaining data from various regional agencies. Bassin, Mittleman, Wypij, Joshipura, and Douglass (2004) compared fluoride exposure measurement to estimate fluoride concentrations of public water systems using secondary data sources (from the state or local level) to those from the 1992 CDC fluoride census. They cautioned researchers to consider limitations of using a secondary data source to estimate fluoride in drinking water, particularly in studies

where exposure to fluoride is the primary exposure of interest. Apart from measurement errors in correctly estimating an individual's exposure to fluoride (especially lifetime exposure), such methods, when analyzed as a person-level exposure, also lead to an ecological fallacy introduced by assuming that fluoride exposure on an individual basis is uniform across the population, and also that the effect of fluoride on every individual is uniform. Alternately, methods that conduct interviews with individuals and also simultaneously assess the fluoride concentrations of drinking water sources are more reliable than ecological exposure assessment. However, as estimates of exposures, these may introduce major measurement errors unless all sources of fluoride exposure are assessed accurately, something that is not possible in retrospective studies. Not only can measurement inaccuracies create difficulties in inference making, but causal inferences from case–control studies may add to problems in conclusions. Control area selection is another controversial issue in some studies assessing fluoride–cancer association. For example, inclusion of cancer mortality from non-fluoridated areas grouped together with fluoridated areas within the past 5 years, and cancer mortality compared between fluoridated areas and the whole United States—including areas with fluoride in the water supplies, are not appropriate control selection policies.

Prospective studies, on the other hand, may be better able to ascertain fluoride exposure by using methods that are common in nutrition epidemiology, such as diet charts and food-frequency questionnaires, to accurately record exposures (Sohn, Noh, & Burt, 2009). Other common measurements of fluoride exposure, such as concentration of fluoride in serum or nail-clippings, may provide good estimates for current fluoride exposure or burden, but would not provide an estimate of the exposure in the past when the case may have originated. "Given the overlap among caries/fluorosis groups in mean fluoride intake and extreme variability in individual fluoride intakes, firmly recommending an 'optimal' fluoride intake is problematic" (Warren, Levy, Broffitt, Cavanaugh, Kanellis, & Weber-Gasparoni, 2009).

Water Fluoridation

The first community program for water fluoridation was instituted in Grand Rapids, Michigan, in 1945. That was followed by water fluoridation in Newburgh, New York (1945) and Evanston, Illinois (1946). Thereafter, several other cities in different countries adopted water fluoridation such as Brantford, Canada (1945); The Netherlands (1953); New Zealand (1954); the United Kingdom (1955); the German Democratic Republic (1959); Indianapolis, Indiana (1951), San Francisco, California (1952), Philadelphia, Pennsylvania (1954), Chicago, Illinois (1956), New York, New York (1965), Dallas, Texas (1966), Detroit, Michigan (1967), Los Angeles, California (1999), Las Vegas, Nevada (2000), Sacramento, California (2000), and San

Antonio, Texas (2002) in the United States. The public water supply in 43 out of the 50 largest cities in the United States is currently fluoridated (Jones, Burt, Petersen, & Lennon, 2005). Other countries having extensive public water fluoridation include: Australia, Brazil, Canada, Chile, Colombia, Hong Kong Special Administrative Region of China, Ireland, Israel, Malaysia, New Zealand, Singapore, and the United Kingdom. Water fluoridation ended in the erstwhile Soviet Union and Eastern Europe with the breakup of the Soviet Union. European countries such as Austria, Belgium, Denmark, France, Germany, Italy, Norway, and Sweden have no water fluoridation.

Early studies conducted in the United States showed that the reduction in dental caries could range from 35 to 60%. Ho Chi Minh City in Vietnam introduced water fluoridation in 1990 after the prevalence of dental caries continued to increase despite the introduction of a school-based dental health program in 1979. Five years of water fluoridation helped to reduce the prevalence of dental caries in 12-year-olds from 84% in 1989 to 78% in 1995, with a mean DMFT of 3.4 in 1990 and 2.7 in 1995 (Quan, 2000). Other studies conducted in communities with fluoridated drinking water in Australia, Britain, Canada, Ireland, New Zealand, and the United States show reductions in dental caries in the range of 15–40% less tooth decay. Systematic reviews provide a more recent update (McDonagh et al., 2000). Murray, Rugg-Gunn, and Jenkins (1991), Newbrun (1989), Ripa (1993), the review of fluorides (PHS, DHHS, 1991), and other reports have summarized the results of early studies conducted worldwide to determine the effectiveness of fluoridation.

> The best available evidence suggests that fluoridation of drinking water supplies does reduce caries prevalence, both as measured by the proportion of children who are caries-free and by the mean change in dmft/DMFT score. The studies were of moderate quality (level B), but of limited quantity. The degree, to which caries is reduced, however, is not clear from the data available. The range of the mean difference in the proportion (%) of caries-free children is –5.0 to 64%, with a median of 14.6% (interquartile range 5.05, 22.1%). The range of mean change in dmft/DMFT score was from 0.5 to 4.4, median 2.25 teeth (interquartile range 1.28, 3.63 teeth). It is estimated that a median of six people need to receive fluoridated water for one extra person to be caries-free (interquartile range of study NNTs 4, 9. The best available evidence from studies following withdrawal of water fluoridation indicates that caries prevalence increases, approaching the level of the low fluoride group. Again, however, the studies were of moderate quality (level B), and limited quantity. The estimates of effect could be biased due to poor adjustment for the effects of potential confounding factors. (McDonagh et al., 2000)

In the United States, an independent, nonfederal Task Force on Community Preventive Services conducting a systematic review of eligible best studies (Group A: before-and-after measurements of caries at the tooth level, in studies with concurrent comparison groups) found that following

starting fluoridation, the median decrease in dental caries experience among children ages 4–17 years during 3 to 12 years of follow-up was 29.1% in studies with concurrent comparison groups (21 study arms) when decay rates were measured before and after water fluoridation. Furthermore, in studies where decay rates were measured after water fluoridation, only a 50.7% decrease occurred during 3 to 12 years of follow up (20 study arms) (Guide to Community Preventive Services, 2002). Following stopping fluoridation, increase in caries during 6–10 years of follow up was 17.9%. The outcome measures in these systematic reviews are percent of persons with caries in primary or permanent dentition, DMFT/dft, or DMFS/dfs indices. Furthermore, the age group studied and year of follow-up also vary from study to study. Nevertheless, the Task Force on Community Preventive Services found *strong* evidence for promoting fluoridation even with the widespread use of other sources of fluoride. The readers should consult the original reports for a better understanding of the methodology used in these systematic reviews.

The difference in caries levels between fluoridated and non-fluoridated communities today in developed countries is lower compared to the studies conducted in the early 1950s. One explanation for this observation is that the decline in dental caries observed only in fluoridated communities in the early 1960s is now seen in non-fluoridated areas as well. Other explanations (Newbrun, 1989; Ripa, 1993) include the increased availability of other forms of fluoride, a lowering of background caries levels, extension of the benefits of water fluoridation through consumption of commercial beverages and foods processed in fluoridated areas, and children who attend schools in fluoridated areas. In the United States and other countries where drinking water fluoridation is practiced, it continues to be an ideal program for fluoride delivery for several reasons. First, the benefits accrue to everyone without active participation because there is no need for an individual to alter behavior. Second, the effect is both systemic and topical, and therefore benefits continue throughout life. Third, the concentration of fluoride in water compared to alternative forms of fluoride delivery is lower and therefore safer. Fourth, the frequency of fluoride exposure is higher when it is present in water and therefore more effective. Finally, the cost-effectiveness is higher compared to other modes of fluoride delivery. The average cost for a community to fluoridate its water is estimated to range from approximately $0.50 a year per person in large communities to approximately $3.00 a year per person in small communities. For most cities, every $1.00 invested in water fluoridation saves $38.00 in dental treatment costs. Table 17.3 demonstrates the relative cost and effectiveness of different mechanisms of fluoride applications.

One principle embodied in public health programs is that small effects reaching large population groups yield dramatic societal benefits. Other potential benefits of water fluoridation are that it has the potential to reduce disparities in dental health. Because children in non-fluoridated communities

TABLE 17.3 Relative Ranking of Fluoride Alternatives by Different Criteria

Fluoride Alternative	*Cost*	*Safety vs Risk*	*Effectiveness*[a]	*Acceptability as a Public Health Alternative*
Public fluoridated water	Low	Low	High	High
Toothpaste	Low	Moderate	High	Moderate
Rinses (school administered)	Low	Moderate	Moderate	High
Supplements	Moderate	High	Moderate	Moderate
Sealants	Moderate	Low	Moderate	Low
Chewing gum (Xylitol)	Moderate	Moderate	Moderate	Moderate
Varnishes	Unknown	Low	Moderate	Low
Salt	Unknown	Not Applicable (N/A)	Moderate	Low
Milk	Unknown	N/A	Moderate	Low

[a]Effectiveness unknown: floss, toothpicks, fluoridated bottled water; effectiveness low: gels.

Adapted from Kimminau, Shepherd, and Starrett, 2000.

receive beverages manufactured in fluoridated communities, they also benefit from direct fluoride exposure.

Dental Fluorosis and Drinking Water Fluoridation

A potential problem with the fluoridation of the public drinking water supply has been the potential of occurrence of dental fluorosis. The University of York Systematic Review suggested that there was some evidence of dental fluorosis in milder forms occurring in fluoridated areas. The significance of such mild forms of dental fluorosis has been debated.

> Dental fluorosis was the most widely and frequently studied of all negative effects. The fluorosis studies were largely cross-sectional designs, with only four before–after designs. Although 88 studies of fluorosis were included, they were of low quality. The mean validity score for fluorosis was only 2.8 out of 8. All, but one, of the studies were of evidence level C. Observer bias may be of particular importance in studies assessing fluorosis. Efforts to control for the effects of potential confounding factors, or reducing potential observer bias were uncommon.
>
> As there may be some debate about the significance of a fluorosis score at the lowest level of each index being used to define a person as "fluorosed," a second method of determining the proportion "fluorosed" was

selected. This method describes the number of children having dental fluorosis that may cause "aesthetic concern."

With both methods of identifying the prevalence of fluorosis, a significant dose–response relationship was identified through a regression analysis. The prevalence of fluorosis at a water fluoride level of 1.0 ppm was estimated to be 48% (95% CI 40 to 57) and for fluorosis of aesthetic concern it was predicted to be 12.5% (95% CI 7.0 to 21.5). A very rough estimate of the number of people who would have to be exposed to water fluoride levels of 1.0 ppm for one additional person to develop fluorosis of any level is 6 (95% CI 4 to 21), when compared with a theoretical low fluoride level of 0.4 ppm. Of these approximately one quarter will have fluorosis of aesthetic concern, but the precision of these rough estimates is low. These estimates only apply to the comparison of 1.0 ppm to 0.4 ppm, and would be different if other levels were compared. (McDonagh et al., 2000)

Design and Analysis Issues

Several methodological challenges not only limit the usefulness of studies related to water fluoridation, but also add to confusion over results and recommendation.

A total of 26 studies of the effect of water fluoridation on dental caries were found. For this objective, the quality of studies found was moderate (no level A studies). A large number of studies were excluded because they were cross-sectional studies and therefore did not meet the inclusion criteria of being evidence level B or above. All but three of the studies included were before–after studies, two included studies used prospective cohort designs, and one used a retrospective cohort design. All before–after studies located by the search were included. The most serious defect of these studies was the lack of appropriate analysis. Many studies did not present an analysis at all, while others only did simple analyses without attempting to control for potentially confounding factors. While some of these studies were conducted in the 1940s and 1950s, prior to the common use of such analyses, studies conducted much later also failed to use methods that were commonplace at the time of the study.

Another defect of many studies was the lack of any measure of variance for the estimates of decay presented. While most studies that presented the proportion of caries-free children contained sufficient data to calculate standard errors, this was not possible for the studies that presented dmft/DMFT scores. Only four of the eight studies using these data provided estimates of variance. (McDonagh et al., 2000)

Fluorides and Other Health Concerns

Several studies have raised important issues concerning association of fluoride (especially from water fluoridation) and hip fracture (e.g., Kurttio, Gustavsson, Vartiainen, & Pekkanen, 1999), osteosarcoma (e.g., Gelberg,

Fitzgerald, Hwang, & Dubrow, 1995), acute fluoride poisoning (e.g., Gessner, Beller, Middaugh, & Whitford, 1994), oral and pharyngeal cancer (e.g., Glattre & Wiese, 1979), and thyroid and other cancers (e.g., Kinlen, 1975; Swanberg, 1953). Most of these studies have demonstrated lack of confounder adjustment (e.g., non-adjustment for sex, age, BMI, calcium intake, non-water fluoride exposure and menopausal status among women in studies of association of hip fracture and water fluoride exposure). Fluoride exposure from public water supply is difficult to ascertain, and studies often assess such exposure ecologically from community water sources (e.g., Kurttio et al., 1999) that may lead to ecological fallacy. Reported statistics from which conclusions are drawn also differ widely. For example, in a fluoride exposure hip fracture study (e.g., Kurttio et al., 1999), most reported odds ratios were close to unity; the reported value of correlation coefficient (0.71) translates to an R-square of 0.5041, implying that only about half the variance of estimated fluoride concentration from geological survey data is associated with actual measured fluoride concentration. Similarly, a systematic review reported that of the 27 studies included in the review about the analysis of water fluoridation and fracture incidence, "10 studies presented crude results only, 12 presented adjusted-effect measures such as relative risks and odds ratios, and five studies presented standardized results. Of these, six studies failed to control for the effect of any possible confounding factors" (McDonagh et al., 2000).

Several studies have tried to examine the association of fluoride exposure to children's intelligence. A recent meta-analysis (Tang, Du, Ma, Jiang, Zhou, 2008) assessed some of these studies and arrived at the conclusion that children living in fluoridated areas in China (only) were five times more likely to have low IQ than those living in non-fluoridated areas. However, the study failed to mention the inclusion and exclusion criteria for the meta-analysis, and how the evidence in each study was judged vis-à-vis exclusion/inclusion criteria. Interestingly, the funnel plot of the meta-analysis clearly demonstrated a potential for publication bias which would mean that smaller and larger studies differed in systematic ways. This issue was, however, not discussed by the authors, and they simply proceeded to make their conclusions ignoring the bias issue. It is likely that relatively loose inclusion and exclusion criteria could have resulted in biases which could potentially invalidate the results. A discussion of meta-analysis and funnel plots is provided in Chapter 4. Several other credibility issues related to association of water fluoridation with IQ and osteosarcoma have been discussed at length by Pollick (2006).

Association between fluorides and several cancers has been published, and causal association between fluorides and cancers are often professed. Analysis of sub-parts of larger study data has given rise to a controversy about association between fluoride and osteosarcoma (Bassin, Wypij, Davis & Mittleman, 2006; Douglass & Joshipura, 2006) as the results from one

sub-analysis (Bassin et al., 2006) was not upheld over complete data analysis (Douglass & Joshiupura, 2006). Animal studies assessing association between high fluoride level in drinking water and cancers were similarly negative (NTP, 1990). Small sample size in studies of osteosarcoma and fluorides are a major deterrent in establishing true associations. The age-adjusted incidence of all types of bone and joint cancers as reported in SEER is 1.0 per 100,000 among men and 0.8 per 100,000 among women with the incidence being slightly higher among Whites than in other racial/ethnic groups; whereas the overall mortality rate is 0.5 per 100,000 (Horner, Ries, Krapcho, Neyman, Aminou et al., 2009). Regional studies obtaining a still smaller sample size do not lend to easy multivariable analysis. For example, the study reported by Bassin et al. (2006) included 103 cases (only 27 had fluoride exposure) and 225 controls (77 had fluoride exposure). Whereas bivariate analysis (t-tests and Chi-square tests) are reported, multivariable adjustments in conditional logistic regression models suggested null results with large values for confidence limit ratios. However, the authors presented sex-specific results showing an effect in men (60 cases; 122 controls) (adjusted OR: 5.43; CI: 1.5, 19.9). The authors did not mention testing for effect measure modification; they chose to present effect estimates for men and women separately without testing for heterogeiety of effects, assuming an effect modification between fluoride and osteosarcoma by sex as they presented sex-specific association assuming a different effect in males and females based on suggestion of possible affect of fluoride in male rates from one study (Bucher, Hejtmancik, Toft, Persing, & Eustis et al., 1991). The conclusion of Bucher et al. (1991) was:

> Although the collective results of this study do not suggest a significant carcinogenic potential for sodium fluoride, in view of the widespread exposure of the population to fluorides from a variety of sources it would appear prudent to re-examine previous animal and human epidemiologic studies, and perform further studies as needed to evaluate more fully any possible association between exposure to fluorides and the occurrence of osteosarcomas of bone.

Some cross-sectional studies assess serum fluoride levels in cancer patients to assess causal association between fluoride exposure and cancers. For example, from a Hospital-based study, Sandhu, Lal, Kundu, and Kharb (2009) have recently suggested that raised fluoride levels in osteosarcoma cases suggest osteoblastic activity in osteosarcoma. They further insinuate a "role" for fluoride in osteosarcoma. Because this study described its analytic approach in one sentence ("The data so obtained was analyzed using one-way ANOVA"), it is difficult to comment on the analytical strategy, especially because of the small sample size and multiple comparisons reported in the study, and because of the complete absence of information about whether they conducted any separate analysis adjusting for other variables.

Also, they failed to provide a description of the sample (for example, age and sex distribution) which makes it difficult to make meaningful conclusions within the framework of the study. However, as in this study, there exists a classical problem that is common when making a causal argument based on cross-sectional studies: That of assuming a temporal sequence between exposure and outcome.

Fluoride has been shown to produce mutations in animal studies at very high concentrations, such as 4.3 ppm (Mihashi & Tsutsui, 1996). However, several types of genetic mutations have been associated with osteosarcoma, such as the RB gene, and genes regulating cell cycle such as P53, cyclins, cyclin dependent kinases, and kinase inhibitors (Kumar, Abbas, Fausto, & Mitchell, 2007), the causes for which are many. In addition, confounding due to various other possible carcinogenic exposures is possible, although not studied (for example, radium).

Control area selection is another controversial issue in some studies assessing fluoride–cancer association. For example:

> Overall, the findings of studies of bone fracture effects showed small variations around the 'no effect' mark. A meta-regression of bone fracture studies also found no association with water fluoridation . . . There is no clear association between water fluoridation and overall cancer incidence and mortality. This was also true for osteosarcoma and bone/joint cancers. Only two studies considered thyroid cancer and neither found a statistically significant association with water fluoridation. Overall, no clear association between water fluoridation and incidence or mortality of bone cancers, thyroid cancer or all cancers was found . . . Interpreting the results of studies of other possible negative effects is very difficult because of the small numbers of studies that met inclusion criteria on each specific outcome, and poor study quality. A major weakness of these studies generally was failure to control for any confounding factors. Overall, the studies examining other possible negative effects provide insufficient evidence on any particular outcome to permit confident conclusions. Further research in these areas needs to be of a much higher quality and should address and use appropriate methods to control for confounding factors. (McDonagh et al., 2000)

In general, it is considered that there exists no firm evidence of association of water fluoridation with adverse health effects of a serious nature. Overall, the evidence for causal association between fluorides and bone fractures and cancers is not very strong, considering methodological problems with these studies. Better designed studies that are able to measure fluoride exposure accurately and those that are designed to infer causality should be able to clarify the association between fluoride exposures and these adverse health outcomes. Even if a causal association between fluorides from water fluoridation and cancer is assumed, from policy perspective, the attributable fraction needs to be considered, which leads to a

situation demanding the need to assess direct costs of fluoridation vs benefits (i.e., how many cases would be saved if water fluoridation was stopped vs benefits of water fluoridation in terms of dental caries reduction at optimal concentration of fluoride). Considering the rarity of osteosarcoma, the attributable fraction would perhaps be exceedingly small and benefits of optimal water fluoridation are known to be high.

Water Fluoridation Policy

In the United States, the main policy for ensuring widespread fluoride exposure is through water fluoridation, which has been categorized as one of the major public health achievements of the 20th century (Burt, Keels, & Heller, 2003). A recent article asked the question: "Is water fluoridation still necessary?" and concluded that:

> At present, fluoridation remains the best tool to combat caries in many countries. Another way to consider the question is to ask, what evidence is there to show fluoridation to be unnecessary in the countries where it is widely practiced? An alternative strategy for preventing dental caries across all social strata in the population has not emerged, while the costs of treatment have not declined . . . . Measuring the impact of interventions to control dental caries is difficult, because it is characterized by a complex interaction of multiple risk factors. Documenting the impact of fluoridation is even more challenging, because the immediate impact is not apparent. Therefore, research should continue to assess its impact and to determine the appropriate level of fluoride in water to balance the benefits of fluoride against the risks of enamel fluorosis in any one country. Similarly, surveillance and research activities should continue to assess the effect of total fluoride exposure. Promising new approaches to eliminate dental caries as a public health problem should be pursued. (Kumar, 2008)

An often-cited longitudinal study conducted by the National Institutes of Dental and Craniofacial Research found that the prevalence of fluorosis in seven areas where the water supply was optimally fluoridated did not increase over a 10-year period (1980–1990), suggesting that fluorosis increases are caused by the availability of fluoride in sources other than drinking water. Burt, Keels, and Heller (2003) used an 11-month break in water fluoridation (September 1990–October 1991) in Durham, North Carolina, to identify the time when developing incisors were most sensitive to development of fluorosis. They examined 1896 children in different age cohorts because "children are thought to be at particular risk of fluorosis if fluoride is ingested during the 'critical period,' the developmental time around late secretion/early maturation at which the unerupted tooth appears to be especially sensitive to fluoride examined." They did not find any caries effects, a finding they attributed to "continuing exposure to fluoride from other sources as well as to generally low levels of caries." The study was uncertain about any

fluorosis-reduction effect. They considered several issues such as age bias effect of children, examiner error, and unusual fluoride exposure effects. However, they concluded that "the break was not long enough to lead to a reduction of fluorosis prevalence, and that the higher prevalence levels seen in children born in 1985–1987 are the result of some unexplained exposure to excess fluoride in their infant years."

Under the Safe Drinking Water Act, the U.S. Environmental Protection Agency (EPA) is required to establish, review, and maintain exposure standards for contaminants in the public drinking water systems that might cause any adverse effects on human health. Included in these standards are the Maximum Contaminant Level Goal (MCLG), the Maximum Contaminant Level (MCL), and the Secondary Maximum Contaminant Level (SMCL). The MCLG is defined as a concentration at which no adverse effects are expected to occur and the margins of safety are judged adequate. Whereas the MCLG is a health goal, MCL is the enforceable standard that is usually set as close to the MCLG as possible, allowing for factors such as treatment technology and costs. The EPA establishes an SMCL for contaminants, which is a guideline for managing drinking water for aesthetic, cosmetic, or technical effects (U.S. Environmental Protection Agency, 2001). These levels are reviewed periodically by EPA in the wake of new scientific evidence.

The National Research Council Committee on Fluoride in Drinking Water of the National Academy of Sciences, USA (2006) reviewed research on various health effects from exposure to fluoride, including studies conducted in the last 10 years to assess the protective role of the EPA's drinking water standard for fluoride—currently, a maximum of 4 milligrams of fluoride per liter of water (4 mg/L). The report concluded that the drinking water fluoride at 4 mg/L does not protect against adverse health effects, but such or higher exposure to fluoride over a lifetime is likely to increase the risk for bone fractures. The committee was unambiguous about its charge and the role of fluoride at the stated concentration of 4mg/L, *which is almost four times the concentration found in drinking water*, should it be interpreted as evidence that the current fluoride level in drinking water leads to any increased risk of fluorosis:

> Many public health agencies and experts endorse adding fluoride to the water as an effective method of preventing tooth decay in communities where natural fluoride levels are low. The "optimal" concentration range of fluoride in drinking water for preventing tooth decay was set at a range of 0.7 to 1.2 mg/L more than 40 years ago by the U.S. Public Health Service. In 2000, it was estimated that approximately 162 million people had artificially fluoridated water. The recommended range for artificial fluoridation is below the EPA standards and was designed for a different purpose, so it is important to note that the safety and effectiveness of the practice of water fluoridation was outside the scope of this report and is not evaluated. This report only evaluates EPA's standards. (EPA, 2001)

The report recommended further research such as exposure assessment at the individual level rather than the community level; population studies of moderate and severe enamel fluorosis in relation to tooth decay and to psychological, behavioral, or social effects; studies designed to clarify the relationship between fluoride ingestion, fluoride concentration in bone, and clinical symptoms of skeletal fluorosis; and more studies of bone fracture rates in people exposed to high concentrations of fluoride in drinking water. Boxes 17.2 and 17.3 enumerate the recommendations by the CDC for using fluoride to prevent and control dental caries in the United States and recommendations by the International Health Agencies, respectively.

BOX 17.2 Recommendations for Using Fluoride to Prevent and Control Dental Caries in the United States

Public Health and Clinical Practice

1. Continue and extend fluoridation of community drinking water.
2. Counsel parents and caregivers regarding use of fluoride toothpaste by young children, especially those <2 years of age.
3. Target mouth-rinsing to persons at high risk.
4. Judiciously prescribe fluoride supplements.
5. Apply high-concentration fluoride products to persons at high risk for dental caries.

Self-Care

1. Know the fluoride concentration in the primary source of drinking water.
2. Use small amounts of fluoride frequently.
3. Supervise use of fluoride toothpaste among children <6 years of age.
4. Consider additional measures for persons at high risk for dental caries.
5. Use an alternative source of water for children under 8 years of age whose primary drinking water contains >2 ppm fluoride.

Consumer Product Industries and Health Agencies

1. Specify the fluoride concentration of bottled water on the bottle label.
2. Promote use of small amounts of fluoride toothpaste by children.
3. Develop a low-fluoride toothpaste for children.
4. Collaborate to educate healthcare professionals and the public.

Further Research

1. Continue metabolic studies of fluoride.
2. Identify biomarkers of fluoride.
3. Reevaluate the method of determining the optimal fluoride concentration of community drinking water.

(Continues)

BOX 17.2 Recommendations for Using Fluoride to Prevent and Control Dental Caries in the United States *(Continued)*

4. Evaluate the effect of fluoride mouth rinse, fluoride supplements, and other modes of delivering fluoride on dental caries.
5. Study the current cost-effectiveness of fluoride modalities.
6. Conduct descriptive and analytical epidemiologic studies.
7. Identify effective strategies to promote adoption of recommendations for using fluoride.

Centers for Disease Control and Prevention. MMWR Recomm Rep 2001; 50(RR-14):1–42.

BOX 17.3 Fluoride Recommendation by International Health Agencies

- The WHO Oral Health Program continues to emphasize that everyone should be encouraged to brush daily with fluoride toothpaste. Where possible, municipal water supply reaching a large population and water fluoridation using fluoride at a concentration of 0.5–1 mg/L should be a method of choice (Petersen & Lennon, 2004).
- WHO, Pan American Health Organization (PAHO), and other international health agencies recommend the introduction of salt fluoridation where water fluoridation cannot be implemented or where water fluoridation cannot be further pursued for philosophical or political reasons. It has been suggested that countries with very high levels of fluoride ingestion, and where risk of severe dental fluorosis or of skeletal fluorosis are high, should maintain a maximum fluoride level of 1.5 mg/L as recommended by WHO Water Quality Guidelines (Petersen & Lennon, 2004; World Health Organization, 2003).
- Furthermore, where sugar consumption is high or increasing, the caries-preventive effects of fluorides need to be enhanced. WHO recommends that every effort must be made to develop affordable fluoride toothpastes for use in developing countries (Petersen & Lennon, 2004).

Other Sources of Fluoride

Fluoride Supplementation

Historically, children in non-fluoridated areas have been advised to take fluoride supplements as an alternative to water fluoridation. Public health programs have incorporated daily fluoride use in congregated settings like schools. Compliance with the prescribed regimen is required to realize max-

TABLE 17.4 ADA Council on Scientific Affairs Recommendations on Fluoride Dosage Schedule

	Fluoride ION Level in Drinking Water (ppm)[a]		
Age	*<0.3 ppm*	*0.3–0.6 ppm*	*>0.6 ppm*
Birth–6 months	None	None	None
6 months–3 years	0.25 mg/day[b]	None	None
3–6 years	0.50 mg/day	0.25 mg/day	None
6–16 years	1.0 mg/day	0.50 mg/day	None

[a]1.0 ppm = 1 mg/liter.

[b]2.2 mg sodium fluoride contains 1 mg fluoride ion.

ADA Council on Access, Prevention, and Interprofessional Relations, 1995.

imum benefits. Many investigators have reported that when children reported taking tablets/drops on an everyday basis, they were at higher risk for dental fluorosis either because the fluoride dose was high or because the peak level exceeds the threshold (National Research Council [NRC], 1993). Therefore, the fluoride supplement schedule has been adjusted downward in many countries. The recommendations by the ADA Council on Scientific Affairs are shown in Table 17.4. As shown in the table, one needs to know the actual level of fluoride in the drinking water. It may require testing for fluoride, especially for those who use water from several sources and independent wells. This can be a formidable challenge in both developed and developing countries. Fluoride supplementation is not recommended in areas where the fluoride in drinking water exceeds 0.6 ppm.

Fluoride Toothpastes

The daily use of commercial toothpastes containing fluoride has been recommended for all persons because it increases the level of fluoride in saliva and plaque. A large number of clinical trials show caries reductions. However, not all fluorides are bioavailable. Only those dentifrices with the ADA Seal of Acceptance should be recommended because it ensures the bioavailibility of fluorides. Several studies show that fluoride dentifrices may contribute to a higher intake of fluoride and subsequent development of dental fluorosis (NRC, 1993; USDHHS, 1991). According to Whitford, the amount of fluoride introduced into the mouth with each brushing ranges from 0.1 to 2.0 mg and the amount swallowed varies from 10 to 100% depending on the age of the child (Whitford, 1996). If these levels of fluoride are ingested in addition to fluoridated water or use of fluoride supplements during the formation of teeth, dental fluorosis can be expected to occur. Therefore, it is recommended that young children use only a "pea-size" (amount or a "smear

layer") of toothpaste and that they expectorate after brushing. Several countries market low-fluoride toothpastes. The trials that tested 250 ppm fluoride vs 1000 ppm fluoride showed conflicting results. Findings from 500 to 550 ppm fluoride toothpastes suggested that these may be as efficacious as 1000 ppm fluoride toothpastes.

