Epidemiology: the foundation of public health

Roger Detels, M.D., M.S.

Abstract

Epidemiology is the basic science of public health, because it is the science that describes the relationship of health or disease with other health-related factors in human populations, such as human pathogens. Furthermore, epidemiology has been used to generate much of the information required by public health professionals to develop, implement, and evaluate effective intervention programmes for prevention of disease and promotion of health, such as the eradication of smallpox, the anticipated eradication of polio and guinea worm disease, and prevention of heart disease and cancer. Unlike pathology, which constitutes a basic area of knowledge, and cardiology, which is the study of a specific organ, epidemiology is a philosophy and methodology that can be applied to learning about and resolving a very broad range of health problems. The “art” of epidemiology is knowing when and how to apply the various epidemiological strategies creatively to answer specific health questions; it is not enough to know what the various study designs and statistical methodologies are. The uses and limitations of the various epidemiological study designs are presented to illustrate and underscore the fact that the successful application of epidemiology requires more than a knowledge of study designs and epidemiological methods. These designs and methods must be applied appropriately, creatively, and innovatively if they are to yield the desired information. The field of epidemiology has been expanding dramatically over the last three decades, as epidemiologists have demonstrated new

uses and variations of traditional study designs and methods. We can anticipate that the scope of epidemiology will expand even more in the future as increasing numbers of creative epidemiologists develop innovative new strategies and techniques.

The chapters in this section present detailed discussions of the principles and methods of epidemiology. In this introductory chapter, I will attempt to define epidemiology, to present ways in which epidemiology is used in the advancement of public health, and, finally, to discuss the range of applications of epidemiological methodologies.

What is epidemiology?

There are many definitions of epidemiology, but every epidemiologist will know exactly what it is that he or she does. Defining epidemiology is difficult primarily because it does not represent a body of knowledge, as does, for example, anatomy, nor does it target a specific organ system, as does cardiology. Epidemiology represents a method of studying a health problem and can be applied to a wide range of problems, from transmission of an infectious disease agent to the design of a new strategy for health care delivery. Furthermore, that methodology is continually changing as it is adapted to a greater range of health problems and more techniques are borrowed and adapted from other disciplines such as mathematics and statistics.

Maxcy, one of the pioneer epidemiologists of the past century, offered the following definition: “Epidemiology is that field of medical science which is concerned with the relationship of various factors and conditions which determine the frequencies and

distributions of an infectious process, a disease, or a physiologic state in a human community" (Lilienfeld 1978). The word itself comes from the Greek *epi*, *demos*, and *logos*; literally translated it means the study (*logos*) of what is upon (*epi*) the people (*demos*). John Last, in the *Dictionary of Epidemiology*, has defined epidemiology as "The study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems." Last's definition underscores that epidemiologists are not concerned only with disease but with "health-related events", and that ultimately epidemiology is committed to control of disease. All epidemiologists, however, will agree that epidemiology concerns itself with populations rather than individuals, thereby separating it from the rest of medicine and constituting the basic science of public health. Following from this, therefore, is the need to describe health and disease in terms of frequencies and distributions in the population. The epidemiologist relates these frequencies and distributions of specific health parameters to the frequencies of other factors to which populations are exposed in order to identify those that may be causes of ill health or promoters of good health. Inherent in the philosophy of epidemiology is the idea that ill health is not randomly distributed in populations, and that elucidating the reasons for this non-random distribution will provide clues regarding the risk factors for disease and the biological mechanisms that result in loss of health.

Because epidemiology usually focuses on health in *human* populations it is rarely able to provide experimental proof in the sense of Koch's postulates, as can often be done in the laboratory sciences. Epidemiology more often provides an accumulation of

increasingly convincing indirect evidence of a relationship between health or disease and other factors. This process, referred to as causal inference (see Hoggett & Greenland, Ch. 6.13), includes considering an observed relationship in terms of its strength, consistency, specificity, temporality, biologic gradient, plausibility, coherence and experimental evidence (Hill 1965).

Although they will differ on the exact definitions of epidemiology, most epidemiologists will agree that they try to characterize the relationships among the agent, the environment, and the host (usually human). The epidemiologist considers health to represent a balance among these three forces, as shown in Figure 1.

Changes in any one of these three factors may result in loss of health. For example, the host may be compromised as a result of treatment with steroids, making him/her more susceptible to agents that do not ordinarily cause disease. On the other hand, a breakdown in the water-supply system may result in an increased exposure of people to agents such as cryptosporidium, as happened in 1993 in Milwaukee, WI (MacKenzie *et al.* 1994). Finally, some agents may become more or less virulent over time--often because of the promiscuous use of antibiotics--thereby disturbing the dynamic balance among agent, host, and environment. Two examples are the cases of acute necrotizing fasciitis caused by *Streptococcus A* (Communicable Disease Surveillance Centre 1994) and the development of multidrug-resistant tuberculosis (Chapman and Henderson 1994).

