2

Causal Concepts

All diseases have a sort of natural life; that is, they begin, grow, attain maturity, decline, and terminate.—William Farr (1862, p. 194)

2.1 NATURAL HISTORY OF DISEASE

Simple Model

The **natural history of disease** refers to the progression of a disease in an individual over time. This includes all disease-related phenomena from before initiation of the disease (the **stage of susceptibility**) until resolution of the disease (the **stage of recovery, disability, or death**).

Figure 2.1 is a schematic of the natural history of a disease when considering the ultimate (last) causal exposure in a simple causal model. In the period following exposure to the agent, the prospective case enters a **stage of subclinical disease** (also called the **preclinical phase**). This corresponds to the time during which the etiologic agent is present within the body but has not yet produced discernible signs or symptoms. This phenomenon was recognized by Jacob Henle more than a century-and-a-half ago when he wrote "the symptoms of disease do not appear directly after the entry of the contagious agent but rather after a certain period, which varies with different contagions" (Henle, 1840, p. 14).

Both infectious and noninfectious agents are characterized by subclinical stages of disease. With infectious diseases, this corresponds to the **incubation period**. With noninfectious diseases, this may be called the **latent period** or (loosely) the **empirical induction period**. In considering a cancer, for example, the latent period corresponds to the time between the ultimate neoplastic transformation that leads to nondifferentiated and uncontrolled cell proliferation and the progression of these changes to a state that produces recognized physiologic disturbances.

Incubation periods vary considerably for agent–disease pairs. Some diseases are characterized by short incubation periods (e.g., cholera has a brief 24- to 48-hour incubation period), others by intermediate incubation periods (e.g., chickenpox has a typical incubation period of 2 to 3 weeks), and still others by extended incubation [e.g., the median incubation period of acquired immunodeficiency syndrome (AIDS) is approximately 10 years]. Table 2.1 lists incubation periods for selected infectious disease agents. Notice that even for a given disease incubation periods can vary considerably. For example, the incubation period for human immunodeficiency virus (HIV) and AIDS ranges from 3
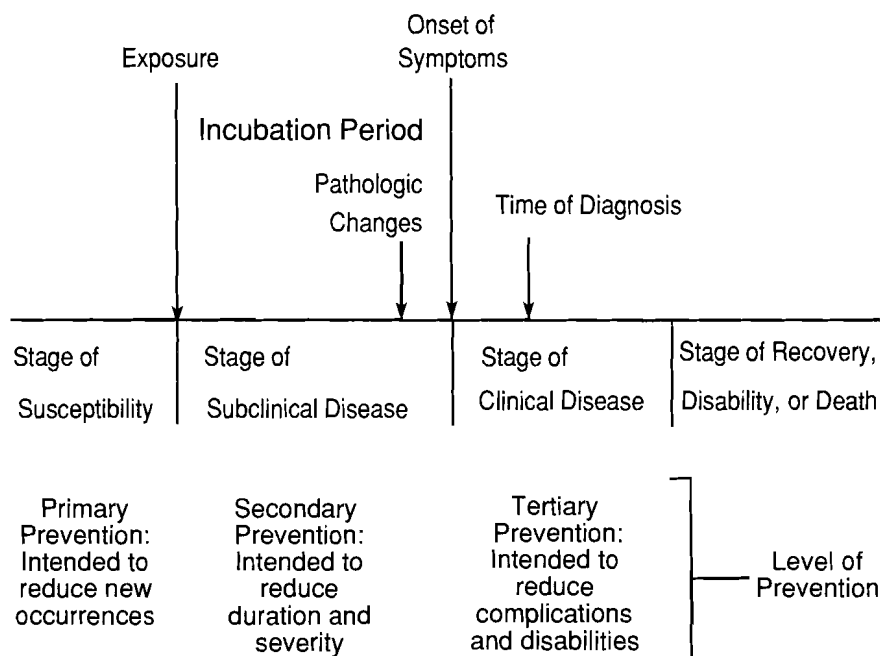

Figure 2.1. Stages in the natural history of disease and prevention, ultimate agent.

TABLE 2.1. Incubation Periods for Selected Infectious Diseases

Disease	Typical Incubation Period[a]
Acquired immune deficiency syndrome	Infection to appearance of antibodies: 1–3 months; median time to diagnosis: approx. 10 years; treatment lengthens the incubation period
Amebiasis	2–4 weeks
Chickenpox	13–17 days
Common cold	2 days
Hepatitis B	60–90 days
Influenza	1–5 days
Legionellosis	5–6 days
Malaria (*Plasmodium vivax* and *P. ovale*)	14 days
Malaria (*P. malariae*)	30 days
Malaria (*P. Falciparum*)	12 days
Measles	7–18 days
Mumps	12–25 days
Poliomyelitis, acute paralytic	7–14 days
Plague	2–6 days
Rabies	2–8 weeks (depends on severity of wound)
Salmonellosis	12–36 hours
Schistosomiasis	2–6 weeks
Staphylococcal food poisoning	2–4 days
Tetanus	3–21 days

Source: Benensen (1990).
[a]The incubation periods of some diseases vary considerably. See the latest edition of *Control of Communicable Diaseases in Man* (Benensen, 1990) for details.

to more than 20 years. The empirical induction period for leukemia caused by exposure to the fallout from the atomic bomb blast in Hiroshima ranged from 2 to more than 12 years (Cobb et al., 1959). The empirical induction period for bladder tumors in industrial dyestuff workers ranges from 5 to more than 40 years (Fig. 2.2). Variability in incubation is due to differences in host resistance, pathogenicity of the agent, exposure-dose, and the availability of cofactors responsible for disease.