Self-Applied Fluoride Rinses or Gels

Fluoride products containing 0.05% neutral sodium fluoride or 0.4% stannous fluoride are available over the counter for home use. Other products containing higher amounts of fluoride are also available as prescription items. These products are indicated for adolescents and adults who are at moderate-to-high risk for caries. Patients who are likely to benefit from these rinses or gels are individuals undergoing orthodontic treatment, adults with exposed root surfaces, cancer patients undergoing head and neck radiation, and patients who are predisposed to decreased salivary flow such as persons taking certain medications. The use of fluoride gels or rinses for children under age 6 should be strongly discouraged because of their inability to control the swallowing reflex (Ripa, 1991).

Fluoride Mouth Rinses

School-based fluoride mouth-rinse programs have been promoted since the late 1970s as an effective measure to control dental caries. This program is targeted toward elementary school children in low-income and non-fluoridated areas as studies in fluoridated areas have shown it to less cost-effective. In general, children rinse with either a 5 ml or 10 ml solution of 0.2% neutral sodium fluoride solution for 1 minute once a week in the classroom. Some Scandinavian countries have used fortnightly rinses (Ripa, 1991).

Professionally Applied Fluoride Products

Fluoride products containing high concentrations of fluoride are available for use in dental offices. These products contain either 2% or 1.1% sodium fluoride, 1.23% acidulated phosphate fluoride (APF), or 8% stannous fluoride. Varnishes containing 5% sodium fluoride have also been approved for desensitizing teeth. APF is the most commonly used topical fluoride in dental offices. Although prior prophylaxis is not required, generally it is applied during recall visits after a prophylaxis. A ribbon of fluoride gel placed in trays is applied to teeth for 4 minutes following drying of the teeth. It has been estimated that most trays hold 5 gms of gel in each section. A substantial amount will be retained if suction devices are not used and patients do not expectorate (LeCompte, 1987; Ripa, 1991).

Considering the potential for exposure to multiple fluoride sources and the lower effectiveness of high concentration of fluoride, the following recommendations have been made for the topical use of fluorides in dental offices:

1. Target topical fluorides to children at risk for caries. Caries-free children in fluoridated areas are not likely to benefit from it.
2. Apply not more than 2 gms of gel per tray or approximately 40% of the tray capacity.
3. Use a saliva ejector during the 4-minute application procedure and have the patient tilt his or her head forward to minimize swallowing.
4. Wipe the teeth to remove excess gel.
5. Instruct patients to expectorate thoroughly for 1–2 minutes following treatment.

Neutral sodium fluoride gels are also available as an alternative to APF to avoid etching of porcelain and composite restorations. The use of pastes containing fluoride in performing prophylaxis is also a common practice. These pastes are not needed when a topical fluoride is also applied. Further, it should be pointed out that neither the practice of using fluoride-containing paste as a substitute for topical application nor the 1-minute application of 1.23% APF is supported by research (Ripa, 1991). The 1-minute application of 1.23% APF is based on laboratory studies that show 50–60% uptake of fluoride within the first minute. The actual caries-inhibitory effect in patients who receive a 1-minute application of 1.23% APF is not known.

Salt Fluoridation

Success in using salt fluoridation has been reported in Colombia, Costa Rica, Jamaica, the Canton of Vaud in Switzerland, and some parts of France and Germany. Jamaica is the only country where virtually all salt destined for human consumption on the island has been fluoridated since 1987 (Estuprnan-Day, Baez, Horowitz, Warpeha, Southerland et al., 2001; Warpeha, Beltran-Aguilar, & Baez, 2001). Fluoride concentrations in salt used around the world range from 90 to 350 mg/kg, although more recent studies have suggested an optimal concentration of around 250 mg/kg (Marthaler, 1983).

Milk Fluoridation

As an alternative to water fluoridation, several other vehicles have been tried to deliver fluorides to the oral environment. All fluoridated milk programs reported until now have been for children. It is conceivable that elderly people in assisted care centers and nursing homes can be a good target for milk fluoridation schemes for control of root caries. Bulgaria is a world leader in milk fluoridation.

Other Fluoride-Containing Products

Fluoride varnishes that contain 22,600 ppm of fluoride have been available in Europe and Canada for more than 15 years as a caries preventive agent. In the United States, these products have been approved for use only recently for desensitizing teeth. Many fluoride-releasing dental restorative materials are available in the market. The release of fluoride from these materials offers the benefits of elevated fluoride levels in the dentin, enamel, plaque, and saliva. This property has the potential to reduce recurrent caries; however, the consideration of this property is secondary in the selection of a restorative material.

Hiiri, Ahovuo-Saloranta, Nordblad, and Mäkelä (2006) compared the occlusal dental caries preventive effect of pit and fissure sealants and fluoride varnishes in children and adolescents using a systematic review. They could not conduct a meta-analysis to compute an overall effect estimate because of clinical and methodological diversity between study designs which did not allow their combination.

> Four studies were eligible for inclusion in the review. Three of the four studies compared the effectiveness of sealants with fluoride varnish application, and one study compared the effectiveness of sealants and fluoride varnish combination with fluoride varnish alone. Results of two studies revealed the effectiveness of pit and fissure sealants to be statistically significantly higher than an application of fluoride varnish every 6 months in preventing occlusal decays of first molars at 23 months (RR 0.74, 95% CI 0.58 to 0.95) and at 9 years follow up (RR 0.48, 95% CI 0.29 to 0.79). One of these studies was classed as at low risk of bias, one of moderate to high risk. One small study at moderate to high risk of bias failed to find a statistically significant difference between sealants and fluoride varnishes. One study of low risk of bias found a statistically significant difference in favor of the sealants and fluoride varnish combination compared with merely fluoride varnish at 24 months follow up with RR 0.36 (95% CI 0.21 to 0.61). The age of children in the included studies was 5 to 9 years. Allocation concealment was classified adequate in two of these four studies. (Hiiri et al., 2006)

Therefore the authors found that for occlusal caries prevention, there existed some evidence that pit and fissure sealants were somewhat superior to fluoride varnish application. This evidence is based on a very small number of studies, and whether this translates to benefits in terms of effectiveness in real-world public health programs is unknown. Well-designed trials should be able to address this issue, although difference in application methods and other local logistic issues may continue to dictate the use of both two methods as alternative choices in different circumstances.

However, taking the real-world population level effectiveness research forward, Ahovuo-Saloranta, Hiiri, Nordblad, Mäkelä, and Worthington (2008) addressed the question: Are pit and fissure sealants effective dental

caries preventive methods in children and adolescents in the population? They conducted systematic review of randomized or quasi-randomized controlled trials of at least 12 months in duration, comparing sealants with no sealant or sealants, and performed a meta-analysis using random-effects model to derive an overall effect estimate (risk ratio).

> Sixteen studies were included in the review; 7 studies provided data for comparison of sealant versus control without sealant and 10 studies for comparison of sealant versus sealant. Five split-mouth studies and one parallel group study with 5 to 10 year old children found a significant difference in favor of second or third generation resin-based sealants on first permanent molars, compared to a control without sealant, with a pooled RR of 0.13 (95% confidence interval (CI) 0.09 to 0.20), 0.22 (95% CI 0.15 to 0.34), 0.30 (95% CI 0.22 to 0.40), and 0.40 (95% CI 0.31 to 0.51) at 12, 24, 36 and 48–54 months follow up, respectively. Further, one of those studies with 9 years of follow up found significantly more caries in the control group compared to resin sealant group; 27% of sealed surfaces were decayed compared to 77% of surfaces without sealant. The results of the studies comparing different sealant materials were conflicting. (Ahovuo-Saloranta et al., 2008)

There seems to be some evidence of the effectiveness of using pit and fissure sealants in preventing dental caries, and it appears to be a recommendable procedure for preventing dental caries in occlusal surfaces of permanent molars. These studies were conducted in areas where well-equipped dental clinics and adequate services were available. Although one may ask the question: Whereas pit and fissure sealants seem to be effective in high risk groups, is this benefit also translated to other lower risk groups, or to settings where dental services are sparse? The number of studies included in the previously mentioned review was small, so there seems to be a case to recommend larger studies in different risk and clinical service availability settings, including poor settings, to determine the effectiveness of pit and fissure sealants and their benefits among the highest risk group (which is usually associated with poverty and low socioeconomic class). A discussion of the role of poverty and other sociocultural factors in determining oral health is presented in Chapter 18.

SECTION

III

Oral Epidemiology in Sociobiological Context

18 Social Epidemiology

Social epidemiology is a relatively newer branch of epidemiology that assesses how social conditions of individuals and populations impact their health; that is, it assesses the social distribution and determinants of health and disease. Social epidemiology examines "both specific features of, and pathways by which, societal conditions affect health" (Krieger, 2001a). A central difficulty in social epidemiology is to explain how socioeconomic conditions are causally linked to health status and outcomes. Health outcomes are predicated upon individual experiences of the biological disease phenomenon and the ability to access a system of peers, support groups, and professionals to address the biological disease. Therefore, even if the fundamental character of a disease process may rest entirely within the biological compartment of the body, its progress and outcome may be a function of the interaction of biological phenomenon and the sociopolitical–cultural environment within which the individual finds oneself. Social capital is defined as the "resources accessed by individuals and groups within a social structure that facilitate cooperation, collective action, and the maintenance of norms" (Fujiwara & Kawachi, 2008). It refers to social relationships and connections (social networks) between individuals that may include neighborhoods, cultural groups, peer-network groups, social support groups, or any association that has the ability to impact an individual's interaction with the (health) system in a meaningful way.

Theories of Social Epidemiology

There are three main theories used in social epidemiology: (1) psychosocial theory views health outcomes as functions of host–agent–environment interactions (i.e., impact of social environment to host resistance to disease);

(2) social production of disease and/or political economy of health views a social continuum with some people benefitting from policies at the cost of others (i.e., it assumes prioritization of capital accumulation over human needs and assesses the impact on health, thereby mainly emphasizing inequalities of health); and (3) ecosocial theory and the related multilevel frameworks views ecological and social factors as intricately intertwined in ways so that they alter health states, with each influencing the other as an ever-dynamic, constantly changing phenomenon. This approach involves nested hierarchies (i.e., humans are nested in population and ecology in spatiotemporal dimensions) as dynamic states assessed employing mathematical modeling to understand the individual/unique phenomenon in relation to external general processes (Krieger, 2001b).

Social epidemiologists have placed a lot of emphasis on theory and its development. In general, a central tenet in social epidemiology is that it is the theory that drives the understanding, explanation, solution development, and intervention implementation about the questions being asked. These theories are therefore derived from different perspectives that may not necessarily be comprehensive or mutually exclusive. These perspectives emphasize certain central tenets in proposing their theses around which their health care world view revolves. Four main perspectives have been identified from which social epidemiology theories originate: the material perspective, the cultural/behavioral perspective, the psychosocial perspective, and the life-course perspective (Sisson, 2007).

The Materialist Perspective (or Explanation)

The materialist perspective (also called the materialist explanation) takes the view that the mere possession of assets such as wealth and education do not impact health inequalities, but access to tangible resources such as food, shelter, fundamental amenities, and services impact health inequalities. This perspective, therefore, focuses on health-impacting factors that are beyond an individual's direct control. Sanders, Slade, Turrell, Spencer, and Marcenes (2006) correctly point out that [oral] "health is responsive to absolute levels of material resource. Not only does income permit access to timely and comprehensive health care, it also provides opportunities for a whole constellation of choices that affect health." However, from the standpoint of robustness of one's financial status (that will permit and allocate expenditure under different line items in an individual's budget), utilization of income may be a function of the robustness of total wealth that an individual possesses and various financial confidence-related factors prevailing at that time. This perspective distinguishes between mere current income and the total financial capacity (wealth) of an individual, thereby recognizing that the two constructs, income and wealth, are not necessarily directly associated. Similarly, however, education opens many doors overall, but higher

education does not necessarily translate to greater health awareness, health literacy, or utilization of preventive measures. Although material resources such as wealth and education are important determinants for access to health-related resources, these are not directly responsible for health inequalities. Development of health policies are modified by perspectives. For example, a perspective that views wealth as a direct proximal factor in determining health status might provide monies in the hand of an individual to be used for health care. On the other hand, a perspective that believes that wealth has an indirect, more distal effect may instead emphasize development of healthcare resources and provide incentives for an individual to utilize those services instead of putting money directly in the hand of the individual for healthcare-related spending. Development of epidemiological variables for assessing causal or predictive relationships between different social factors and their position in relation to proximity to outcomes will therefore be slightly different, depending upon the theoretical basis applied to construction of the variable (e.g., current/annual income vs assessment of total household wealth).

The Cultural/Behavioral Perspective

The cultural/behavioral perspective takes the view that people's health-related choices and attitudes are predicated upon their sociocultural backgrounds. For example, persons from a lower socioeconomic background may make poorer health behavior choices leading to their poorer health. Whereas cultural background impacts one's health-related practices, in the era of increased globalization, cross-cultural influences and adaptations are common. Easy access to electronic media in the form of radio, television, and the Internet has placed a large amount of available health-related information into people's hands. The mere existence of traditional cultural features that are conducive to good health need not necessarily be retained over cultural evolutionary pressures, as many cultural anthropologists state anecdotally. Behaviors acquired over a period of time are not easily transferred to the next generation. Therefore, policies and interventions based on a cultural/behavioral perspective may either have a shorter time horizon over which to demonstrate changes or should include mechanisms that will ascertain integration of intervention-related practices into the cultural pattern of the social group in question.

The Psychosocial Perspective

The psychosocial perspective suggests the view that psychological stresses develop as a function of the socioeconomic position of people, and these stresses are fundamentally responsible for health inequalities. According to this perspective, people from lower socioeconomic backgrounds face

greater challenges in life and experience a greater number of negative life events that directly elevate their psychological stresses, leading to poorer health status and health outcomes. The negative life events leading to elevated psychosocial stresses include less social support, less control at work, less job security, and living in communities with lower levels of trust and higher levels of crime and antisocial behavior compared to individuals from higher socioeconomic groups (Sisson, 2007).

> There are two mechanisms through which stress could influence health: the direct and indirect models (Elstad, 1998). The etiological basis of the direct model postulates that stress leads to the development of ill health by triggering a specific chain of events that leads to the development of specific diseases, or by having a general negative effect on the body, reducing resilience and increasing vulnerability to disease (Kelly, Hertzman, & Daniels, 1997). The indirect model proposes that people experiencing higher levels of psychosocial stress are more likely to make behavioral or lifestyle choices that are damaging to health (Elstad, 1998). (Sisson, 2007)

The Life-Course Perspective

The life-course perspective encompasses the materialist, psychosocial, and cultural/behavioral perspectives. It views health status as a product of present and past living conditions. Therefore, each event in the past through the present, over an individual's life course, impacts the current health status. For example, if a person lived a life of abject poverty till he or she was 20 years old, and then suddenly discovered unlimited wealth, then such a person may incur bad health in several ways. First, the poverty phase of the life would have been a situation of deprivation leading to poor immune status, perhaps exposure to several disease conditions prevailing in poor surroundings where he or she lived, and nutritional deficiencies, poor oral hygiene, poor education, extreme physical and psychological hardships, and so on. Thereafter, at 20 years of age, the exposure to sudden wealth may have propelled the person to spend on high-fat and sugary foods, excessive alcohol, tobacco, and other expensive health menaces. Therefore, the health status of this person at 40 years of age would be impacted not just by his or her wealthy current living conditions (and perhaps poor choices due to lack of education, awareness, and so on), but also the fundamental health compromises acquired in the early stages of life dictated by the then-prevailing extreme poverty. Influences of exposure to events in the life-course perspective may be either a one-time event, continuous (described by a straight or curved line in fixed, incremental cumulative or decreasing doses—the *accumulation model*), or episodic (in multiple different time points, few or more in frequency, or dependent on certain thresholds of life conditions—the *critical periods/latent effects model*; the effects of such exposures to life conditions may or may not be cumulative). Life-course epidemiology is defined as "the

study of long-term effects on chronic disease risk of physical and social exposures during gestation, childhood, adolescence, young adulthood, and later adult life to understand causal links between exposures and outcomes taking into consideration the importance of time (duration) and timing in the disease development" (Kuh & Ben-Shlomo, 2004).

> This suggests that exposures in the beginning of life play a role in initiating disease processes before the disease manifests as overt pathology. The specific period in the life stage when an exposure occurs, known as the timing effect, may be also important in understanding its later effects on the aetiology of chronic disease. A life-course approach to studying chronic disease aetiology is not merely a collection of longitudinal data or the use of a particular study design or analytical method. Rather, the unique feature of this approach is a theoretical framework which assumes and tests a temporal ordering of exposure variables and their interrelationship with a specific outcome. (Nicolau, Thomson, Steele, & Allison, 2007)

Poverty

Defining poverty has been an ongoing activity, perhaps since the origins of civilization. Poverty is defined and measured in different ways by different entities depending upon their objective of defining poverty. Whereas the commonest, colloquial image of poverty is lack of money, the World Bank uses several wide-ranging concepts to define poverty as a multidimensional construct. Therefore, poverty is any or a combination of the following: hunger, lack of shelter, being sick and not being able to see a doctor, not having access to school, not knowing how to read, not having a job, fear for the future, living one day at a time, losing a child to illness brought about by unclean water, powerlessness, and lack of representation and freedom—any of these or a combination of these. Poverty is a function of community development and may be defined at the global, country, community, family, or individual level (World Bank, 2007).

Poverty may be described in absolute or relative terms. *Absolute poverty* compares the standards against a set reference standard that may be consistent across all countries; for example, the percent of the population of a country that is able to eat food to provide the amount of minimum calories required to live. In contrast, *relative poverty* is defined in reference to a certain artificially defined threshold (which changes over time or due to changes in policies practiced by political authorities, societal norms, and values); for example, income below a certain fixed threshold, a fixed level of consumption, or expenditure. Therefore, if a person earns above the designated threshold, the person is not poor. Such a threshold is usually set by governments and is called the poverty line. Although comparison across time using a fixed poverty line is possible, the relative value of money (i.e., its purchasing power) changes and people may be defined as living "above"

the poverty line, but their real living standards may not be much different from those designated to live at or below the poverty line. Criteria for defining the poverty line are always controversial, and open to political machinations. Information on consumption and income is obtained through sample surveys, with which households are asked to answer detailed questions on their spending habits and sources of income. Such surveys are conducted more or less regularly in most countries. These sample survey data-collection methods are increasingly being complemented by participatory methods, where people are asked what their basic needs are and what poverty means for them. Interestingly, new research shows a high degree of concordance between poverty lines based on objective and subjective assessments of needs. Description of income and associated social disparities utilize relative poverty terms when they compare differences between income groups, social groups, or other groups defined on common shared characteristics.

Measuring Poverty

Because poverty is a multidimensional construct, its measurement is complex. The commonest form of measuring poverty is by using the income factor alone. Poverty measured in this way is often called "income-poverty." Another common method used to measure poverty is based on consumption levels. A person is considered poor if his or her consumption or income level falls below some minimum level necessary to meet basic needs. When estimating poverty worldwide, the same reference poverty line should be used and expressed in a common unit across countries to allow comparability. Therefore, for the purpose of global aggregation and comparison, the World Bank uses reference lines set at $1.00 and $2.00 per day (more precisely $1.08 and $2.15 in 1993 Purchasing Power Parity terms). It has been estimated that in 2001, some 1.1 billion people had consumption levels below $1.00 a day and 2.7 billion lived on less than $2.00 a day. These figures are lower than earlier estimates, indicating that some progress has taken place in alleviating poverty, but poverty still remains too high in terms of human suffering (World Bank, 2007).

At an individual/family level, poverty is traditionally measured as based on annual income or annual household income. The U.S. government defines a poverty line annually for different households depending on the number of persons in the household. A generalized form of the poverty line, the Federal Poverty Level (FPL), is also published. The FPL is used as a guideline to develop arbitrary categorical levels such as 100%, 200%, and 300% of FPL, and so on. By themselves, these categories have no causal significance, but they are easy to construct and allow fixed frames of reference across different studies and across time spans. NHANES reports income under different heads such as categorized annual household and another

constructed variable—Poverty Income Ratio (PIR). For NHANES, NCHS suggests that PIR is "the best income variable to use when comparing data over time because it is 'relatively' standardized for inflation and other factors. However, the method of calculation has been changed slightly over time. The primary reporting categories are 0.000–0.999 (below poverty) and 1.000 and above (at or above poverty)" (National Center for Health Statistics [NCHS], 1996). In oral health research, assessment of poverty is often done by other methods such as participation in USDA food assistance program (Special Supplemental Nutrition Program for Women, Infants, and Children [WIC]), Food Stamp Program, and School Lunch and Breakfast Programs).

New Directions in Poverty Measurement

While much progress has been made in measuring and analyzing income poverty, other dimensions of poverty have not been examined and measured adequately. Currently, comparable and high-quality social indicators for education, health, and access to services and infrastructure are often used. Newer indicators are needed to track other dimensions of poverty—for example, risk, vulnerability, social isolation, and development of and access to social capital. Aggregate indices that integrate various dimensions of poverty would be useful. Still other indicators at the level of each dimension are needed when it may not make sense to aggregate the various dimensions into one index. This will increase the numbers of poverty indicators, social support indicators, social exclusion indicators, and barriers to development and freedom. The poor are generally viewed as a homogeneous group of people united by their lack of resources (Sen & Dreze, 1999). However, "the poor should not be regarded as an undifferentiated mass, but that one should rather identify particular groups which have been struck by a catastrophic imbalance between needs and resources. Small-holders, farm laborers, tenant farmers and herdsmen may well all be poor, but the ways in which they are affected by famine can differ greatly" (Erikson, 1998).

Development and Freedom

Social development implies a reduction in poverty, unemployment, and inequality. Development is a process of expanding human freedom, suggesting the expansion of freedom to be the primary end as well as the principal means of development. Development includes a constitutive role (involves substantive freedom [e.g., related to starvation, morbidity, mortality] to enrich life). An instrumental role in development (involves emancipation through rights, opportunities, and entitlements) (Sen, 1999). Freedom has multiple dimensions assessed through types of freedom as defined by Amartya Sen (1999). Freedoms are of the following types: political (to determine who governs), economic (opportunities for use of economic resources), social opportunities (e.g., involves education, health), transparency guaran-

tees (transparent working system and ethics), and protective security (provision of social safety net). Poor people are usually deprived of each of these freedom types (Sen & Dreze, 1998).

Poor people may exist in an informal economy. The best-known economic effect of the informalization process is to reduce the costs of labor substantially. Existence in an informal economy leaves poor people out of any real or potential official measures (such as enforceable healthcare requirements or measures), which could be a cornerstone of primary prevention strategies. Poor people in such situations get into a double jeopardy: loss of official healthcare benefits that may be provided, and loss of personal capacity to purchase health care due to very low income levels. Poor people may live a marginalized existence with a blocked path for a hopeful future for themselves or their children. Social marginalization perpetuates their vicious cycle of poverty. They may have virtually no control over their physical, economic, social, or emotional environment, leading to ghettoization and a trapped existence. Viewed strictly from a financial angle, the overall span of their productive period is small, and even if some of them entertain plans to break the system, their financial viability remains in question because they have no accessible opportunities (Castells & Portes, 1989).

Race and Ethnicity

Traditionally, race has been defined based on skin color and external phenotypical expressions apparent in physical features of human beings. Social ostracizing and racial prejudice has been a hallmark of human history across the world. Perhaps, as in recent history, the end of apartheid in South Africa followed by substantial political changes marked an era of renewed interest in reassessing race in an effort to understand its importance or irrelevance in human disease. In the recent past, there was a "near consensus that racial categories were too poorly defined, too historically tainted, and too tied up with all manner of social prejudice to play much of any positive role in etiologic biomedical research" (Kauffman & Cooper, 2008). However, with the advent of advanced techniques in molecular biology and understanding from the Human Genome Project, the debate about the biological basis of the definition of race has resurfaced. For example, using the HapMap data, Barreiro et al. suggested a role of natural selection in modern human differentiation (Barreiro, Laval, Quach, Patin, & Quintana-Murci, 2008) by showing that in several genes, positive selection increased differentiation within gene regions to result in local adaptation of human populations. Genes for which this phenomenon was observed included those for skin pigmentation and hair development, immune response to pathogens, DNA repair and replication, sensory functions, and metabolic pathways, among others. However, they found that negative selection reduced population differentiation at the level of amino acid change, particu-

larly in disease-related genes (Barreiro et al., 2008). It has been suggested that "this work will serve as a stepping stone for further research in identifying candidate genes for disease, many of which may be tied to ethnicity" (Norrgard & Schultz, 2008). However, the scope that biological advancement has is to forever remove the idea of racial superiority and inferiority and instill an idea of variability in human populations as an acceptable norm devoid of prejudice. How do we reach this goal?

In epidemiological research, the variable of race may be used to classify people based on their nationality or geographical location that may have environmental, economical, political, cultural, or other shared exposures that may be linked to disease occurrence. Such classifications may help in disease prediction and program planning to implement interventions. Variation in genotype prevalence may occur to the extent that existence of a certain genotype may occur in one population and another population may be completely devoid of it (e.g., CCR5-Δ32 and APOE-ε4 alleles). However, in order to be truly compatible with a "racial" theory, genotype prevalence must exactly or near exactly match with racial groupings and must also be substantially expressed differently in other racial groups. Current understanding points away from such possibilities due to the huge variations among humans being discovered by the HuGE project: "The molecular techniques that are newly available to epidemiologists provide a much greater specificity of exposure assessment, but do not rescue a paradigm from its logical failings" (Kaufman & Cooper, 2008).

Swan et al. (2006) assessed the ways in which race is classified in cancer epidemiology literature. They found the following four common ways: (1) grouping race data according to the "any mention" method—respondents are categorized in each of the race groups by their self-report; (2) individuals are included in a group only if they report a single race—the U.S. 2000 Census used this method; (3) ethnicity is treated as race—therefore, people reporting Hispanic/Latino origin are tabulated as Latino and no other racial categorization is done; and (4) grouping a combination of race and Hispanic/Latino ethnicity, but classifies individuals in the group with which they most identify (Latino is used as a race category). Although not reported by Swan et al., use of a person's family name or surname is also used to classify by race, especially in death certificates. Therefore, respondents who identify themselves as Latino and not as American Indian/Alaskan Natives (AIAN) are classified as Latino. Thus, those AIANs who would otherwise have been counted in that group are no longer counted as AIANs by this definition.

Misclassification of race is a common problem in epidemiological studies. For example, terms such as *Latino* or *Hispanic* are arbitrary terms that do not represent any specific cultural, biological, or other grouping criteria. The entire continent of South America (most Latinos will have origins in South America) is a multiethnic, multicultural, and multiracial aggregation of people, and coalescing them into one group is arbitrary. Perhaps such catego-

rization may have political utility, but is not very useful in epidemiological research. Misclassification of race among AIANs has been a matter of concern and debate. For example, many AIANs, especially those residing in California and the Southwest, have Spanish surnames, often a historical remnant of ancestors who were slaves of missions, and they are frequently counted as Latino when a surname is used as a proxy for race/ethnicity. Also, AIAN race may not be recorded on medical records (e.g., at hospitals or health clinics), possibly as a result of incorrect assignment by clerks or because there is no option for recording American Indian or Alaskan Native on intake forms. In addition, self-identifications may change when a tribe formerly "unrecognized" becomes federally recognized by Congress or if tribal enrollment ordinances change. Furthermore, errors are made on birth certificates or death certificates, both of which can be used to obtain population data (Swan et al., 2006).

It has been suggested that the American Indian and Alaskan Native race is underreported on death certificates and other health-related data sets in Washington State and elsewhere. Misclassification of race/ethnicity occurs in health-related databases most often because the recorded information is based on observation by physicians, coroners or medical examiners, or other healthcare workers rather than on patient self-reports or reports by close relatives. Errors and biases on death certificates in Washington State persist. Methods to reduce misclassification can improve data quality and enhance efforts to measure and reduce racial/ethnic health disparities (Stehr-Green, Bettles, & Dee Robertson, 2002).

Another "racial/ethnic" category used commonly in the United States is "Asian/Pacific Islanders," which combines people from origins in China, Mongolia, Japan, the Indian subcontinent, South East Asia, South Asia, and the far east countries such as Java, Sumatra, and Papua New Guinea—clearly a huge and extremely diverse population. In the era of great globalization efforts, more suitable, and perhaps more representative and sensitive classification for race/ethnicity needs to be developed to serve epidemiological needs better and obtain more valid results from studies conducted globally.

Oral Health Disparities

Disparity is the occurrence of inequality or differences between two or more groups. It is a dispassionate term that merely records the existence of a situation. However, political orientation, social awareness, and the call for higher citizenry imparts passion, direction, and meaning to the term, leading to scientific activism that necessitates action beyond mere recording of existing differences. The National Institutes of Health defines health disparities as the "differences in the incidence, prevalence, mortality, and burden of diseases and other adverse health conditions that exist among specific population groups in the United States" (NIH, 2002). The pro bono nature of health

sciences dictates finding solutions to resolve disparities, establish health justice, and aspire for health for all with equal opportunity to access the health system in an effort to improve upon current health status aiming at greater goals for health status for populations. For example, the *Healthy People 2010* has two major overall goals: (1) to increase the span of healthy life; and (2) to eliminate health disparities across categories of gender, race or ethnicity, education or income, disability, geographic location, and sexual orientation (ODPHP, USDHHS, 2009).

Another example that the field of health disparities has come to mean more than mere documentation of existence of disparities is that while the USDHHS (2000) documented that disparities have been observed in oral health outcomes by gender, race or ethnicity, education, income, disability, geographic location, and sexual orientation, and that despite improvements in some oral health status indicators, the burden of disease is not evenly distributed across all segments of societies; the National Institute for Dental and Craniofacial Research (NIDCR) responded with setting up a Health Disparities Research Program with "a plan to eliminate craniofacial, oral and dental health disparities" (NIDCR, 2007).