The epidemiologist uses another triad to study the relationship of agent, host, and environment: time-place-person. Using various epidemiological techniques described in subsequent chapters, the epidemiologist describes disease or disease factors occurring in the population in terms of characteristics of time (for example, trends, outbreaks, etc.), place (the geographic area in which the disease is occurring), and person (the characteristics of the affected individuals; e.g., age, gender, etc.) to elucidate the causative agent, the natural history of the disease, and the environmental factors that increase the likelihood of the host acquiring the disease. With this information the epidemiologist is able to suggest ways to intervene in the disease process to either prevent disease or death.

Epidemiology has been described as the “art of the possible”. Because epidemiologists work with human populations, they are rarely able to manipulate events or the environment as can the laboratory scientist. They must, therefore, exploit situations as they exist naturally to advance knowledge. They must be both pragmatic and realistic. They must realize both the capabilities and limitations of the discipline. Morris has said that the “epidemiologic method is the only way to ask some questions..., one way of asking others and no way at all to ask many” (Morris 1975). The art of epidemiology is to know both when epidemiology is the method of choice and when it is not, and how to use it to answer the question.

Applying the epidemiological method to resolve a health question successfully can be compared to constructing a memorable Chinese banquet. It is not enough to have the

best ingredients and to know the various Chinese cooking methods. The truly great Chinese chef must be able to select the appropriate ingredients and cooking methods to bring out the flavors of each individual dish and, further, must know how to construct the correct sequence of dishes to excite the palate without overwhelming it. They create a memorable banquet by adding their creative genius to the raw ingredients and the established cooking methods. Similarly, it is not enough for the epidemiologist to know the various strategies and methods of epidemiology; the innovative epidemiologist must be able to apply them creatively to obtain the information needed to understand the natural history of the disease. It is not enough to know what a cohort study is; the epidemiologist must know when the cohort design is the appropriate design for the question at hand, and then must apply that design appropriately and creatively. These skills make epidemiology more than a methodology. It is this opportunity for creativity and innovation that provides excitement for the practitioner and makes the successful practice of epidemiology an art.

For example, Imagawa and colleagues identified probable transient HIV-1 infection in men, implying clearance of the virus by the immune system of the men, by focusing their viral isolation studies on the relatively few HIV-1-antibody-negative homosexual men who had many different sexual partners (Imagawa *et al.* 1989). A simple cohort study of antibody-negative individuals would have required a cohort of thousands of men rather than the 133 studied. The effects of passive smoking were demonstrated by cohort studies of non-smoking family members of smokers and in nursing students by comparing the reported symptoms in roommates of smokers and non-smokers who kept

diaries of their symptoms. The roommates of the smokers had a 1.8 greater risk of episodes of phlegm than roommates of non-smokers (Schwartz and Zeger 1990). Colley *et al.* (1974), Tager *et al.* (1979), and Tashkin *et al.* (1984) demonstrated that children of smokers had lower levels of lung function than children of non-smokers. All of these investigators used traditional study designs, but demonstrated their creativity by applying that design to those specific populations which were most likely to reveal a relationship if it existed.

Epidemiologic studies rarely provide “proof” of a causal relationship. Thus, there is continuing debate among epidemiologists about what constitutes adequate criteria for inferring a causal relationship from epidemiological studies (Rothman 1988). Hill suggested the following criteria for establishing a causal relationship: strength of association (statistical probability and risk ratio), consistency of findings across multiple studies, specificity of the relationship, temporality (outcome follows causation), biologic gradient (a dose-response relationship), plausibility, coherence (consistency with prior knowledge), experimental evidence, and analogy (relationship hypothesized is similar to that in known relationships) (Hill 1965). Susser has added to these criteria the ability of the observed relationship to correctly predict other relationships (Rothman 1988). The debate goes on, but the principle is the same: epidemiologic studies seldom provide “proof” of a causal relationship in the sense of Koch's postulates, but may be used to reveal a possible relationship and build a convincing case that this relationship is causal.

Uses of epidemiology in support of public health

Epidemiology is the basic science of public health because it is the health science that describes health and disease in *populations* rather than in individuals, information essential for the formulation of effective public health initiatives to prevent disease and promote health in the community. I have taken the liberty of updating the “Functions of Epidemiology”, first expounded by Morris (Morris, 1957) and Breslow and Detels (Holland et al., 2007). They are:

1. Describe the spectrum of the disease. Disease represents the end-point of a process of alteration of the host's biological systems. Although many disease agents are limited in the range of alterations they can initiate, others, such as measles, can cause a variety of disease end-points. For example, the majority of infections with rubeola (the measles virus) result in the classical febrile, blotchy rash-disease, but the rubeola virus can also cause generalized haemorrhagic rash and acute encephalitis. Years after initial infection, measles can also cause subacute sclerosing panencephalitis (SSPE), a fatal disease of the central nervous system.