The **stage of clinical disease** begins with a patient's first symptoms and ends with resolution of the disease through recovery, disability, or death. Depending on host factors, speed of the disease process, and efficacy of the health-care system, the lag between the onset of symptoms and diagnosis may be considerable. Therefore, it is the onset of symptoms and not the time of diagnosis that marks the beginning of the clinical stage of disease.

Understanding the natural history of a disease is essential when studying its epidemiology. For example, AIDS can only be understood after identifying the multifarious stages of HIV infection (Fig. 2.3). Exposure to HIV is followed by an acute response that may be accompanied by unrecognized flulike symptoms. Although prospective cases do not exhibit detectable antibodies until approximately 6 weeks following the initial infection, they can still be infectious during this acute phase. A lengthy incubation period ensues during which CD4+ lymphocyte counts decline and the patient is free from symptoms. The risk of developing AIDS is low during the initial years of infection but increases over time as the immune response is progressively destroyed. AIDS then expresses itself in many different ways (e.g., opportunistic infections, various cancers, dementia, wasting syndrome).
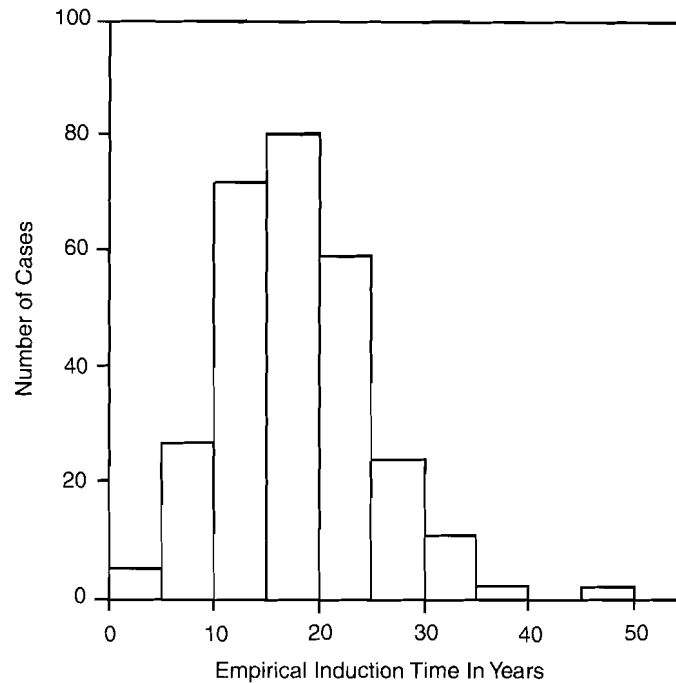
Figure 2.2. *Distribution of empirical induction period of urinary bladder tumors in industrial dyestuff workers. (Source: Case et al., 1954, p. 85; reproduced with permission of BMJ Publishing Group.)*

Primary, Secondary, and Tertiary Prevention

Disease prevention efforts can be classified according to the stage of disease at which they are applied (Fig. 2.1). **Primary prevention** is directed toward the stage of susceptibility, before the pathogen establishes itself in the body. The goal of primary prevention is to prevent the disease from occurring in order to reduce its incidence in the community. Examples of primary prevention include needle-exchange programs to prevent the spread of HIV, various vaccination programs, and smoking prevention programs.

Secondary prevention is directed toward the subclinical stage of disease, in people who carry the agent but are not yet symptomatic. That goal of secondary prevention is to either delay the emergence of disease or reduce its severity once it emerges. Treating asymptomatic HIV-infected patients with antiretroviral agents to delay the onset of AIDS is a form of secondary prevention. Screening for cervical cancer via Pap smears is also a form of secondary prevention.

Tertiary prevention is directed toward the clinical stage of disease. The aim of tertiary prevention is to prevent or minimize the progression of the disease or its sequelae. Screening people with diabetes for diabetic retinopathy to promptly treat the progression of blindness is a form of tertiary prevention. Tertiary prevention is especially import for preventing disability and containing the costs of health care associated with chronic degenerative conditions.

Multiple Causal Factors

A central tenet of modern epidemiology is that *every cause is interdependent on other causal factors*. That is, diseases are caused by the cumulative effect of a number of factors.
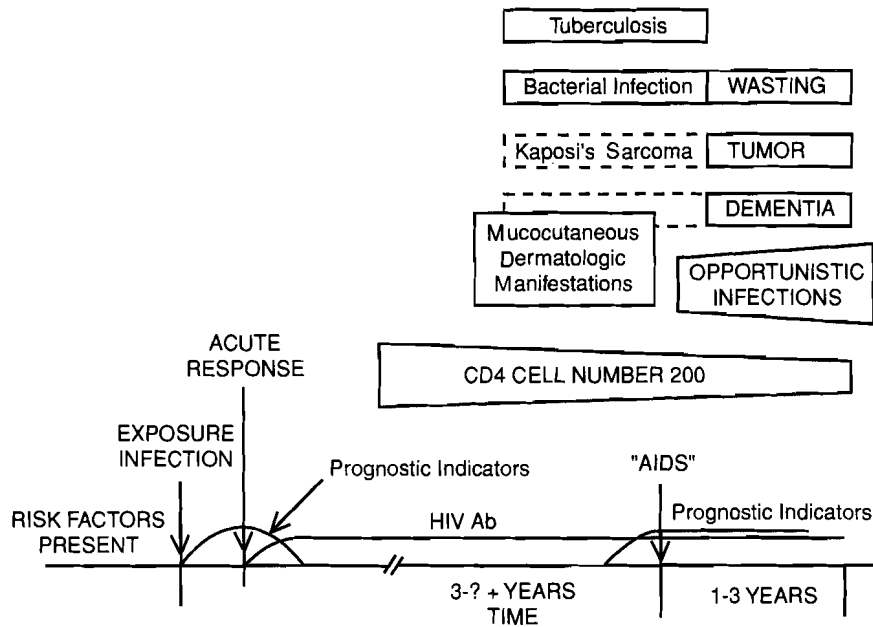
Figure 2.3. Natural history and progression of HIV infection and AIDS. (Source: Cotton, D. J. (1995). The natural history of HIV infection: implications for clinical trials. In D. M. Finkelstein & D. A. Schoenfeld (Eds.). AIDS Clinical Trials, p. 8; Copyright © 1995 by Wiley-Liss, Inc. Reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.)