NIDCR identified the following underserved and disadvantaged population subgroups that fit within the research activities supported by the Health Disparities Research Program to eliminate/ reduce disparities (NIDCR, 2007):

- All race/ethnic populations with health disparities including African American, Hispanic (Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish culture or origin regardless of race), American Indian or Alaskan Native, Asian or Pacific Islanders
- Low income rural (e.g., Appalachian) or urban dwellers
- Special-needs populations (e.g., physically or mentally disabled)
- People living with HIV/AIDS
- The elderly
- Homebound and institutionalized individuals

Health disparities may occur due to a variety of reasons such as poverty, unequal access to health care, lower educational attainment, racism, social networks, physical location (neighborhoods), incarceration, and social stigma. Some of these issues are products of system-wide factors, whereas others may be more due to group or personal attitudes. Health disparities are commonly measured as within- and between-group differences, and of the groups or sub-groups selected as a reference. Measurement of "total disparity" involves evaluating the distribution of health among all individuals in a population without regard to their social group membership. Measurement of "disproportionality" implies a "disproportionate share" or

an "unequal burden" of health status in groups. Disproportionality implies that it is unfair that some groups experience more ill health than others, thereby adding a social justice angle to disparity measurement. Another measure of health disparity is "social-group disparity," which measures differences between individuals from social groups and does not distinguish differences between social groups (e.g., racial, ethnic). Social-group disparity is the basis for one of the overarching goals in *Healthy People 2010* stated earlier (ODPHP, USDHHS, 2009).

Existing Disparities in Oral Health

Oral health disparities brought to attention by the first ever report on oral health by the Surgeon General of the United States in the year 2000 focused on the profound oral health disparities existing in this country (USDHHS, 2000). It has been shown that even after accessing the system, there occur substantial differences in clinical conditions, awareness of treatment options, treatment discussions, treatment recommendations, and treatment received by different socio-economic-ethnic groups (Kressin, 2005). Oral health-related disparities may be assessed as disparities in oral health status by race, ethnicity, income, gender, neighborhood characteristics, and age; oral health service access; and oral health service provision, and were reviewed recently (Chattopadhyay, 2008). Box 18.1 outlines some of the salient points related to existing oral health disparities.

BOX 18.1 Summary of Existing Oral Health Disparities

Existing Disparities in Oral Health Status

- **Dental Caries:** Childhood caries highly prevalent among racial/ethnic minorities, and poor, rural, immigrant groups. Sequelae of dental caries also follows similar pattern. Lack of oral health is compounded by lack of access to health care.
- **Tooth Loss:** Race and socioeconomic disadvantage are strong determinants of tooth loss. African Americans are more likely to receive tooth extractions.
- **Periodontal Disease:** Persons in lower socioeconomic position and minorities have more periodontal disease, more advanced periodontal disease, and faster progression of disease. African Americans with at least an annual dental check-up had fourfold higher odds of established periodontitis compared to Whites (OR: 3.64, CI: 1.43, 9.24). Complex periodontal

BOX 18.1 Summary of Existing Oral Health Disparities *(Continued)*

treatment needs among persons in lower socioeconomic class are much greater compared to those in higher socioeconomic classes.

- **Oral Cancer:** Incidence rates among African American and white women are similar; mortality and survival rates for African American women are consistently lower than white women. Incidence rates among African American men are substantially higher than white counterparts. These rates have started to decline among African American men. Mortality and survival rates among African American men are substantially lower compared to white counterparts. Incidence, mortality, and survival among all men are worse than among women.
- **Orofacial Pain:** Lower socioeconomic position is associated with greater self-reported dental pain.
- **Persons with Special Care Needs:** Persons with special needs have poorer oral health, poorer access to healthcare system, lesser insurance coverage, and greater need for social and nursing support compared to those who do not have special care needs.
- **Elderly Persons:** Dental caries burden among non-institutionalized elderly African Americans are high. Poorer elderly persons with special needs and those living in rural areas have greater unmet needs compared to other elderly Americans.

Disparities in Oral Health Service Access

- **Service Use:** Service use of poorer persons and minorities is lower compared to other groups. Percentage of Americans visiting a dentist has remained constant for more than a decade. Odds of orthodontic treatment (OR = 8.7) is greater among those with a dental visit in the past year compared to those without a visit. African Americans and those with unknown race are less likely to receive root canal therapy compared to Whites.
- **Geographic and Neighborhood Issues:** Oral health status of persons living in rural areas is lower than those living in urban areas. Compared to urban dwellers, rural dwellers have fewer teeth, fewer annual dental visit reports, and greater unmet dental needs. Similar patterns exist for those living in poorer neighborhoods compared to well-to-do neighborhoods.
- **Provision of Services:** Experienced practitioners show more variation in periodontal decision making; clinicians linked to a training center share common treatment philosophy compared to those outside; Medicaid enrollees find it difficult to locate a dentist and receive dental care.

Adapted from Chattopadhyay, 2008.

Inequalities between groups as evidenced by the association of periodontal disease to race, ethnicity, education, and income continue to be pervasive in the United States over the years. For example, Borrell and Crawford (2008) reported that the prevalence of periodontitis was 3.6%, with black people (7.2%) exhibiting significantly higher prevalence than Mexican Americans (4.4%) and white people (3.0%). After adjusting for selected sociodemographic characteristics, black adults, those with less than a high school education, and those with low income were 1.94 (95% CI 1.46–2.58), 2.06 (95% CI 1.47–2.89), and 1.89 (95% CI 1.18–3.04) times more likely to have periodontitis than white people, those with more than a high school diploma, and those with high income, respectively.

Despite overall decreases in oral cancer incidence and mortality rates, the disparities associated with socially disadvantaged groups have continued to exist. From 1975 through 2004, age-adjusted incidence rates, age-adjusted mortality rates, and survival rates from oral and pharyngeal cancers are higher for men than for women, being highest for black men. Five-year relative survival rates for patients diagnosed during the period 1995 to 2001 were higher for whites than for blacks and lowest for black males (Morse & Kerr, 2006). Interestingly, the survival rates of oral cancer have remained fairly stable over the decades. Similar results have been reported from Kentucky (Miller, Henry, & Rayens 2003) as well as from Florida (Tomar, Loree, & Logan, 2004), North Carolina (Elter, Patton, & Strauss, 2005), and elsewhere. Table 18.1 shows the stages of oral and pharyngeal cancer at presentation from the SEER data in 5-year groups every 10 years. Whereas the overall trends are similar for blacks and whites, the greater proportion of earlier stage presentation among whites over blacks stands out. Although the proportion of regional spread presentations increased for all, the proportion of increase is greater for white men compared with black men (8% vs 6%, respectively) from the period of 1985–1989 to the period of 1996–2000.

Generally, oral cancer-related disparities in the population, particularly in racial/ethnic groups and historically underserved low-socioeconomic and rural populations, reflect at least three controlling factors: (1) decreased access to care due to limited availability of medical practitioners within many rural communities and/or lack of available financial resources for the care; (2) genetic contribution to increased risk, particularly in racial/ethnic minorities; and (3) behavioral attitudes regarding the expectations, needs, and importance of oral health care (Chattopadhyay, 2008). The Surgeon General's Report clearly enunciated the importance of oral health to general health and emphasized a similar lack of oral health improvement in subsets of the U.S. population, including racial/ethnic minorities and historically underserved populations (e.g., rural and socioeconomically disadvantaged communities) (USDHHS, 2000).

Oral cancer disparities between whites and African Americans are well documented. A SEER (1973–2002) data-based study found that African

TABLE 18.1 Stage Distribution by Race and Gender for Oral Cavity and Pharynx Cancer, All Ages

	White			*Black*		
	1975–1979	*1985–1989*	*1996–2003*	*1975–1979*	*1985–1989*	*1996–2003*
Percent Localized	41%	40%	36%	26%	18%	20%
Percent Regional	37%	43%	50%	49%	56%	60%
Percent Distant	12%	9%	9%	17%	19%	16%
Percent Unstaged	10%	8%	5%	8%	7%	4%
	White Men			*Black Men*		
	1975–1979	*1985–1989*	*1996–2003*	*1975–1979*	*1985–1989*	*1996–2003*
Percent Localized	41%	39%	34%	25%	15%	16%
Percent Regional	37%	44%	52%	51%	57%	63%
Percent Distant	12%	9%	9%	17%	22%	17%
Percent Unstaged	10%	8%	4%	7%	6%	4%
	White Women			*Black Women*		
	1975–1979	*1985–1989*	*1996–2003*	*1975–1979*	*1985–1989*	*1996–2003*
Percent Localized	40%	41%	42%	30%	26%	29%
Percent Regional	39%	42%	43%	46%	53%	54%
Percent Distant	11%	8%	8%	16%	13%	12%
Percent Unstaged	9%	9%	6%	8%	8%	5%

Source: Horner, M.J., Ries, L.A.G., Krapcho, M., Neyman, N., Aminou, R. Howlader, N., et al., (eds). (2009). *SEER cancer statistics review 1975–2006.* Bethesda, MD: National Cancer Institute. Retrieved August 12, 2009, from http://seer.cancer.gov/csr/1975_2006

Americans had a significantly higher proportion of cancer, mainly in the tongue, that had spread to a regional node or to a distant site at diagnosis than whites; a significantly higher proportion of tongue cancer that were >4 cm in diameter at time of diagnosis. Black men in particular experienced lower 5-year relative survival rates than white men for tongue cancer (Shiboski, Schmidt, & Jordan, 2007). One possible explanation for the lower survival among blacks may be a difference in access to, and utilization of, healthcare services associated with unemployment and lack of insurance.

Costs associated with oral cancer tend to be high, and therefore oral cancer-related disparities between income, insurance, and employment groups are expected. However, if determined efforts are to be launched to eliminate such disparities, then accurate data related to the costs of oral cancer treatment are needed. Although there have been a number of innovations for the treatment of squamous cell carcinoma of head and neck in the past 5 years, the lack of economic data complicates the task of evaluating the effectiveness of these new interventions. In this time of mounting concerns over healthcare costs, more emphasis on economic data is clearly warranted (Menzin, Lines, & Manning, 2007). Head and neck cancer patients incur greater costs at all levels during their treatment. In a study, Lang et al. (2004) found that the mean rates (%) for a variety of resources used by head and neck cancer patients were higher compared to controls without head and neck cancers: rate of hospitalization (82% vs 55%), mean number of hospitalizations (2.5 vs 1.4), mean number of inpatient days (24 vs 12), rate of skilled nursing care use (22% vs 13%), mean number of days of skilled nursing care (9 vs 5), rate of home healthcare use (48% vs 26%), and rate of hospice care use (14% vs 3%), all respectively. Overall, substantial variation in direct and indirect costs occur by treatment modality. The most cost-saving scenario is one that prevents head and neck cancers altogether (Menzin et al., 2007). Studies assessing cost–benefit issues related to oral cancers (and cancers in general) are very few, and an understanding of the role of economic factors is just beginning to occur (Nadler, Eckert, & Neumann, 2006).

Predictors of Oral Health Disparities

Life-course events of individuals since their childhood, especially their childhood socioeconomic position, predicts oral health disparities well (Thomson et al., 2004). This study found that those in a lower socioeconomic position at 5 years of age had greater caries experience by the time they reached 26 years of age even after controlling for childhood oral health. These persons were more likely to have experienced tooth loss and periodontitis.

Barriers to better oral health that exhibit early in the life of a person, if they persist, may continue to act in similarly predictable ways. This may

lead to a cumulative increase in disease burden and the ability to utilize the system and evoke a response from it. Responsiveness to new dental symptoms may play a significant role in initiating and/or maintaining oral health disparities. For example, in one study, persons belonging to low socioeconomic strata of the society were found to be more likely to report dental symptoms and were less likely to take steps to address these (Gilbert, Duncan, & Shelton, 2003). A recent study demonstrated that there was a discrete threshold of income below which oral health deteriorated, suggesting that the benefit to oral health of material resources occurs mostly at the lower end of the full socioeconomic distribution (Sanders et al., 2006). Neighborhood characteristics may impact oral health disparity prediction; in such assessments, specific area-based measures of socioeconomic status are valuable in documenting these inequalities and may be more meaningful than composite area-based indices of socioeconomic status (Armfield, 2007).

Steps to Eliminate Oral Health Disparities

Oral health disparities result from the complex interaction of multiple factors. Some suggestions to tackle oral health disparities at different levels include developing networks comprising community- and faith-based organizations, local and national government health institutions, clinical service providers, researchers, and immigrant-service and advocacy organizations; direct involvement of community representatives in community-based participatory research; increasing minority-oriented research activities; increasing workforce diversity; a military-style oral healthcare system; and modifying clinical timings to accommodate work hours and other time commitments of those with poorer oral health status. A common suggestion is to increase the number of underrepresented minority dentists in the workforce under the assumptions that such professionals will work in areas with greater proportions of people from their cultural/ethnic/racial backgrounds and that people prefer to be treated by providers from the same race/ethnicity background as themselves. A randomized trial, however, found that most people did not have a preference on the race/ethnicity of their provider (Bender, 2007).

Eliminating disparities in oral health requires enhanced efforts at preventing disease, promoting health, and delivering appropriate care. It also requires a thorough understanding of the lower use of already available effective preventive and treatment services, and additional interventions to address the identified causes. Boxes 18.2, 18.3, and 18.4 outline the guiding principles on developing oral health interventions; that is, the local and national upstream actions to promote oral health and the summary of steps different states are taking to address oral health disparities.

BOX 18.2 Guiding Principles on Developing Oral Health Interventions

Interventions should have the following properties and characteristics:

- **Empowerment of Communities:** Interventions should enable individuals and communities to exert more control over the personal, socioeconomic, and environmental factors affecting their oral health.
- **Participatory Approach:** Key stakeholders should be encouraged to be actively involved in all stages of planning, implementing, and evaluating interventions and programs.
- **Holistic:** Interventions should adopt a broad approach focusing upon the common risks and conditions that determine oral and general health.
- **Intersectoral Integration:** Partnerships working across all relevant agencies and sectors are essential to ensure that oral health improvement is placed upon the wider public health agenda and should be actively involved.
- **Equity Establishment:** The need to focus action on addressing oral health inequalities should be of paramount importance in the planning of interventions and programs.
- **Evidence-Based Approach:** Existing knowledge of effectiveness and good practice should be the basis for developing future oral health improvement interventions and programs.
- **Sustainable Programming:** Achieving long-term improvements in oral health that can be maintained by individuals and communities is crucial.
- **Multistrategy Approach:** Tackling the underlying determinants of oral health requires a combination of complementary actions such as healthy public policies, community development, and environmental change.
- **Appropriate Program Evaluation:** Sufficient resources and appropriate methods should be directed toward the evaluation and monitoring of oral health interventions and programs.

Adapted from Watt, 2007.

BOX 18.3 Examples of Local and National Upstream Actions to Promote Oral Health

Local Level

- Encourage schools to become part of the Health Promoting Schools Network.
- Develop oral health and nutrition policies in preschools and nurseries.
- Encourage sales of subsidized toothbrushes and toothpastes through community clinics.

BOX 18.3 Examples of Local and National Upstream Actions to Promote Oral Health *(Continued)*

- Encourage nurseries and schools to provide subsidies on healthy snacks and drinks.
- Encourage the engagement of community action groups in oral health projects.
- Support development of local infant feeding policies and ensure oral health messages are included.
- Encourage development of oral health policies in older people's residential homes and care centers.

National Level

- Support regulation on content and timing of television advertisements promoting children's foods and drinks.
- Encourage tighter legislation on food labeling and food claims on products affecting health and oral health.
- Encourage greater availability of sugar-free pediatric medicines.
- Support removal of Value Added Tax (VAT) and other taxes on fluoride toothpastes and toothbrushes.
- Support legislation on community water fluoridation.
- Support food and nutrient standards for school meals and other foods and drinks sold in schools.
- Encourage safety standards for school play areas and other leisure facilities for injury protection.
- Support legislation on wearing of seat belts, helmets, and mouth guards.

Adapted from Watt, 2007.

BOX 18.4 Steps Being Taken by States to Assess and Address Oral Health Disparities

- Initiating studies to identify and document existing disparities in the population.
- Developing goals, and plan a roadmap to address oral health disparities.
- Identifying current infrastructural capacity and plan for future expanded goals to address oral health disparities.
- Initiating oral health surveillance systems to monitor the oral health condition of the population and also to monitor disparities.

(Continues)

BOX 18.4 Steps Being Taken by States to Assess and Address Oral Health Disparities *(Continued)*

- Expanding and maintaining oral disease preventive services through programs such as water fluoridation, dental sealants, health education, and school-based programs through fixed and mobile clinics.
- Developing oral health prevention programs for adults and children in remote areas through various centering initiatives.
- Developing tobacco habit prevention programs.
- Finding alternatives for improving access to oral health care through increasing program, service, and insurance coverage.
- Maintaining regular oral health surveillance system. Using existing self-report surveys and news surveys where necessary.

Adapted from Chattopadhyay, 2008.

Developing Measurement Instruments and Indices

Developing an index for measuring abstract ideas has been an important activity in social epidemiology, and oral health-related quality of life measures are good examples of such efforts. To develop an instrument for measurement or an index, a series of essential steps include:

- Developing the theoretical framework for the abstract idea underlying the index
- Developing a conceptual model on which the index will be based
- Identifying dimensions and potential variables/factors of value that may contribute to the index
- Testing and finalizing the model and components
- Formulating the index and developing a scoring method for using the index
- Fine-tuning and finalizing the index and scoring
- Evaluating the index, testing (validity, reliability)

Each potentially important and available factor may be tried sequentially for fit into a single working model. These factors may be grouped into conceptual domains and subdomains, and may be summed up in some ways using suitable weights to form an index. Depending upon their primacy in the chain of sequence or position in a causal chain of events, the individual factors may be fundamental components or effectors factors that may be grouped into functionally similar subdomains (see Figure 18.1). In

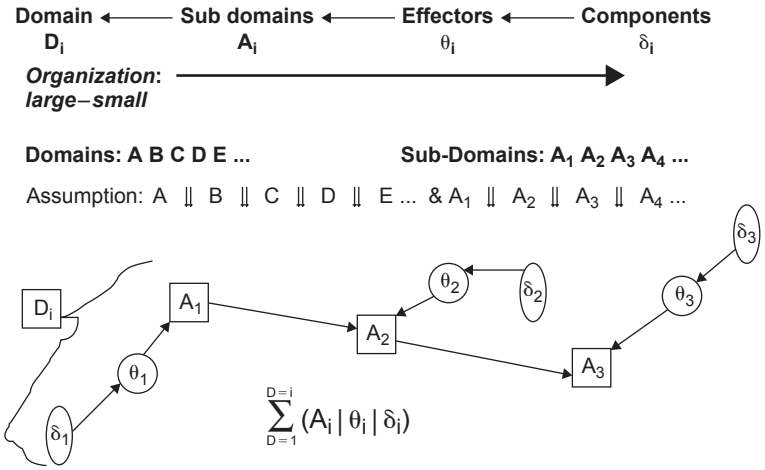


FIGURE 18.1 Conceptual diagram for developing an index. Smaller individual factors are grouped into larger sequential groups. The factors and organized groups should be independent of each other. These grouped factors are finally summed up in some manner to lead to a final calculated statistic that functions as an index (sometimes simply stated as a total raw score or a weighted score of summed component scores across domains).

conceptualizing indices, the primary assumption is that the subdomains are independent of each other, and the domains are independent of each other. Such models must consider causal and confounding issues (Dawid, 2002; Greenland & Brumback, 2002; Hernan, Hernandez-Diaz, Werler, & Mitchell, 2002), key measurement issues involved in quantifying an abstract idea (Falqueto, Lima, Borges, & Barreto, 2004) and tying these pieces into a consistent measurable whole (Slade, 2002). Once these indices and measuring instruments are developed, they need to be carefully tested for validity, consistency, and reliability.

Factor Analysis and Principal Component Analysis

In order to bring these variables together to contribute to form a measurement instrument or an index, data sets need to be reduced to manageable levels by removing redundancies and specifying structures (relationships between domains, subdomains, effectors, and components) clearly. Factor analysis (FA) and principal component analyses (PCA) are two different but related commonly used methods to achieve these goals. If the variances of the errors in FA can be assumed to be the same for all errors, then the two methods become the same (PCA is a type of FA). However, because this assumption is not always true, the use of FA and PCA as interchangeable terms should be avoided. PCA is essentially used for converting multiple

correlated variables into a smaller number of uncorrelated variables and minimizing variability in data. This also helps to support the independence assumption.

FA describes variability among observed variables that are modeled as linear combinations of the unobserved factors (and error terms) to derive interdependencies among factors that help in reducing the data set. The two main applications of FA are to reduce the data by reducing the number of variables and to detect structure in the relationships between variables to help classify them. Therefore two or more variables may be combined into a single factor. *Confirmatory* FA helps to test specific hypotheses about structure of factors in a sample, whereas *correspondence* FA helps to analyze two-way and multi-way associations between variables. FA can be used for several purposes, such as:

- Reducing a large number of variables to a smaller number of factors for efficient modeling
- Identifying clusters and correlated variables
- Improving statistical efficiency by using fewer tests over larger multiple testing scenarios
- Validating an index by assessing load of its constituting variables on factors

Propensity Score

The use of propensity scores in social epidemiology studies, instrument design, and index development is increasing compared to earlier times. Studies may be subject to biases; the use of propensity scores to reduce selection bias was mentioned briefly in Chapter 5. Propensity score, first introduced by Rosenbaum and Rubin (1983), is a conditional probability measured as the probability that a subject (or factor or variable) is assigned to a particular condition in a study given a set of known background information or covariates. For example, it is the probability of being assigned to the exposure group vs control group in a study. Overall, as groups of subjects have similar propensity scores, they are expected to have similar values of all covariates. Propensity scores reduce dimensionality of variables and may be understood as summarization all of the covariate information about exposure group selection and conversion from vector to a scalar quantity. Comparison groups can be matched on their propensity scores. This method uses the counterfactual argument framework discussed in Chapter 3. Therefore, groups can be matched on the probability of being exposed even if in reality the group was not exposed. "This is what randomization does, except it (typically) forces all subjects to have a true propensity of exposure equal to 0.05 and works on unmeasured confounders too" (Oakes & Johnson, 2006).

Propensity scores can be estimated using logistic regression, and they may be used for matching, stratification, or as a covariate in a regression

analysis with careful attention to principles of parsimony, regression diagnostics, residual analysis and multicollinearity among variables. However, propensity score methods should not be viewed as a panacea to matching, bias, and confounding problems. They do not account for unobserved factors; their application situations other than binary situations are in early stages of development; misspecification can lead to erroneous inferences, especially if it does not take into account the reasons for which the factors were allocated to their treatment groups; missing data on important variables may lead to missing propensity scores; and decisions of handling missing data may impact propensity scores (Oakes & Johnson, 2006).

Samples matched on propensity scores also need statistical methods for matched samples. Appraising the use of propensity score in medical literature, Austin (2008) pointed out that very few studies used the method appropriately for matching and suggested five guidelines for design, analysis, and reporting of studies employing propensity score matching:

1. Strategy for creating matched pairs should be explicitly stated and justified.
2. Sampling mechanism (with or without replacement) should be explicitly described.
3. Baseline characteristics of the groups should be explicitly described.
4. Differences in distributions of baseline characteristics of groups should be appropriately assessed and described.
5. Appropriate analytical methods for matched data should be used.

Oral Health-Related Quality of Life

Because of its complex nature, no standard definition of quality of life (QoL) exists, although some have been attempted. One way to define QoL is to attribute a degree to which a person enjoys the important possibilities of life. Quality of one's life is a direct function of the state of one's health. However, measuring QoL is a difficult process because of the nature of the phenomenon of health. A definition of health is difficult. The WHO has defined health as not merely the absence of disease and infirmity but a state of complete social physical and mental well-being. However, the problem in quantifying health or QoL arises because these ideas are abstract, multidimensional, complex, ambiguously defined, subjective, dynamic, in a state of flux, and interpretable according to a population's contemporary contextual sociocultural–political–economic attributes. Furthermore, an individual's own perceptions societal norms change over time. Several studies have demonstrated that oral health status impacts a person's QoL. Apart from the usual difficulties in measuring QoL, measuring the impact of oral health-related factors on QoL has the added difficulty of gleaning out only the effects of oral health-related factors from the plethora of factors that impact QoL. Such oral factors and their impact on QoL are together identified

under the paradigm of oral health-related QoL. QoL is affected by oral health status because it impacts a person's chewing ability, speech, looks, taste, self-esteem, self-image, confidence, and social interaction.

Overall there are two aspects of oral health influencing QoL measures: (1) quality of one's general life that is compromised due to poor oral health, and (2) oral health-related quality of life (OHRQoL) issues. Such impacts and assessments are, to an extent, functions of cultural attributes, societal norms, and the psychological orientation of individuals. Such perceptions of status and health are real regardless of whether they are subject to psychological impacts. Sanders et al. (2006) suggest that "missing teeth, while self-reported, is still an objective condition as are household income and family composition." The same is true for any health state.

Several instruments have been developed to measure OHRQoL, as outlined in Table 18.2. Naito et al. (2006) reviewed the literature associated with QoL and OHRQoL to assess the association between the two and reported that some studies found a significant association. They further reported that reduced health-related QoL was found among patients with temporomandibular disorder. Similarly, people with craniomandibular and cervical spinal pain and poor oral health had substantial QoL impairment. Other oral factors that contributed to impaired QoL included dissatisfaction with teeth and mouth, xerostomia, and edentulousness.

OHRQoL measures have been used in national surveys in the United Kingdom and Australia, although their use in the United States has been limited to some surveys asking a single question on this area (Rozier & Pahel, 2008). The National Oral Health Surveillance System does not report any measures of OHRQoL issues either. However, an assessment of OHRQoL in large epidemiological surveys indicates that poorer scores of OHRQoL are associated with fewer teeth, diseased teeth, untreated disease, unmet dental needs, occasional and episodic visit to seek dental care, and non-white racial/ethnic characteristic (Slade, 2002). Currently new OHRQoL instruments are being developed and targeted to children.

OHRQoL instruments can be an effective tool in program planning and outcome evaluation. Planning is a systematic approach to defining the problem, setting priorities, developing specific goals and objectives, and determining alternative strategies and a method of implementation. Program evaluation can be effected by asking program-critical questions such as: Has the objective been attained (effectiveness)? How much has the attainment of the objective cost us (efficiency)? How does the cost compare to anticipated costs? Has priority been given to the most useful strategy for the attainment of the objectives, and is the strategy acceptable (appropriateness)? Has the program addressed the overall health problem, or was it directed at only a part of it (adequacy)? Did the program equitably address the needs of all segments of the population? OHRQoL may also form an important part of outcome evaluation (i.e., a systematic way to assess the extent to which a program, service, or initiative has achieved its intended results). The pur-

TABLE 18.2 Outline of Commonly Used OHRQoL Instruments

Abbreviation	*Dimensions Measured*	*Number of Questions*	*Domains*
DHI	Pain, worry, conversation	3	3
GOHAI	Chewing, eating, social contacts, appearance, pain, worry, self-consciousness	12	3
DIP	Appearance, eating, speech, confidence, happiness, social life, relationships	25	5
OHIP-49 (Variants: OHIP-14; -20; -G; -G5; -G21, COHIP)	Function, pain, physical disability, psychological disability, social disability, handicap	49	7
SOHSI	Chewing, speaking, symptoms, eating, communication, social relations	42	8
OHQoLI	Oral health, nutrition, self-related oral health, overall quality of life	15	6
DIDL	Comfort, appearance, pain, daily activities, eating	36	5
OHRQoL	Daily activities, social activities, conversation	3	3
OIDP (Variant: Child-OIDP)	Performance in eating, speaking, oral hygiene, sleeping, appearance, emotion	10	3
OHQoL-UK®	Eating, appearance, speech, general health, breath, comfort/relaxation, sleep, confidence, worry, mood, personality, social life, romantic relationships, smiling, work, finance	16	3

pose of an outcome-based evaluation is to obtain information about the effectiveness of the particular service activities, strategies, or efforts in order to support planning and improvement of services. Outcome evaluation provides several benefits such as tracking progress, making decisions and/or improving the quality of initiatives and efforts, creating accountability, and marketing successful efforts.

However, application of these and other existing instruments have been attempted in epidemiological studies, but not in regular clinical practice. "Virtually no work has been done to investigate applications of OHRQoL in assessing the standard of clinical care for patients" (Rozier & Pahel, 2008). Such applications, when they start being attempted, have the potential for improving quality of care, patient satisfaction, quality of research, and public health practice.

19 Bioethics in Oral Health

Medical ethics, or bioethics, is the study of moral issues in the fields of medical treatment and research. It is also used to describe ethical issues in the life sciences and the distribution of scarce medical resources (McGee & Caplan, 2007). Bioethics has grown as a discipline over a decade and today bioethicists play important roles in clinical decision making, research studies, legislature, professional organizations, and community activities; and in academia, government, nongovernmental organizations, and industry. Ethics is a branch of philosophy that deals with leading a good life and guiding right/moral conduct in life. Applied ethics is a discipline of philosophy that attempts to apply ethical theory to real-life situations. In clinical settings, one can understand applied ethics by assessing the nature of activity involved. For example, medical ethics deals with interactions between patients and providers; research ethics deals with the conduct of research studies, analytical methodology, drawing of inferences and publication of results; and public health ethics deals with relations between institutions and populations. In reality, the distinction between medical, research, and public health ethics becomes blurred, especially in epidemiological activities. Bioethics deals with the application of ethical principles to clinical care and involves moral, legal, political, biomedical research, and life sciences technologies (Viens & Singer, 2008).

Ethical principles practiced in one region (such as a country) may or may not be applicable in another region depending upon social norms, religious orientation, political philosophy, and cultural attributes of the region. Cultural norms are intricately linked to the understanding of truth telling, decision making (competency and capacity), and end-of-life care within different ethnic groups (Berger, 1998). Increasing globalization has also impressed upon the scientific community the need to develop a strong interface between ethics and clinical practice and research. The need to find a common platform for ethical conduct has become very important with

globalized research collaborations. Global ethics deals with the ethical and moral questions that arise because of the globalization phenomenon.

Several developments have contributed to the growth and importance of bioethics in daily medical, research, and public health practice. These developments include the philosophical issues stemming from the extension of life, the need to define the start and end of life, the point of death and handling of "brain-dead" persons, the need for extensive end-of-life care, assessing the role of higher citizenship in the developing world, redefining the role and need for a placebo in clinical trials, the emphasis on Quality of Life (QoL), and the notion that QoL is an important outcome that should be pursued. The political and social development of nation states and their increasing democratization has positively influenced human awareness of the sense of equality and justice that increases the scope and need for ethical conduct in clinical practice and research across the world.