   Various types of epidemiological studies have been used to elucidate the spectrum of disease resulting from many agents and conditions. For example, cohort studies have been used to document the role of high blood pressure as a major cause of stroke, myocardial infarct, and chronic kidney disease. For rare diseases such as SSPE and multiple sclerosis, case-control studies have been useful to identify the role of the rubeola virus (Alter 1976; Detels *et al.* 1973). Knowing the spectrum of disease that can result from specific infections and

conditions allows the public health professional to design more effective intervention strategies: for example, education, screening, and treatment programmes to reduce the prevalence of high blood pressure will also reduce the incidence of myocardial infarct, stroke, and chronic kidney disease (Hypertension Detection and Follow-up Program Cooperative Group, 1979).

2. Describe the natural history of disease. Epidemiological studies can be used to describe the natural history of disease, to elucidate the specific alterations in the biological system in the host and to improve diagnostic accuracy. For example, cohort studies of individuals who were infected with HIV, the “AIDS virus”, revealed that a drop in the level of T-lymphocytes having the CD4 marker was associated with being infected with HIV, and that a further decline in CD4 cells was associated with developing clinical symptoms and AIDS (Detels *et al.* 1987; Polk *et al.* 1987). This observation stimulated immunologists to focus their research on the interaction of the immune system and HIV. From a clinical perspective, clinicians can target HIV-antibody-positive individuals who have declining CD4 cells for prophylactic treatment when it is most likely to be effective. Epidemiology can also be used to describe the impact of treatment on the natural history of disease. For example, a cohort study design was used to demonstrate the public health effectiveness of combined highly active retroviral therapy on reducing both the incidence of AIDS and extending survival of those who already had the disease (Detels *et al.*, 1998). Thus, describing the natural history of AIDS among both treated and untreated individuals has assisted researchers to focus their studies and clinicians to use the limited treatment

modalities available more effectively (Phair *et al*. 1992). The field of “clinical epidemiology” applies research on the natural history of disease to improving the diagnostic accuracy of physicians in their clinical practice (Sackett *et al.* 1991).

3. Community diagnosis. Epidemiological surveys are often used to establish the morbidity and mortality from specific diseases, allowing efficient use of limited public health funds for control of those diseases having the greatest negative impact on the health of the community. For example, an epidemiological survey in one area of China identified the epidemic of HIV due to plasma donations in villages (Wu & Detels, 1995; Wu et al., 2001; Ji et al., 2006). The use of disability-adjusted years (DALYs) has allowed quantification of the importance of non-lethal conditions on the public’s health (Murray, 1994).
4. Describe the clinical picture of a disease. Epidemiological strategies can identify who is likely to get a disease such as capillariasis, the characteristic symptoms and signs, the extent of the epidemic, the risk factors, and the causative agent, and can help to determine the effectiveness of treatment and control efforts (Detels et al., 1969).
5. Identify factors that increase or decrease the risk of acquiring disease. Having specific characteristics increases the probability that individuals will or will not develop disease. These `risk factors' may be social (smoking, drinking), genetic (ethnicity), dietary (saturated fats, vitamin deficiencies), and so on. Knowing these risk factors can often provide public health professionals with the necessary tools to design effective programmes to intervene before disease occurs. For example, descriptive, cross-sectional, case-control, cohort, and

intervention studies have all shown that smoking is the biggest single risk factor for ill health, because it is a major risk factor for cardiovascular disease, chronic respiratory disease, and many cancers (for example, of the lung, nasopharynx, and bladder). Thus, smoking is the leading cause of disability and death in developed countries, if not the world. Health education campaigns and other strategies to stop or reduce smoking, based on these epidemiologic studies, are now a major public health activity in most countries of the world.

6. Identify precursors of disease and syndromes. High blood pressure, a treatable condition, has been identified through case-control studies and cohort studies as a precursor to heart disease, stroke, and kidney disease (Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure, 1997).
7. Test the efficacy of intervention strategies. A primary objective of public health is to prevent disease through intervention in the disease process. But a vaccine or other intervention programme must be proven to be effective before it is used in the community. Double-blind placebo-controlled trials are a necessary step in developing an intervention programme, whether that programme is administration of a new vaccine, a behavioural-intervention strategy to stop smoking or a community intervention study to lower heart disease. Although it may be argued that injection of a saline placebo is no longer considered ethical, a proven vaccine, such as polio, can often be used as a placebo for a trial of a new vaccine for a different disease, as was used for trials of rubella vaccines in Taiwan (Detels *et al.* 1969). Widespread use of an intervention not subjected to

epidemiological studies of efficacy may result in implementation of an ineffective intervention programme at great public expense and may actually result in greater morbidity and mortality because of an increased reliance on the favoured but unproven intervention and a reduced use of other strategies which are thought to be less effective but which are actually more effective.