Even infectious disease agents do not act alone. Two people exposed identically to the same infectious agent experience different symptoms—sometimes no symptoms at all—depending on various agent, host, and environmental contributors (see Section 2.3). When considering multiple causal factors, a sophisticated view of "incubation" is needed. This is aided by considering separate induction and latent periods. The **induction period** of a causal factor is the time between its causal action and initiation of the disease (Fig. 2.4). The **latent period** of a disease is the time between disease initiation and disease detection. The combined induction and latent periods is the **empirical induction period** (Rothman, 1981).

Consider as an example the relation between dietary fats and myocardial infarction. Excessive dietary fat consumption exerts one of its causal actions through atherosclerosis. However, the effects of this causal action may not be seen until a blood clot gets lodged in the affected coronary vessel. Many years may pass between the causal action of coronary
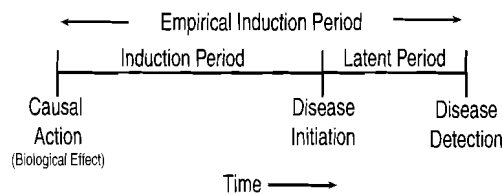

Figure 2.4. Induction period, latent period, and empirical induction period.
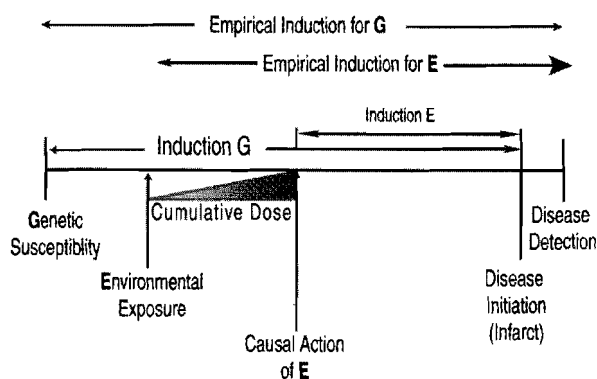
Figure 2.5. *Induction periods for genetic and environmental (excessive dietary fats) for myocardial infarct.*

narrowing and the infarct that initiates myocardial ischemia. In the meantime, other causal factors may contribute to the disease process (e.g., hypertension, obesity, sedentary lifestyle, genetic factors), each exerting its causal action, directly or indirectly.

Conceptualizing the induction period for a chronic exposure is particularly challenging when a cumulative dose is involved in its biological effect. Figure 2.5 is a schematic of the natural history of a thromboembolic-induced myocardial infarct. Let G represent genetic susceptibility to coronary atherosclerosis, and let E represent exposure to excessive dietary fats. To simplify matters, only these two causal factors and the ultimate infarct that initiates the disease are considered. Assume G exerts its causal action at conception and E exerts its causal action after an arbitrary exposure dose is achieved. Notice that G and E have different induction periods. For G, the induction is from conception to initiation of myocardial ischemia. For E, the induction is from the point of atherosclerosis to infarction. Because the effects of the infarct are acute, the latent period between infarction and symptoms is typically short. (The diagram is not drawn to scale to allow for labeling.)

2.2 SPECTRUM OF DISEASE AND "THE ICEBERG"

Disease known to the general practitioner represents only the tip of the iceberg—John Last (1963, p. 30)

Spectrum of Disease

A particular disease can display a broad range of manifestations and severities, ranging from silent in some people to progressive and fulminating in others. When considering infectious diseases, this range of manifestations is referred to as the **gradient of infection**. When considering noninfectious diseases, the range of manifestations is referred to as the **spectrum of disease**. For example, HIV infection ranges from inapparent, to mild disease (e.g., AIDS-related complex), to severe disease (e.g., wasting syndrome). Coronary artery disease exists in the following forms: asymptomatic atherosclerosis, compromised cardiac circulation resulting in transient ischemia and angina, myocardial infarction, and death.

The Iceberg Metaphor

The spectrum of some illnesses has been liked to an iceberg, in that like an iceberg, the bulk of the problem may be hidden from view (Last, 1963; Fig. 2.6). This phenomenon applies to chronic diseases, injuries, and infectious diseases and may be either quantitative or qualitative in exhibited differences. Uncovering disease that might otherwise be below "sea level" by screening and early detection may allow for better disease control. Consider:

- For every successful suicide attempt there are many more unsuccessful attempts and a still larger number of people with depressive illness that might be severe enough to have them wish to end their lives. With appropriate treatment, depressives with suicidal tendencies would be less likely to have suicidal ideation and be less likely to attempt suicide.
- Reported cases of AIDS represents only the tip of HIV infections. With proper antiretroviral therapy, clinical illness may be delayed and transmission averted.
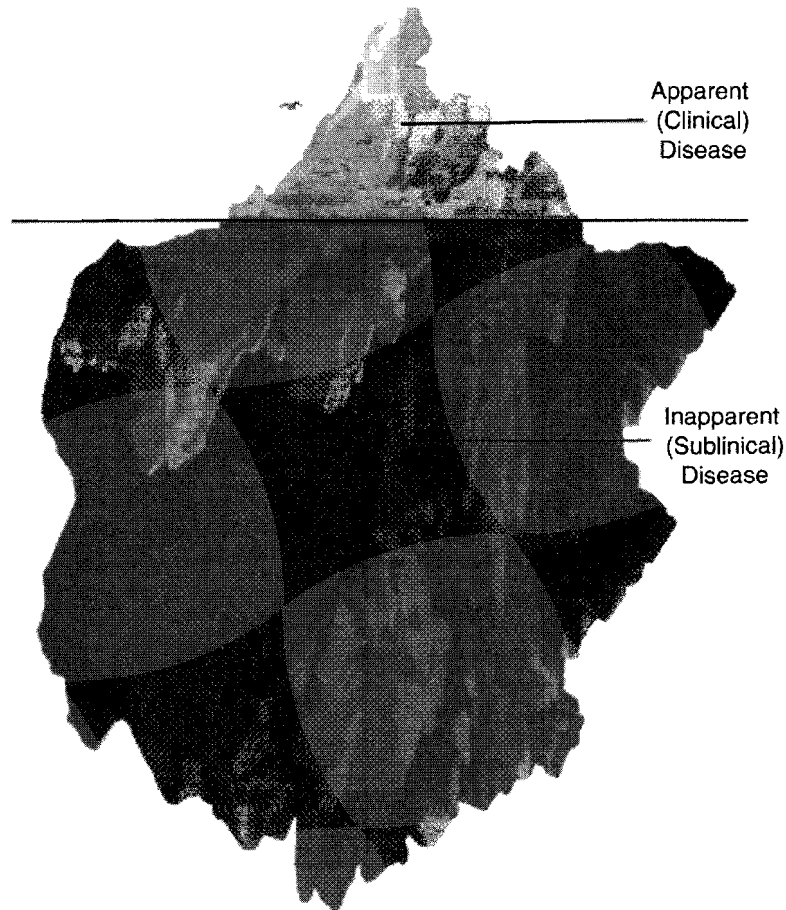

Apparent
(Clinical)
Disease

Inapparent
(Sublinical)
Disease
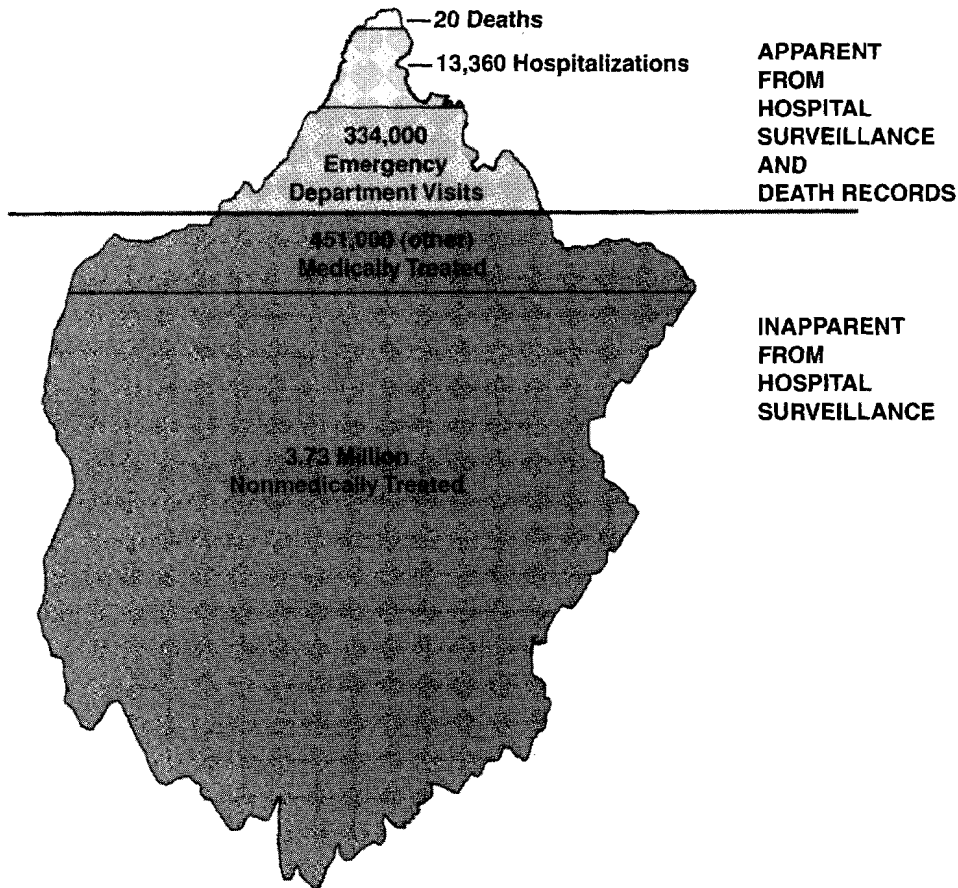
Figure 2.6. *Iceberg metaphor.*

Figure 2.7. *Approximate number of dog bite injuries in the United States annually. (Based on Weiss et al., 1998).*

- Serious dog bite injuries often go undetected. For each fatal dog bite there are about 670 dog bite hospitalizations, 16,000 emergency department visits, 21,000 medical visits to other clinics, and 187,000 nontreated bites (Weiss et al., 1998; Fig. 2.7). With effective recognition, animal control programs can be put into place to prevent dog bite injuries.

Screening for disease below the surface of normal detectability can be part of an effective disease control program.

2.3 CAUSAL CONCEPTS

One reason that the notion of "cause" is so important is that it carries suggestions of relations at a deeper level of understanding than the direct observation under study.—Cox (1986, p. 963)

What Is a Cause?