Landmark events in the development of ethical principles to guide medical and public health practice and health research include the Nuremberg code, Belmont Report, Title 45 protection of human subjects of the Code of Federal Regulations, World Medical Association Declaration of Helsinki, and Ethical Issues in Epidemiological/Experimental Research and Clinical Trial Issues.

Principles of Bioethics

The general principles on which the practice of bioethics is based include autonomy of individuals and respect for others; beneficence, meaning doing good; non-malfeasance, meaning doing no harm; justice; trust in relationships; veracity; fidelity; avoidance of killing; gratitude; and reparation. Ethical issues are complex and are not always conducive to dichotomous good/bad and right/wrong types of outcomes. These principles should not be viewed as regimented individual principles working in absence of other principles, but rather as that final ethical actions are usually outcomes of interactions of several ethical principles.

- **Autonomy:** The principle of autonomy addresses the rights of individuals and emphasizes respect for others. It deals with rights to privacy, freedom of choice, acceptance of responsibility for one's actions, respect for people's decisions based on informed consent even if those are opposed to one's own beliefs, and not unduly influencing those who have different beliefs and cultural attributes.
- **Beneficence:** This means doing good and implies improving outcomes to maximum possible. As standards improve, the beneficence principle becomes an engine for continuous improvement in action and outcomes.

- **Non-malfeasance:** This means doing no harm. Whereas beneficence is a positive-action principle, non-malfeasance is passive and guides when actions can be taken—if actions harm people, then this principle considers those actions to be wrong even if there is a reasonable probability of a beneficial outcome at a later point. In such situations, the principle of autonomy may be invoked to allow the person involved to make a choice after informed consent. This principle is evident in recruiting participants for clinical trials and other research studies, and must be carefully adhered to so as to allow for conduct of trials and studies in an ethical manner. Paternalistic actions must be guided by principles of autonomy and non-malfeasance. Balancing potential benefits and harms is a crucial activity in several situations, especially when clear evidence or guidelines do not exist. Such situations require interactions of several ethical principles before decisions can be made and actions can be taken. In a clinical situation, balancing the professionally determined ideal treatment plan, patients' lack of information restricting capacity for informed consent, and financers' stipulations are common situations where potential benefits and harms need to be balanced to arrive at a consensus plan.
- **Justice:** This principle deals with moral correctness and the sense of fairness to help people achieve what they truly deserve and are entitled to. The ethics of justice deals with moral choices through a measure of rights of the people involved and chooses the solution that seems to damage the fewest number of people. The understanding of what people deserve may be subject to interpretation at times, but entitlement is usually clearly stated in legal and policy terms. Whereas justice is not always easily apparent in individual situations, the principle takes up center stage at population levels, especially in public policy and public health practice. Justice, distribution of health care, and prudent use of scarce resources are critical in ensuring the health of any country's population. Overall, there are two basic ideas that are elaborated in several bioethical principles that guide the distribution of scarce health resources: individual well-being and resolution of interpersonal conflicts in a fair manner. These two basic ideas have resulted in six important ethical principles that guide the practice of bioethics:

1. *The Impartiality Principle:* Reasons for action are agent-neutral, rather than agent-relative, and apply to all agents equally; particularly, reasons for actions should be evaluated from an impartial perspective,

rather [than] from an agent-relative perspective that accords special weight to the personal aspects of agents' life.

2. *The Well-Being Principle:* There is a reason to protect and enhance the well-being of persons.
3. *The Equal Chance Principle:* In resolving interpersonal conflicts of well-being, there is a reason to accord equal weight to the well-being of each person, by following two hierarchical subprinciples: (a) there is a reason to distribute benefits or inevitable costs between all persons equally, so that each would get the maximum possible (roughly) equal benefit or bear the minimum possible (roughly) equal loss—provided that the benefit or reduction of cost for each person is significant; or (b) when this is impossible, there is a reason to give each person the highest possible *equal chance* to be preferred.
4. *The Importance Principle:* The strength of the reason provided by the Well-Being Principle depends on the importance of the interest at stake and the conjectured probabilities concerning the possible effects of the considered action or inaction on it (positively or negatively). Assuming equal probabilities, the more important is the interest, the stronger is the reason to protect it. In resolving interpersonal conflicts, there is, therefore, a reason to prefer the person who would otherwise suffer the most severe harm or the person who could be benefitted most significantly.
5. *The Substantial Difference Principle:* The reason provided by the Importance Principle prevails over the reason provided by the Equal Chance Principle if, assuming that all relevant probabilities are equal, there is a substantial gap in the importance of the competing interests.
6. *The Principle of Fairness: Responsibility:* When an interpersonal conflict requires a choice between the well-being of individuals, there is a reason to prefer a person who is not responsible for the existence of the conflict to a person who is and a person who is less responsible to a person who is more responsible. (Segev, 2005)

Distributive justice deals with moral choices and responsibilities of the people involved, and it chooses the solution that seems to damage the fewest number of people. It is concerned with appropriate and just allocation of goods in a society. Allegiance to specific distribution theory determines one's interpretation of justice. A *utilitarian distribution* perspective attempts to increase the "net good" of the society; a *libertarian* perspective places more emphasis on freedom, whereas an *egalitarian* perspective professes provision of equal opportunities to people to meet their needs and improve their lives. The subscription of a society to one of the distributive justice perspectives determines the scope of the activities under the purview of the actors involved in justice-related programs and issues. For example, globally, it is assumed that the rich should contribute to the welfare of the poor as an extension of the Good Samaritan Principle (if one can prevent a very grave

harm from befalling another person, at no more than a trivial cost to oneself, then one is obligated to prevent the harm; that is, to help the other person). The role of rich nations in helping poor nations is also viewed through the perspective of distributive justice. Systematization of distributive justice and saving the global economy is expected to become a centerpiece of discussion and action as countries take steps to protect the global economy and develop strategies to save their own economies as well. Ethical and political dilemmas include balancing local and global economic needs without compromising beliefs in political economy while serving the obligations of distributive justice.

Other Ethical Principles

Several other ethical principles, such as trust in relationships, veracity, fidelity, avoidance of killing, gratitude, and reparation, impact healthcare activities and decision making, although those are not often explicitly mentioned in texts.

Ethical Handling of Information

Collecting, processing, and assessing information is a critical activity in oral health clinical practice, research, and public health activities. Information flow is a two-way process where the clinician/researcher/public health practitioner conveys information to the patient/research participant and/or the public, and at the same time, assesses the responses to questions and dialogues that are generated. Several issues such as consent, capacity, disclosure, voluntary action, veracity of statements, and confidentiality of information being sought and disclosed require firm guidance to comply with the ethical principles mentioned above.

Capacity

Capacity is the ability of a person to understand, effectively process information, make rational decisions based on the information, and assess consequences of those decisions. In clinical settings, undergoing investigations; accepting a treatment plan; understanding drug and procedure effects, side effects, and anticipated outcomes; and the need for follow-up and preventive measures are common situations where a patient's capacity is a key factor in decision making and compliance to therapy. Research participants also need to understand clearly the nature of the study and their expected outcomes, provision, or withholding of treatment; experimental nature of the studies; and the possibility and consequences of receiving placebo or experimental drugs or procedures. Mostly, the patient/study participant's ca-

pacity is not explicitly assessed, and he or she tends to "trust" the clinician's judgment. The moral responsibility of the treating clinician/researcher in such situations increases. All efforts should be made to ascertain that the informed consent is provided by a capable party. If needed, participants and patients should be informed further, enabled, and then assessed for their capacity to understand information being provided to them.

Informed Consent

Informed consent, which implies authorization by the patient to medical interventions and procedures in an educated manner, is influenced by capacity, disclosure, and voluntary action. *All consent is not informed consent*. Informed consent may be given explicitly in the form of written or oral statements, as opposed to implied consent where the patient indicates willingness but does not provide explicit statements. The mere fact of having signed consent documents does not preclude careful adherence to the consent *process*. However, informed consent is neither a necessary nor sufficient condition for ethical conduct of clinical research (Emanuel, Wendler, & Grady, 2000). Informed consent is required by the law, is endorsed by major international professional bodies, and also guides proper ethical conduct of research and care giving. Perhaps there are two exceptions for informed consent: (1) where patients voluntarily waive their rights and hand over these rights to other persons or the clinician, and (2) where the principle of "therapeutic privilege" is applicable; that is, when disclosure of information would cause harm to the patient such as serious physical harm, psychological distress, or emotional harm. In both these situations, great care must be taken to ascertain that the underlying exceptional conditions truly exist because these exceptions are subject to abuse.

Confidentiality

Confidentiality implies that all information that the patient/study participant shares will remain private and not be used for purposes other than for which the information is sought; and the information will not be disclosed to persons other than those who are required to have the information for the purpose of clinical management of the patient or research study. Use of electronic records for collecting, storing, and analyzing data has made it possible for information to be accessed easily and quickly over wide geographic areas. This has necessitated changes in the law to protect health data. The Health Information Portability and Accountability Act (HIPAA), was enacted by the U.S. Congress in 1996 (USDHHS, 2009). The "Privacy Rule" of HIPAA, which was implemented on April 14, 2003, regulates the use and disclosure of certain information (Protected Health Information [PHI]) held

by "covered entities." Covered entities include healthcare clearinghouses, employer-sponsored health plans, health insurers, and medical service providers that engage in certain transactions. Healthcare research also comes under the purview of HIPAA.

Disclosure

Disclosure implies the faithful and complete conveying of important information pertinent to the study: treatment type, medications, standard of care, alternatives and their rights, current state of evidence, and expected and certainty of outcomes to the patient/study participant in a manner understandable to him or her. Ambiguous language, complex terminology, or vague description should be avoided.

Veracity

Truth telling requires that the clinician/researcher is open, forthright, candid, and honest with the patient/study participant about all aspects of the study that may impact the patient/study participant. Veracity promotes trust in the relationship, helps the patient in making good decisions, and may impact compliance and clinical outcomes. A trusting relationship between a research team and study participant also minimizes logistic issues and increases participant compliance, facilitating the outcomes for the study as well as the participant. Recall bias may be related to the conscious or unconscious modification of truth. Therefore, it is a good idea to include other questions and tests in the study instrument that can verify the authenticity of statements made by study participants.

Voluntary Information

The voluntary release of information is a central aspect in the ethical conduct of treatment or research. Coercion of any nature (direct/indirect, overt/covert) should be avoided. Voluntariness stems from the study participant's/patient's right to decide freely without pressure. Voluntary information is generally assumed to be truthful, informed, and free from any pressures. Although a patient's medical condition or socioeconomic–political status may impact one's freedom to choose, these conditions are outside the control of the treating clinician or researcher, thus precluding ethical implications for the clinician/researcher. Clinicians and researchers should ascertain that the patient/study participant is agreeing to the intervention/study voluntarily and should refrain from any manipulation, however subtle, to move the patient toward a direction of the clinician/researcher's predilection.

Research and Publication Ethics

Ethical conduct of research and publishing of scientific work is fundamental to the progress of science. Research involving human subjects and research involving animals is carefully assessed for ethical conduct and needs clearance from an Institutional Review Board (IRB) of the institution where the research is being conducted. In collaborative work spanning across different institutions, the IRBs of one or all of the institutions may be involved in ensuring the ethical conduct of research. Most established research institutions in the world have their IRBs or ethics committees assess the research being conducted in their institution. However, such control is not available in several countries. With the growth of multinational clinical trials, it has become important to ascertain that proper IRB clearances are sought for the research to be ethically conducted and the results to be considered valid.

Research Ethics

Clinical research has grown from small individual scientists' clinics and laboratories to large and expensive trials run by rich industries, and along with this growth, economic and political stakes have increased. This has resulted in the use of incentives to recruit researchers, participants, and other support staff in the conduct of clinical research. Whereas in clinical practice setups, the clinician is obligated toward the patient, in research, this obligation may come in direct confrontation with incentives for research leading to complex conflict-of-interest situations. Such conflicts have to be declared, because these may jeopardize the ethical conduct of the research and must be approved by the IRBs before persons with conflicts of interest can conduct the research work. The Council for International Organizations of Medical Sciences (CIMOS) developed the International Ethical guidelines for Biomedical Research Involving Human Subjects in 2002 and defined research:

> The term "research" refers to a class of activity designed to develop or contribute to generalizable knowledge. Generalizable knowledge consists of theories, principles or relationships, or the accumulation of information on which they are based, that can be corroborated by accepted scientific methods of observation and inference. In the present context "research" includes both medical and behavioural studies pertaining to human health. Usually "research" is modified by the adjective "biomedical" to indicate its relation to health . . . .
>
> Research involving human subjects includes: studies of a physiological, biochemical or pathological process, or of the response to a specific intervention—whether physical, chemical or psychological; in healthy subjects or patients; controlled trials of diagnostic, preventive or therapeutic

> measures in large groups of persons, designed to demonstrate a specific generalizable response to these measures against a background of individual biological variation; studies designed to determine the consequences for individuals and communities of specific preventive or therapeutic measures; and studies concerning human health-related behavior in a variety of circumstances and environments. (CIMOS, 2002)

The U.S. Department of Health and Human Services defines research as "systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge . . . constitute research for purposes of this policy, whether or not they are conducted or supported under a program which is considered research for other purposes" (DHHS, 2005).

> Research subject to regulation, and similar terms are intended to encompass those research activities for which a federal department or agency has specific responsibility for regulating as a research activity, (for example, Investigational New Drug requirements administered by the Food and Drug Administration). It does not include research activities which are incidentally regulated by a federal department or agency solely as part of the department's or agency's broader responsibility to regulate certain types of activities whether research or non-research in nature (for example, Wage and Hour requirements administered by the Department of Labor).
>
> Human subject means a living individual about whom an investigator (whether professional or student) conducting research obtains (1) Data through intervention or interaction with the individual, or (2) Identifiable private information.
>
> Intervention includes both physical procedures by which data are gathered (for example, venipuncture) and manipulations of the subject or the subject's environment that are performed for research purposes. Interaction includes communication or interpersonal contact between investigator and subject. Private information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects. (DHHS, 2005)

Conducting health research requires that all investigators involved in the research activities should be trained in human subject research and certified by appropriate training agencies, usually the local IRBs. Although such basic training updates researchers' knowledge about the historical origins of ethics conduct of research and a general description of when and how to approach the IRBs, there is increasing recognition that ethics train-

ing for researchers requires more effort, and perhaps should be viewed as an ongoing activity.

> Many believe that informed consent makes clinical research ethical. However, informed consent is neither necessary nor sufficient for ethical clinical research. Drawing on the basic philosophies underlying major codes, declarations, and other documents relevant to research with human subjects, we propose seven requirements that systematically elucidate a coherent framework for evaluating the ethics of clinical research studies: (1) value—enhancements of health or knowledge must be derived from the research; (2) scientific validity—the research must be methodologically rigorous; (3) fair subject selection—scientific objectives, not vulnerability or privilege, and the potential for and distribution of risks and benefits, should determine communities selected as study sites and the inclusion criteria for individual subjects; (4) favorable risk-benefit ratio—within the context of standard clinical practice and the research protocol, risks must be minimized, potential benefits enhanced, and the potential benefits to individuals and knowledge gained for society must outweigh the risks; (5) independent review—unaffiliated individuals must review the research and approve, amend, or terminate it; (6) informed consent—individuals should be informed about the research and provide their voluntary consent; and (7) respect for enrolled subjects—subjects should have their privacy protected, the opportunity to withdraw, and their well-being monitored. Fulfilling all seven requirements is necessary and sufficient to make clinical research ethical. These requirements are universal, although they must be adapted to the health, economic, cultural, and technological conditions in which clinical research is conducted. (Emanuel et al., 2000)

Publication Ethics

Complete data fabrication in oral health research, although uncommon, has been reported and led to the retraction of a series of high-profile articles published in major journals (Kotzin, 2007). "The scientific community has a duty to warn people to ignore an article containing faked data and must try to prevent inadvertent citation of it" which may be accomplished by publishing a retraction and linking it to indices that cite the article in question, and by verifying the integrity of other articles published by the author of the article in question (Sox & Rennie, 2006).

> Maintaining the integrity of the scientific literature requires governmental institutions that have the authority to investigate and punish guilty scientists and requires that research institutions investigate alleged fraud. It requires journal editors to issue a retraction when they learn that their journal has published a tainted article. It requires research institutions to accept their responsibility to investigate every article published by a scientist who has published even one fraudulent article. Finally, it requires authors to take pains to avoid citing retracted articles and to issue a correction when they inadvertently cite a retracted article. (Sox & Rennie, 2006)

The Committee on Publication Ethics (COPE) "aims to define best practice in the ethics of scientific publishing and to assist authors, editors, editorial board members, readers, owners of journals and publishers" (the COPE website can be accessed at http://publicationethics.org/). COPE has developed a series of flowcharts dealing with special situations that may arise at different steps in the lifecycle of a scientific article and potential solutions to ethical issues that may occur or have already occurred in scientific publications. These free, downloadable flowcharts provide guidelines about how to respond to situations such as duplicate publications, plagiarism, and fabricated data. These flowcharts also help in addressing situations arising out of changes in authorship such as corresponding author requests the addition of an extra author before publication; corresponding author requests removal of an author before publication, requests the addition of an extra author after publication, requests the removal of an author after publication, a suspected guest, ghost, or gift authorship; and advice on how to spot authorship problems. Guides for reviewers reviewing manuscripts for scientific publications include situations when a reviewer suspects an undisclosed conflict of interest in a submitted manuscript, a reader suspects an undisclosed conflict of interest in a published article, anyone suspects an ethical problem with a submitted manuscript, and anyone suspects reviewer misconduct. The flowcharts also describe how COPE handles complaints against editors.

When a problem arises in an already published article, journals retract the article. Retractions are complicated actions and may take time and tedious logistic planning to ensure that all relevant links are updated with the information. Extensive publication across several journals of high reputation involving ethically questionable research is usually uncommon, but such events have occurred involving oral health research (Curfman, Morrissey, & Drazen, 2006).

> For the National Library of Medicine (NLM) to label something a retraction, the notice must be cited on a numbered page in a journal indexed in MEDLINE and generally, the retraction notice must appear in the same journal title that published the retracted article. Only statements that are specifically labeled retraction or withdrawal are considered to be retractions. If the statement is headed "Questionable Science" or something similar, it is labeled as a "comment" by NLM. Comments are substantive articles, letters, or editorials that challenge, refute, support, or expand upon a previously published article.
>
> Since 1984, NLM has played a major role in informing the users of MEDLINE of indexed journal articles that have been subsequently revealed as fraudulent. Far fewer than 1% of more than 600 000 articles indexed annually are retracted. However, the potential impact can be great if inaccurate information forms the basis for subsequent research or is used in the treatment of patients. By NLM's definition, a retraction states that an ar-

ticle previously published, was based on deliberately falsified or unsubstantiated data. (Kotzin, 2007)

Clinical and Research Ethics in a Globalized Era

The major oral health practice and research-related associations that have committees dealing with ethics emphasize ethics in dental practice and research and disseminate such information include IADR, AADR, ADA, WHO, and PAHO. Ethical issues faced by dentists and oral health researchers across the world include quality of care and adjustment to ever-increasing needs for standard of care; the need for, scope, and extent of advertising; intraprofessional relationships; interprofessional relationships; and self-regulation practices and conflicts that may arise between professional interests and interests of the community at large (e.g., the role of dental therapists in increasing access to oral health care in the United States). Dental therapists can become valued members of the dental team throughout the world, helping to improve access to care and reducing existing disparities in oral health (Nash et al., 2008); patient autonomy; conflicts with patients; justice; intraprofessional relationships; and financial transactions.

Global ethics as a concept tries to provide a foundation and platform for ethical interaction between a variety of people and their various levels of organization in the society across the world. Currently it is a complex and ill-defined set of ideological and philosophical claims about the de facto and ideal norms that guide the interactions of people and communities around the world (Boyle, 2004). Although ethical principles as a guide to human interaction within and across countries have generally existed for a long time and are a cornerstone of human development, the notion that global ethics deals with the moral questions that arise from globalization is relatively new. Globalization has increased the ease and frequency of the interaction between people separated by space and cultural attributes. Living in different parts of the world has also increased people's awareness of events and access to information about events across the world almost instantly. It no longer takes weeks for information to travel through the mail or through controlled media, although the information may be filtered out according to the needs of ruling powers in different areas. At the same time, because any individual is free to contribute and assimilate information in un-regulated form, the quality of such information may potentially be in question.

As research activities grow to encompass different centers across different countries, new ethical challenges will emerge. These challenges will require new regulations across different countries and cultures in tune with ethical principles outlined above. Currently, dental laboratory work is often outsourced and shipped to far-off places in the world. As this phenomenon

takes greater hold, standardization of product fabrication process and quality, material handling, and labor laws will play an important role in the globalized ethical conduct of clinical activities and research. Another challenge is to address ethical conflicts that may arise out of patients getting treatment in different parts of the world. There will be a need to also standardize the quality of care across different areas of the world as "medical tourism" increases.

20 Systems Thinking

Most of the major threats to the health of the public arise from a complex mix of behavioral, economic, political, social, and biological factors. The efforts to translate knowledge concerning the preventability of premature death and chronic illness are often hindered by a host of barriers that arise in attempts to intervene to prevent disease. These barriers include the intransigence of risk factors in populations, inadequate screening, lack of access to health services and coordination of care, limitations on treatments and compliance with treatment, the delay from diagnosis to treatment, and a lack of understanding of the leverage points of the healthcare delivery system that are most effective in applied public health. These barriers are characteristic of health problems that have been termed *policy resistant* (Sterman, 2006). Policy resistance is the tendency for interventions to be defeated by the system's response to the intervention itself. For example, increased drop-out rates due to severe side effects of drugs in clinical trials and high emergency room use for minor problems by the uninsured may be viewed as examples of policy resistance.

Systems Science

As a field of study, systems science is interdisciplinary in nature. It studies all complex systems including societies; biological phenomenon; and financial, economic, and political systems and their interactions. It finds its theoretical moorings in systems theory, dynamical systems theory, control theory, communication theory, social systems theory, and cybernetics. Systems science brings together diverse fields such as medicine, engineering, public health and the social sciences to interact with and impact each other. Increasingly, applications for systems science are being found in a variety of

fields such as public health systems, systems biology, systems dynamics, systems ecology, systems engineering, and systems psychology. It is, in many ways, the reverse of reductionism; that is, it seeks to integrate individually defined or functioning units into the larger complex hierarchical system observed in reality with all interacting forces that influence action in myriad ways to produce real-world outcomes. Whereas the reductionist paradigm is very useful in discerning individual-level action under a variety of environmental conditions that can be manipulated in different ways, lending it to relatively easy experimental control, systems science recognizes that in real-world situations, individual entities respond to a variety of stimuli to produce an outcome that is a net result of the action of multiple forces, and "ideal" conditions do not exist. Therefore, systems science seeks factors that are modifiable in a real-world situation (leverage points), to optimize the overall system output toward desired goals.

While there have been calls to employ a systems approach in health sciences, its use has been limited. The *systems dynamics* approach was developed by Forester in the mid 1950s and has been employed in numerous fields including business, engineering, and transportation. Employing systems thinking and systems dynamics to address problems in health sciences, especially public health, is a more recent innovation, although several early studies have demonstrated promise for its use—efforts have been undertaken in diabetes, cardiovascular diseases, cancer, and AIDS (Homer & Hirsch, 2006). Whereas Milstein et al. (2007) developed a community health model, Hirsh and Immediato (1999) created the Healthcare Microworld. The Healthcare Microworld is a consortium of three modules: (1) organization and management of healthcare delivery systems; (2) strategies for improving the health of a community; and (3) enablers for participants to combine the previous two modules and act as managers of a health plan who can invest in a combination of delivery system improvements and health interventions to see which ones offer the best results. These models highlight how well-intentioned attempts at integrated delivery systems have often failed when providers focus on the "bottom line" rather than on the creation of integrated systems, and illustrate how long-term results can be obtained from investments in programs that decrease social and behavioral health risks. Further, the clustering of risk factors for many of the chronic diseases, especially oral diseases, strongly argues for a systems approach and the development of a common research agenda for these "syndemics"—diseases that arise from common risk factors such as smoking, alcohol consumption, lack of physical activity, and inadequate diet (Hirsh & Immediato, 1999; Milstein et al., 2007). Equally importantly, development and progress of most diseases are also modified by nonbiologic factors that take root in the sociocultural–politico–legal arena such as socioeconomic position, access to health care, decision making about immediacy of need to access a healthcare professional, cultural attributes, and law and policy formulation.

A comprehensive understanding of the resultant outcome of these varieties of forces toward health outcomes is possible only through complete systems level approaches as opposed to reductionist approaches.

A number of lessons have been learned from the early experiences in the application of systems models in public health. First, the process of the development of system dynamic models for public health problems is an iterative one that requires conceptual thought, initial model building, testing and validation, sensitivity analysis, further refinement, and continued development. It is apparent that the development of these models is highly influenced by the expertise and the knowledge base of the research team. Consequently, it has been argued that multidisciplinary expertise is required, reflecting both the substantive area of inquiry as well as a familiarity with the methods for analysis. System dynamics models need to employ group model-building techniques that take advantage of a range of expertise to address the "policy-resistant" health problem. In oral health care, this usually requires the input of subject matter experts such as different dental specialists, oral epidemiologists, statistical modelers, and other clinicians who understand current procedures regarding diagnosis, treatment, and oral health systems. Models can be limited by the subject matter experience of those involved in the process, and therefore the strength of any model is subject to the limitations for the designers as much as the data upon which it is based (Jones, Homer, Murphy, Eissen, Milstein, et al., 2006; Mabry, Olster, Morgan, & Abrams, 2008).

Second, models need to be constructed based on real data but need to remain flexible to incorporate expertise and "experience-based" data for those components of the model for which there are data gaps. For example, with oral cancer screening, there are limited data to assess the sensitivity and specificity of oral screening procedures and how often they are or should be performed. As there are limited data for the recommendation for screening for oral cancers, there are limited incentives to do so. Indeed, the development of the models may potentially highlight some of the "gaps" in the public health system for addressing the policy-resistant problem of interest. There is a "complex jurisdiction" for the identification and treatment of oral diseases. Screening, early detection and prevention of diseases, and prophylaxis may best be performed by community providers such as dentists or family medicine practitioners. However, the linkages for referral from these community providers to dental clinics that can manage complex cases may be more nebulous than for other medical problems. For example, nonavailability of dentists is often cited as an important reason for not accessing dental services when needed by poor persons. The development of systems models may highlight these gaps, identify potential solutions and modifiable factors, and improve the understanding of systems processes.

Third, there are several potential benefits of using systems models over traditional analytic methods in epidemiology. Despite the recognition that

chronic diseases (e.g., cancer, dental caries, periodontal disease) arise from multiple interacting factors, commonly used epidemiologic models are fundamentally linear (e.g., linear and logistic regression, survival analysis) and often neglect the consideration of interaction effects or nonlinear effects. These models do not typically reflect feedback processes, time delays, and stocks and flows, and are not constructed to model the dynamic processes of changing incidence and survival. The ability to look at multiple levels of causation and feedback loops in a causal model of disease and survival is more tractable with systems dynamic modeling. These models can further our understanding of the time frames appropriate for a consideration of intervention effects and can incorporate secular changes that may occur such as the introduction of a new screening technique, treatment procedures, or changes in the prevalence of risk factors.

Finally, explicit testing of commonly held assumptions can be undertaken using sensitivity analysis and simulation methods. We commonly assume that diseases will be cured and disease-free survival times will be increased if diseases are detected at earlier stages and effective treatments are obtained in a timely manner. However, it is difficult to assess at which point in the continuum of the initiation and progression of diseases—screening, detection, and treatment—in populations is the most effective leverage point for influencing the best clinical outcomes. System dynamic models can provide answers under different scenarios given the ease at which sensitivity analyses can be undertaken. Under differing scenarios, there may be adequate access to care and availability of the latest treatment protocols. However, delays in obtaining treatment, inadequate insurance coverage, or the absence of a coordinated care system may lead to transitions to later stages of disease, making the clinical states more complex, requiring more sophisticated and expensive management plans, and therefore shortening survival times and worsening outcomes. In addition, simple analysis of aggregate survival rates do not indicate whether these assumptions are the same in all age, gender, race, or the same insurance status, or in groups with a genetic susceptibility to the outcome. Development of a more nuanced analysis on susceptibilities and differences in population subgroups can be modeled using a systems approach.

Modeling with Systems Approach

As suggested by Mabry et al. (2008), systems modeling are problem-focused and outcomes oriented. These begin with a complex but clearly defined health problem and work backwards from the problem to identify the multiple causal pathways and feedback/feed-forward loops that will lead to development of the most powerful and efficient set of interventions to address the problem. This brings to problem solving a perspective in which the problem space is conceptualized as a system of interrelated component

parts (i.e., the "big picture"—for a similar model for diabetes risk, see Figure 3 of Mabry et al. [2008]). From the literature review, investigators may identify all factors associated with the problem and develop a series of mental models presumably acting at different levels of a system for oral diseases. This system is viewed as a coherent whole, while the relationships among the components are also recognized and seen as critical to the system because they give rise to the emergent properties of the system. Emergent properties are those properties that can only be seen at the system level and are not attributes of the individual components themselves. A general approach to systems modeling may include the following sequential steps:

1. Preparation of mental models
2. Preparation of model structures and calibration of models. Defining stocks and flows (see later part of this section for definitions of common terms in systems modeling)
3. Development of a baseline behavior model
4. Intervention tests with each model
5. Sensitivity analyses using different theoretically defined scenarios ranging from "best-case" to "worst-case" scenarios. Comparison with standard one-level modeling approaches (e.g., linear regression, logistic regression) and multilevel approaches may also be undertaken

Comparison of systems models with other types of modeling strategies will permit comprehensive assessment for a chain of population flows beginning when a person becomes at risk for disease and continuing through initial onset, diagnosis, and progression to the final outcome of disease or death. Such breadth of scope allows the systems model to anticipate nonlinear changes in variables, such as the incidence rate, that narrower models would miss. Systems models specify how population groups accumulate in several health states associated with specific diseases (e.g., early/late presentation, delayed/missed diagnoses/treatment, prognosis differences, complications, and death) along with the rates at which people flow from one state to another. A systems approach does not take a static view of dynamic population or time in its analysis and recognizes that events do not end at censoring.