Although an intervention such as a vaccine may have been demonstrated to have efficacy in double-blind trials, it may fail to provide protection when used in the community. Double-blind trials may demonstrate the `biological efficacy' of the vaccine; but if the vaccine is not acceptable to the majority of the public, they will refuse to be vaccinated, and the “public health effectiveness” of the vaccine will be very low. For example, the typhoid vaccine provided some protection against small infecting inocula, but the frequency of unpleasant side-effects with the whole cell vaccines and the need for multiple injections in the past influenced many people against being vaccinated (Chin, 2000).

Another problem of inferring public health efficacy from small vaccine trials is that volunteers for vaccine trials may not be representative of the general public which needs to be protected against a specific disease. Thus broad-based intervention trials also need to be carried out, to demonstrate the acceptability and public health effectiveness of a vaccine or other intervention to the population in need of protection.

Since there are adverse side-effects associated with any vaccine, ongoing evaluations of the cost-benefit relationship of specific vaccines are important. By comparing the incidence of smallpox with the incidence of adverse side-effects

from the smallpox vaccine, Lane *et al.* (1969) demonstrated that more disease resulted from routine use of the vaccine in the U.S. than by transmission from imported cases.

There are several epidemiological strategies that can be used for ongoing evaluation of intervention programmes. Serial cross-sectional studies can be used to determine if there has been a change in the prevalence of disease or of indicators of health status over time. The cohort design can be used to compare incidence of disease in comparable populations receiving and not receiving the prevention programme. The case-control design can be used to determine if there are differences in the proportion of cases and non-cases who had the intervention programme.

8. Investigate epidemics of unknown aetiology. Epidemiological strategies were used to establish the extent, cause, modes of transmission, and risk factors for Ebola haemorrhagic fever, which first occurred in the Congo in 1976 (Feldmann et al., 1996; eMedicine.com, 2005), and capillariasis in the Philippines in the mid-1960s (Detels et al., 1969).
9. Evaluate public health programmes. Departments of health are engaged in a variety of activities to promote the health of the community, ranging from vaccination programmes to clinics for the treatment of specific diseases. Ongoing evaluation of such programmes is necessary to assure that they continue to be cost-effective. Periodic review of routinely collected health statistics can provide information about the effectiveness of many programmes. For those programmes for which relevant statistics are not routinely available, cohort studies and serial

cross-sectional studies of the incidence and changing prevalence of the targeted disease in the populations which are the intended recipients of these programmes can measure whether the programmes have had an impact and are cost effective. For example, most countries have established STD clinics, but studies in Thailand (Prempree et al., 2007) and Beijing (Zhao et al., 2007) have demonstrated that the majority of patients with STDs do not attend the government STD clinics.

10. Elucidate mechanisms of disease transmission. Understanding the mechanisms of disease transmission can suggest ways in which public health professionals can protect the public by stopping transmission of the disease agent. Epidemiological studies of the various arboviral encephalitides have incriminated certain species of mosquitoes as the vectors of disease and specific animals as the reservoirs for the viruses. For example, public health efforts in California to prevent western equine encephalitis have concentrated on control of the mosquito vector and vaccination of horses, which are a reservoir of the virus. Although an effective vaccine for smallpox had been available for almost two hundred years, eradication of the disease was not achieved until the recognition that the low infectivity of varicella virus and the relatively long incubation for development of smallpox could be used to develop a strategy of surveillance for cases, with identification and immediate vaccination of all susceptible contacts (containment). Using this containment strategy based on epidemiologic principles, smallpox was eradicated through a worldwide effort in less than ten years (Fenner *et al.* 1988).
11. Elucidate the molecular and genetic determinants of disease progression. Epidemiological strategies helped to elucidate the changes in the human immune response by CD4 and CD8 cells that accompany infection by disease agents (Detels et al., 1983; Fahey et al., 1984; Ho et al., 1995), and genetic factors (such as CCR5 absence or heterozygosity) that prevent HIV infection and slow progression of HIV disease (Liu *et al*., 2004).

From the examples given above, it should be clear that epidemiology functions as the backbone or core of evidence-based public health practices, as well as a key strategy for evaluating the effectiveness of both clinical and public health interventions.

Applications of epidemiology

Specific epidemiological study designs are used to achieve specific public health goals. These goals range from identifying a suspected exposure-disease relationship to establishing that relationship, to designing an intervention to prevent it, and, finally, to assessing the effectiveness of that intervention. The usual sequence of study designs in the identification and resolution of a disease problem are:

- Ecologic studies
- Cross-sectional (prevalence) surveys
- Case-control studies
- Cohort studies
- Experimental studies

There are, however, many exceptions to the application of this sequence of study designs, depending on such things as the prevalence and virulence of the agent and the nature of the human response to the agent.

The earliest suspicion that a relationship exists between a disease and a possible causative factor is frequently obtained from observing correlations between exposure

and disease from existing data such as mortality statistics and surveys of personal or national characteristics. These can be correlations observed across geographical areas (ecological studies) or over time, or a combination of both. Many of the initial epidemiological investigations into chronic bronchitis used vital statistics data, particularly data on mortality. Case-control studies identified smoking as a possible causal factor for chronic bronchitis. Subsequent prevalence studies confirmed the relationship, as have cohort studies. Finally, a decline in respiratory symptoms of chronic bronchitis and a concurrent, but slower decline in lung function, has been observed in individuals who cease smoking (Colley 1992).