Definition The *Oxford English Dictionary* (second edition) defines a cause as "that which produces an effect," while *Webster's New Collegiate* (9th edition) defines a cause as "something that brings about an effect or a result." The 18th-century British philosopher David Hume (1772, Section VII) defined a cause as

> ... an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words where, if the first object had not been, the second never had existed.

Thus, a causal factor is an event, an act, a condition, or a state of nature that "initiates or permits, alone or in conjunction with other causes, a sequence of events, resulting in an effect" (Rothman, 1976, p. 588). Some epidemiologists use the term *determinant* interchangeably with "cause," defining a determinant as "any factor . . . that brings about change for better or worse in a health condition" (Susser, 1973, p. 3). In addition, modern epidemiologic definitions of disease causation incorporate an essential element of time: A cause of disease is "an event, condition or characteristic that preceded a disease without which the disease event either would not have occurred at all or would not have occurred until some later time" (Rothman & Greenland, 1998, p. 8). If not for the causal factor, the disease would not have occurred or would have occurred later in the life of the individual. On a population basis, we expect that a change in the level of a causal factor will be accompanied by an increase (or decrease, since in some instances causal factors are preventive) in the incidence of disease, *caeteris parabus* (all other things being equal). We also expect that if the causal factor can be eliminated or diminished, the frequency of disease or its severity will decline.

Necessity and Sufficiency In epidemiology, the term *exposure* has developed to mean any condition considered as a possible cause of disease. An exposure is said to be **necessary** when it always precedes the disease. An exposure (or more, realistically, a set of exposures) is considered **sufficient** when disease becomes inevitable. Four possibilities exist:

	Exposure E is	
	Necessary	Sufficient
1	+	+
2	+	−
3	−	+
4	−	−

1. Causal exposure E is both **necessary and sufficient** if E and disease D are always present together and E acting alone inevitably leads to D ($E \rightarrow D$). Rarely, if ever, are single exposures necessary and sufficient to cause disease. A possible exception to this is in the case of a homozygous genetic abnormality, such as seen in Tay-Sachs disease. But even here, a case can be made that other factors were required before the disease became manifest. Two heterozygous recessive Tay-Sachs parents needed to mate, the recessive alleles needed to be contributed by both parent, genetic

and prenatal screening needed to be avoided, and so on. One can nearly always find other contributing components to a causal mechanism.

2. Causal exposure E is **necessary but not sufficient** if E is always present when disease D occurs but D does not always occur in the presence of E. A cause is necessary when it is part of the definition of the disease, as with infectious phenomena. For example, *Mycobacterium* is necessary for tuberculosis. However, the tubercular bacterium is not always sufficient to cause the disease—it is possible for a person to carry *Mycobacterium* in his or her body yet remain asymptomatic. This implies that complementary factors (F) are required for disease to become manifest ($E + F \to D$). In the case of tuberculosis, complementary factors include familial exposure, immunosuppression, genetic susceptibility, poor nutrition, overcrowding, failure to diagnose and treat during the asymptomatic stages of disease, multiple drug resistance, and so on.

3. Causal exposure E is **not necessary but is sufficient** for disease D if D can occur in the absence of E, but D always occurs in its presence ($E \to D$; $F \to D$). For example, Down syndrome always leads to mental retardation, but other perinatal factors may also cause this outcome. As was the case in scenario 1, a case can be made that sufficiency of a single factor is rare in epidemiology (e.g., contributing causes of Down syndrome include older maternal age, lack of screening, consanguineous marriages, and teratogenic factors).

4. Explanatory factor E is **neither necessary nor sufficient** in causing disease D if E may or may not precede D. This implies that additional factors (F) must accompany E in causing disease ($E + F \to D$; $F + G \to D$; etc.). For example, cigarette smoking is neither necessary nor sufficient to cause lung cancer—even under the most extreme conditions, lung cancer will affect only a fraction of smokers, and lung cancer can occur in the absence of smoking. Most causal factors in modern epidemiology fall into this category.

Indispensable Part of a Causal Mechanism   So how then do we think about nonnecessary, nonsufficient causal factors? Briefly, each cause is viewed as a ***causal component*** of a multifactored ***causal mechanism***. *For example, high serum cholesterol, while neither necessary nor sufficient to cause a myocardial infarct, is an indispensable part (component) of myocardial infarction in many individuals. This multicomponent view of causality contrasts with the now antiquated "unifactor" (single agent) conception of disease etiology proposed by Jakob Henle in 1849 and later modified by his student Robert Koch in 1882. The following criteria form the basis of the **Henle–Koch postulates**:*

1. The agent must be present in each and every case of the disease (i.e., the agent is *necessary*).

2. The agent can occur in no other disease as a fortuitous or nonpathogenic parasite (i.e., the agent is *specific*).

3. The agent must be isolated from the body of an infected host in pure culture and must be capable of repeatedly causing the disease anew in a susceptible host (i.e., the agent is *sufficient*).

Limitations of Koch's postulates in explaining both infectious and noninfectious diseases are now widely recognized. Even Koch did not regard his postulates as strict criteria, realizing almost immediately that many bacterial agents could not be expected to fulfill criterion 2 and criterion 3—he was aware of the various carrier states in which the agent was nonpathogenic and recognized that exposure to the cholera bacillus was not sufficient by itself to produce cholera in all individuals. It is ironic that Jakob Henle's grandson, Werner Henle, helped establish the causal relationship between Epstein–Barr virus and mononucleosis without fulfilling a single one of his grandfather's postulates (Evans, 1978). Nevertheless, the epidemiologist must be aware that many people fall into the habit of using the one-to-one correspondence between cause and effect that is typical of the Koch–Henle postulates.