In systems modeling, four basic common terms are regularly used. *Stocks* are the basic building blocks of the model and represent anything that accumulates (e.g., prevalence/incidence rates in populations, clinician knowledge and skills, and cumulative outcomes such as survival cases and cause-specific outcomes). Stocks are tangible, countable, and physical accumulations. Stocks can also be used to represent the degree of nonphysical accumulations such as clinician knowledge and patient awareness. *Flows* represent activities that lead to inputs and outputs to stocks such as disease occurrence; population migration; transformation from acute to chronic disease, i.e., precancer to cancer; and occurrence of metastasis. Flows change

the magnitude of stocks in the system. *Connectors* transmit information to regulate flows between different components. Connectors can connect into flows or converters but never into stocks (but they can affect both input and output flows). Magnitudes of stocks are affected only by flows. For example, if patients are not treated, the prevalence pool stocks will increase, and survival stocks will decrease. *Converters* contain equations that generate an output value during each time interval of a simulation. For example, increasing/decreasing complex disease detection success rates may be transformed by another variable, such as a clinician continuing education program, in the model. Converters can also be used to store constant values; for example, a fixed success rate for disease outcomes depending on clinical criteria.

The modeling process involves three main stages:

1. **Constructing a model:** The model defines stocks and then constructs links to variables that affect the size of the stocks. These are usually direct inputs or outputs modeled using flows. The process also involves adjustment of the magnitude of flows by converters using links or being affected by the size of stocks.
2. **Parameterizing the model:** Once the model is constructed, relationships among the model elements may be quantified using both linear and nonlinear relationships derived from the literature and investigators' understanding of incidence rates, outcome types and rates, and survival rates and their determining factors, along with biological parameters of diseases.
3. **Exploring model dynamics:** Once the models are parameterized, investigators may examine the model outputs by generating outputs in tabular and/or graphical form to explore defined quantitative and/or qualitative outcomes and by manipulating the parameters to perform a simulation-based sensitivity analysis.

Data Sources

Systems modeling is an ecologic multilevel approach and may use data from a variety of sources that can be linked. For example, data can be aggregated (with different contributed variables) from disease registries (patient demographics, disease-specific information, prognosis, treatment types and outcomes); the U.S. Census Bureau (demographics, population growth, and death rates); and the National Health and Nutrition Examination Survey (general biological markers, prevalence of important comorbidities). Similarly, useful information available from research literature and other surveys such as the National Health Interview Survey, the Medical Expenditure Panel Survey, the Behavioral Risk Factor Surveillance System, and additional surveys that may exist, or specific surveys that may be un-

dertaken for the problem being addressed, may be incorporated into the data set used for modeling. While pooling data across different data sources, appropriate reweighting of the data must be done by keeping in mind the different survey-sampling techniques for the included surveys from which data is incorporated in the systems modeling data set. These weights may be determined, for example, using census data (or other problem-specific data) to stratify by several variables such as gender, race/ethnicity, age, socioeconomic position (using income, education, and employment as indicators), and other variables as needed.

Although systems modeling is viewed as an ecologic multilevel approach, it can address any complex system, including human disease etiology process and molecular and genetic epidemiology problems. For example, in this paradigm, addressing molecular etiopathology, personal individual-level factors appear at an "ecologic" level compared to tissue/cellular/subcellular-level factors.

Systems Modeling Software

STELLA® and Archemedes® are available software for systems modeling. The Archimedes Model is a full-scale simulation model of human physiology, diseases, behaviors, interventions, and healthcare systems. By using advanced methods of mathematics, computing, and data systems, the Model enables managers, administrators, and policy makers to be better informed and to make smarter decisions than has previously been possible. To build an Archimedes-like model, investigators would need studies of the underlying physiological processes from which data will be generated (Schlessinger & Eddy, 2002).

The systems modeling software helps in the development of models and provides a large number of opportunities to explore by asking "What if?" and watching what happens, which is a key element in sensitivity analysis and also in developing a simulation for potential future real and hypothetical scenarios. A simulation allows developing relations between stocks, flows, connectors, and converters in such a way that the values for each parameter can represent the inputted data that can be correctly modeled. Parameter values in these models can be varied in response to different hypothetical scenarios to develop system-level model-based simulation sensitivity analyses. As investigators run the model, the software standard numerical methods adopts the system of equations that comprises the model. Simulations in software are done using Euler's method; second-order Runge–Kutta; and fourth-order Runge–Kutta algorithms (iThink/STELLA, 2008). In numerical analysis, the Runge–Kutta methods are an important family of implicit and explicit iterative methods for the approximation of solutions of ordinary differential equations by using a trial step at the midpoint of an interval to cancel out lower-order error terms.

Modeling Outcome and Significance

Systems analysis permits mapping and modeling, simulation and analysis, and assessment of effects through various communication tools within the software. This approach allows investigators to:

1. Develop a vision of the "big picture."
2. Jump the gap between theory and the real world.
3. Simulate a system over time.
4. Enable creative change in systems.
5. Clearly communicate system inputs and outputs and demonstrate outcomes.

It is expected that systems modeling will lead to a deeper understanding of intransigent problems associated with disease states and health systems, thus enabling further breakthroughs in the understanding and reduction of the burden and suffering of the major important diseases that afflict nations. Systems analytical method as part of systems science, because of its unique ability to consider simultaneously both the whole system and its individual parts, is capable of producing solutions that take into account a broad range of factors pertinent to the problem under consideration; for instance, genetic-to-environmental, cellular-to behavioral, and biological-to-social systems approaches. These have proven extremely valuable for unlocking complex, multidimensional health issues and for transforming this knowledge into effective interventions that can fundamentally change population health (Mabry et al., 2008).

Systems Biology

Systems biology views organisms as an interacting system of genes, proteins, and biochemical reactions that are key elements in the organism's lifecycle, determining its form and function. It views interactions in living cells and their organized forms as complex information processing systems. The primary goal of systems biology is to study them and investigate how they behave. "It is increasingly being appreciated that the biological system is much more than the sum of its parts! Biological systems have been described as complex information processing systems and the primary goal to study them is to investigate how they behave" (Chattopadhyay, 2009). Biological systems adapt to unstructured environments and are able to learn from experience (Iyengar, 1998). Systems biology has been defined as "a comprehensive quantitative analysis of the manner in which all the components of a biological system interact functionally over time" (Aderem, 2005). It integrates chemistry, computer science, engineering, life sciences, mathematics, physics, and other sciences to provide understanding of biological

processes. Biological systems are generally complex, open systems that exhibit four important properties:

- *Emergence:* Properties of biological systems occur due their existence as a system and are not inherent in or integral to their parts. These cannot be projected from the individual properties of the parts to the system.
- *Irreducibility:* The system cannot be broken into sections of component parts without loss of identity and function as the system.
- *Modularity:* Biological systems are a functional entity of interacting parts that exist as a unit with a common function.
- *Robustness:* Biological systems are sturdy and strong in form, i.e., they maintain their phenotypical form in the face of adverse environmental insults, genetic variations, and random variability.

In etiological research, epidemiological and statistical models may be combined with computational models to develop deeper understanding of spatio–structural–functional attributes of biological systems in order to understand the mechanisms and pathways involved in biological functioning. Model development is the key factor in a systems biology approach—model development faces several challenges such as confounding by several factors, functional redundancy, biases, information errors, random errors, and processes effecting examined factors. To minimize errors, the Institute for Systems Biology (2008) suggests the following principles for developing good systems models:

1. Global approaches should be taken to data collection and analyses.
2. Information derived from diverse data types should be integrated.
3. Mathematical and statistical modeling is essential to the quantitative analysis of a system's properties.
4. Biology should drive technology which, in turn, makes better biology possible.
5. Systems biology research should create an interactive interdisciplinary scientific culture.
6. The results of research should be freely disseminated.

Computer software has been developed to address systems biology questions that may be useful in studies dealing with biomics, fluxomics, genomics, glycomics, interactomics, metabolomics, proteomics, and transcriptomics. Commonly used software includes Archimedes, Cell Designer, Cellerator, DBsolve, Dynetica, E-Cell, Gepasi, Jarnac/ Jdesigner, NetLogo, Stella, StochSim, and Virtual cell. They may used standard modeling languages such as Systems Biology Markup Language (SBML), Cell Markup Language (CellML), Systems Biology Graphical Notation (SBGN), and use

and report to biological interaction databases such as AfCS-nature Signaling Gateway, Biomolecular Object Network Databank, The Database of Interacting Proteins, EcoCyc, Kyoto Encyclopedia of Genes and Genomes, Signaling Transduction Knowledge Environment, NCI-Nature Pathway Interaction Database, and the Biological Interaction database for Protein-nucleic Acid (Chattopadhyay, 2009).

Factors underlying health states encompass interrelated workings of several biological, behavioral, economic, and social factors. Whereas a reductionist view of disease processes and their outcomes is useful in understanding fundamental mechanisms, they fail to explain interrelatedness between parts and their interactions to produce disease, and they limit the choices for diagnosis and therapeutic applications (Chattopadhyay, 2009). The systems biology approach attempts to answer questions about health determinants in a comprehensive quantitative and qualitative analysis of the manner in which all the components of a biological system interact functionally over time.

References

Chapter 1

Agency for Healthcare Research and Quality. (2000, March). Outcomes research fact sheet. AHRQ Publication No. 00-P011. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved July 29, 2009 from http://www.ahrq.gov/clinic/outfact.htm

Beck, J. D. (1998). Risk revisited. *Community Dent Oral Epidemiol*, *26*, 220–225.

Beck, J. D., Offenbacher S. (2005, November). Systemic effects of periodontitis: epidemiology of periodontal disease and cardiovascular disease. *J Periodontol*, *76*(11 Suppl), 2089–2100.

Burke, L. B. (2001, January/February). U.S. regulation of pharmaceutical outcomes research. *Value Health*, *4*(1), 5–7.

Burt, B. A. (2001). Definitions of risk. *Journal of Dental Education*, *65*(10), 1007–1008.

Caufield, P. W., Li, Y., & Dasanayake, A. (2005, May). Dental caries: an infectious and transmissible disease. *Compend Contin Educ Dent*, *26*(5 Suppl 1), 10–16.

Centers for Disease Control and Prevention (CDC). (2008). National oral health surveillance system. Retrieved September 24, 2008, from http://www.cdc.gov/nohss/

Defraia, E., Baroni, G., & Marinelli A. (2006, May). Dental arch dimensions in the mixed dentition: a study of Italian children born in the 1950s and the 1990s. *Angle Orthod*, *76*(3), 446–451.

Denzin, N. K., & Lincoln, Y. S. (1994). Introduction: entering the field of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1–17). Thousand Oaks, CA: Sage.

Heidari, E., Dickinson, C., Wilson, R., & Fiske, J. (2007, January). Verifiable CPD paper: oral health of remand prisoners in HMP Brixton, London. *Br Dent J*, *202*(2), E1.

Heng, C. K., & Morse, D.E. (2002, Winter). Dental caries experience of female inmates. *J Public Health Dent*, *62*(1), 57–61.

Hill, A.B. (1965). The environment and disease: association or causation? Proceedings of the Royal Society of Medicine, *58*, 295–300.

John F. Kennedy Presidential Library and Museum. *Day of affirmation address* (news release text version). Retrieved May 26, 2009, from http://www.jfklibrary.org/Historical+Resources/Archives/Reference+Desk/Speeches/RFK/Day+of+Affirmation+Address+News+Release.htm

Joshipura, K. (2002, June).The relationship between oral conditions and ischemic stroke and peripheral vascular disease. *J Am Dent Assoc*, *133*(Suppl), 23S–30S.

Kshirsagar, A. V., Offenbacher, S., Moss, K. L., Barros, S. P., & Beck, J. D. (2007). Antibodies to periodontal organisms are associated with decreased kidney function. The Dental Atherosclerosis Risk in Communities study. *Blood Purif*, *25*(1),125–132.

Kurttio, P., Gustavsson, N., Vartiainen, T., & Pekkanen, J. (1999). Exposure to natural fluoride in well water and hip fracture: a cohort analysis in Finland. *Am J Epidemiol*, *150*(8), 817–824.

Lamster, I. B., Lalla, E., Borgnakke, W. S., & Taylor, G. W. (2008, October). The relationship between oral health and diabetes mellitus. *J Am Dent Assoc*, *139*(Suppl), 19S–24S.

Last, J. M., (Ed.). (2001). *A dictionary of epidemiology* (4th ed.). New York: Oxford University Press.

Lindsten, R., Ogaard, B., Larsson, E., & Bjerklin, K. (2002, October). Transverse dental and dental arch depth dimensions in the mixed dentition in a skeletal sample from the 14th to the 19th century and Norwegian children and Norwegian Sami children of today. *Angle Orthod*, *72*(5), 439–448.

Meyer, M. S., Joshipura, K., Giovannucci, E., & Michaud, D. S. (2008, November). A review of the relationship between tooth loss, periodontal disease, and cancer. *Cancer Causes Control*, *19*(9), 895–907.

Naidoo, S., Yengopal, V., & Cohen, B. (2005, February). A baseline survey: oral health status of prisoners—Western Cape. *SADJ*, *60*(1), 24–27.

Offenbacher, S., Boggess, K. A., Murtha, A. P., Jared, H. L., Lieff, S., McKaig, R. G., et al. (2006, January). Progressive periodontal disease and risk of very preterm delivery. *Obstet Gynecol*, *107*(1), 29–36.

Pitiphat, W., Joshipura, K. J., Gillman, M. W., Williams, P. L., Douglass, C. W., Rich-Edwards, J. W. (2008, February). Maternal periodontitis and adverse pregnancy outcomes. *Community Dent Oral Epidemiol*, *36*(1), 3–11.

Ruma, M., Boggess, K., Moss, K., Jared, H., Murtha, A., Beck, J., et al. (2008, April). Maternal periodontal disease, systemic inflammation, and risk for preeclampsia. *Am J Obstet Gynecol*, *198*(4), 389e1–5.

Sharma, B. R. (2007, January). Gender identity disorder and its medico-legal considerations. *Med Sci Law*, *47*(1), 31–40.

Sisson, K. L. (2007). Theoretical explanations for social inequalities in oral health. *Community Dent Oral Epidemiol*, *35*, 81–88.

United States Department of Health and Human Services (USDHHS). (2000). *Oral health in America: a report of the surgeon general—executive summary*. Rockville, MD: U.S. Department of Health and Human Services, National Institute of Dental and Craniofacial Research, National Institutes of Health.

World Health Organization (WHO). (2008a). *What is oral health?* Retrieved May 5, 2008, from http://www.who.int/mediacentre/factsheets/fs318/en/index.html

World Health Organization (WHO). (2008b). *International classification of diseases (ICD)*. Retrieved May 5, 2008, from http://www.who.int/classifications/icd/en/

Xiong, X., Buekens, P., Fraser, W. D., Beck, J., & Offenbacher, S. (2006, February). Periodontal disease and adverse pregnancy outcomes: a systematic review. *BJOG*, *113*(2), 135–143.

Chapter 2

Amstutz, R. D., & Rozier, R. G. (1995, June). Community risk indicators for dental caries in school children: an ecologic study. *Community Dent Oral Epidemiol*, *23*(3), 129–137.

Burt, B. A., Ismail, A. I., Morrison, E. C., & Beltran, E. D. (1990, May). Risk factors for tooth loss over a 28-year period. *J Dent Res*, *69*(5), 1126–1130.

Chattopadhyay, A. (2003, May). HIV-associated oral disease: prevalence, incidence and role of secretory leukocyte protease inhibitor. PhD dissertation. University of North Carolina at Chapel Hill, NC, USA.

Chattopadhyay, A., Caplan, D. J., Slade, G. D., Shugars, D. C., Tien, H-C., & Patton, L. L. (2005). Incidence of oral candidiasis and oral hairy leukoplakia in HIV-infected adults in North Carolina. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, *99*(1), 39–47.

Edgar, W. M., & Higham, S. M. (1995, November). Role of saliva in caries models. *Adv Dent Res*, *9*(3), 235–238.

Haneuse, S., & Wakefield, J. (2008, March). Geographic-based ecological correlation studies using supplemental case–control data. *Stat Med*, *27*(6), 864–887.

Hennessy, S., Bilker, W. B., Berlin, J. A., & Strom, B. L. (1999). Factors influencing the optimal control-to-case ratio in matched case–control studies. *Am J Epidemiol*, *149*, 195197.

Kelsey, J., Whittemore, A. S., Evans, A. S., & Thompson, W. D. (1996). *Methods in observational epidemiology* (2nd ed.). New York: Oxford University Press.

Kirakozova, A., & Caplan, D. J. (2006, August). Predictors of root canal treatment in teeth with full coverage restorations. *J Endod*, *32*(8), 727–730.

National Institute of Health. (2008). Obstetrics and Periodontal Therapy (OPT) Study. Retrieved on July 8, 2009, from http://clinicaltrials.gov/ct2/show/NCT00066131?spons=%22National+Institute+of+Dental+and+Craniofacial+Research+(NIDCR)%22&spons_ex=Y&rank=34

National Institute of Mental Health. (2008, September). NIMH Multisite HIV/STD Prevention Trial for African American couples group. Methodological overview of an African American couple-based HIV/STD prevention trial. *J Acquir Immune Defic Syndr*, *49*(Suppl 1), S3–S14.

Patton, L. L., McKaig, R. G., Strauss, R. P., & Eron, J. J., Jr. (1998, September). Oral manifestations of HIV in a southeast USA population. *Oral Dis*, *4*(3), 164–169.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Russell, J. S., Littlewood, S. J., Blance, A., & Mitchell, L. (2008, September). The efficacy of a plasma arc light in orthodontic bonding: a randomized controlled clinical trial. *J Orthod*, *35*(3), 202–209.

Salway, R., & Wakefield, J. (2008, January). A hybrid model for reducing ecological bias. *Biostatistics*, *9*(1), 1–17.

Sreenivasan, P. K., Furgang, D., Markowitz, K., McKiernan, M., Tischio-Bereski, D., Devizio, W., et al. (2009, June). Clinical anti-microbial efficacy of a new zinc citrate dentifrice. *Clin Oral Investig*, *13*(2), 195–202.

Sun, X. W., Sun, G. L., & Xiao, L. J. (2008, February). Evaluation of gingival inflammation related to different retraction agents. *Hua Xi Kou Qiang Yi Xue Za Zhi*, *26*(1), 53–55 [Article in Chinese].

Wakefield, J., & Shaddick, G. (2006, July). Health-exposure modeling and the ecological fallacy. *Biostatistics*, *7*(3), 438–455.

Chapter 3

Gillies, D. (1998). The Duhem thesis and the Quine thesis. In M. Curd & J. A. Cover (Eds.), *Philosophy of science: the central issues* (pp. 302–319). New York: Norton.

Hill, A. B. (1965). The environment and disease: association or causation? Proceedings of the Royal Society of Medicine, *58*, 295–300.

Phillips, C. V., & Goodman, K. J. (2006, May 26). Causal criteria and counterfactuals; nothing more (or less) than scientific common sense. *Emerg Themes Epidemiol*, *3*, 5.

Rothman, K. J. (2002). *Epidemiology—an introduction*. New York: Oxford University Press.

Rothman, K. J. & Greenland, S. (2005). Causation and causal inference in epidemiology. *Am J Public Health*, *95*(Suppl 1), S144–S150.

Surgeon General. (1964). Advisory committee of the USPHS: smoking and health. PHS Publication No 1103, Washington, DC.

Chapter 4

Altman, D. G., Machin, D., Bryant, T. N., & Gardner, M. J. (2000). *Statistics with confidence* (2nd ed.). London: BMJ Books.

Bonner, B. C., Clarkson, J. E., Dobbyn, L., & Khanna S. (2006). Slow-release fluoride devices for the control of dental decay. *Cochrane Database of Systematic Reviews*, Issue 4. Art. No.: CD005101. DOI: 10.1002/14651858.CD005101.pub2.

Burt, B. A., Keels, M. A., & Heller, K. E. (2003, January). Fluorosis development in seven age cohorts after an 11-month break in water fluoridation. *J Dent Res*, *82*(1), 64–68.

Chattopadhyay, A., Caplan, D. J., Slade, G. D., Shugars, D. C., Tien, H-C., & Patton, L. L. (2005). Risk indicators for oral candidiasis and oral hairy leukoplakia in HIV-infected adults. *Community Dent Oral Epidemiol*, *33*, 35–44.

Chattopadhyay, A., & Ray, J. G. (2008, February). AgNOR cut-point to distinguish mild and moderate epithelial dysplasia. *J Oral Pathol Med*, *37*(2), 78–82.

Cochrane Collaboration. (2002). Combined studies—weighted averages. Retrieved July 8, 2009, from http://www.cochrane-net.org/openlearning/HTML/mod 12-3.ht

Goldstein, G. R. (2002, January). What is evidence based dentistry? *Dent Clin North Am*. *46*(1), 1–9, v.

Keenan, J. V., Farman, A. G., Fedorowicz, Z., & Newton, T. (2005). Antibiotic use for irreversible pulpitis. *Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: CD004969. DOI: 10.1002/14651858.CD004969.pub2.

Kobberling, J., Trampisch, H. J., & Windeler, J. (Eds.). (1990). Memorandum for the evaluation of diagnostic measures. *J Clin Chem Clin Biochem*, *28*, 873–879.

Moher, D., Tetzlaff, J., Tricco, A. C., Sampson, M., & Altman, D. G. (2007). Epidemiology and reporting characteristics of systematic reviews, *PLoS Med*, *4*(3), e78.

National Center for Health Statistics. (1996). NHANES III references and reports. Atlanta, GA: Centers for Disease Control and Prevention.

Poole, C. (2001). Low p-values or narrow confidence intervals: which are more durable? *Epidemilogy*, *12*, 291–294.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Yeung, A., Hitchings, J. L., Macfarlane, T. V., Threlfall, A., Tickle, M., & Glenny, A. M. (2005). Fluoridated milk for preventing dental caries. *Cochrane Database of Systematic Reviews*, Issue 3. Art. No.: CD003876. DOI: 10.1002/14651858 .CD003876.pub2.

Chapter 5

Beck, J. D., Caplan, D. J., Preisser, J. S., & Moss, K. (2006). Reducing the bias of probing depth and attachment level estimates using random partial-mouth recording. *Community Dent Oral Epidemiol*, *34*, 1–10.

Fenwick, J., Needleman, I. G., & Moles, D. R.(2008, September). The effect of bias on the magnitude of clinical outcomes in periodontology: a pilot study. *J Clin Periodontol*, *35*, 775–782.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*,153–161.

Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik 43*, 172–198. English translation by J. A. Wheeler & H. Zurek. Quantum Theory and Measurement. Princeton University Press, 1983.

Kingman, A., Susin, C., & Albandar, J. M. (2008). Effect of partial recording protocols on severity estimates of periodontal disease. *J Clin Periodontol*, *35*, 659–667.

Lee, J. Y., Rozier, R. G., Norton, E. C., & Vann, W. F., Jr. (2005). Addressing selection bias in dental health services research. *J Dent Res*, *84*, 942–946.

Leslie, S. & Thiebaud P. (2007). Using propensity scores to adjust for treatment selection bias. Paper 184-2007; SAS Global Forum 2007. Retrieved July 08, 2009, from http://www2.sas.com/proceedings/forum2007/184-2007.pdf

Longford, N.T. (1999). Selection bias and treatment heterogeneity in clinical trials. *Statistics in Medicine*, *18*, 1467–1474.

Miller, K. D., Rahman, Z. U., & Sledge, G.W., Jr. (2001). Selection bias in clinical trials. *Breast Dis*, *14*, 31–40.

Rosenthal, R. & Jacobson, L. (1992). *Pygmalion in the classroom* (Expanded ed.). New York: Irvington.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). Modern epidemiology (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Shelton, B. J., Gilbert, G. H., Lu, Z., Bradshaw, P., Chavers, L. S., & Howard, G. (2003, June 30). Comparing longitudinal binary outcomes in an observational oral health study. *Stat Med*, *22*, 2057–2070.

Wang, Y., Localio, R., & Rebbeck, T. R. (2005). Bias correction with a single null marker for population stratification in candidate gene association studies. *Hum Hered*, *59*, 165–175.

Chapter 6

Chattopadhyay, A., Caplan, D. J., Slade, G. D., Shugars, D. C., Tien, H-C., Patton, L. L. (2005). Risk indicators for oral candidiasis and oral hairy leukoplakia in HIV-infected adults. *Community Dent Oral Epidemiol*, *33*, 35–44.

Greenland, S., & Brumback, B. (2002). An overview of relations among causal modelling methods. *Int J Epidemiol*, *31*(5), 1030–1037.

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology, 10*(1), 37–48.

Hernán, M. A., Hernández-Díaz, S., Werler, M. M., & Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol. 155*(2), 176–184.

Merchant, A. T. & Pitiphat W. (2002). Directed acyclic graphs (DAGs): an aid to assess confounding in dental research. *Community Dent Oral Epidemiol*, *30*(6), 399–404.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. New York: Cambridge University Press.

Robins, J. M. (2001). Data, design, and background knowledge in etiologic inference. *Epidemiology*, *12*(3), 313–320.

Rothman, K. J., Greenland, S., Lash, T. L. (2008). *Modern epidemiology* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Sadler, T. W. (2006). *Langman's medical embryology* (10th ed.). Baltimore, MD: Lippincott Williams & Wilkins.

Weinberg, C. R. (1993). Toward a clearer definition of confounding. *Am J Epidemiol*, *137*(1), 1–8.

Chapter 7

Baier, D., & Wernecke, K. (2003). Studies in classification, data analysis, and knowledge organization. Innovations in classification, data science, and information systems. Proceedings of the 27th Annual Conference of the Gesellschaft für Klassifikation e.V., Brandenburg University of Technology, Cottbus, March 12–14, 2003.

Centers for Disease Control. (1993). IMP1.DOC. Retrieved July 08, 2009, from http://ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHANES/NHANESIII/7A/imp1/imp1-acc.pdf

Chattopadhyay, A., Arevalo, O., & Sohn W. (2008, April). Understanding measurement of dental diseases and research participation in practice set-up. *Dent Clin North Am*, *52*(2), 367–386.

Chattopadhyay, A., Souza, T., & Arevalo, O. A. (2009). Electronic health records—development, adoption and utility for academic and practice based research. In

A. Dasakalaki (Ed.), *Handbook of research on dental computing and applications: advanced techniques for clinical dentistry*. Hershey, PA: IGI Publishers.

Hartley, A., Ho, K. F., McConkey, C., & Geh, J. I. (2005, October). Pathological complete response following preoperative chemoradiotherapy in rectal cancer: analysis of phase II/III trials. *Br J Radiol*. *78*(934), 934–938.

Wang, T., Zhu, G., & Keen, K. J. (2003). Genome-wide linkage analysis of the tracking of systolic blood pressure using a mixed model. *BMC Genet*, *4*(Suppl 1), S88. DOI: 10.1186/1471-2156-4-S1–S88.

World Health Organization (WHO). (2006). Electronic health records: manual for developing countries. Manila, Philippines: Author.

Chapter 8

Bonate, P. L. (2000). *Analysis or pretest-posttest designs*. Boca Raton, FL: Chapman & Hall/CRC.

Chattopadhyay, A., Caplan, D.J., Slade, G.D., Shugars, D.C., Tien, H-C, Patton, L.L. (2005). Incidence of oral candidiasis and oral hairy leukoplakia in HIV-infected adults in North Carolina. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, *99*(1), 39–47.

Goodman, S. N. (1999a). Towards evidence-based medical statistics. I. The p-value fallacy. *Ann Intern Med*, *130*, 995–1004.

Goodman, S. N. (1999b). Towards evidence-based medical statistics. II. The Bayes factor. *Ann Intern Med*, *130*, 1005–1013.

Killeen, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological science: a journal of the American Psychological Society/APS*, *16*(5), 345–353.

Poole, C. (2001, May). Low p-values or narrow confidence intervals: which are more durable? *Epidemiology*, *12*(3), 291–294.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Selke, T., Bayarri, M. J., & Berger, J. (2001). Calibration of p-values for testing precise null hypotheses. *Am Stat*, *55*, 62–71.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical data analysis using the SAS system* (2nd ed.). Cary, NC: SAS Institute Inc.

Chapter 9

Bower, E. & Scambler, S. (2007). The contributions of qualitative research towards dental public health practice. *Community Dent Oral Epidemiol*, *35*, 161–169.

Carvalho, R. M., Costa, L. R., & Marcelo, V. C. (2008, November). Brazilian dental students' perceptions about medical emergencies: a qualitative exploratory study. *J Dent Educ*, *72*(11), 1343–1349.

Dunning, D. G., Durham, T. M., Lange, B. M., & Aksu, M. N. (2008). The state of the art in evaluating the performance of assistant and associate deans as seen by deans and assistant and associate deans. *J Dent Educ*, *72*(4), 458–471.

Einstein, A., Podolsky, B., & Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys Rev*, *47*, 777.

Eisner, E. W. (1996). Yes, but is it research? The conversation continues: should a novel count as a dissertation in education? Chicago: AERA Session 5.25.

Ellis, C. (1997). Evocative autoethnography: writing emotionally about our lives. In W. G. Tierney & Y. S. Lincoln (Eds.), *Representation and the text* (pp. 115–142). Albany, NY: SUNY Press.

Freeman, R., & Stevens, A. (2008, October). Nursing caries and buying time: an emerging theory of prolonged bottle feeding. *Community Dent Oral Epidemiol*, *36*(5), 425–433. Epub 2008 Apr 21.

Gardner, K. M., & Aleksejuniene, J. (2008, November). Quantitative and qualitative analysis of student feedback on eportfolio learning. *J Dent Educ*, *72*(11), 1324–1332.

Hertlein, K. M., Lambert-Shute, J., & Benson, K. (2004, September). Postmodern influence in family therapy research: reflections of graduate students. *The Qualitative Report*, *9*(3), 538–561.

Holmes, D. C., Boston, D. W., Budenz, A. W., & Licari, F. W. (2003, December). Clinical curriculum for the twenty-first century. *J Dent Educ*, *67*(12), 1299–1301.

Hymes, K.B., Cheung, T., Greene, J. B., Prose, N.S., Marcus, A., Ballard, H., et al. (1981). Kaposi's sarcoma in homosexual men: a report of eight cases. *Lancet*, *2*, 598–600.

Merriam, S. B. (Ed.). (2002, January). *Qualitative research in practice: examples for discussion and analysis*. San Francisco, CA: John Wiley & Sons.

Muglali, M., Koyuturk, A. E., & Sari, M. E. (2008, Spring). Deciduous tooth extraction and folkloric beliefs in the Samsun region. *J Hist Dent*., *56*(1), 24–30.