Although this is the usual sequence in which the various epidemiological study designs are applied, there are exceptions to this sequence. Furthermore, all study designs are not appropriate to answer all health questions. The usual applications of each of the different epidemiological study designs and the limitations of each are, therefore, presented briefly below and in greater depth in subsequent chapters.

Ecologic studies (see Volume 2, Chapter 6.2)

The use of existing statistics to correlate the prevalence or incidence of disease in groups or populations to the frequency or trends over time of suspected causal factors in specific localities has often provided the first clues that a particular factor may cause a specific disease. These epidemiological strategies, however, document only the co-occurrence of disease and other factors in a population; the risk factors and the disease may not be occurring in the same people within the population. These types of

descriptive studies are inexpensive and relatively easy to do, but the co-occurrence observed may be due merely to chance. For example, the incidence of both heart disease and lung cancer has increased concurrently with the prevalence of automatic washing machines in the US. Few people, however, would attribute the increase in these two diseases to the use of automatic washing machines. Thus, ecologic studies must be interpreted with caution. Nonetheless, they often reveal important relationships and can provide a strong rationale for undertaking more expensive analytic studies.

Cross-sectional/prevalence surveys (see Volume 2, Chapters 6.3, 6.4, 6.17)

Cross-sectional/prevalence surveys establish the frequency of disease and other factors in a community. Since they require the collection of data, however, they can be expensive. They are useful to estimate the number of people in a population who have disease and can also identify the difference in frequency of disease in different subpopulations. This descriptive information is particularly useful to health administrators who are responsible for developing appropriate and effective public health programmes. Cross-sectional studies can also be used to document the co-occurrence of disease and suspected risk factors not only in the population but also in specific individuals within the population. The cross-sectional study design is useful to study chronic diseases such as multiple sclerosis, which have a reasonably high prevalence, but an incidence that is too low to make a cohort study feasible (Detels *et al.* 1978). On the other hand, they are not useful for studying diseases that have a very low prevalence, such as subacute sclerosing panencephalitis or variant Creutzfeldt-Jakob disease. Cross-sectional studies are subject to problems of respondent bias,

recall bias, and undocumented confounders. Further, unless historical information is obtained from all the individuals surveyed, the time-relationship between the factor and the disease is not known. Further, prevalence surveys identify people who have survived to that time point with disease and, thus, under-represent people with a short course of disease.

The cross-sectional study design is used in two special types of studies: field studies and surveillance. Field studies are usually investigations of acute outbreaks which require immediate identification of the causative factors if effective public health interventions are to be implemented in a timely fashion. Surveillance is the monitoring of disease or health-related factors over time and uses serial cross-sectional surveys to observe trends. Surveillance is important to identify diseases that are becoming an increasing public health problem, to assure that diseases already brought under control remain under control, and to evaluate the impact of public health intervention strategies.

Case-control studies (see Volume 2, Chapter 6.5)

The case-control study compares the prevalence of suspected causal factors between individuals with disease and controls. If the prevalence of the factor is significantly different in cases than it is in controls, this factor may be associated with the disease. Although case-control studies can identify associations, they do not measure risk. An estimate of relative risk, however, can be derived by calculating the odds ratio. Case-control studies are often the analytic study design used initially to investigate a suspected association. Compared to cohort and experimental studies, they are usually

relatively cheap and easy to do. Cases can often be selected from hospital patients and controls either from hospitalized patients with other diseases or by using algorithms or formulas for selecting community (neighborhood) or other types of controls. Selection bias, however, is often a problem, especially when using either hospitalized cases or controls. The participants are seen only once, and no follow-up is necessary. Although time sequences can often be established retrospectively for factors elicited by interview, they usually cannot be for laboratory test results. Thus, an elevation in factor B may either be causally related, or it may be a result of the disease process and not a cause. Furthermore, factors elicited from interview are subject to recall bias; for example, patients are often better motivated to recall events than controls because they are concerned about their disease. The case-control study is particularly useful for exploring relationships noted in observational studies. A hypothesis, however, is necessary for case-control studies. Relationships will be observed only for those factors studied. Case-control studies are not useful for determining the spectrum of health outcomes resulting from specific exposures, since a definition of a case is required in order to do a case-control study. On the other hand, case-control studies are the method of choice for studying rare diseases. Case-control studies are often indicated when a specific health question needs to be answered quickly.