Causal Pie Model

Rothman's (1976) causal pie (sufficient/component cause) model helps clarify how necessary and nonnecessary component causes contribute to disease occurrence on a population level. Figure 2.8 is a causal pie model representing a particular disease. Two causal mechanism are possible, each involving multiple causal components. Factor A in this example is a necessary component cause, since it always precedes the disease. Factors B, C, and D are nonnecessary component causes.

A **causal complement** of a factor is the factor or set of factors that completes a given causal mechanism. In Figure 2.8, the causal complements of A are $(B + C)$ and D; the causal complement of factor B is $(A + C)$; the causal complement of factor C is $(A + C)$; the causal complement of factor D is A. Factors that work together in a given causal mechanism are said to be **interdependent** or **interact causally**. For example, when a person develops an infectious disease, the infectious agent interacts with lack of immunity to cause disease. Falling or some form of trauma interacts with osteoporosis to cause hip fractures in the elderly. Smoking interacts with genetic susceptibility and other environmental carcinogens to cause lung cancer. Dietary factors interact with lack of exercise, genetic susceptibility, and arterial thromboembolism to cause a heart attack. Thus, causal interdependencies (interactions) have direct health relevance.

Sufficient Causal Mechanism 1   Sufficient Causal Mechanism 2
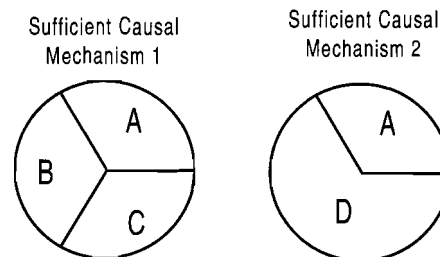
A, B, C, and D are component causes.
A is a necessary component cause.

Figure 2.8. Two mechanisms of disease represented with causal pies.

The causal pie model highlights the fact that individual **risk** is an all-or-none phenomenon. In a given individual, either a causal mechanism is completed or it is not. In fact, the notion of risk is useful only when an individual is regarded as a member of a recognizable subgroup of the population. For example, we may speak of the risk of lung cancer in 50- to 55-year-old men who have smoked one pack of cigarettes per day for 30 years. We can then say the risk in that particular population subunit is the expected *proportion* in the group that develops lung cancer.

Causal pies can also be used to illuminate how the effects of a causal exposure depend on the prevalence of its causal complements in the population. The effect of phenylketanines, for instance, depends on the prevalence of the inborn error of catabolism marked by the absence of phenylalanine hydroxylase known as phenylketonuria. The effect of the sickle-cell trait depends on the prevalence of the falciparum malaria protozoan and female *Anopheles* mosquitos in the environment. The effects of falls in the elderly depends on the prevalence of osteoporosis in the group.

Hogben's (1933) discussion of **yellow shank disease in chickens** provides a memorable example of how the population effects of an agent cannot be separated from the prevalence of its causal complements in the population (MacMahon & Pugh, 1970). The trait of yellow shank in poultry is a condition expressed only in certain genetic strains of fowl when fed yellow corn. A farmer with a susceptible flock who switches from white corn to yellow corn will perceive the disease as entirely environmental. A farmer who feeds only yellow corn to a flock with multiple strains of chickens, some of which are susceptible to the yellow shank condition, will perceive the condition as entirely genetic. In fact, the effects of yellow corn cannot be separated from the genetic makeup of the flock, and the effect of the genetic makeup of the flock cannot be separated from the presence of yellow corn in the environment. To ask whether yellow shank disease is environmental or genetic is like asking whether the sound of a faraway drum is caused by the drum or the drummer—one does not act without the other—there is an interdependence (causal interaction).

Causal Web

To ascertain the cause of cholera, we must consider it not only in individual cases, but also in its more general character as an epidemic.—John Snow (1849, p. 746)

The **causal web** is a metaphor that emphasizes the interconnectedness of casual components in a population. We speak of direct causes and indirect causes comprising causal webs. **Direct causes** are proximal to pathogenic events. **Indirect causes** are distal ("upstream") from pathological events. For example, occlusion of a coronary artery via a blood clot is a direct cause of a myocardial infarction, whereas social and environmental factors that lead to hyperlipidemia, obesity, a sedentary lifestyle, arteriosclerosis, and coronary stenosis are indirect causes. In considering a given disease, indirect and direct causes form a hierarchical causal web, often with reciprocal relations among factors. A casual web model is shown in Figure 2.9.

Levels of cause in a causal web may be classified as:

- **Macrolevel** (including social, economic, and cultural determinants)
- **Individual-level** (including personal, behavioral, and physiological determinants)
- **Microlevel** (including organ system, tissue, cellular, and molecular determinants)
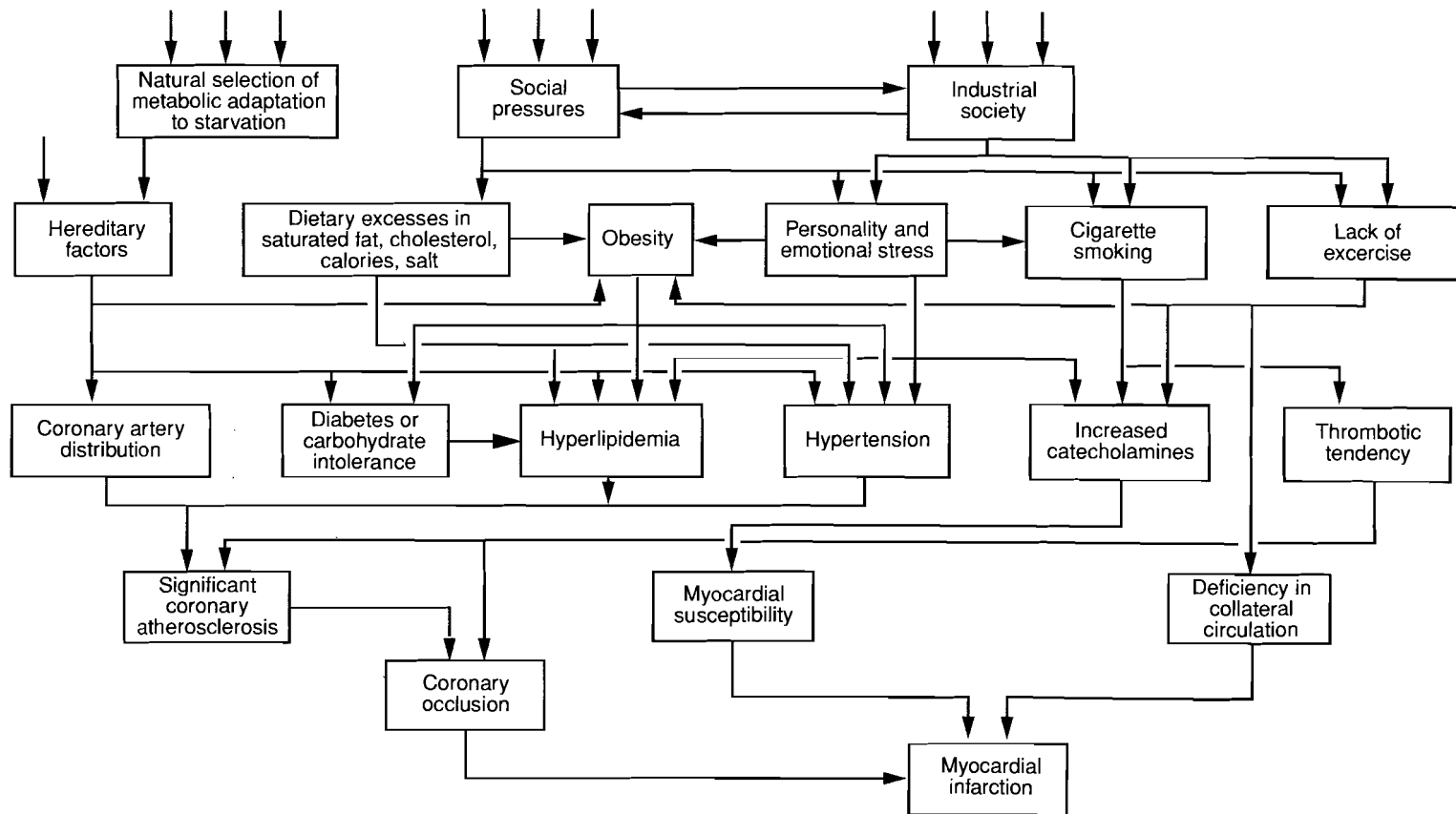
Figure 2.9. *Causal-web model for myocardial infarction. (Source: Friedman, G. D. (1974).* Primer of Epidemiology. *New York: McGraw-Hill; reproduced with the permission of the McGraw-Hill Companies.)*

Figure 2.10 considers these levels of cause for early childhood mortality in non-industrialized countries. In this example, the macrolevel encompasses broad social, economic, and cultural conditions that lead to a paucity of food, shelter, and sanitation. Individual-level causes include child-care practices that expose children to pathogens, malnutrition, and dehydration. Microlevel causes include the immediate pathophysiologic interaction between malnutrition and the pathogenic respiratory and gastrointestinal agents that ultimately lead to death (Millard, 1994).

The relative contribution of these various levels of study in epidemiology and public health have been the subject of considerable and sometimes contentious debate, with advocates for each level of study claiming particular and general benefits to their way of addressing problems. In practice, however, advocating one or another level may hinder achieving the most practical strategy for preventing a given disease. Maintaining fragmented methods of research into the various levels of disease occurrence can only obstruct understanding and ultimately delay effective prevention measures (Savitz, 1997).

Agent, Host, and Environment

Causal components can be classified as agent, host, or environmental factors (Fig. 2.11). **Agents** are biological, physical, and chemical factors whose presence, absence, or relative amount (too much or too little) are necessary for disease to occur (Table 2.2). **Host factors** include personal characteristics and behaviors, genetic predispositions, and immunologic and other susceptibility-related factors that influence the likelihood or severity of disease. **Environmental factors** are external conditions other than the agent that contribute to the disease process. Environmental factors can be physical, biologic, or social in nature.

Multiple agent, host, and environmental factors are viewed in interdependent ecological terms. Over time, it is possible for causal and preventive factors to form an **epidemiologic**
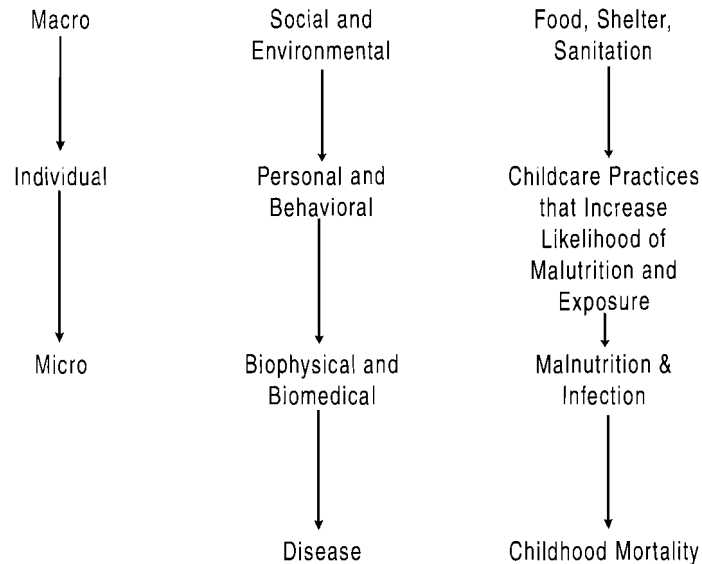
Figure 2.10. Macro-, individual-, and microlevels of cause for early childhood mortality in nonindustrialized countries.
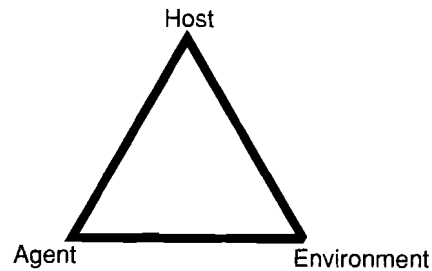
Figure 2.11. *Agent, host, and environment triad.*