Nordenram, G., Norberg, A., & Bischofberger, E. (1994). Ethical aspects of dental care for demented patients. Methodological considerations. *Swed Dent J*, *18*(4), 155–164.

Richardson, L. (1994). Writing: a method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 923–948). Thousand Oaks, CA: Sage.

Chapter 10

Agency for Healthcare Research and Quality (AHRQ). (2008). Medical expenditure panel survey. Retrieved July 08, 2009, from http://www.meps.ahrq.gov/mepsweb/

Barmes, D. E. & Infirri, J. S. (1977, January). WHO activities in oral epidemiology. *Community Dent Oral Epidemiol*, *5*(1), 22–29.

Beltrán-Aguilar, E. D., Barker, L. K., Canto, M. T., Dye, B. A., Gooch, B. F., Griffin, S. O., et al. (2005, August 26). Centers for Disease Control and Prevention. Surveillance for dental caries, dental sealants, tooth retention, edentulism, and enamel fluorosis—United States, 1988–1994 and 1999–2002. *MMWR Surveill Summ*, *54*(3), 1–43.

Caplan, D. J., Slade, G. D., & Gansky, S. A. (1999, Winter). Complex sampling: implications for data analysis. *J Public Health Dent*, *59*(1), 52–59.

Centers for Disease Control and Prevention. (2008). National Oral Health Surveillance System. Retrieved September 24, 2008, from http://www.cdc.gov/nohss/

Chattopadhyay, A., Arevalo, O., & Sohn, W. (2008, April). Understanding measurement of dental diseases and research participation in practice set-up. *Dent Clin North Am*, *52*(2), 367–386.

Chattopadhyay, A., Caplan, D. J., & Slade, G. D. (2009). Professional charges not reimbursed to dentists in the US: evidence from Medical Expenditure Panel Survey, 1996. Community Dent Health [Epub ahead of print].DOI. 10.1922/CDH_2312Chattopadhyay07

Chattopadhyay, A., Slade, G. D., & Shugars, D. A. (2003). Charges for oral health care during a period of economic growth in the US: 1987–1996. *J Public Health Dent*, *63*(2), 104–111.

Kalton, G. (1983). *Introduction to survey sampling*. Thousand Oaks, CA: Sage.

National Center for Health Statistics (NCHS). (2008). Surveys and data collection systems. Retrieved July 08, 2009, from http://www.cdc.gov/nchs/surveys.htm

Slade, G. D., & Beck, J. D. (1999, Spring). Plausibility of periodontal disease estimates from NHANES III. *J Public Health Dent*, *59*(2), 67–72.

Thacker, S. B., & Berkelman, R. L. (1988). Public health surveillance in the United States. *Epidemiol Rev*, *10*, 164–190.

Winn, D. M., Johnson, C. L., & Kingman A. (1999, Spring). Periodontal disease estimates in NHANES III: clinical measurement and complex sample design issues. *J Public Health Dent*, *59*(2), 73–78.

World Health Organization (WHO). Oral health information systems. Retrieved September 24, 2008, from http://www.who.int/oral_health/action/information/surveillance/en/print.html

Chapter 11

American Society of Health-System Pharmacists. (1998). Suggested definitions and relationships among medication misadventures, medication errors, adverse drug events, and adverse drug reactions. *Am J Health-Syst Pharm*, *55*, 165–166.

Blackwelder, W. C. (2004). Current issues in clinical equivalence trials. *J Dent Res*. *83* Spec No C:C113–C115.

Boehm, I., & Bauer, R. (2000, February). Low-dose methotrexate controls a severe form of polyarteritis nodosa. *Arch Dermatol*. *136*(2), 167–169.

Bondemarka, L., Holmb, A. K., Hansenc, K., Axelssond, S., Mohline, B., Brattstromf, V., et al. (2007). Long-term stability of orthodontic treatment and patient satisfaction: a systematic review. *Angle Orthodontist*, *77*, 181–191.

Brazzelli, M., McKenzie, L., Fielding, S., Fraser, C., Clarkson, J., Kilonzo, M., et al. (2006). Systematic review of the effectiveness and cost-effectiveness of Heal-Ozone® for the treatment of occlusal pit/fissure caries and root caries. *Health Techy Assess*, *10*, 1–96.

Canner, P. L., Huang, Y. B., & Meinert, C. L. (1981a, September). On the detection of outlier clinics in medical and surgical trials: I. Practical considerations. *Control Clin Trials*, *2*(3), 231–240.

Canner, P. L., Huang, Y. B., & Meinert, C. L. (1981b, September). On the detection of outlier clinics in medical and surgical trials: II. Theoretical considerations. *Control Clin Trials*, *2*(3), 241–252.

CHMP Working Party on Efficacy of Medicinal Products. (1995). Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products. Note for guidance III/3630;92-EN. *Statistics in Med*, *14*, 1658–1682.

Ernst, E. (2007, May). Placebo: new insights into an old enigma. *Drug Discov Today*, *12*(9–10), 413418.

Food and Drug Administration. (1995). Guideline for industry. The extent of population exposure to assess clinical safety: for drugs intended for long-term treatment of non-life-threatening conditions. Retrieved July 08, 2009, from www .fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/ Guidances/ucm073083.pdf

Food and Drug Administration. (1998). Guidance for institutional review boards and clinical investigators 1998 update. Retrieved July 08, 2009, from http:// www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/ GuidancesInformationSheetsandNotices/ucm115664.htm

Food and Drug Administration (2009). Federal Food, Drug, and Cosmetic Act. Retrieved on July 08, 2009, from http:// www.fda.gov/opacom/laws/fdcact/ fdcact1.htm

Hartzema, A. G., Porta, M. S., & Tilson, H. H. (1987, September). Introduction to pharmacoepidemiology. *Drug Intell Clin Pharm*, *21*(9), 739–740.

Hickel, R., Roulet, J. F., Bayne, S., Heintze, S. D., Mjör, I. A., Peters, M., et al. (2007, March). Recommendations for conducting controlled clinical studies of dental restorative materials. *Clin Oral Investig*, *11*(1), 5–33.

Hujoell, P. P., Cunha-Cruz, J., Banting, D. W., & Loesche, W. J. (2006). Dental flossing and interproximal caries: a systematic review. *J Dent Res*, *85*, 298–305.

Kaitin, K. I. (1995). Worthwhile persistence—the process of drug development. *Odyssey*, *1*(3), 17–22.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S., (Eds.). (2000). To err is human: building a safer health system. Institute of Medicine. Washington, DC: National Academy Press.

McLaren, J. F., Burney, R. O., Milki, A. A., Westphal, L. M., Dahan, M. H., & Lathi, R. B. (2008). Effect of methotrexate exposure on subsequent fertility in women undergoing controlled ovarian stimulation. *Fertil Steril*, [epub ahead of print] DOI:10.1016/j.fertnstert.2008.07.009

Moore, G. E. (1965, April 19). Cramming more components onto integrated circuits. *Electronics*, *38*, 8.

Nicholson, J. W., & Czarnecka B. (2006). The clinical repair of teeth using direct filling materials: engineering considerations. *Proc IMECHE*, *220*, 635–645.

National Institutes of Health (NIH). (2008). Glossary of clinical trials terms. Retrieved July 08, 2009, from http://clinicaltrials.gov/ct2/info/glossary

Porta, M. S. & Hartzema, A. G. (1987, September). The contribution of epidemiology to the study of drugs. *Drug Intell Clin Pharm*, *21*(9), 741–747.

Robinson, P. G., Deacon, S. A., Deery, C., Heanue, M., Walmsley, A. D., Worthington, H. V., et al. (2005, April). Manual versus powered toothbrushing for oral health. *Cochrane Database Syst Rev*, *18*(2), CD002281.

Scully, C., & Bagan, J. V. (2004, July 1). Adverse drug reactions in the orofacial region. *Crit Rev Oral Biol Med*, *15*(4), 221–239.

Strom, B. L. (1994). What is pharmacoepidemiology? Chapter 1. In B. L. Strom (Ed.), *Pharmacoepidemiology* (2nd ed., pp. 3–13). New York: Churchill Livingstone.

Temple, R. (1982). Government viewpoint of clinical trials. *DIA*, 16, 10–17.

Tredwin, C. J., Scully, C., & Bagan-Sebastian, J. V. (2005, July). Drug-induced disorders of teeth. *J Dent Res*, *84*(7), 596–602.

Tulloch, J. F. C., Antczak-Bouckoms, A. A., & Tuncay, O. C. (1989). A review of clinical research in orthodontics. *Am J Orthod Dentofacial Orthop*, *95*, 499–504.

Valania, M. (2008). Quality control and assurance in clinical research—a system of checks and examinations that helps ensure the quality of clinical trials. Applied clinical trials online. Retrieved November 15, 2008, from http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/CRO%2FSponsor+Articles/Quality-Control-and-Assurance-in-Clinical-Research/ArticleStandard/Article/detail/310811

Walker, A. M. (2001, August/September). Pattern recognition in health insurance claims databases. *Pharmacoepidemiol Drug Saf*, *10*(5), 393–397.

World Medical Organization. (1996, December 7). Declaration of Helsinki. *British Medical Journal*, *313*(7070), 1448–1449.

Chapter 12

Anderson, D., Yu, T. W., McGregor, D. B. (1998). Comet assay responses as indicators of carcinogen exposure. (Review). *Mutagenesis*, *13*, 539–555.

Andrieu, N., & Goldstein, A. M. (1998). Epidemiologic and genetic approaches in the study of gene-environment interaction: an overview of available methods. *Epidemiol Rev*, *20*, 137–147.

Bacanu, S. A., Devlin, B. & Roeder, K. (2002). Association studies for quantitative traits in structured populations. *Genet Epidemiol*, *22*(1), 78–93.

Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., et al. (2000). Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*, *406*, 536–540.

Carr, K. M., Bittner, M., & Trent, J. M. (2003). Gene-expression profiling in human cutaneous melanoma. *Oncogene*, *22*, 3076–3080.

Collins, A.R. (2004). The comet assay for DNA damage and repair. (Review). *Molecular Biotechnology*, *26*, 249–261.

Dawes, C., & Wood, C. M. (1973). The contribution of oral minor mucous gland secretions to the volume of whole saliva in man. *Arch Oral Biol*, *18*, 337–342.

Devlin, B. & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, *55*(4), 997–1004.

Dorak, M. T. (2007). Statistical analysis in HLA and disease association studies. Retrieved July 08, 2009, from www.dorak.info/hla/stat.html

Dorak, M. T. (2009). Genetic epidemiology. Retrieved July 08, 2009, from http://www.dorak.info/epi/genetepi.html

Dowsett, S. A., Archila, L., Foroud, T., Eckert, G. J., & Kowolik, M. J. (2002). The effect of shared genetic and environmental factors on periodontal disease parameters in untreated adult siblings in Guatemala. *Jour of Periodontology*, *73*, 1160–1168.

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA*, *95*, 14863–14868.

Fairbairn, D. W., Olive, P. L., & O'Neill, K. L. (1995). The Comet assay: a comprehensive review. (Review). *Mutat Res*, *339*, 37–59.

Fanland, J., & Ren, Y. (2006). Statistical analysis of DNA microarray data in cancer research. *Clin Cancer Res*, *12*, 4469–4473.

Frank, R. & Hargreaves, R. (2003). Clinical biomarkers in drug discovery and development. *Nature Reviews Drug Discovery*, *2*, 566–580.

Goldstein, A. M., Falk, R. T., Korczak, J. F., & Lubin, J. H. (1997). Detecting gene-environment interactions using a case–control design. *Genetic Epidemiol*, *14*, 1085–1089.

Helmerhorst, E. J., & Oppenheim, F. G. (2007). Saliva: a dynamic proteome. *J Dent Res*, *86*(8), 680–693.

Hershberg, R., Lipatov, M., Small, P. M., Sheffer, H., Niemann, S., Homolka, S., et al. (2008). High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS Biol*, *6*, e311. DOI: 10.1371/journal .pbio.0060311

Hu, S., Wang, J., Meijer, J., Ieong, S., Xie, Y., Yu, T., et al. (2007). Salivary proteomic and genomic biomarkers for primary Sjögren's dyndrome. *Arthritis and Rheumatism*, *56*, 3588–3600.

Human Genome Project. (2008). Facts about genome sequencing. Retrieved July 08, 2009, from http://www.ornl.gov/sci/techresources/Human_Genome/faq/seqfacts.shtml

Janes, H., & Pepe M. (2006). The optimal ratio of cases to controls for estimating the classification accuracy of a biomarker. *Biostatistics*, *7*(3), 456–468. DOI: 10.1093/biostatistics/kxj018

Khoury, M., & Flanders, D. (1996). Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case–control studies with no controls. *Am J Epidemiol*, *144*, 207–213.

Khoury, M. J., & Wagener, D. K. (1995). An epidemiologic evaluation of the use of genetics to improve the predictive value of disease risk factors. *Am J Hum Genet*, *56*, 835–844.

Levin, B. R., Lipsitch, M., & Bonhoeffer, S. (1999). Population biology, evolution, and infectious disease: convergence and synthesis. *Science*, *283*, 806–809.

McKee, J. K. (2005). A genetic model of dental reduction through the probable mutation effect. *Am J of Phys Anthro*, *65*, 231–241.

Messana, I., Inzitari, R., Fanali, C., Cabras, T., & Castagnola, M. (2008). Facts and artifacts in proteomics of body fluids. What proteomics of saliva is telling us. *J Sep Sci*, *31*, 1948–1963.

Millikan, R. (2001). Re: population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Ints*, *93*, 156.

Moller, P., Knudsen, L. E., Loft, S., & Wallin, H. (2000, October). The comet assay as a rapid test in biomonitoring occupational exposure to DNA-damaging agents and effect of confounding factors. *Cancer Epidemiol Biomarkers Prev*, *9*(10), 1005–1015.

Morton, N. E. (1997). Genetic epidemiology. *Ann Hum Genet*, *61*, 1–13.

Morton, N. E., & Collins, A. E. (1995). Statistical and genetic aspects of quality control for DNA identification. *Electrophoresis*, *16*(9), 1670–1677.

National Human Genome Research Institute. (2008). Methods for discovering and scoring single nucleotide polymorphisms. Retrieved July 08, 2009, from http://www.genome.gov/10001029

Östling, O. & Johanson, K. J. (1984). Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochem Biophys Res Commun*, *123*, 291–298.

Pan, Z., Trikalinos, T. A., Kavvoura, F. K., Lau, J., & Ioannidis, J. P. (2005) Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. *PLoS Med*, *2*(12), e334. DOI: 10.1371/journal.pmed.0020334

Quackenbush, J. (2001). Computational analysis of microarray data. *Nat Rev Genet*, *2*, 418–427.

Rabkin, C. S., & Rothman, N. (2001). Introduction to molecular epidemiology. In M. Carrington & A. R. Hoelzel (Eds.), *Molecular epidemiology a practical approach* (pp. xvii–xix). Oxford: Oxford University Press.

Rojas, E., Lopez, M. C., & Valverde, M. (1999). Single cell gel electrophoresis. Methodology and applications. (Review). *J Chromatogr B Biomed Sci Appl*, *722*, 225–254.

Rothman, N., Wacholder, S., Caporaso, N. E., Garcia-Closas, M., Buetow, K., & Fraumeni, J. F., Jr. (2001). The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens. *Biochim Biophys Acta*, *1471*(2), C1–C10.

Schulte, P. A. (1993). A conceptual and historical framework for molecular epidemiology. In P. A. Schulte & F. P. Perera (Eds.), *Molecular epidemiology: principles and practices* (pp. 3–44). San Diego, CA: Academic Press.

Schwartz, J. L. (2000). Biomarkers and molecular epidemiology and chemoprevention of oral carcinogenesis. *Crit Rev Oral Biol Med*, *11*, 92.

Sham, P., Bader, J. S., Craig, I., O'Donovan, M., & Owen, M. (2002, November). DNA pooling: a tool for large-scale association studies. *Nat Rev Genet*, *3*(11), 862–871.

Singh, N. P., McCoy, M. T., Tice, R. R., & Schneider, E.L. (1988). A simple technique for quantitation of low levels of DNA damage in individual cells. *Exp Cell Res*, *175*, 184–191.

Siqueira, W. L., Salih, E., Wan, D. L., Helmerhorst, E. J., & Oppenheim, F. G. (2008). Proteome of human minor salivary gland secretion. *J Dent Res*, *87*(5), 445–450.

Stodolsky, M. (2008). Whose genome was sequenced in the public (HGP) and private projects? Retrieved July 08, 2009, from http://www.ornl.gov/sci/techresources/Human_Genome/faq/seqfacts.shtml

Stringer, C. B., Humphrey, L. T., & Compton, T. (1997). Cladistic analysis of dental traits in recent humans using a fossil outgroup. *J Human Evolution*, e*32*, 389–402.

Wacholder, S., Rothman, N., & Caporaso, N. (2000). Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst*, *92*, 1151–1158.

Wang, J. (2008). Computational biology of genome expression and regulation—a review of microarray bioinformatics. *J Environ Path, Toxicol, and Oncol*, *27*(3), 157–179.

Weeraratna, A. T., Nagel, J. E., DeMello-Coelho, V., & Taub, D. D. (2004). Gene expression profiling: from microarrays to medicine. *J Clin Immunol*, *24*, 213–224.

Yang, Q. & Khoury, M. (1997). Evolving methods in genetic epidemiology. III. Gene environment interaction in epidemiologic research. *Epidemiol Rev*, *19*, 33–43.

Chapter 13

American Academy of Pediatric Dentistry. (2006). Policy on use of a caries-risk assessment tool (CAT) for infants, children, and adolescents. Retrieved July 08,

2009, from http://www.aapd.org/media/Policies_Guidelines/P_CariesRiskAssess.pdf

Ansai, T., Tahara, A., Ikeda, M., Katoh, Y., Miyazaki, H., & Takehara, T. (2000). Caries risk in Japanese pre-school children. *Pediatric Dent*, *22*, 377–380.

Assaf, A.V., de Castro Meneghim, M., Zanin, L., Tengan, C., & Pereira, A. C. (2006). Effect of different diagnostic thresholds on dental caries calibration—a 12-month evaluation. *Community Dent Oral Epidemiol*, *34*, 213–219.

Bader, J. D., Shugars, D. A., & Bonito, A. J. (2001a). Systematic reviews of selected dental caries diagnostics and management methods. *J Dent Educ*, *65*, 960–968.

Bader, J. D., Shugars, D. A., & Bonito, A. J. (2001b, December). A systematic review of selected caries prevention and management methods. *Community Dent Oral Epidemiol*, *29*(6), 399–411.

Bader, J. D., Shugars, D. A., & Bonito, A. J. (2002, Fall). A systematic review of the performance of methods for identifying carious lesions. *J Public Health Dent*, *62*(4), 201–213.

Beck, J. D., & Drake, C. W. (1997, Spring). Do root lesions tend to develop in the same people who develop coronal lesions? *J Public Health Dent*, *57*(2), 82–8.

Beighton, D., Brailsford, S., Samaranayake, L. P., Brown, J. P., Ping, F. X., Grant-Mills, D., et al. (2004, March). A multi-country comparison of caries-associated microflora in demographically diverse children. *Community Dent Health*, *21*(1 Suppl), 96–101.

Bratthall, D., & Hänsel Petersson, G. (2005). Cariogram—a multifactorial risk assessment model for a multifactorial disease. *Community Dent Oral Epidemiol*, *33*, 256–264.

Broadbent, J. M., & Thomson, W. M. (2005). For debate: problems with the DMF index pertinent to dental caries data analysis. *Community Dent Oral Epidemiol*, *33*, 400–409.

Caplan, D. J., Slade, G. D., Biesbrock, A. R., Bartizek, R. D., McClanahan, S. F., & Beck, J. D. (1999). A comparison of increment and incidence density analyses in evaluating the anticaries effects of two dentifrices. *Caries Res*, *33*, 16–22.

Cawson, R. A. & Odell, E. W. (2008). *Cawson's Essentials of Oral Pathology and Oral Medicine*. Oxford, UK: Churchill Livingstone, Inc.

Chattopadhyay, A., Arevalo, O., & Sohn, W. (2008, April). Understanding measurement of dental diseases and research participation in practice set-up. *Dent Clin North Am*, *52*(2), 367–386.

Denny P. C., Denny P. A., Takashima J., Galligan J., & Navazesh M. (2007). A novel caries risk test. *Ann N Y Acad Sci*, *1098*, 204–215.

Hannigan, A., O'Mullane, D. M., Barry, D., Schafer, F., & Roberts, A. J. (2000). A caries susceptibility classification of tooth surfaces by survival time. *Caries Res*, *34*, 103–108.

Harris, R., Nicoll, A. D., Adair, P. M., & Pine, C. M. (2004, March). Risk factors for dental caries in young children: a systematic review of the literature. *Community Dent Health*, *21*(1 Suppl), 71–85.

Hujoel, P. P., Isokangas, P. J., Tiekso, J., Davis, S., Lamont, R. J., DeRouen, T. A., et al. (1994). A re-analysis of caries rates in a preventive trial using Poisson regression models. *J Dent Res*, *73*, 573–579.

International Caries Detection and Assessment System Coordinating Committee. (2005). Rationale and evidence for the International Caries Detection and Assess-

ment System (ICDAS II). Retrieved July 08, 2009, from http://www.dundee.ac.uk/dhsru/docs/Final%20ICDAS%20II%20criteria%20document.doc

Ismail, A. I. (2004). Visual and visuo-tactile detection of dental caries. *J Dent Res*, *82*, C56–C66.

Ismail, A. I., Sohn, W., Tellez, M., Amaya, A., Sen, A., Hasson, H., et al. (2007). The International Caries Detection and Assessment System (ICDAS): an integrated system for measuring dental caries. *Community Dent Oral Epidemiol*, *35*, 170–178.

Johnson, M. F. (2004). The role of risk factors in the identification of appropriate subjects for caries clinical trials: design considerations. *J Dent Res*, *83*(Spec No C), C116–C118.

Katz, R. V. (1996, Winter). The RCI revisited after 15 years: used, reinvented, modified, debated, and natural logged. *J Public Health Dent*, *56*(1), 28–34.

Lawrence, H. P., Beck, J. D., Hunt, R. J., & Koch, G. G. (1996, October). Adjustment of the M-component of the DMFS index for prevalence studies of older adults. *Community Dent Oral Epidemiol*, *24*(5), 322–331.

Manton, D. J., & Messer, L. B. (2007, March). The effect of pit and fissure sealants on the detection of occlusal caries in vitro. *Eur Arch Paediatr Dent*, *8*(1), 43–48.

Messer, L. B. (2000, March). Assessing caries risk in children. *Aust Dent J*, *45*(1), 10–16.

Muller-Bolla, M., Lupi-Pégurier, L., Tardieu, C., Velly, A. M., & Antomarchi, C. (2006). Retention of resin-based pit and fissure sealants: a systematic review. *Community Dent Oral Epidemiol*, *34*, 321–336.

Offner, G. D, & Troxler, R. F. (2000, December). Heterogeneity of high-molecular-weight human salivary mucins. *Adv Dent Res*, *14*, 69–75.

Park, W. K., Chung, J. W., Kim, Y. K., Chung, S. C., & Kho, H. S. (2006, October). Influences of animal mucins on lysozyme activity in solution and on hydroxyapatite surfaces. *Arch Oral Biol*, *51*(10), 861–869.

Powell, L. V., Leroux, B. G., Persson, R. E., & Kiyak, H. A. (1998). Caries risk in the elderly. *Community Dent Oral Epidemiol*, *26*, 170–176.

Sanchez-Figueras, A., Jr. (2003). Occlusal pit-and-fissure caries diagnosis: a problem no more. A science-based diagnostic approach using a laser-based fluorescence device. *Compend Contin Educ Dent*, *24*(5 Suppl), 3–11.

Shuler, C. F. (2001). Inherited risks for susceptibility of dental caries. *J Dent Educ*, *65*, 1038–1045.

Silva, D. G., Stevens, R. H., Macedo, J. M., Hirata, R., Pinto, A. C., Alves, L. M., et al. (2009). Higher levels of salivary MUC5B and MUC7 in individuals with gastric diseases who harbor Helicobacter pylori. *Arch Oral Biol*, *54*(1), 86–90.

Stookey, G. K. (2005, October). Quantitative light fluorescence: a technology for early monitoring of the caries process. *Dent Clin North Am*, *49*(4), 753–770.

Stookey, G. K., Jackson, R. D., Zandona, A. G., & Analoui, M. (1999, October). Dental caries diagnosis. *Dent Clin North Am*, *43*(4), 665–677.

Tranæus, S., Shi, X-Q., & Angmar-Månsson, B. (2005). Caries risk assessment: methods available to clinicians for caries detection. *Community Dent Oral Epidemiol*, *33*, 265–273.

Turssi, C. P., Lima, R. Q., Faraoni-Romano, J. J., & Serra, M. C. (2006, December). Rehardening of caries-like lesions in root surfaces by saliva substitutes. *Gerodontology*, *23*(4), 226–230.

U.S. Department of Health and Human Services. (2000). *Oral health in America: a report of the surgeon general*. Rockville, MD: U.S. Department of Health and Human Services, National Institute of Dental and Craniofacial Research, National Institutes of Health.

Walls, A. W., Silver, P. T., & Steele, J. G. (2000, February). Impact of treatment provision on the epidemiological recording of root caries. *Eur J Oral Sci*, *108*(1), 3–8.

Wickström, C., & Svensäter, G. (2008, June). Salivary gel-forming mucin MUC5B—a nutrient for dental plaque bacteria. *Oral Microbiol Immunol*, *23*(3), 177–182.

Chapter 14

American Academy of Periodontology. (1989). Consensus report. M. Nevins, W. Becker, & K. Kornman (Eds.), Proceedings of the World Workshop in Clinical Periodontics. Chicago: American Academy of Periodontology.

Armitage, G. C. (2002). Classifying periodontal diseases—a long-standing dilemma. *Periodontology 2000, 30*, 9–23.

Armitage, G. C. (2004). Periodontal diagnosis and classification of periodontal diseases. *Periodontology 2000*. *34*, 9–21.

Axtelius, B., SoÈderfeldt, B., & AttstroÈm, R. (1999). A multilevel analysis of factors affecting pocket probing depth in patients responding differently to periodontal treatment. *J Clin Periodontol*, *26*, 67–76.

Beck, J. D., Caplan, D. J., Preisser, J. S., & Moss, K. (2006). Reducing the bias of probing depth and attachment level estimates using random partial-mouth recording. *Community Dent Oral Epidemiol*, *34*, 1–10.

Blicher, B., Joshipura, K., & Eke, P. (2005). Validation of self-reported periodontal disease: a systematic review. *J Dent Res*, *84*(10), 881–890.

Borrell, L. N., Burt, B. A., & Taylor, G. W. (2005). Prevalence and trends in periodontitis in the USA: from the NHANES III to the NHANES, 1988 to 2000. *J Dent Res*, *84*(10), 924–930.

Cohen, M. E. (1993, November). Bursts of periodontal destruction and remission, percolation phase shifts, and chaos. *J Periodontal Res*, *28*(6 Pt 1), 429–436.

Dietrich, T., Stosch, U., Dietrich, D., Kaiser, W., Bernimoulin, J. P., & Joshipura, K. (2007, July). Prediction of periodontal disease from multiple self-reported items in a German practice-based sample. *J Periodontol*, *78*(7 Suppl), 1421–1428.

Dye, B. A., & Thornton-Evans, G. (2007, July). A brief history of national surveillance efforts for periodontal disease in the United States. *J Periodontol*, *78*(7 Suppl), 1373–1379.

Eke, P. I. (2005). Public health implications of periodontal infections in adults. Conference proceedings. *J Public Health Dent*, *65*, 56–65.

Eke, P. I., & Genco, R. J. (2007, July). CDC Periodontal Disease Surveillance Project: background, objectives, and progress report. *J Periodontol*, *78*(7 Suppl), 1366–1371.

Genco, R. J., Falkner, K. L., Grossi, S., Dunford, R., & Trevisan, M. (2007, July). Validity of self-reported measures for surveillance of periodontal disease in two western New York population-based studies. *J Periodontol*, *78*(7 Suppl), 1439–1454.

Gilbert, G. H., & Litaker, M. S. (2007, July). Validity of self-reported periodontal status in the Florida dental care study. *J Periodontol*, *78*(7 Suppl), 1429–1438.

LaVange, L. M., & Koch, G. G. (2007, July). Statistical projection of clinical subsample estimates to a survey population. *J Periodontol*, *78*(7 Suppl), 1400–1406.

Löe, H., & Silness, J. (1963). Periodontal disease in pregnancy. I. Prevalence and severity. *Acta Odontol Scand*, *21*, 533–551.

Miller, K., Eke, P. I., & Schoua-Glusberg, A. (2007, July). Cognitive evaluation of self-report questions for surveillance of periodontitis. *J Periodontol*, *78*(7 Suppl), 1455–1462.

Oringer, R. J., & Williams, R. C. (2000, October). Assessment and management of periodontal infections: a medical-surgical approach. *Compend Contin Educ Dent*, *21*(10A), 906–914, 916–918.

Page, R. C., Eke, P. I. (2007, July). Case definitions for use in population-based surveillance of periodontitis. (Review). *J Periodontol*, *78*(7 Suppl), 1387–1399.

Slade, G. D. (2007, July). Interim analysis of validity of periodontitis screening questions in the Australian population. *J Periodontol*, *78*(7 Suppl), 14631470.

Slots, J. (2007). Herpesviral-bacterial synergy in the pathogenesis of human periodontitis. *Curr Opin Infect Dis, 20,* 278–283.

Slots, J., Saygun, I., Sabeti, M., & Kubar, A. (2006, August). Epstein-Barr virus in oral diseases. *J Periodontal Res*, *41*(4), 235–244.

Taylor, G. W., & Borgnakke, W. S. (2007, July). Self-reported periodontal disease: validation in an epidemiological survey. *J Periodontol*, *78*(7 Suppl), 1407–1420.

Tomar, S. L. (2007, July). Public health perspectives on surveillance for periodontal diseases. (Review). *J Periodontol*, *78*(7 Suppl), 1380–1386.

Tu, Y. K., Gilthorpe, M. S., Griffiths, G. S., Maddick, I. H., Eaton, K. A., & Johnson, N. W. (2004a, January). The application of multilevel modeling in the analysis of longitudinal periodontal data—part I: absolute levels of disease. *J Periodontol*, *75*(1), 127–136.

Tu, Y. K., Gilthorpe, M. S., Griffiths, G. S., Maddick, I. H., Eaton, K. A., & Johnson, N. W. (2004b, January). The application of multilevel modeling in the analysis of longitudinal periodontal data—part II: changes in disease levels over time. *J Periodontol*, *75*(1), 137–145.

Chapter 15

Arbes, S. J., Jr, Olshan, A. F., Caplan, D. J., Schoenbach, V. J., Slade, G. D., & Symons, M. J. (1999, December). Factors contributing to the poorer survival of black Americans diagnosed with oral cancer (United States). *Cancer Causes Control*, *10*(6), 513–523.

Bagan, J. V., & Scully, C. (2008, February). Recent advances in oral oncology 2007: epidemiology, aetiopathogenesis, diagnosis and prognostication. *Oral Oncol*, *44*(2), 103–108.

Berwick, M. & Vineis, P. (2000a). Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. *J Natl Cancer Inst*, *92*(11), 874–897.

Berwick, M. & Vineis, P. (2000b). RESPONSE re: markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. *J Natl Cancer Inst*, *92*(18), 1537.

Blot, W. J., McLaughlin, J. K., Winn, D. M., Austin, D. F., Greenberg, R. S., Preston-Martin, S., et al. (1998). Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer Res*, *48*, 3282–3287.

Brennan, J. A., Boyle, J. O., Koch, W. M., Goodman, S. N., Hruban, R. H., Eby, Y. J., et al. (1995). Association between cigarette smoking and mutations of the p53 gene in squamouscell carcinoma of the head and neck. *N Engl J Med*, *332*, 712–717.

Brennan, P., Lewis, S., Hashibe, M., Bell, D. A., Boffetta, P., Bouchardy, C., et al. (2004). Pooled analysis of Alcohol dehydrogenase genotypes and head and neck cancer: a huge review. *Am J Epidemiol*, *59*, 1–16.

Burzynski, N. J., Rankin, K.V., Silverman, S., Jr., Scheetz, J. P., & Jones, D. L. (2002, Summer). Graduating dental students' perceptions of oral cancer education: results of an exit survey of seven dental schools. *J Cancer Educ*, *17*(2), 83–84.

Bykov, V. J., Jansen, C. T., & Hemminki, K. (1998). High levels of dipyrimidine dimers are induced in human skin by solar-simulating UV radiation. *Cancer Epidemiol Biomarkers Prev*, *7*, 199–202.

Calabrese, L., Bruschini, R., Ansarin, M., Giugliano, G., De Cicco, C., Ionna, F., et al. (2006, December). Role of sentinel lymph node biopsy in oral cancer. *Acta Otorhinolaryngol Ital*, *26*(6), 345–349.

Castellsagué, X., Quintana, M. J., Martínez, M. C., Nieto, A., Sánchez, M. J., Juan, A., et al. (2004). The role of type of tobacco and type of alcoholic beverage in oral carcinogenesis. *Int J Cancer*, *108*, 741–749.

Chang, F., Syrjänen, S., Kellokoski, J., & Syrjänen, K. (1991). Human papilloma virus (HPV) infections and their associations with oral disease. *J Oral Pathol Med*, *20*, 305–317.

Chang, F., Syrjänen, S., Nuutinen, J., Kärjä, J., & Syrjänen, K. (1990). Detection of human papillomavirus (HPV) DNA in oral squamous cell carcinomas by in situ hybridization and polymerase chain reaction. *Arch Dermatol Res*, *282*, 493–497.

Chattopadhyay, A. (1989). Epidemiologic study of oral cancer in eastern India. *Indian J Dermatol*, *34*, 59–65.

Cheng, L, Sturgis, E. M., Eicher, S. A., Spitz, M. R., & Wei, Q. (2002). Expression of nucleotide excision repair genes and the risk for squamous cell carcinoma of the head and neck. *Cancer*, *94*(2), 393–397.

Coleman, M. P., Gatta, G., Verdecchia, A., Estève, J., Sant, M., Storm, H., et al. (2003). EUROCARE working group. EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century. *Ann Oncol*, *14*(Suppl 5Lv), 128149.

Daly, M. B. (1993). The chemopreventive of cancer: directions for the future. *Cancer Epidemiol Biomark Prev*, *2*, 509–512.

D'Silva, N. J., & Ward, B. B. (2007). Tissue biomarkers for diagnosis and management of oral squamous cell carcinoma. *Alpha Omegan*, *100*(4), 182–189.

Elting, L. S., Avritscher, E. B., Cooksley, C. D., Cardenas-Turanzas, M., Garden, A. S., & Chambers, M. S. (2008, January). Psychosocial and economic impact of cancer. *Dent Clin North Am*, *52*(1), 231–252.

Epstein, J. B., Gorsky, M., Cabay, R. J., Day, T., & Gonsalves, W. (2008, June). Screening for and diagnosis of oral premalignant lesions and oropharyngeal squamous cell carcinoma: role of primary care physicians. *Can Fam Physician*, *54*(6), 870–875.

Franceschi, S., Talamini, R., Barra, S., Barón, A. E., Negri, E., Bidoli, E., et al. (1990). Smoking and drinking in relation to cancers of the oral cavity, pharynx, larynx, and esophagus in northern Italy. *Cancer Res*, *50*, 6502–6507.

Franceschi, S., Favero, A., Conti, E., Talamini, R., Volpe, R., Negri, E., et al. (1999a). Food groups, oils and butter, and cancer of the oral cavity and pharynx. *Br J Cancer*, *80*, 614–620.

Franceschi, S., Levi, F., Conti, E., Talamini, R., Negri, E., Dal Maso, L., et al. (1999b). Energy intake and dietary pattern in cancer of the oral cavity and pharynx. *Cancer Causes Control*, *10*, 439–444.

Franco, E. L., Kowalski, L. P., Oliveira, B. V., Curado, M. P., Pereira, R. N., Silva, M. E., et al. (1989). Risk factors for oral cancer in Brazil: a case–control study. *Int J Cancer*, *43*, 992–1000.

Goode, E. L., Ulrich, C. M., & Potter, J. D. (2002). Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol Biomark Prev, II,* 1513–1530.

Greenspan, D., D' Villiers, E. M., Greenspan, J. S., D'Souza, Y. G., & Zur, H. H. (1988). Unusual HPV types in oral warts in association with HIV infection. *J Oral Pathol*, *17*, 482–488.

Greenwald, P., Kelloff, G. C., Boone, C. W., & McDonald, S. S. (1995). Genetic and cellular changes in colorectal cancer: proposed targets of chemopreventive agents. *Cancer Epidemiol Biomark Prev*, *4*, 691–702.

Gupta, P. C., Murti, P. R., & Bhonsle, R.B. (1996). Epidemiology of cancer by tobacco products and the significance of TSA. *Crt Rev Toxicol*, *26*, 183–198.

Ha, P. K., & Califano, J. A. (2004). The role of human papillomavirus in oral carcinogenesis. *Crit Rev Oral Biol Med*, *15*, 188–196.

Hagensee, M. E., Cameron, J. E., Leigh, J. E., & Clark, R. A. (2004, July). Human papilloma virus infection and disease in HIV-infected individuals. *Am J Med Sci*, *328*(1), 57–63.

Hashibe, M., Brennan, P., Strange, R. C., Bhisey, R., Cascorbi, I., Lazarus, P., et al. (2003). Meta- and pooled analysis of GSTM1, GSTT1, GSTP1, and CYAP1A1 genotypes and the risk of head and neck cancer. *Cancer Epidemiol Biomarkers Prev*, *12*, 1509–1517.

Hemminki, K., Xu, G., & Le Curieux, F. (2000). Re: markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. *J Natl Cancer Inst*, *92*(18), 1536–1537.

Herrero, R., Castellsagué, X., Pawlita, M., Lissowska, J., Kee, F., Balaram, P., et al. (2003). IARC multicenter oral cancer study group. Human papilloma virus and oral cancer: the International Agency for Research on Cancer multicenter study. *J Natl Cancer Inst*, *95*, 1772–1783.

Horner, M. J., Ries, L. A. G., Krapcho, M., Neyman, N. Aminou, R., Howlader, N., et al. (eds). (2009). *SEER cancer statistics review 1975–2006*. Bethesda, MD: National Cancer Institute. Retrieved August 12, 2009, from http://seer.cancer.gov/csr/1975–2006/

Hsieh, L. L., Chien, H. T., Chen, I. H., Liao, C. T., Wang, H. M., Jung, S. M., et al. (2003). The XRCC1 399Gln polymorphism and the frequency of p53 mutations in Taiwanese oral squamous cell carcinomas. *Cancer Epidemiol Biomarkers Prev*, *12*(5), 439–443.

Huang, B., Valentino, J., Wyatt, S. W., & Gal, T. J. (2008, August). Incidence of oral cavity and pharynx cancer in Kentucky. *J Ky Med Assoc*, *106*(8), 355–360.

International Agency for Research on Cancer. (1991). *Monographs on the evaluation of carcinogenic risks to humans. Coffee, tea, mate, methylxanthines, and methylglyoxal*. Volume 51. Lyon: IARC Publications.

International Agency for Research on Cancer. (2003). *Handbooks of cancer prevention. Fruits and vegetables*. Volume 8. Lyon: IARC Publications.

Jemal, A., Siegel, R., Ward, E., Murray, T., Xu J., Smigal, C., et al. (2006). Cancer statistics. *CA Cancer J Clin*, *56*, 106–130.

Kansky, A. A., Poljak, M., Seme, K., Kocjan, B. J., Gale, N., Luzar, B., et al. (2003). Human papillomavirus DNA in oral squamous cell carcinomas and normal oral mucosa. *Acta Virol*, *47*, 11–16.

Katz, D. L. (2001). Clinical epidemiology and evidence-based medicine fundamental principles of clinical reasoning and research. Thousand Oaks, CA: Sage.

Kenady, D. E., Arnold, S. M., Regine, W. F., & Mentzer, R. M., Jr. (2002, November). Oral/pharyngeal squamous carcinoma: treatment strategies. *J Ky Med Assoc*, *100*(11), 488–494.

Kujan, O., Glenny, A. M., Duxbury, J., Thakker, N., & Sloan, P. (2005). Evaluation of screening strategies for improving oral cancer mortality: a Cochrane systematic review. *J Dent Educ*, *69*, 255–265.

Lambropoulos, A. F., Dimitrakopoulos, J., Frangoulides, E., Katopodi, R., Kotsis, A., & Karakasis, D. (1997). Incidence of human papillomavirus 6, 11, 16, 18 and 33 in normal oral mucosa of a Greek population. *Eur J Oral Sci*, *105*, 294–297.

Lavelle, C. L. B. & Scully, C. (2005). Criteria to rationalize population screening to control oral cancer. *Oral Oncology*, *41*, 11–16.

Lee, J. W., Soung, Y. H., Kim, S. Y., Nam, H. K., Park, W. S., & Nam, S. W. (2005). Somatic mutations of EGFR gene in squamous cell carcinomas of head and neck. *Clin Cancer Res*, *11*, 2879–2882.

McLeod, N. M. H., Saeed, N. R., & Ali, E. A. (2005). Oral cancer: delays in referral and diagnosis persist. *British Dental Journal*, *198*, 681–684.

Merchant, A., Husain, S. S., Hosain, M., Fikree, F. F., Pitiphat, W., Siddiqui, A. R., et al. (2000). Paan without tobacco: an independent risk factor for oral cancer. *Int J Cancer*, *86*, 128–131.

Mignogna, M. D., Fedele, S., Russo, L.Lo., Ruoppo, E., & Muzio, L.Lo. (2001). Oral and pharyngeal cancer: lack of prevention and early detection by health care providers. *European J of Cancer Prevent*, *10*, 381–383.

Miller, C. S., & White, D. K. (1996). Human papillomavirus expression in oral mucosa, premalignant conditions, and squamous cell carcinoma: a retrospective review of the literature. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, *82*, 57–68.

Oh, T. J., Kim, C. J., Woo, S. K., Kim, T. S., Jeong, D. J., Kim, M. S., et al. (2004). Development and clinical evaluation of a highly sensitive DNA microarray for detection and genotyping of human papilloma viruses. *J Clin Microbiol*, *42*, 3272–3280.

Olshan, A. F., Watson, M. A., Weissler, M. C., & Bell, D. A. (2002). XRCC1 polymorphisms and head and neck cancer. *Cancer Lett*, *178*(2), 181–186.

Page, D. L. (1994). The epidemiology of tumor markers in breast cancer management: prognostic markers. *Cancer Epidemiol Biomark Prev*, *3*, 101–104.

Patton, L. L., Ashe, T. E., Elter, J. R., Southerland, J. H., & Strauss, R. P. (2006). Adequacy of training in oral cancer prevention and screening as self-assessed by physicians, nurse practitioners, and dental health professionals. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, *102*, 758–764.

Patton, L. L., Epstein, J. B., & Kerr, A. R. (2008, July). Adjunctive techniques for oral cancer examination and lesion diagnosis: a systematic review of the literature. *J Am Dent Assoc*, *139*(7), 896–905.

Prime, S. S., Thakker, N. S., Pring, M., Guest, P. G., & Paterson, I. C. (2001). A review of inherited cancer syndromes and their relevance to oral squamous cell carcinoma. *Oral Oncol*, *37*, 1–16.

Ragin, C. C. R., Modugno, F., & Gollin, S. M. (2007). The epidemiology and risk factors of head and neck cancer: a focus on human papillomavirus. *J Dent Res*, *86*(2), 104–114.

Remmerbach, T. W., Brinckmann, U. G., Hemprich, A., Chekol, M., Kuhndel, K., & Liebert, U. G. (2004). PCR detection of human papillomavirus of the mucosa: comparison between MY09/11 and GP5+/6+ primer sets. *J Clin Virol*, *30*, 302–308.

Ries, L. A. G., Melbert, D., Krapcho, M., Stinchcomb, D. G., Howlader, N., Horner, M. J., et al. (Eds). (2008). *SEER cancer statistics review 1975–2005*. Bethesda, MD: National Cancer Institute. Retrieved July 08, 2009, from http://seer.cancer.gov/csr/1975_2005/, based on November 2007 SEER data submission, posted to the SEER website.

Robinson, P. N., & Mickelson, A. R. (2006, April). Early diagnosis of oral cavity cancers. *Otolaryngol Clin North Am*, *39*(2), 295–306.

Sancho-Garnier, H., & Theobald, S. (1993). Black (air cured) and blond (flue cured) tobacco and cancer risk II: pharynx and larynx cancer. *Eur J Cancer*, *29A*, 273–276.

Sankaranarayanan, R., Mathew, B., Jacob, B. J., Thomas, G., Somanathan, T., Pisani, P., et al. (2000). Early findings from a community-based, cluster randomized, controlled oral cancer screening trial in Kerala, India. The Trivandrum Oral Cancer Screening Study Group. *Cancer*, *88*, 664–673.

Schwartz, J. L. (2000). Biomarkers and molecular epidemiology and chemoprevention of oral carcinogenesis. *Crit Rev Oral Biol Med*, *11*, 92–122.

Sciubba, J. J. (2001). Oral cancer. The importance of early diagnosis and treatment. *Am J Clin Dermatol*, *2*(4), 239–251.

Scott, S. E., Grunfeld, E. A., & McGurk, M. (2006). Patient's delay in oral cancer: a systematic review. *Community Dent Oral Epidemiol*, *34*, 337–343.

Silverman, S., Jr. (1998). *Oral cancer* (4th ed.). American Cancer Society. St. Louis, MO: Mosby.

Sterman, J. D. (2006, March). Learning from evidence in a complex world. *Am J Public Health*, *96*(3), 505–514. [Epub 2006, January 31.]

Surveillance, Epidemiology, and End Results (SEER) Program. (2008). Retrieved from http://www.seer.gov. Seer*Stat Database: Incidence—SEER 17 Regs Limited-Use + Hurricane Katrina Impacted Louisiana Cases, Nov 2007 Sub (1973–2005 varying), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2008, based on the November 2007 Submission.

Syrjänen, K., Syrjänen, S., Lamberg, M., Pyrhönen, S., & Nuutinen, J. (1983). Morphological and immunohistochemical evidence suggesting human papillomavirus (HPV) involvement in oral squamous cell carcinogenesis. *Int J Oral Surg*, *12*, 18–24.

Terai, M., Hashimoto, K., Yoda, K., & Sata, T. (1999). High prevalence of human papillomaviruses in the normal oral cavity of adults. *Oral Microbiol Immunol*, *14*, 201–205.

U.S. Preventive Services Task Force. (2004a, February). Screening for oral cancer: recommendation statement. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved July 08, 2009, from http://www.ahrq.gov/clinic/3rduspstf/oralcan/oralcanrs.htm

U.S. Preventive Services Task Force. (2004b, November). *Screening for oral cancer: a brief evidence update for the U.S. Preventive Services Task Force*. Rockville, MD: Agency for Healthcare Research and Quality.

Chapter 16

Axéll, T., & Henricsson, V. (1985). The occurrence of recurrent aphthous ulcers in an adult Swedish population. *Acta Odontol Scand*, *43*, 121–125.

Beck, J. D., Eke, P., Heiss, G., Madianos, P., Couper, D., Lin, D., et al. (2005). Periodontal disease and coronary heart disease: a reappraisal of the exposure. *Circulation*, *112*, 19–24.

Boggess, K. A., Lieff, S., Murtha, A. P., Moss, K., Beck, J., & Offenbacher, S. (2003). Maternal periodontal disease is associated with an increased risk for preeclampsia. *Obstet Gynecol*, *101*, 227–231.

Caliskan, M. K. (1993). Success of pulpotomy in the management of hyperplastic pulpitis. *Int Endod J*, *26*(2), 142–148.

Caplan, D. J., Reams, G., & Weintraub, J. A. (1999). Recommendations for endodontic referral among practitioners in a dental HMO. *J Endod*, *25*(5), 369–375.

Caplan, D. J., & Weintraub, J. A. (1997). Factors related to loss of root canal filled teeth. *J Public Health Dent*, *57*(1), 31–39.

Carinci, F., Pezzetti, F., Scapoli, L., Martinelli, M., Avantaggiato, A., Carinci, P., et al. (2003). Recent developments in orofacial cleft genetics. *J Craniofac Surg*, *14*(2), 130–143.

Chattopadhyay, A., Caplan, D. J., Slade, G. D., Shugars, D. C., Tien, H-C., & Patton, L. L. (2005a). Incidence of oral candidiasis and oral hairy leukoplakia in HIV-infected adults in North Carolina. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, *99*, 39–47.

Chattopadhyay, A., Caplan, D. J., Slade, G. D., Shugars, D. C., Tien, H-C., & Patton, L. L. (2005b). Risk indicators for oral candidiasis and oral hairy leukoplakia in HIV-infected adults. *Community Dent Oral Epidemiol*, *33*, 35–44.

Chattopadhyay, A., & Chatterjee, S. (2007). Epidemiology of recurrent aphthous ulcers among adults in the United States. *Community Dentistry and Oral Epidemiology*, *35*, 152–159.

Chattopadhyay, A., & Patton, L. L. (2007). Risk indicators for jointly occurring oral candidiasis and oral hairy leukoplakia. *AIDS Patient Care and STDs*, *21*, 825–832.

Chattopadhyay, A., & Ray, J. (2007, December). Successful management of pulpal involvement using pulpotomy in adults. *Clinical Dentistry*, *1*, 26–29.

Conner, D. A., Caplan, D. J., Teixeira, F. B., & Trope, M. (2007, November). Clinical outcome of teeth treated endodontically with a nonstandardized protocol and root filled with resilon. *J Endod*, *33*(11), 1290–1292.

Cvek, M., Mejare, I., & Andreasen, J. (2004, October). Conservative endodontic treatment of teeth fractured in the middle or apical part of the root. *Dent Traumatol*, *20*(5), 261–269.

Dasanayake, A. P. (1998). Poor periodontal health of the pregnant woman as a risk factor for low birth weight. *Ann Periodontol*, *3*, 206–212.

Ebersole, J. L., Cappelli, D., Mott, G., Kesavalu, L., Holt, S. C., & Singer, R. E. (1999, October). Systemic manifestations of periodontitis in the non-human primate. *J Periodontal Res*, *34*(7), 358–362.

Eppley, B. L., van Aalst, J. A., Robey, A., Havlik, R. J., & Sadove, A. M. (2005). The spectrum of orofacial clefting. *Plast Reconstr Surg*, *115*(7), 101e–114e.

Friedewald, V. E., Kornman, K. S., Beck, J. D., Genco, R., Goldfine, A., Libby, P., Offenbacher, S., et al. (2009). The American Journal of Cardiology and Journal of Periodontology editors' consensus: periodontitis and atherosclerotic cardiovascular disease. *Am J Cardiol*, *104*(1), 59–68.

Greenspan, D., Gange, S. J., Phelan, J. A., Navazesh, M., Alves, M. E., MacPhail, L. A., et al. (2004). Incidence of oral lesions in HIV-1-infected women: reduction with HAART. *J Dent Res*, *83*, 145–150.

Herzberg, M. C. & Meyer, M. W. (1996, October). Effects of oral flora on platelets: possible consequences in cardiovascular disease. *J Periodontol*, *67*(10 Suppl), 1138–1142.

Hunter, K. D., Gibson, J., Lockhart, P., Pithie, A., & Bagg, J. (1998). Fluconazole-resistant candida species in the oral flora of fluconazole-exposed HIV-positive individuals. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, *85*, 558–564.

Jeffcoat, M. K., Geurs, N. C., Reddy, M. S., Cliver, S. P., Goldenerg, R. L., & Hauth, J. C. (2001). Periodontal infection and preterm birth: results of a prospective study. *J Am Dent Assoc*, *132*, 875–880.

Jeffcoat, M. K., Hauth, J. C., Geurs, N. C., Reddy, M. S., Cliver, S. P., Hodgkins, P. M., et al. (2003). Periodontal disease and preterm birth: results of a pilot intervention study. *J Periodontol*, *74*, 1214–1218.

Johansen, A. M., Wilcox, A. J., Lie, R. T., Andersen, L. F., & Drevon, C. A. (2009). Maternal consumption of coffee and caffeine-containing beverages and oral clefts: a population-based case-control study in Norway. *Am J Epidemiol*, *169*(10), 1216–1222.

Joshipura, K. (2002). The relationship between oral conditions and ischemic stroke and peripheral vascular disease. *JADA*, *133*, 23S–30S.

Joshipura, K., Wand, H. C., Merchant, A. T., & Rimm, E. B. (2004). Periodontal disease and biomarkers related to cardiovascular disease. *J Dent Res*, *83*(2), 151–155.

Katz, M. H., Greenspan, D., Westenhouse, J., Hessol, N. A., Buchbinder, S. P., Lifson, A. R., et al. (1992). Progression to AIDS in HIV-infected homosexual and bisexual men with hairy leukoplakia and oral candidiasis. *AIDS*, *6*, 95–100.

Kirakozova, A. & Caplan, D. J. (2006, August). Predictors of root canal treatment in teeth with full coverage restorations. *J Endod*, *32*(8), 727–730. [Epub 2006, June 9.]

Kolker, J. L., Damiano, P. C., Bentler, S. E., Flach, S. D., Armstrong, S. R., Caplan, D. J., et al. (2006, Winter). The cost-effectiveness of large amalgam and crown restorations over a 10-year period. *J Public Health Dent*, *66*(1), 57–63.

Kolokotronis, A., Kioses, V., Antoniades, D., Mandraveli, K., Doutsos, I., & Papanayotou, P. (1994). Immunologic status in individuals infected with HIV oral candidiasis and hairy leukoplakia. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, *78*, 41–46.

Krapels, I. P., Vermeij-Keers, C., Müller, M., de Klein, A., Steegers-Theunissen, R. P. (2006). Nutrition and genes in the development of orofacial clefting. *Nutr Rev*, *64*(6), 280–288.

Kshirsagar, A. V., Offenbacher, S., Moss, K. L., Barros, S. P., & Beck, J. D. (2007). Antibodies to periodontal organisms are associated with decreased kidney function. *Blood Purif*, *25*, 125–132.

Lamster, I. B., Lalla, E., Borgnakke, W. S., & Taylor, G. W. (2008). The relationship between oral health and diabetes mellitus. *J Am Dent Assoc*, *139*, 19S–24S.

Lopez, N. J., Smith, P. C., & Gutierrez, J. (2002). Periodontal therapy may reduce the risk of preterm low birth weight in women with periodontal disease: a randomized controlled trial. *J Periodontol*, *73*, 911–924.

Manau, C., Echeverria, A., Agueda, A., Guerrero, A., & Echeverria, J. J. (2008). Periodontal disease definition may determine the association between periodontitis and pregnancy outcomes. *J Clin Periodontol*, *35*, 385–397.

Mattila, K., Vesanen, M., Valtonen, V., Nieminen, M., Palosuo, T., Rasi, V., et al. (2002). Effect of treating periodontitis on C-reactive protein levels: a pilot study. *BMC Infectious Diseases*, *2*, 30–33.

McDougal, R. A., Delano, E. O., Caplan, D., Sigurdsson, A., & Trope, M. (2004). Success of an alternative for interim management of irreversible pulpitis. *JADA*, *135*, 1707–1712.

Menezes, R., Bramante, C. M., Letra, A., Carvalho, V. G., & Garcia, R. B. (2004, September). Histologic evaluation of pulpotomies in dog using two types of mineral trioxide aggregate and regular and white Portland cements as wound dressings. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, *98*(3), 376–379.

Meyer, M. S., Joshipura, K., Giovannucci, E., & Michaud, D. S. (2008). A review of the relationship between tooth loss, periodontal disease, and cancer. *Cancer Causes Control*, *19*, 895–907.

Mithani, S. K., Mydlarz, W. K., Grumbine, F. L., Smith, I. M., & Califano, J. A. (2007). Molecular genetics of premalignant oral lesions. *Oral Diseases*, *13*, 126–133.

Offenbacher, S., Katz, V., Fertik, G., Collins, J., Boyd, D., Maynor, G., et al. (1996). Periodontal infection as a possible risk factor for preterm low birth weight. *J Periodontol*, *67*, 1103–1113.

Offenbacher, S., Lieff, S., Boggess, K. A., Murtha, A. P., Madianos, P. N., Champagne, C. M., et al. (2001). Maternal periodontitis and prematurity. Part I. Obstetric outcome of prematurity and growth restriction. *Ann Periodontol*, *6*, 164–174.

Office of Public Health Genomics, CDC. (2007). HuGE Reviews. Retrieved July 08, 2009, from http://www.cdc.gov/genomics/hugenet/reviews.htm

Ortiz-Posadas, M. R., Vega-Alvarado, L., & Maya-Behar, J. (2001). A new approach to classify cleft lip and palate. *Cleft Palate Craniofac J*, *38*(6), 545–550.

Patton, L. L., McKaig, R. G., Eron, J. J., Jr, Lawrence, H. P., & Strauss, R. P. (1999). Oral hairy leukoplakia and oral candidiasis as predictors of HIV viral load. *AIDS*, *13*, 2174–2176.

Patton, L. L., & Van der Horst, C. (1997). Oral infections and other manifestations of HIV disease. *Infectious Dis Clin North Am*, *13*, 879–900.

Reibel, J. (2003). Significance of clinical, histopathological, and molecular biological characteristics. *Crit Rev Oral Biol Med*, *14*, 47–62.

Reit, C., Bergenholtz, G., Caplan, D., & Molander, A. (2007, April). The effect of educational intervention on the adoption of nickel-titanium rotary instrumentation in a public dental service. *Int Endod J*, *40*(4), 268–274. [Epub 2007, February 8.]

Rivera-Hildago, F., Stulman, J. D., & Beach, M. M. (2004). The association of tobacco and other factors with recurrent aphthous stomatitis in an U.S. adult population. *Oral Dis, 10*(6), 335–345.

Santini, A. (1986, December). Long-term clinical assessment of pulpotomies with calcium hydroxide containing ledermix in human permanent premolars and molars. *Acta Odontol Pediatr*, *7*(2), 45–50.

Saremi, A., Nelson, R. G., Tulloch-Reid, M., Hanson, R. L., Sievers, M. L., Taylor, G. W., et al. (2005). Periodontal disease and mortality in type 2 diabetes. *Diabetes Care*, *28*(1), 27–32.

Shugars, D. C., Adesanya, M., Diehl, S. R., Redman, R. S., Malley, K. J., Silva, R. G., et al. (2007). Self-awareness of premalignant oral lesions among United States veterans. *Journal of Spec Care Dentist*, *27*, 87–94.

Slade, G. D., Offenbacher, S., Beck, J. D., Heiss, G., & Pankow, J. S. (2000, January). Acute-phase inflammatory response to periodontal disease in the US population. *J Dent Res*, *79*(1), 49–57.

Tessier, P. (1976). Anatomical classification facial, cranio-facial and latero-facial clefts. *J Maxillofac Surg*, *4*(2), 69–92.

Triantos, D., Porter, S. R., Scully, C., & Teo, C. G. (1997). Oral hairy leukoplakia: clinico-pathologic features, pathogenesis, diagnosis, and clinical significance. *Clin Infect Dis*, *25*, 1392–1396.

Tsai, C., Hayes, C., & Taylor, G. W. (2002). Glycemic control of type 2 diabetes and severe periodontal disease in the U.S. adult population. *Community Dent Oral Epidemiol*, *30*(3), 182–192.

Tumbarello, M., Tacconelli, E., Caldarola, G., Morace, G., Cauda, R., & Ortona, L. (1997). Fluconazole resistant oral candidiasis in HIV-infected individuals. *Oral Dis*, *3*, S110–S112.

Vieira, A. R. (2006). Association between the transforming growth factor alpha gene and nonsyndromic oral clefts: a HuGE review. *Am J Epidemiol*, *163*(9), 790–810.

Warnakulasuriya, S., Johnson, N. W., & van der Waal, I. (2007). Nomenclature and classification of potentially malignant disorders of the oral mucosa. *J Oral Pathol Med*, *36*, 575–580.

Warnakulasuriya, S., Reibel, J., Bouquot, J., & Dabelsteen, E. (2008). Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. *J Oral Pathol Med*, *37*, 127–133.

World Health Organization. (1973). Report of a meeting of investigators on the histological definition of precancerous lesions. Geneva: Author.

Wyszynski, D. F., Sárközi, A., & Czeizel, A. E. (2006). Oral clefts with associated anomalies: methodological issues. *Cleft Palate Craniofac J*, *43*(1), 1–6.

Yazdy, M. M., Honein, M. A., Rasmussen, S. A., & Frias, J. L. (2007). Priorities for future public health research in orofacial clefts. *Cleft Palate Craniofac J*, *44*(4), 351–357.

Chapter 17

ADA Council on Access, Prevention, and Interprofessional Relations. (1995). Caries diagnosis and risk assessment. A review of preventive strategies and management. *JADA, 126*(Suppl), 1S–24S.

Ahovuo-Saloranta, A., Hiiri, A., Nordblad, A., Mäkelä, M., & Worthington, H. V. (2008). Pit and fissure sealants for preventing dental decay in the permanent teeth of children and adolescents. *Cochrane Database Syst Rev*, *8*(4), CD001830.

Bassin, E. B., Mittleman, M. A., Wypij, D., Joshipura, K., & Douglass, C. W. (2004). Problems in exposure assessment of fluoride in drinking water. *J Public Health Dent*, *64*(1), 45–49.

Bassin, E. B., Wypij, D., Davis, R. B., & Mittleman, M. A. (2006). Age-specific fluoride exposure in drinking water and osteosarcoma (United States). *Cancer Causes Control*, *17*(4), 421–428.

Brunelle, J. A. & Carlos, J. P. (1990). Recent trends in dental caries in U.S. children and the effect of water fluoridation. *J Dent Res*, *69*, 723–727.

Bucher, J. R., Hejtmancik, M. R., Toft, J. D., II, Persing, R. L., Eustis, S. L, & Haseman, J. K. (1991). Results and conclusions of the National Toxicology Program's rodent carcinogenicity studies with sodium fluoride. *Int J Cancer*, *48*(5), 733–737.

Burt, B. A., Eklund, S. A., & Lewis, D. W. (1992). *Dentistry, Dental Practice, and the Community*. Philadelphia, PA: W. B. Saunders.

Burt, B. A, Keels, M. A, & Heller, K. E. (2003). Fluorosis development in seven age cohorts after an 11-month break in water fluoridation. *J Dent Res*, *82*(1), 64–68.

Centers for Disease Control and Prevention. (2001). *MMWR Recomm Rep*, *50*(RR-14), 1–42.

Clarkson, J. (1989). Review of terminology, classifications, and indices of developmental defects of enamel. *Adv Dent Res*, *3*, 104–109.

Dawid, A. A. P. (2002). Influence diagrams for causal modeling and inference. *In Stat Rev*, *2*, 161–189.

Den Besten, P. K. (1999). Mechanism and timing of fluoride effects on developing enamel. *J Public Health Dent*, *59*, 247–251.

Douglass, C. W., & Joshipura, K. (2006). Caution needed in fluoride and osteosarcoma study. *Cancer Causes Control*, *17*(4), 481–482.

Eggen, K. H. & Rölla, G. (1983). Surface properties of fluoride treated hydroxyapatite as judged by interactions with albumin and lysozyme. *Scand J Dent Res*, *91*, 347–350.

Ekstrand, J. & Oliveby, A. (1999). Fluoride in the oral environment. *Acta Odontol Scand*, *57*, 330–333.

Estuprnan-Day, S. B., Baez, B., Horowitz, H., Warpeha, R., Sutherland, B., & Thamer, M. (2001). Salt fluoridation and dental caries in Jamaica. *Community Dent Oral Epidemiol*, *29*, 247–252.

Evans, R. W., & Darvell, B. W. (1995). Refining the estimate of the critical period for susceptibility to enamel fluorosis in human maxillary central incisors. *J Public Health Dent*, *55*, 238–249.

Falqueto, J., Lima W., Borges, P. S. S, & Barreto, J. M. (2004). The measurement of artificial intelligence—an IQ machine. Retrieved July 08, 2009, from www.inf.ufsc.br/~13c/artigos/Falqueto01.pdf

Fejerskov, O., Richards, A., & Den Besten, P. (1996). The effect of fluoride on tooth mineralization. In O. Fejerskov, J. Ekstrand, & B. Burt (Eds.), *Fluoride in dentistry* (pp. 112–146). Copenhagen: Munksgaard.

Friedman, M., Solouki, S., Gurevitz, S., Gedalia, I., & Onisi, M. (1984). Fluoride concentrations, its uptake by hydroxyapatite and on dissolution rate. *Clin Prev Dent*, *6*, 20–22.

Geddes, D. A. M., & Bowen, W. H. (1990). Summary of session III: fluoride in saliva and dental plaque. *J Dent Res*, *69*, 637.

Gelberg, K., Fitzgerald, E., Hwang, S., & Dubrow, R. (1995). Fluoride exposure and childhood osteosarcoma: a case–control study. *Am J Pub Health*, *85*, 1678–1683.

Gessner, B. D., Beller, M., Middaugh, J. P., & Whitford, G. M. (1994). Acute fluoride poisoning from a public water system. *N Engl J Med*, *330*(2), 95–99.

Greenland, S. & Brumback, B. (2002). An overview of relations among causal modeling methods. *Int J Epidemiol*, *31*, 1030–1037.

Guide to Community Preventive Services. (2002). Preventing dental caries: community water fluoridation. Retrieved July 08, 2009, from http://www.thecommunityguide.org/oral/fluoridation.html

Hernan, M. A., Hernandez-Diaz, S., Werler, M. M., & Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*, *155*, 176–184.

Hiiri, A., Ahovuo-Saloranta, A., Nordblad, A., & Mäkelä, M. (2006). Pit and fissure sealants versus fluoride varnishes for preventing dental decay in children and adolescents. *Cochrane Database Syst Rev*, *8*(4), CD003067.

Horner, M. J., Ries, L. A. G., Krapcho, M., Neyman, N., Aminou, R., Howlader, N., et al (eds). SEER Cancer Statistics Review, 1975–2006, National Cancer Institute. Bethesda, MD, Based on November 2008 SEER data submission, posted to the SEER web site, 2009. Available at http://seer.cancer.gov/csr/1975_2006/

Horowitz, H. S. (1986). Indexes for measuring dental fluorosis. *J Public Health Dent*, *46*, 179–183.

Jones, S., Burt, B. A., Petersen, P. E., & Lennon, M. A. (2005, September). The effective use of fluorides in public health. *Bulletin of the WHO*, *83*(9), 670–676.

Kimminau, K. S., Shepherd, M. D., & Starrett, B. E. (2000, September). *Alternatives to standard community water fluoridation* (Pub. No. 00-101). Topeka, KS: Kansas Health Institute.

Kinlen, L. (1975). Cancer incidence in relation to fluoride level in water supplies. *Br Dent J*, *138*, 221–4.

Kumar, J. V., Swango, P., Haley, V., & Green, E. (2000). Intra-oral distribution of dental fluorosis in Newburgh and Kingston, New York. *J Dent Res*, *79*(7), 1508–1513.

Kumar, J. V. (2008, July 1). Is water fluoridation still necessary? *Adv Dent Res*, *20*(1), 8–12.

Kumar, V., Abbas, A. K., Fausto, N., Mitchell, R. N. (2007). Robbins basic pathology (8th ed.). Philadelphia, PA: Saunders Elsevier.

Kurttio, P., Gustavsson, N., Vartiainen, T., & Pekkanen, J. (1999). Exposure to natural fluoride in well water and hip fracture: a cohort analysis in Finland. *Am J Epidemiol. 15; 150*(8): 817–24.

Larsen, M. J., Kirkegaard, E., & Poulsen, S. (1987). Patterns of dental fluorosis in a European country in relation to the fluoride concentration of drinking water. *J Dent Res*, *66*, 10–12.

Lazzari, E. P. (1976). *Dental biochemistry* (2nd ed.). Philadelphia, PA: Lea & Febiger.

LeCompte, E. J. (1987). Clinical application of topical fluoride products—risks, benefits, and recommendations. *J Dent Res*, *66*, 1066–1071.

Lennon, M. A., Whelton, H., O'Mullane, D., & Ekstrand, J. (2005). Fluoride. In *Nutrients in drinking water, sanitation and health protection and the human environment* (pp. 180–186) Geneva: World Health Organization. Retrieved July 29, 2009, from http://www.who.int/water_sanitation_health/dwq/nutrientsindw.pdf

Manji, F., Baelum, V., Fejerskov, O., & Gemert, W. (1986). Enamel changes in two low-fluoride areas of Kenya. *Caries Res*, *20*, 371–380.

Marthaler, T. (1983). Practical aspects of salt fluoridation. *SSO Schweiz Monatsschr Zahnheilkd, 93*(12), 1197–1214.

McDonagh, M., Whiting, P., Bradley, M., Cooper, J., Sutton, A., Chestnutt, I., et al. (2000). A Systematic Review of Public Water Fluoridation. York, UK: University of York, NHS Centre for Reviews and Dissemination.

Mihashi, M. & Tsutsui, T. (1996). Clastogenic activity of sodium fluoride to rat vertebral body-derived cells in culture. *Mutat Res*, *368*(1), 7–13.

Murray, J. J., Rugg-Gunn, A. J., & Jenkins, G. N. (1991). *Fluorides in caries prevention* (3rd ed.). Oxford: Wright, Butterworth, Heinemann.

National Research Council. (1993). *Health effects of ingested fluoride*. Washington, DC: National Academy Press.

National Research Council Committee on Fluoride in Drinking Water. (2006). *Flouride in drinking water: a scientific review of EPA's standards*. Washington, DC: The National Academy of Sciences.

NTP: National Toxicology Program (1990). Toxicology and carcinogenesis of sodium fluoride in F344 TN rats and B 6C3FL mice. *Technical Report Services*, *393*, NIH Publication No. 90-2848.

Newbrun, E. (1989). Effectiveness of water fluoridation. *J Public Health Dent, 49*(5), 279–289.

Oliveby, A., Ekstrand, J., & Lagerlöf, F. (1987). Effect of salivary flow rate on salivary fluoride clearance after use of a fluoride containing chewing gum. *Caries Res*, *21*, 393–401.

Petersen, P. E., & Lennon, M. A. (2004). Effective use of fluorides for the prevention of dental caries in the 21st century: the WHO approach. *Community Dent Oral Epidemiol*, *32*, 319–321.

Pollick, H. F. (2006). Concerns about water fluoridation, IQ, and osteosarcoma lack credible evidence. *Int J Occup Environ Health*, *12*(1), 91–94.

Pretty, I. A., Tavener, J. A., Browne, D., Brettle, D. S., Whelton, H., & Ellwood, R. P. (2006). Quantification of dental fluorosis using fluiorescence imaging. *Caries Res*, *40*, 426–434.

Public Health Service, Department of Health and Human Services (PHS, DHHS). (1991). Review of fluoride: benefits and risks report of the ad hoc subcommittee on fluoride of the committee to coordinate environmental health and related programs. Retrieved July 08, 2009, from http://health.gov/environment/ReviewofFluoride/

Reynolds, E. C., & Riley, P. F. (1981). Kinetics of hydroxyapatite dissolution during incubation with sucrose and S. mutans and its application to cariogenicity testing. *Caries Res*, *15*, 501–507.

Quan, D. T. H. (2000). Dental caries status of primary teeth after 4 years of fluoridation in Ho Chi Minh City. Dinh T. N. (Ed). *Proceedings of Asia conference of oral health promotion for school children 2001*. Tokyo: World Health Organization.

Ripa, L. W. (1991). A critique of topical fluoride methods (dentifrices, mouth rinses, operator-, and self-applied gels) in an era of decreased caries and increased fluorosis prevalence. *J Public Health Dent*, *51*, 23–41.

Ripa, L. W. (1993). A half-century of community water fluoridation in the United States. Review and Commentary. *J Public Health Dent*, *53*, 17–44.

Rozier, R. G. (1994). Epidemiologic indices for measuring the clinical manifestations of dental fluorosis: overview and critique. *Adv Dent Res*, *8*, 39–55.

Russell, A. L. (1961). The differential diagnosis of fluoride and non-fluoride enamel opacities. *J Public Health Dent*, *21*, 143–146.

Sandhu, R., Lal, H., Kundu, Z. S., & Kharb, S. (2009). Serum fluoride and sialic acid levels in osteosarcoma. *Biol Trace Elem Res*. DOI: 10.1007/s12011-009-8382-1

Singh, K. A., & Spencer, J. A. (2004). Relative effects of pre- and post-eruption water fluoride on caries experience by surface type of permanent first molars. *Community Dent Oral Epidemiol*, *32*, 435–446.

Slade, G. D. (ed). (1997). *Measuring oral health quality of life*. Chapel Hill, NC: University of North Carolina, Department of Dental Ecology, School of Dentistry.

Sohn, W., Noh, H., & Burt B. A. (2009). Fluoride Ingestion Is Related to Fluid Consumption Patterns. *J Public Health Dent*, Epub ahead of print DOI: 10.1111/j.1752-7325.2009.00133.x

Swanberg, H. (1953). Fluoridation of water and its relation to cancer. *Miss Valley Med J*, *75*, 125–128.

Tang, Q. Q., Du, J., Ma, H. H., Jiang, S. J., & Zhou, X. J. (2008, Winter). Fluoride and children's intelligence: a meta-analysis. *Biol Trace Elem Res*, *126*(1–3), 115–120.

ten Cate, J. M. (1999). Current concepts on the theories of the mechanism of action of fluoride. *Acta Odontol Scand*, *57*, 325–329.

Twetman, S., Lindqvist, L., & Sund, M. L. (1986). Effect of human lysozyme on 2-deoxyglucose uptake by Streptococcus mutans and other oral microorganisms. *Caries Res, 20*(3), 223–229.

U.S. Department of Health and Human Services (USDHHS). (1991). Review of fluoride benefits and risks: report of the Ad Hoc Subcommittee on Fluoride of the Committee to Coordinate Environmental Health and Related Programs. Washington, DC: U.S. Department of Health and Human Services, Public Health Service.

U.S. Environmental Protection Agency. (2001). *Fluoride in Drinking Water*.Washington, DC: National Academies Press.

Warpeha, R., Beltran-Aguilar, E., & Baez, R. (2001). Methodological and biological factors explaining the reduction in dental caries in Jamaican school children between 1984 and 1995. *Pan American Journal of Public Health*, *10*, 37–44.

Warren, J. J., Levy, S. M., Broffitt, B., Cavanaugh, J. E., Kanellis, M. J., & Weber-Gasparoni K. (2009). Considerations on optimal fluoride intake using dental fluorosis and dental caries outcomes—a longitudinal study. *J Public Health Dent, 69*(2), 111–115.

Whitford, G. M. (1996). The metabolism and toxicity of fluoride. In G. M. Whitford & Lussi, A., *Monographs in Oral Science* (Vol. 16). Basel, Switzerland: S. Karger AG.

World Health Organization (WHO). (2003). *Diet, nutrition and the prevention of chronic diseases*. WHO Technical Report Series 916. Geneva: Author.

Chapter 18

Armfield, J. A. (2007). Socioeconomic inequalities in child oral health: a comparison of discrete and composite area-based measures. *J Pub Hlth Dent*, *67*, 119–125.

Austin, P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med*, *27*(12), 2037–2049.

Barreiro, L. B., Laval, C., Quach, H., Patin, E., & Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nature Genetics, 40,* 340–345.

Bender, D. J. (2007, June). Patient preference for a racially or gender-concordant student dentist. *J Dent Educ*, *71*(6), 726–745.

Borrell, L. N. & Crawford, N. D. (2008). Social disparities in periodontitis among United States adults 1999–2004. *Community Dent Oral Epidemiol*, *36*, 383–391.

Castells, M. & Portes, A. (1989). World underneath: the origins, dynamics and effects of the informal economy. In A. Portes, M. Castells, & L. A. Benton (Eds.), *The informal economy* (pp. 11–33), Baltimore, MD: The Johns Hopkins University Press.

Chattopadhyay, A. (2008). Oral health disparities in the United States. *Dental Clin N Am*, *52*, 297–318.

Dawid, A. A. P. (2002). Influence diagrams for causal modeling and inference. *Stat Rev*, *2*,161–189.

Elstad, J. (1998). The psycho-social perspective on social inequalities in health. *Sociol Health Illn*, *20*, 598–618.

Elter, J. R, Patton, L. L., & Strauss, R.P. (2005). Incidence rates and trends for oral and pharyngeal cancer in North Carolina: 1990–1999. *Oral Oncol*, *41*(5), 470–479.

Erikson, R. (1998, December 10). The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1998. Presentation Speech by Professor Robert Erikson of the Royal Academy of Sciences. Retrieved July 08, 2009, from http://nobelprize.org/nobel_prizes/economics/laureates/1998/presentation-speech.html

Falqueto, J., Lima, W., Borges, P. S. S., & Barreto, J. M. (2004). The measurement of artificial intelligence—an IQ machine. Retrieved July 02, 2009, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.15.2723

Fujiwara, T., & Kawachi, I. (2008). Social capital and health: a twin study. *Am J Prev Med*, *35*, 139–144.

Gilbert, G. H., Duncan, R. P., & Shelton, B. J. (2003, December). Social determinants of tooth loss. *Health Serv Res*, *38*(6 Pt 2), 1843–1862.

Greenland, S. & Brumback, B. (2002). An overview of relations among causal modeling methods. *Int J Epidemiol*, *31*, 1030–1037.

Hernan, M. A., Hernandez-Diaz, S., Werler, M. M., & Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*, *155*, 176–184.

Kauffman, J. S., & Cooper, R. S. (2008). Race in epidemiology: new tools, old problems. *Annals Epi*, *18*, 119–123.

Kelly, S., Hertzman, C., & Daniels, M. (1997). Searching for biological pathways between stress and health. *Annu Rev Public Health*, *18*, 437–462.

Krieger, N. (2001a). A glossary for social epidemiology. *J Epidemiol Community Health, 55*(10), 693–700.

Krieger, N. (2001b). Theories for social epidemiology in the 21st century: an ecosocial perspective. *Int J Epidemiol, 30,* 668–677.

Kressin, N. R. (2005, September). Racial/ethnic disparities in health care: lessons from medicine for dentistry. *J Dent Educ*, *69*(9), 998–1002.

Kuh, D., & Ben-Shlomo, Y. (2004). *The life course approach to chronic disease epidemiology* (2nd ed.). Oxford: Oxford University Press.

Lang, K., Menzin, J., Earle, C. C., Jacobson, J., & Hsu, M. A. (2004). The economic cost of squamous cell cancer of the head and neck: findings from linked SEER-Medicare data. *Arch Otolaryngel Head Neck Surg, 130*(11), 1269–1278.

Menzin, J., Lines, L. M., & Manning, L. N. (2007). The economics of squamous cell carcinoma of the head and neck. *Curr Opin Otolaryngol Head Neck Surg*, *15*(2), 68–73.

Miller, C. S., Henry, R. G., & Rayens, M. K. (2003). Disparities in risk of and survival from oropharyngeal squamous cell carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, *95*(5), 570–575.

Morse, D. E. & Kerr, A. R. (2006). Disparities in oral and pharyngeal cancer incidence, mortality and survival among Black and White Americans. *J Am Dent Assoc, 137*(2), 203–212.

Nadler, E., Eckert, B., & Neumann, P. J. (2006). Do oncologists believe new cancer drugs offer good value? *Oncologist*, *11*, 90–95.

Naito, M., Yuasa, H., Nomura, Y., Nakayama, T., Hamajima, N., & Hanada, N. (2006). Oral health status and health related quality of life: a systematic review. *J. Oral Sci*, *48*, 1–7.

National Center for Health Statistics. (1996). *Analytic and reporting guidelines: the third national health and nutrition examination survey, NHANES III (1988–1994)*. Hyattsville, MD: National Center for Health Statistics Centers for Disease Control and Prevention.

National Institute for Dental and Craniofacial Research. (2007). *Health disparities program. A plan to eliminate craniofacial, oral and dental health disparities*. Retrieved July 08, 2009, from http://www.nidcr.nih.gov/Research/Extramural/ClinicalResearch/HealthDisparitiesProgram.htm

National Institutes of Health. (2002). *Strategic research plan and budget to reduce and ultimately eliminate health disparities*, *Vol. 1, Fiscal Years 2002–2006*. Retrieved July 08, 2009, from http://ncmhd.nih.gov/our_programs/strategic/pubs/VolumeI_031003EDrev.pdf

Nicolau, B., Thomson, W. M., Steele, J. G., & Allison, P. J. (2007). Life-course epidemiology: concepts and theoretical models and its relevance to chronic oral conditions. *Community Dent Oral Epidemiol*, *35*, 241–249.

Norrgard, K., & Schultz, J. (2008). Using SNP data to examine human phenotypic differences. Nature Education 1(1). Retrieved on July 08, 2009, from http://www.nature.com/scitable/topicpage/Using-SNP-Data-to-Examine-Human-Phenotypic-706

Oakes, J. M., & Johnson, P. J. (2006). Propensity score method matching for social epidemiology. In J. M. Oakes & J. S. Kaufman (Eds.). *Methods in Social Epidemiology*. San Francisco, CA: John Wiley & Sons.

Office of Disease Prevention and Health Promotion, US Department of Health and Human Services. *Healthy People*. Retrieved June 29, 2009, from http://www.healthypeople.gov

Rosenbaum, P. R., & Rubin, D.,B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Rozier, G. R., & Pahel, B. T. (2008). Patient- and population-reported outcomes in public health dentistry: oral health related quality of life. *Dent Clin N Am, 52,* 345–365.

Sanders, A. E., Slade, G. D., Turrell, G., Spencer, A. J., & Marcenes, W. (2006). The shape of the socioeconomic-oral health gradient: implications for theoretical explanations. *Community Dent Oral Epidemiol*, *34*, 310–319.

Sen, A. (1999). Development as freedom [Introduction] and The ends and means of development [Chapter 2]. In *Development as freedom*. New York: Anchor Books.

Sen, A., & Dreze, J. (1999). *The Amartya Sen and Jean Dreze omnibus*. New Delhi: Oxford University Press.

Shiboski, C. H., Schmidt, B. L., & Jordan, R. C. K. (2007). Racial disparity in stage at diagnosis and survival among adults with oral cancer in the U.S. *Community Dent Oral Epidemiol*, *35*, 233–240.

Sisson, K. L. (2007). Theoretical explanations for social inequalities in oral health. *Community Dent Oral Epidemiol*, *35*, 81–88.

Slade, G. D. (2002). Assessment of oral health related quality of life. In M. R. Inglehart & R. A. Bagramian (Eds.), *Oral health-related quality of life* (pp. 29–46). Chicago: Quintessence Publishing.

Stehr-Green, P., Bettles, J., & Dee Robertson, L. (2002). Effect of racial/ethnic misclassification of American Indians and Alaskan Natives on Washington State death certificates, 1989–1997. *Am J Public Health*, *92*, 443–444.

Swan, J., Breen, N., Burhansstipanov, L., Satter, D. E., Davis, W. W., McNeel, T., et al. (2006). Cancer screening and risk factor rates among American Indians. *Am J Public Health*, *96*, 340–350.

Thomson, W. M., Poulton, R., Milne, B. J., Caspi, A., Broughton, J. R., & Ayers, K. M. (2004, October). Socioeconomic inequalities in oral health in childhood and adulthood in a birth cohort. *Community Dent Oral Epidemiol*, *32*(5), 345–353.

Tomar, S. L., Loree, M., Logan, H. (2004). Racial differences in oral and pharyngeal cancer treatment and survival in Florida. *Cancer Causes Control*, *15*(6), 601–609.

U.S. Department of Health and Human Services. (2000). *Oral health in America: a report of the surgeon general*. Rockville, MD: U.S. Department of Health and Human Services, National Institute of Dental and Craniofacial Research, National Institutes of Health.

Watt, R. G. (2007). From victim blaming to upstream action: tackling the social determinants of oral health inequalities. *Community Dent Oral Epidemiol*, *35*, 1–11.

World Bank. (2007). *Understanding poverty*. Retrieved July 08, 2009, from http://go.worldbank.org/K7LWQUT9L0

Chapter 19

Berger, J. (1998). Culture and ethnicity in clinical care. *Archives of Internal Medicine*, *158*, 2085–2090.

Boyle, J. (2004). Natural law and global ethics. In M. J. Cherry (Ed.), *Natural law and the possibility of a global ethics* (pp.1–15). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Council for International Organizations of Medical Sciences (CIOMS). (2002). International ethical guidelines for biomedical research involving human subjects. Retrieved July 08, 2009, from http://www.cioms.ch/frame_guidelines_nov_2002.htm

Curfman, G. D., Morrissey, S., & Drazen, J. M. (2006). Retraction: Sudbø, J., et al. DNA content as a prognostic marker in patients with oral leukoplakia. *N Engl J*

Med, *344*, 1270–1278 and Sudbø, J., et al. (2004). The influence of resection and aneuploidy on mortality in oral leukoplakia. *N Engl J Med*, *350*, 1405–1413. *N Engl J Med*, *355*(18), 1927.

Department of Health and Human Services (DHHS). (2005). Code of Federal Regulations. Title 45. Public welfare. Department of Health and Human Services. Part 46. Protection of human subjects. Retrieved July 08, 2009, from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm#46.102

Emanuel, E. J., Wendler, D., & Grady, C. (2000). What makes clinical research ethical? *JAMA*, *283*, 2701–2711.

Kotzin, S. (2007). From Darsee to Sudbø: NLM's role in the retraction process. *Michael*, *4*, 19–26.

McGee, G., & Caplan, A. (2007). An introduction to bioethics—scopes of the field and a brief history of bioethics. In G. McGee & A. Caplan (Eds.). *Bioethics for beginners.* Retrieved July 08, 2009, from http://www.bioethics.net/articles.php?viewCat=3&articleId=1

Nash, D. A., Friedman, J. W., Kardos, T. B., Kardos, R. L., Schwarz, E., et al. (2008, April). Dental therapists: a global perspective. *Int Dent J*, *58*(2), 61–70.

Segev, R. (2005). Well-being and fairness in the distribution of scarce health resources. *J Med Philos*, *30*, 231–260.

Sox, S. C., & Rennie, D. (2006). Research misconduct, retraction, and cleansing the medical literature: lessons from the Poehlman case. *Ann Intern Med*, *144*, 609–613.

U.S. Department of Health and Human Services (USDHHS). (2009). Understanding HIPAA Privacy. In USDHHS. Health Information Privacy. Retrieved on July 08, 2009, from http://www.hhs.gov/ocr/privacy/hipaa/understanding/index.html

Viens, A. M., & Singer, P. A. (2008). Introduction. In P. A. Singer & A. M. Viens, (Eds.). *The Cambridge textbook of bioethics*. Cambridge, UK: Cambridge University Press.

Chapter 20

Aderem, A. (2005). Systems biology: its practice and challenges. *Cell*, *121*(4), 511–513.

Chattopadhyay, A. (2009). Systems approach to understanding oral diseases (in press). In A. Daskalaki (Ed.). *Informatics in oral medicine: advanced techniques in clinical and diagnostic technologies*. Hershey, PA: IGI Publishers.

Hirsh, G. B., & Immediato, S. C. (1999). Microworlds and generic structures as resources for integrating care and improving health. *System Dynamics Review*, *15*(3), 315–330.

Homer, J. B., & Hirsch, G. B. (2006, March). System dynamics modeling for public health: background and opportunities. *Am J Public Health*. *96*(3), 452–458

Institute for Systems Biology. Premise of systems biology. (2008). Retrieved July 08, 2009, from http://www.systemsbiology.org/Systems_Biology_in_Depth/Premise_of_Systems_Biology.

iThink/STELLA 9 Technical Documentation. (2008). *Software reference guide*. Lebanon, NH: Isee Systems.

Iyengar, S. S. (1998). *Computer modeling and simulations of complex biological systems.* Boca Raton, FL: CRC Press.

Jones, A. P., Homer, J. B., Murphy, D. L., Essien, J. D., Milstein, B., & Seville, D. A. (2006, March). Understanding diabetes population dynamics through simulation modeling and experimentation. *Am J Public Health*, *96*(3), 488–494. [Epub 2006, January 31.]

Mabry, P. L., Olster, D. H., Morgan, G. D., & Abrams, D. B. (2008, August). Interdisciplinarity and systems science to improve population health: a view from the NIH Office of Behavioral and Social Sciences Research. *Am J Prev Med*, *35*(2 Suppl), S211–S224.

Milstein, B., Jones, A., Homer, J. B., Murphy, D., Essien, J., & Seville, D. (2007, July). Charting plausible futures for diabetes prevalence in the United States: a role for system dynamics simulation modeling. *Prev Chronic Dis*, *4*(3), A52. [Epub 2007, June 15.]

Schlessinger, L. & Eddy, D. M. (2002). Archimeds: a new model for simulating health care systems—the mathematical formulation. *J Biomed Inform, 35*(1), 37–50.

Sterman, J. D. (2006, March). Learning from evidence in a complex world. *Am J Public Health*, *96*(3), 505–514. [Epub 2006, January 31.]

Index

Boxes, figures, and tables are indicated by b, f, and t following the page number.

Amit Chattopadhyay

Oral Health Epidemiology

PRINCIPLES AND PRACTICE

As a result of scientific advancements and changing demographics in the United States and around the world, people of all ethnic groups and nationalities are retaining their teeth longer. Today's oral health professionals must therefore be prepared to make educated and scientifically reasoned choices addressing a wide range of oral diseases for patients of all ages, and for ambulatory as well as non-ambulatory patients across all demographic profiles.

As the first text of its kind, ***Oral Health Epidemiology: Principles and Practice*** explores the full spectrum of epidemiological and translational clinical research including fundamental mechanisms of human disease, therapeutic intervention, clinical trials, and oral epidemiology. Topics that are unique to oral health, such as the frequent use of split-mouth design on oral research, crossover techniques and clustered nature of caries, periodontal and other dental disease data, are all thoroughly addressed.

KEY FEATURES:

- Thoroughly explores clinical–translational research and the special needs of oral health study designs that are applicable across all specialties in dentistry.
- Serves as a basic guide to advanced techniques such as bioinformatics, genetics, molecular biology, computer simulation, and biostatistics that are now used regularly in oral health research.
- Prepares the reader to design studies, translate the findings to practice, and conduct logical critique of scientific literature.

Jones and Bartlett Publishers
40 Tall Pine Drive
Sudbury, MA 01776
978-443-5000
info@jbpub.com
www.jbpub.com

ISBN: 978-0-7637-5409-9
90000
9 780763 754099