Cohort studies (see Munoz, Volume 2, Chapter 6.6)

Cohort studies follow defined groups of people without disease to identify risk factors associated with disease occurrence. Cohort studies have the advantage of establishing the temporal relationship between an exposure and a health outcome, and, thus, they

measure risk directly. Because the population studied is often defined on the basis of its known likelihood of exposure to suspected factors, cohort studies are particularly suitable for investigating health hazards associated with environmental or occupational exposures. Further, cohort studies will measure more than one outcome of a given exposure and, therefore, are useful for defining the spectrum of disease resulting from exposure to a given factor. Occasionally a cohort study is done to elucidate the natural history of a disease when a group can be identified that has a high incidence of disease but in which specific risk factors are not known. Although this cohort is not defined on the basis of a known exposure, questions are asked and biological specimens are collected from which exposure variables can be identified concurrently or in the future. Unfortunately, cohort studies are both expensive and time consuming. Unless the investigator can define a cohort in which risk factors were measured at some time in the past and has assurance that the cohort has been completely followed up for disease outcome in the interim (historical cohort), the cohort design can take years to decades to yield information about the risks of disease resulting from exposure to specific factors. Assuring that participants remain in a cohort study for such long periods of time is both difficult and expensive. Further, the impact of those who drop out of follow-up must be taken into account in the analysis and interpretation of cohort studies. Finally, exposures may vary over time, complicating the analysis of their impact. Because of the cost and complexity of cohort studies, they are usually done only after descriptive, cross-sectional, and/or case-control studies have suggested a causal relationship. The size of the cohort to be studied is dependent, in part, on the anticipated incidence of the disease resulting from the exposure. For diseases with a very low incidence, population-

based cohort studies usually are not feasible, either in terms of the logistics or of the expense of following very large numbers of people, or both. Cohort studies establish the risk of disease associated with exposure to a factor, but do not “prove” that the factor is causal. The observed factor merely may be very closely correlated with the real causative factor or may even be related to the participants' choice to be exposed.

A variant of the cohort study which has become popular is the “nested case-control study” (Gange *et al.* 1997). Cases which arise from a cohort study are compared to individuals followed in the cohort who have not developed disease using the usual case-control analytic strategies. The advantage of this type of study is that the exposure variables are collected before knowledge of the outcome and, therefore, are unlikely to be tainted by recall bias.

Experimental studies (see Volume 2, Chapters 6.7-6.10)

Experimental studies differ from cohort studies because it is the investigator who makes the decision about who will be exposed to the factor based on the specific design factors to be employed (e.g., randomization, matching, etc.). Therefore, confounding factors such as choice that may have led to the subjects being exposed in the cohort studies are usually not a problem in experimental studies. Because epidemiologists usually study human populations, there are few opportunities for an investigator to deliberately expose participants to a suspected factor. On the other hand, intervention studies of randomly assign individuals (Chapter 6.7) or communities (Chapters 6.7-6.10) to receive or not receive an intervention programme that demonstrates a subsequent

reduction in a specific health outcome in the intervention group do provide strong evidence, if not proof, of a causal relationship. Because of the serious implications of applying an intervention that may alter the biological status of an individual or the sociopolitical behaviour of a community, intervention studies should not be undertaken until the probability of a causal or risk relationship has been well established using the other types of study designs.

Meta-analysis (see Volume 2, Chapter 6.14)

Because individual epidemiologic studies rarely provide proof of causation and results of different studies can vary for a number of reasons, including small sample size, a recent trend has been to combine similar studies to increase the power of the analysis. This strategy for data synthesis is known as “meta-analysis”. It has been especially helpful in studying diseases with a low incidence or where similar studies have given conflicting results.

Methodologic Issues (see Volume 2, Chapters 6.12 and 6.13)

Epidemiologic studies, because they deal with humans, are subject to problems such as bias (deviation of results from truth) due to the strategies of recruiting participants, or to differential recall among persons with and without disease and confounding due to factors which are associated with both the exposure variables and outcome variable under study. In the last several decades many new techniques have been developed to reduce the effect of these factors which can influence the outcome of a study and, in some instances, can cause apparent relationships to be observed which are, in fact,

false.

Summary

Epidemiology is the core science of public health because it defines health and disease in human populations, describes disease etiology, and evaluates public health control efforts. Epidemiology achieves these goals through a variety of strategies and methods. Epidemiology is a dynamic science, and is continually evolving new strategies and methods in support of public health goals.

References

Alter M. (1976). Is multiple sclerosis an age-dependent host response to measles? *Lancet* 1(7957), 456-7.

Brookmeyer, R. and Damiano, A. (1989). Statistical methods for short-term projections of AIDS incidence. *Statistics in Medicine*, **8**, 23-34.

Chapman, S.W. and Henderson, H.M. (1994). New and emerging pathogens multiply resistant *Mycobacterium tuberculosis. Current Opinion in Infectious Diseases*, **7**, 231-7.

Chin, J. (2000). *Control of communicable diseases manual*. American Public Health Association, Washington, D.C.

Chin, J., Sato P., and Mann, J. (1990). Projections of HIV infections and AIDS cases to the year 2000. *Bulletin of the World Health Organization*, **68**, 1-11.

Colley, J.R.T., Holland W.W., and Corkhill R.T. (1974). Influence of passive smoking and parental phlegm on pneumonia and bronchitis in early childhood. *Lancet*, **2**(7888), 1031-4.

Colley, J.R.T. (1991). Major public health problems; respiratory system. In *Oxford textbook of public health* (2nd edn) (ed. W.W. Holland, R. Detels, and G. Knox), Vol. 3, pp. 227-48. Oxford University Press.

Communicable Disease Surveillance Centre (1994). Invasive group A streptococcal infections in Gloucestershire. *Communicable Disease Report* (England/Wales), **4**, 97-100.

Detels R, Fahey JL, Schwartz K, Greene RS, Visscher BR, Gottlieb MS. Relation between sexual practices and T-cell subsets in homosexually active men. *Lancet* 1983;**1**(8325):609-611

Detels, R., Grayston, J.T., Kim, K.S.W., Chen, K.P., Gale, J.L., Beasley, R.P., and Gutman, L. (1969). Prevention of clinical and subclinical rubella infection: efficacy of three HPV-77 derivative vaccines. *American Journal of Diseases of Children*, **118**, 295-300.

Detels R, Gutman L, Jaramillo J, Zerrudo E, Banzon T, Valera J, Murrell KD, Cross J, Dizon JJ. An epidemic of intestinal capillariasis in man: A study in a barrio in northern Luzon. *Am J Trop Med Hyg* 1969;**18**(5):676-682

Detels, R., McNew, J., Brody, J.A., and Edgar, A.H. (1973). Further epidemiological studies of subacute sclerosing panencephalitis. *Lancet*, **819**, 11-14.

Detels R, Muñoz A, McFarlane G, Kingsley LA, Margolick J, Giorgi J, Schrager LK, Phair JP. Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV infection duration. *JAMA* 1998;**280**(17):1497-1503

Detels R, Phair JP, Saah AJ, Rinaldo CR, Muñoz A, Kaslow RA, Seminara D, Schrager L, Vermund S. Recent scientific contributions to understanding HIV/AIDS from the Multicenter AIDS Cohort Study. *J Epidemiol [Japan]* 1992;**2**(2):S11-S19

Detels, R., Visscher, B.R., Haile, R.W., Malmgren, R.M., Dudley, J.P., and Coulson, A.H. (1978). Multiple sclerosis and age at migration. *American Journal of Epidemiology*, **108**, 386-93.

Detels, R., Visscher, B.R., Fahey, J.L., Sever, J.L., Gravell, M., Madden, D.L., Schwartz, K., Dudley, J.P., English, P.A., Powers, H., Clark, V.A., and Gottlieb, M.S. (1987). Predictors of clinical AIDS in young homosexual men in a high-risk area. *International Journal of Epidemiology*, **16**, 271-6.

Detels, R., Munoz, A., McFarlane, G., Kingsley, L.A., Margolick, J.B., Giorgi, J., Schrager, L.K., and Phair, J.P. (1998). Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV duration. *Journal of the American Medical Association*, **280** (17), 1497-1503.

Fahey J, Prince H, Weaver M, Groopman J, Visscher B, Schwartz K, Detels R. Quantitative changes in T helper or T suppressor/cytotoxic lymphocyte subsets that distinguish acquired immune deficiency syndrome from other immune subset disorders. *Am J Med* 1984;**76**:95-100

eMedicine.com. Ebola virus. November 2003. eMedicine.com, Inc. 2005.

Feldmann H, Slenczka W, Klenk HD. Emerging and reemerging of filoviruses. *Arch Virol* 11 Suppl:77-100, 1996.

Fenner, F., Henderson, D.A., Arita, I., Jezek, Z., and Ladnyi, I.D. (1988). *Smallpox and its eradication*. World Health Organization, Geneva.

Gange, S., Munoz, A., Schrager, L.K., Margolick, J.B., Giorgi, J.V., Saah, A., Rinaldo, C.R., Detels, R., and Phair, J.P. (1997). Design of nested studies to identify factors related to late progression of HIV infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retroviruses*, **15** (suppl):S5-S9.

Hill, A.B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, **58**, 295-300.

Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virion and CD4 lymphocytes in HIV-1 infection. Nature 1995; 373:123-126.

Hoggett K, Greenland S. Validity and bias in epidemiologic research. *Oxford Textbook of Public Health*, chapter 6.13, 5th ed. Oxford University Press, Oxford, England, 2008.

Holland, W.W., Olsen, J., and Florey, C.D.V. (eds) (2007). The Development of Modern Epidemiology: Personal Reports From Those Who Were There. Oxford University Press, Oxford, England, pgs 177-178.

Hypertension Detection and Follow-up Program Co-operative Group. (1979). Five-year findings of the hypertension detection and follow-up program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension. *Journal of the American Medical Association*, **242**, 2562-71.

Imagawa, D.T., Lee, M.H., Wolinsky, S.M., Sano, K., Morales, F., Kwok, S., Sninsky, J.J., Nishanian, P.G., Giorgi, J., Fahey, J.L., Dudley, J., Visscher, B.R., and Detels, R. (1989). Human immunodeficiency virus type 1 infection in homosexual men who remain seronegative for prolonged periods. *New England Journal of Medicine*, **320**, 1458-62.

Ji G, Detels R, Wu Z, Yin Y. Correlates of HIV infection among former blood/plasma donors in rural China. *AIDS* 2006; 20(4):585-591.

Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (1997). The Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Arch Intl Med* 157:2413-46.

Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR. The Multicenter AIDS Cohort Study: Rationale, organization, and selected characteristics of the participants. *Am J Epidemiol* 1987;**126**(2):310-318

Lane, J.M., Ruben, F.L., Neff, J.M., and Millar, J.D. (1969). Complications of smallpox vaccination, 1968: national surveillance in the United States. *New England Journal of Medicine*, **281**, 1201-8.

Lilienfeld, D.E. (1978). Definitions of epidemiology. *American Journal of Epidemiology*, **107**, 87-90.

Liu C, Carrington M, Kaslow RA, Gao X, Rinaldo CR, Jacobson LP, Margolick JB, Phair J, O'Brien SJ, Detels R. (2004). Lack of associations between HLA class II alleles and resistance to HIV-1 infection among white, non-Hispanic homosexual men. *Journal of Acquired Immune Deficiency Syndromes*, **37(2),** 1313-1317.

Mac Kenzie WR, Hoxie NJ, Proctor ME, Gradus MS, Blair KA, Peterson DE, et al. A massive outbreak in Milwaukee of cryptosporidium infection transmitted through the

public water supply. *N Engl J Med* 1994;331:161-7.

Morris, J.N. (1957). *Uses of epidemiology*. Churchill Livingstone, London.

Morris, J.N. (1975). *Uses of epidemiology* (3rd edn). Churchill Livingstone, London.

Murray, C. J. (1994), Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bulletin of the World Health Organisation* 72:429-45.

Phair, J., Jacobson, L., Detels, R., Rinaldo, C., Saah, A., Schrager, L., and Muñoz, A. (1992). Acquired immune deficiency syndrome occurring within 5 years of infection with human immunodeficiency virus type-1: the Multicenter AIDS Cohort Study. *Journal of Acquired Immune Deficiency Syndromes*, **5**, 490-6.

Polk, B.F., Fox, R., Brookmeyer, R., Kanchanaraksa, S., Kaslow, R., Visscher, B., Rinaldo, C., and Phair, J. (1987). Predictors of the acquired immunodeficiency syndrome developing in a cohort of seropositive homosexual men. *New England Journal of Medicine*, **316**, 61-6.

Prempree P, Detels R, Ungkasrithongkul M, Meksawasdichai S, Panthong S, Ungpanich V. The sources of treatment of sexually transmissible infections in a rural community in central Thailand. *Sex Hlth* 2007, **4(1)**, 17-9.

Rothman, K.J. (ed.) (1988). *Causal inference*. Epidemiology Resources, Chestnut Hill, Massachusetts.

Sackett, D.L., Haynes, R.B., Buyatt, G.H., and Tugwell, P. (1991). *Clinical Epidemiology, a basic science for clinical medicine*. Little, Brown and Company, London.

Schwartz, J. and Zeger, S. (1990). Passive smoking, air pollution, and acute respiratory symptoms in a diary of student nurses. *American Review of Respiratory Disease*, **141**, 62-7.

Sullivan, C.B., Visscher, B.R., and Detels, R. (1984). Multiple sclerosis and age at exposure to childhood diseases and animals: cases and their friends. *Neurology*, **34**, 1144-8.

Tager, I.B., Weiss, S.T., Rosner, B., and Speizer, F.E. (1979). Effect of parental cigarette smoking on the pulmonary function of children. *American Journal of Epidemiology*, **110**, 15-26.

Tashkin, D.P., Clark, V.A., Simmons, M., Reems, C., Coulson, A.H., Bourque, L.B., Sayre, J.W., Detels, R., and Rokaw, S. (1984). The UCLA Population Studies of Chronic Obstructive Respiratory Disease: VII. relationship between parental smoking and children's lung function. *American Review of Respiratory Disease*, **129**, 891-7.

Taylor, J.M.G. (1989). Models for the HIV infection and AIDS epidemic in the United States. *Statistics in Medicine*, **8**, 45-58.

Wu Z, Detels R. HIV-1 infection in commercial plasma donors in China. *Lancet* 1995;**346**:61-62

Wu Z, Rou K, Detels R. Prevalence of HIV infection among former commercial plasma donors in rural Eastern China. *Hlth Policy Planning* 2001;**16**(2):41-46

Zhao G, Detels R, Gu F, Li D, Li X, Li Y, Zhao K. The distribution of people seeking STD services in the various types of health care facilities in Chao Yang District, Beijing, Chinal. Sex Transm Dis, 2007.

(Note: same figure as in third edition)

Figure 1. 1. The triadic relationship between agent, host, and environment in epidemiology