homeostasis. Epidemics may arise when the relative influence of factors are thrown out of balance. For example, an epidemic may arise from any of the following:

- Introduction of a new agent into the population
- Increases in the ability of an agent to survive in the environment
- Increases in an agent's ability to infect the host (**infectivity**)
- Increases in the ability of the agent to cause disease once inside the host (**pathogenicity**)
- Increases in the severity of the disease caused by the agent once it has established itself in the host (**virulence**)
- Increases in the proportion of susceptibles in the population
- Environmental changes that favor growth
- Environmental changes that favor transmission of the agent
- Environmental changes that compromise host resistance

Causal forces can strengthen, weaken, or cancel-out each other, tipping the epidemiologic balance in favor of or against disease. This principal of epidemiologic balance applies to infectious and noninfectious agents, and to various environmental and host factors. As an example, consider the ecologic balance between agent, host, and environmental factors associated with **sulfur oxide air pollution** and overall morbidity (U.S. Department of Health, Education, and Welfare, 1967). In this example, high atmospheric levels of sulfur oxide pollution are traced to industrial pollution. Meteorologic conditions (e.g., climatic inversions) that favor retention of pollutants in the ecosphere have demonstrable effects on increasing morbidity and mortality, with the adverse effects

TABLE 2.2. Types of Disease-Causing Agents

Biologic	Chemical	Physical
Helminths	Nutritive (deficiencies and	Heat
Protozoan	excesses)	Light
Fungi	Poisons	Radiation
Bacteria	Drugs	Noise
Rickettsia	Allergens	Vibration
Viral		Objects
Prion		

Host
- Host genetics
- Stage of infection
- Antiretroviral therapy
- Reproductive tract infections
- Cervical ectopy
- Male circumcision
- Contraception
- Menstruation and pregnancy

Agent
- HIV subtype (A, B, C, D, E)
- Phenotypic differences
- Genotypic differences
- Antiretroviral drug resistance

Environment
- Social norms
- Avg rate of sex-partner change
- Local prevalence/probability of exposure
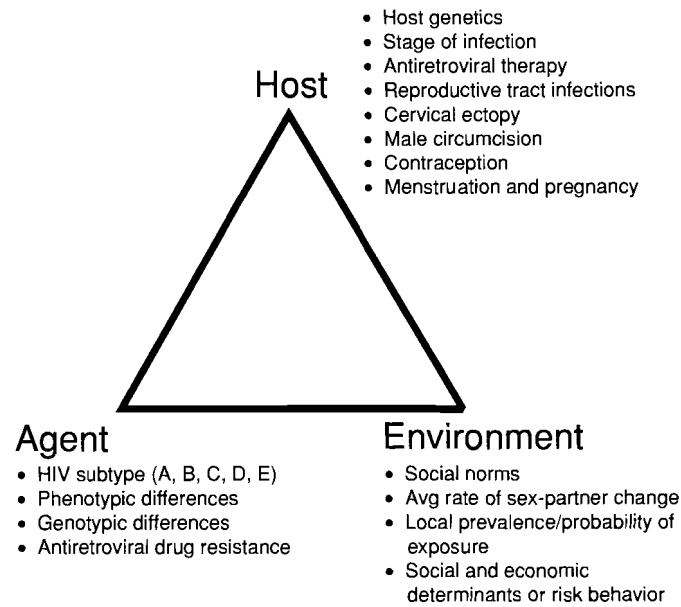- Social and economic determinants or risk behavior

Figure 2.12. Agent, host, and environmental factors associated with the sexual transmission of HIV. (Source: Royce et al., 1997).

of pollution concentrated in individuals with pre-existing cardiac and respiratory disease (Munn, 1970, p. 95). Thus, morbidity and mortality are linked to interdependencies between agent (e.g., sulfur dioxide pollution), host (compromised cardiopulmonary function), and environmental (meteorologic) conditions.

The **sexual transmission of HIV** in a population can also be viewed in ecological terms (Fig. 2.12). Agent factors that influence HIV transmission include the prevalence of the agent in the environment and the phenotype of the agent. Examples of host factors include the coexistence of reproductive tract infections (especially genital ulcers), availability of antiretroviral therapies that decrease the HIV load in the population, sexual behaviors, and contraceptive methods. Environmental factors include the rate of sexual partner exchange, presence of unregulated commercial sex facilities, presence of "crack houses," sexual norms, and so on (Royce et al., 1997).

2.4 EPIDEMIOLOGIC VARIABLES

I keep six honest serving men
(They taught me all I know);
Their names are what and why and when
And how and where and who.

—Rudyard Kipling

Epidemiology *describes* the distribution of disease and disease determinants according to the variables of person, place, and time. Stallones (1980) suggests the following axiom and corollaries which serve as a foundation for this practice: