



© Enculescu Marian Vladut/Shutterstock

## Chapter 10

# Epidemiologic Issues in the Interface With Public Health Policy

## 10.1 Introduction

As reflected in the definition of epidemiology introduced in the first chapter of this textbook, the ultimate goal of epidemiology is to apply knowledge about the distribution and determinants of health-related states to the control of health problems. The translation of study findings into the practice of public health, however, is not an easy task. Like clinicians, policy makers typically grade the quality of the evidence to decide whether it is strong enough to support implementing a program or service (**EXHIBIT 10-1**). Randomized clinical trials are considered as providing the best level of evidence,<sup>1,2</sup> but their application to study questions relevant to health policy is frequently limited by ethical or feasibility concerns, as pointed out by Frieden.<sup>3</sup> Moreover, even decisions based on results from such trials are often difficult as underscored by the inconsistent results from large clinical trials on the effectiveness of mammography as a screening tool for breast cancer.<sup>4,5</sup>

### EXHIBIT 10-1 Levels of evidence.

Grade	Level of evidence	Description of level
A	1a	Systematic review (with homogeneity) of randomized clinical trials
	1b	Individual RCT (with narrow confidence interval)
	1c	“Natural experiments,” consisting of, interventions with dramatic effects (e.g., streptomycin for tuberculosis meningitis; insulin for diabetes)
	2a	Systematic review (with homogeneity) of cohort studies
	2b	Individual cohort study or randomized clinical trial of lesser quality (e.g., with < 80% follow-up)
B	2c	Outcomes research (based on existing records)
	3a	Systematic review (with homogeneity) of case-control studies
	3b	Individual case-control study

C	4a	Temporal (before–after) series with controls and cohort-and case-control studies of lesser quality
	4b	Temporal (before–after) series without controls
D	5	Expert opinion without explicit critical appraisal or not based on logical deduction

Data from Canadian Task Force on the Periodic Health Examination. *Canadian Guide to Clinical Preventive Health Care*. Ottawa, Ontario, Canada: Health Canada; 1994<sup>6</sup>; U.S. Preventive Services Task Force. *Guide to Clinical Preventive Services*. Concord, NH: U.S. Prevention Services Task Force, U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion; 2003<sup>7</sup>; Bigby M, Szklo M. Evidence-based dermatology. In: Freedberg IM, Eisen AZ, Wolff K, et al., eds. *Fitzpatrick's Dermatology in General Medicine*. 6th ed. New York, NY: McGraw-Hill Medical Publishing Division; 2003:2302.<sup>8</sup>

The problems associated with experimental evidence are compounded when nonexperimental study designs are used, as confounding and bias are more likely to occur in these studies. However, it should be emphasized that, notwithstanding problems in inferring causality, observational studies have made important contributions to public health throughout history. Indeed, epidemiology has played a major role in shaping public health policy and prevention, with examples spanning from Snow's classic 19th-century cholera studies leading to the closing of the Broad Street water pump<sup>6</sup> to the United States Preventive Services Task Force recommendations for the primary prevention of cardiovascular diseases by promotion of physical activity and a healthful diet.<sup>7</sup> Additional, more recent examples of policies directly influenced by observational data include fat content labeling of processed foods, air pollution standards, and prohibition of indoor smoking. Relatively recent observational strategies that attempt to strengthen the generation of causal inference in observational studies are the use of instrumental variables (which include Mendelian randomization) and propensity score matching/adjustment (see [Chapter 7, Section 7.5](#)).

Other challenges to the inferential process leading to policy recommendations are common to both experimental and nonexperimental studies, such as lack of consistency across studies, particularly in the presence of weak associations or modest effectiveness.

In this chapter, some epidemiologic issues related to the use of exposure–outcome association data in the development of policy recommendations are discussed. Descriptions of Rothman's causality model<sup>9</sup> and Hill's guidelines to infer causality<sup>10</sup> in the context of their application to the decision-making process are also part of this chapter. Other relevant topics, such as weak associations and homogeneity among studies, are discussed along with the causality guidelines. It should be emphasized that this chapter does not elaborate on many issues that are of interest to health policy students and experts, such as the influence of politics on policy or the role of public health or other agencies. Instead, it tries to emphasize the relevance to prevention of topics that are largely of specific concern to those involved in epidemiologic teaching and research. Although many of the examples discussed in this chapter refer to primary prevention, several concepts discussed here are also applicable to secondary prevention and clinical practice.

## 10.2 Causality: Application to Public Health and Health Policy

---

Inferring whether an association is causal is key to the use of epidemiologic findings in primary prevention and other interventions that aim at modifying the probability of the outcomes of interest. An in-depth discussion of the different models of causal inference is beyond the scope of this book and can be found elsewhere.<sup>11</sup> For practical purposes, the inductive process of prediction—which consists of generalizing results obtained in one or more studies to different target or reference populations—remains the premier approach that public health professionals and policy makers use.

Criteria to define causality were pioneered by Koch as an effort to identify biologic agents responsible for infectious diseases.<sup>12,13</sup> Koch's postulates required that the organism had to be recovered from each case of disease, that the organism had to be cultured *in vitro* from each case, that reinoculation of the purified organism had to cause disease in another host, and that the organism had to be reisolated from the latter. The validity of Koch's paradigm, as expressed by his postulates, has been shown for several infectious diseases. In contradistinction to Koch's paradigm focusing on single causal agents, however, almost a century later, MacMahon, Pugh, and Ipsen<sup>14</sup> proposed the concept of "web of causation" as a way to emphasize the importance of multiple causes of disease. As stated persuasively by Gordis,<sup>15</sup> risk factors in isolation are rarely either sufficient or necessary to cause disease. Even when necessary causes are present, they are usually not sufficient to produce disease—a fact that is particularly true for conditions such as tuberculosis and stomach cancer that in certain populations are rare manifestations of common exposures (*Mycobacterium tuberculosis* and *Helicobacter pylori*, respectively).

The web of causality for stomach cancer sharply underscores the

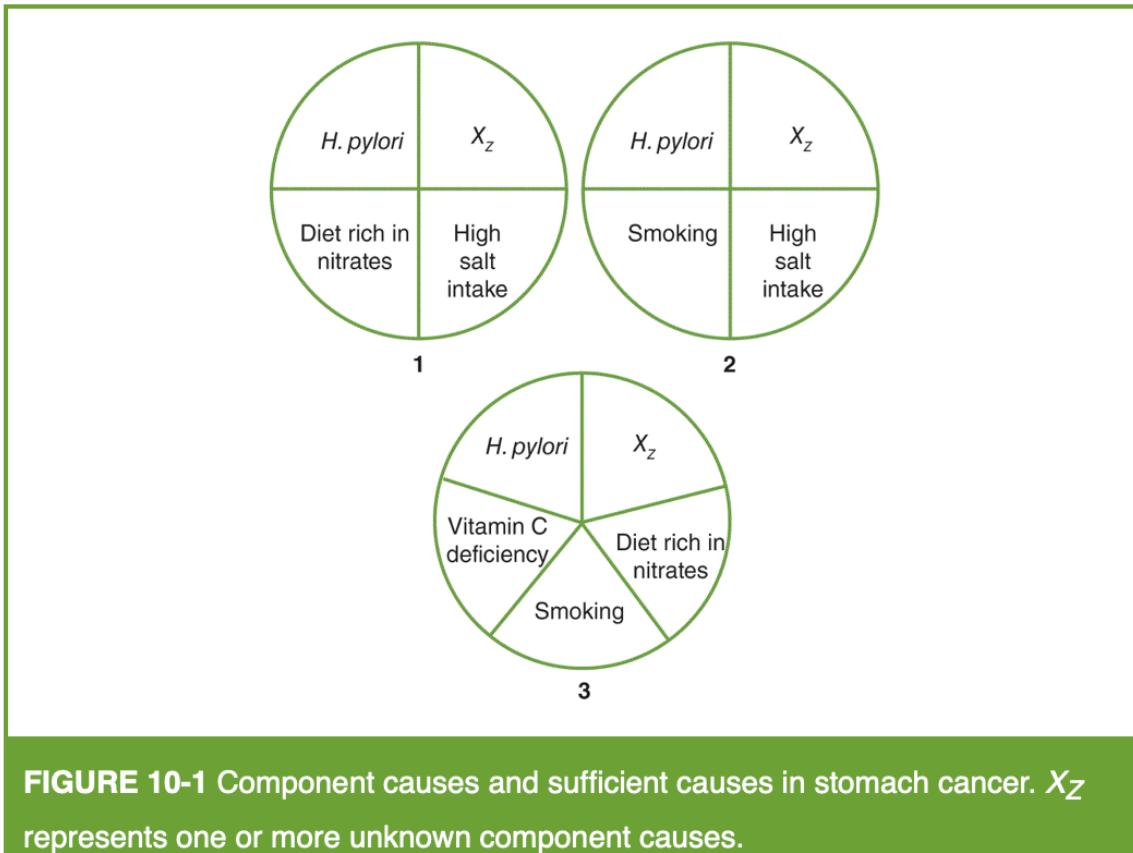
interconnections between multiple risk factors. Although *H. pylori* infection appears to be a necessary causal agent for noncardia gastric cancer,<sup>16</sup> its high prevalence and the relative rarity of this neoplasm strongly suggest involvement of other factors, which may include, for example, smoking and exposure of gastric mucosa to *N*-nitroso compounds. A possible chain of causality may start with low socioeconomic status and household crowding resulting in exposure to *H. pylori*. A subsequent event may be the ingestion of nitrate-rich foods, such as cured meats; these, in turn, are reduced to nitrites by bacteria found in human saliva and the stomach, the growth of which is facilitated by a change in acidity brought about by smoking and excessive salt intake. Nitrites then react with secondary amines found in several ingested foods (such as pork-based products) to form *N*-nitroso carcinogenic compounds. A potent inhibitor of this reaction is vitamin C; thus, its deficiency may be yet another factor that contributes to the formation of these carcinogens and, thus, to the web of causation in noncardia gastric cancer.<sup>17</sup>

### 10.2.1 Rothman's Causality Model

The pathogenesis of stomach cancer underscores the importance of assessing a constellation of risk factors, or *component causes*, acting jointly to form what Rothman has named a *sufficient cause*, defined as “a set of minimal conditions and events that inevitably produce disease.”<sup>9(p11)</sup> These conditions or events can act either simultaneously or sequentially, such as in the case of initiators and promoters of cancer.

On the basis of Rothman’s model, several *sufficient causes* can be postulated for stomach cancer, which are represented in pie graphs of *component causes* (FIGURE 10-1). Each of the complete circles in Figure 10-1 represents a sufficient cause composed of a constellation of component causes. Because the sum of the known component causes of stomach cancer used in these hypothetical examples may not be sufficient to complete a sufficient cause constellation, a variable  $X_Z$ —which represents one or more unknown factors acting as component causes—

has been added to each complete circle in the figure. In the figure,



**FIGURE 10-1** Component causes and sufficient causes in stomach cancer.  $X_Z$  represents one or more unknown component causes.

1. *Sufficient cause a*, formed by the component causes *H. pylori* (a necessary component cause), diet rich in nitrates, high salt intake, and  $X_Z$  (Figure 10-1, part 1).
2. *Sufficient cause b*, formed by *H. pylori*, smoking, high salt intake, and  $X_Z$  (Figure 10-1, part 2).
3. *Sufficient cause c*, formed by *H. pylori*, vitamin C deficiency, smoking, diet rich in nitrates, and  $X_Z$  (Figure 10-1, part 3).

*H. pylori*, assumed to be a necessary cause, appears as a component cause in all sufficient cause constellations.

When considering Rothman's sufficient causes in the context of preventive activities, the following two issues should be pointed out:

1. Elimination of even a single component cause in a given sufficient cause constellation is useful for preventive purposes, as it will by definition remove the "set of minimal events and conditions" that form that sufficient cause. Consider, for

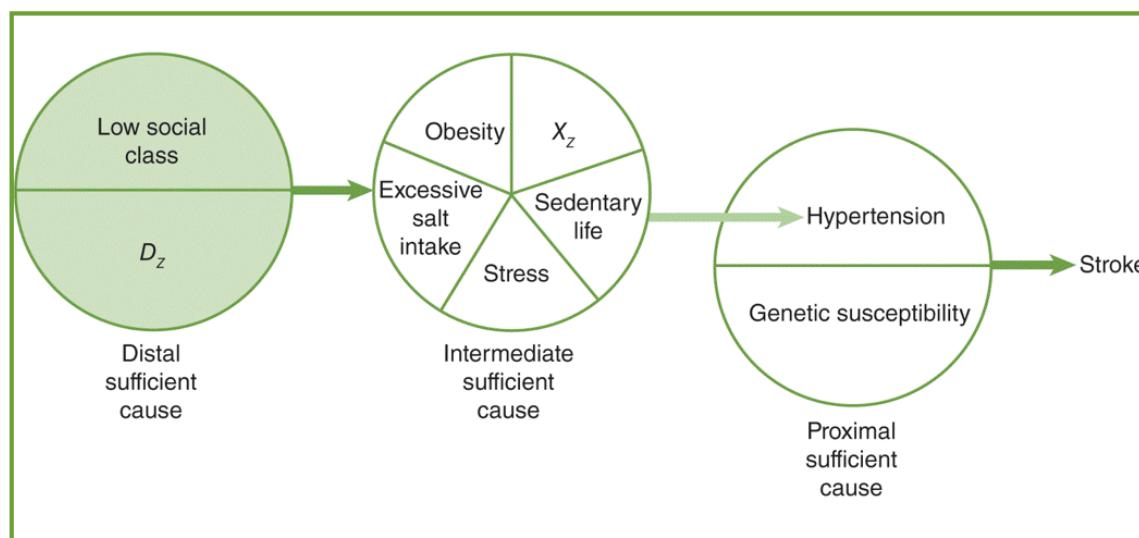
instance, Figure 10-1, sufficient causes 1 and 2: If salt intake were not high, stomach cancer would not occur even if the necessary cause (*H. pylori*) and the other component causes (*H. pylori*, diet rich in nitrates, and  $X_Z$  in sufficient cause a and *H. pylori*, smoking, and  $X_Z$  in sufficient cause b) were present. In other words, all sufficient causes containing the component cause “high salt intake” would cease to exist. This notion is supported by the fact that, as pointed out previously, although the prevalence of *H. pylori* is very common in certain populations (expressed as a percentage), the incidence of stomach cancer in these same populations is fairly rare (expressed as per 100,000).<sup>18</sup>

2. As aptly stated by MacMahon, Pugh, and Ipsen<sup>14(p18)</sup> several decades ago, “to effect preventive measures, it is not necessary to understand causal mechanisms in their entirety.” This important concept is exemplified by Snow’s recommendations pertaining to London’s water supply many years before Pasteur’s discoveries and by Casal and Goldberger’s discovery of the nutritional deficiency nature of pellagra well before the actual vitamin involved was discovered.<sup>19,20</sup> Other examples of instances in which identification of an epidemiologic chain amenable to prevention interventions preceded the discovery of the actual causal factor have been discussed by Wynder.<sup>19</sup> Thus, prevention of only a single component cause, without any knowledge about the other component causes forming a sufficient cause, would be enough to decrease the risk of the disease, as all sufficient causes of which the component cause is part of would disappear.

## 10.2.2 Proximate, Intermediate, and Distal (Upstream) Causes and Prevention

Several researchers<sup>21,22</sup> have criticized epidemiology’s modern

tendency toward reductionism, with a primary focus on proximate component causes, particularly those related to biologic markers of risk. These reductionistic approaches tend to be in tune with clinically oriented, “high-risk” strategies for prevention. In contrast, the study of more upstream causes may provide clues for the development of prevention strategies at the level of the total target population. As argued by Rose,<sup>23,24</sup> the population-wide approach based on distal causes—for example, those related to social determinants of health—might be the most effective prevention strategy for the total population. An example is stroke prevention by either hypertension prevention or treatment. A hypothetical model representing the chain of causality for stroke is proposed in **FIGURE 10-2**. In addition to the distal and proximal sufficient causes, this model recognizes that there may be an intermediate sufficient cause. Likewise, it considers the time sequence of distal, intermediate, and proximal causes. In this example, the joint presence of the component causes low social class and  $D_Z$  is conceptualized as a distal sufficient cause, which in turn results in the intermediate sufficient cause formed by obesity, excessive salt intake, stress, sedentary life, and  $X_Z$ . This intermediate sufficient cause is responsible for a proximal component cause, hypertension, which along with genetic susceptibility, is part of the proximal sufficient cause of stroke. As previously, the subscript  $Z$  represents component causes needed to complete each sufficient cause constellation above and beyond known component causes.



**FIGURE 10-2** Component causes and sufficient causes in stroke.  $D_Z$  represents one or more unknown distal component causes;  $X_Z$  represents one or more unknown intermediate component causes.

Using this example, a typical high-risk preventive strategy may, for example, focus on a proximate component cause—severe hypertension—which would be identified and treated. Although the relative risk of stroke associated with severe hypertension is high, the prevalence of this type of hypertension is much lower than that of prehypertension plus moderate hypertension, which, notwithstanding their weaker association with stroke, are related to a much higher attributable risk in the population (**FIGURE 10-3**). **EXHIBIT 10-2** shows that, although the relative risk of stroke associated with stage 2 hypertension is very high (4.0), its attributable risk is less than that associated with prehypertension, notwithstanding the latter's much lower relative risk (1.5); this is because the prevalence of prehypertension (50%) is much higher than that of stage 2 hypertension (approximately 5%).<sup>24</sup>

**EXHIBIT 10-2** Relative risks and attributable risks for severe (stage 2) hypertension and prehypertension.

$$AR_{pop} = \frac{\text{Prevalence}_{RF} (RR - 1.0)}{\text{Prevalence}_{RF} (RR - 1.0) + 1.0}$$

Stage 2 hypertension (SBP = 160+ or DBP = 100+ mm Hg) and stroke

- Relative risk ~4.0
- Prevalence ~5%

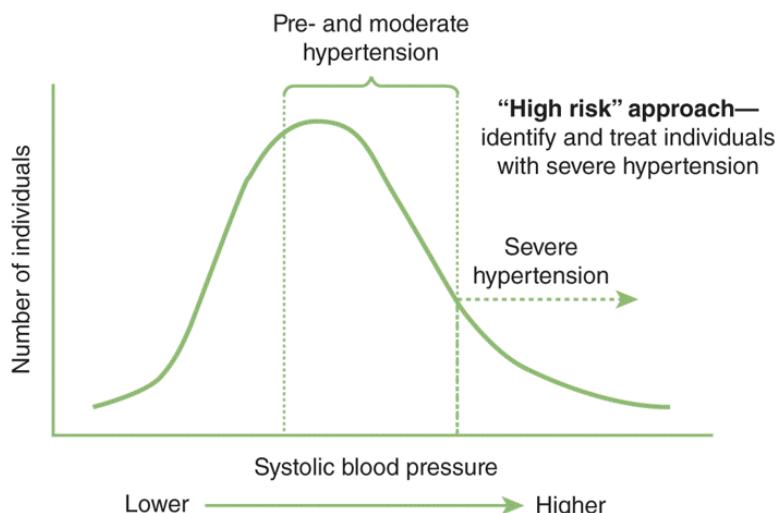
$$AR_{POP} = \frac{0.05 (4.0 - 1.0)}{0.05 (4.0 - 1.0) + 1.0} \times 100 = 13 \%$$

Prehypertension (SBP 120–139 or DBP 80–98 mm Hg) and stroke

- Relative risk ~1.5
- Prevalence ~50%

$$AR_{POP} = \frac{0.50 (1.5 - 1.0)}{0.50 (1.5 - 1.0) + 1.0} \times 100 = 20 \%$$

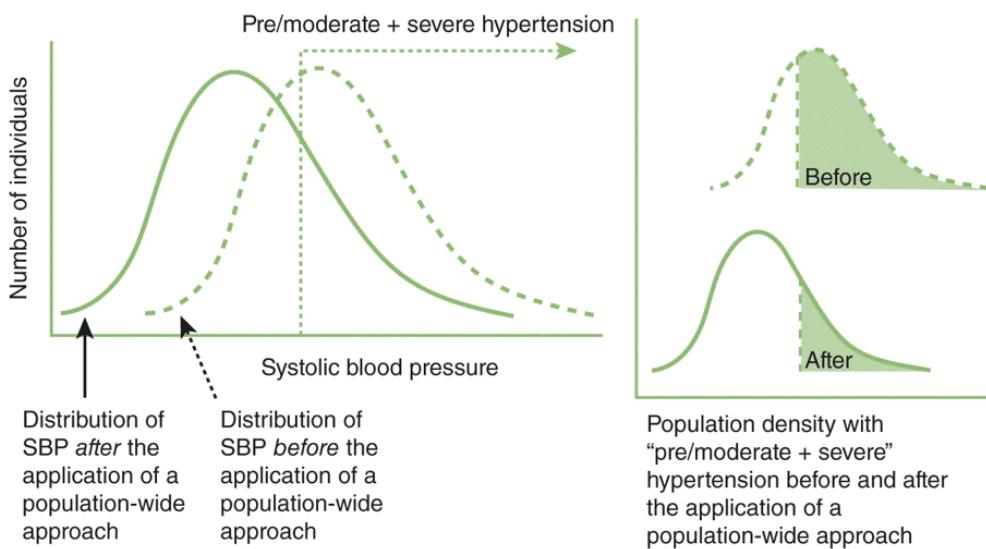
Based on Chobanian AV, Bakris GL, Black HR, et al. The seventh report of the Joint National Committee on prevention, detection, evaluation and treatment of high blood pressure: the JNC 7 report. *J Am Med Assoc.* 2003;289:2560-2572.<sup>25</sup>



**FIGURE 10-3** The “high-risk” approach focuses on individuals with severe hypertension. Although the relative risk for stroke is high in individuals with severe hypertension (compared to those with normal blood pressure), the prevalence of severe hypertension in the total population is low, and thus the associated population attributable risk is low. Most cases of stroke originate among those with pre- and moderate hypertension, the prevalence of which is

high.

The use of a population-wide strategy, consisting of primary prevention by intervention on the distal or intermediate component causes represented in [Figure 10-2](#), could shift the entire blood pressure distribution curve to the left. Examples of this strategy include an improvement in the socioeconomic status of the target population, regulating salt content in processed foods, or promoting the development of urban environments and public transportation options that encourage residents' physical activity. In turn, the prevalence of both prehypertension and moderate/severe hypertension in the total population would decrease ([FIGURE 10-4](#)), resulting in a decrease in stroke incidence of a greater magnitude than that achieved by the high-risk approach. It has been estimated, for example, that a 33% decrease in average salt intake in the population at large would result in a 22% reduction in the incidence of stroke; in comparison, even if all hypertensive patients were identified and successfully treated, this would reduce stroke incidence by only 15%.<sup>26</sup> Similar decreases in all other modifiable hypertension component causes would obviously be expected to have an even greater impact on stroke and coronary heart disease (CHD) incidence than salt reduction alone. Thus, when distal or even intermediate component causes are known, primary prevention based on these causes is generally more effective than intervention directed toward proximal causes. Another example of the advantages of the population-based approach is provided by smoking prevention; whereas smoking cessation is important, its effectiveness is known to be much lower than that of programs that aim at preventing smoking initiation in the total population, such as increase in taxation of cigarettes and prohibition of indoor smoking.<sup>27</sup> For example, it has been estimated that, of the reduction in smoking rates since 2013 in Michigan, United States, 72% was due to increases in taxes and smoke-free air laws and only 26% to cessation treatment policies (2% was due to a decrease in youth access).<sup>28</sup>



**FIGURE 10-4** Distribution of systolic blood pressure (SBP) before and after the application of a population-wide approach. The prevalence of both moderate and severe hypertension decreases, resulting in a decrease of the population attributable risk.

Because screening aims at identifying individuals who already have the disease, it can be regarded as the embodiment of the high-risk approach. Taking this approach even further is selective screening, based on two or more steps.<sup>29</sup> In a two-step screening, the first step is the identification of high-risk individuals—for example, those with a given trait T—and the second, the application of the screening test. In the example illustrated in TABLE 10-1, 100,000 women aged 50–64 years undergo selective screening for incident breast cancer based first on identification of those with a family history (53% of this population) who, in a second phase, are referred to a mammographic exam. Based on published data, the sensitivity and specificity for the discrimination of incident breast cancer are assumed to be, respectively, 0.54 and 0.53 for family history<sup>30</sup> and 0.93 and 0.99 for mammography.<sup>31</sup> At the end of this two-step screening approach, the overall (“program”) sensitivity is estimated to be 0.50. If mammography had been applied to the total population (and not just to those with a positive family history), approximately 11 true cases would have been missed (obtained by multiplying the total number of true cases by the complement of the

sensitivity). With the two-step approach exemplified in [Table 10-1](#), 64 additional cases are missed (i.e.,  $69 + 6 - 11$ ), underscoring the loss of sensitivity resulting from this “high–high-risk” strategy. This example highlights the notion that the main reason for using a high-risk strategy is related to cost-effectiveness rather than to effectiveness alone.

**TABLE 10-1** Program validity of a “high-risk” approach in incident breast cancer screening in a population of 100,000 women aged 50–64 years. “High-risk” is defined by the presence of a family history of breast cancer, with sensitivity = 0.54 and specificity = 0.53.\* Those with a family history undergo mammography, with sensitivity = 0.93 and specificity = 0.99.<sup>†</sup> Yearly incidence of breast cancer is assumed to be about 150/100,000.

Step 1. Identification of individuals with a positive family history			
Family history	Disease present	Disease absent	Total
Present	81	52,920	53,001
Absent	69	46,930	46,999
Total	150	99,850	100,000
	Sensitivity = 0.54	Specificity = 0.53	
Step 2. Mammography in those with a positive family history			
Test	Disease present	Disease absent	Total
Positive	75	529	604
Negative	6	52,391	52,397
Total	81	52,920	53,001
	Sensitivity = 0.93	Specificity = 0.99	

Program’s sensitivity =  $75 \div 150 = 0.50$  (also calculated as the product of the two sensitivity values, or  $0.54 \times 0.93 = 0.50$ )

Program’s specificity =  $[46,930 + 52,391] \div 99,850 = 0.995$  (also calculated as the complement of the product of the complements of the specificities:  $1 - [1 - 0.53] \times [1 - 0.99] = 0.995$ )

\*Data from Hartmann LC, Sellers TA, Frost MH, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med*. 2005;353:229-237.<sup>30</sup>

<sup>†</sup>Data from Mushlin AI, Kouides RW, Shapiro DE. Estimating the accuracy of screening mammography: a meta-analysis. *Am J Prev Med.* 1998;14:143-153.<sup>31</sup>

### 10.2.3 Social Determinants

Epidemiology as a formal discipline started in 19th-century Great Britain, where a heavy emphasis was placed on the importance of the unequal distribution of morbidity and mortality by social class. The recent renewed interest in the study of social determinants of health and disease—which had previously peaked in the 1960s and 1970s<sup>32,33</sup>—has focused on a multilevel causality framework, whereby the exclusive emphasis on proximal causes (e.g., smoking in relation to coronary thrombosis, hypertension as a cause of stroke) has been changed to reflect the interdependence between proximal and more distal or upstream (ecological) variables (e.g., high serum cholesterol levels resulting from difficult access to healthy foods, which in turn is at least partly determined by social class) (see [Chapter 1, Section 1.3](#)). The model subscribed to by social epidemiologists favors neither a reductionistic (proximal cause-oriented) nor a purely ecological, socially determined approach (based on upstream causes) but rather a consideration of both types of component causes in the search for sufficient causes.<sup>22</sup>

A growing methodological interest in the interface between individual-level (usually proximal) and group-level (usually distal or intermediate) variables has led to the development of analytical strategies that take into consideration both types of variables. Excellent summaries of these strategies can be found in the literature.<sup>34,35</sup>

### 10.2.4 Hill's Guidelines

The so-called Hill's criteria, which have been referred to more properly by Gordis as “guidelines,”<sup>15</sup> were originally published as part of the first Surgeon General's Report on Smoking and Health<sup>36</sup> and comprise a series of aspects of a statistical association that, when present, may strengthen the inference that the statistical association is also causal;<sup>10</sup>

however, with the exception of “temporality” (discussed later), failure to satisfy any or even most of these guidelines does not necessarily constitute evidence that the association is not causal.<sup>15,37</sup>

Notwithstanding the renewed interest in other models of causality over the past few years (see, for example, Greenland and Brumback<sup>38</sup>), Hill’s guidelines remain the cornerstone of causal inference for the practical epidemiologist and health policy expert. The overarching implicit questions that these guidelines seek to address are whether confounding and bias are reasonable alternative explanations for an observed statistical association and, if not, whether a cause–effect relationship can be inferred. What follows is an attempt to expand the discussion of Hill’s guidelines and its related issues of meta-analysis and publication bias, with a particular emphasis on consistency of associations across studies. Another issue related to the process of causal inference, sensitivity analysis, is also briefly discussed.

1. *Experimental evidence:* Because randomized trials in humans usually offer the best protection against confounding and bias, they are widely regarded as the gold standard for determining causal associations; thus, they are thought to provide the highest level of evidence for developing recommendations pertaining to preventive and clinical interventions ([Exhibit 10-1](#)).<sup>1,2</sup> In epidemiologic or public health research, however, random allocation is often neither feasible (e.g., when studying social class as a determinant) nor ethically acceptable (e.g., when studying the consequences of a potentially harmful environmental exposure); as a result, these trials are typically limited to assessing interventions that are expected to have a beneficial effect. Internal validity of a randomized trial is optimal when single interventions (e.g., a drug or a vaccine) are studied; from this viewpoint, it may be considered the epitome of reductionism in epidemiology. In addition to the problems related to open trials of effectiveness, such as poor compliance and “crossovers,” results of randomized trials can be erroneously

generalized. A possible example is the Finnish trial of smokers, which could not confirm experimentally the results of observational studies suggesting that a diet rich in beta-carotene and alpha-tocopherol reduced lung cancer incidence.<sup>39</sup> Although this trial's results may have accurately expressed these nutrients' lack of efficacy, they may have alternatively reflected the fact that, without considering their complex relationships (including possible interactions) with other dietary components, intake of beta-carotene or alpha-tocopherol may not be enough to influence a sufficient cause of lung cancer. In other words, the trial likely provided the response to the question asked by its authors (i.e., that using these nutrients as simple pills is not effective in preventing cancer in smokers); however, this response should not necessarily lead to the inference that alpha-tocopherol and/or beta-carotene are not protective if consumed in their natural states as part of a healthy diet.

2. *Temporality:* The presence of the right temporal sequence, “possible cause → possible effect,” *per se* does not constitute proof that the first event caused the second. Thus, for example, the fact that a given viral infection occurring in early life (e.g., measles) precedes a chronic disease (e.g., degenerative arthritis) cannot be said to constitute strong evidence of the former causing the latter. On the other hand, of all the guidelines by which to judge whether the relationship is causal, the demonstration that the exposure preceded the outcome under investigation is the only one that, if not met, eliminates the possibility of a causal connection. Yet it is often difficult to establish temporality in epidemiologic studies, particularly when assessing diseases with long subclinical phases and insidious onsets, such as chronic renal disease or chronic lymphocytic leukemia. A special type of temporal bias, discussed in [Chapter 4, Section 4.4.2](#), is “reverse causality,” which occurs when the presumed outcome (disease) is responsible for the occurrence of the exposure of interest. Thus, for example, a chronic disease

may go undiagnosed for years with a single symptom (e.g., a moderate loss of appetite) that may result in the hypothesized exposure (e.g., an exposure related to a change in diet). Although case-control studies are more amenable to this bias, it may also occur in cohort studies when ascertainment of the disease onset is difficult and the diagnosis is based on symptomatic disease, which may occur long after the subclinical onset of the disease.

3. *Strength of the association:* The rationale for using strength of the association as a guideline to infer causality is that it is more difficult to explain away a stronger than a weaker association on the basis of confounding or bias. Thus, the relationship of lung cancer to active smoking, expressed by a high relative risk, is more likely to be causal than that of CHD to environmental tobacco smoking (passive smoking), for which the relative risk is estimated at between 1.1 and 1.5.<sup>40</sup> As for Hill's guidelines in general (with the exception of temporality), however, observation of a weak association (e.g., one characterized by a relative risk less than 2.0) does not negate the possibility of causality. For public health purposes, careful consideration of whether a weak association is causal is justified by the possibility that it may result in a high population attributable risk if the exposure prevalence is high (see [Chapter 3, Section 3.2.2](#)). As an example, a relative risk of CHD related to environmental tobacco smoke of 1.2–1.3 and an exposure prevalence of 26%<sup>41</sup> would result in a population attributable risk of about 13%, which given the very large absolute number of new coronary events is far from negligible.<sup>42</sup>

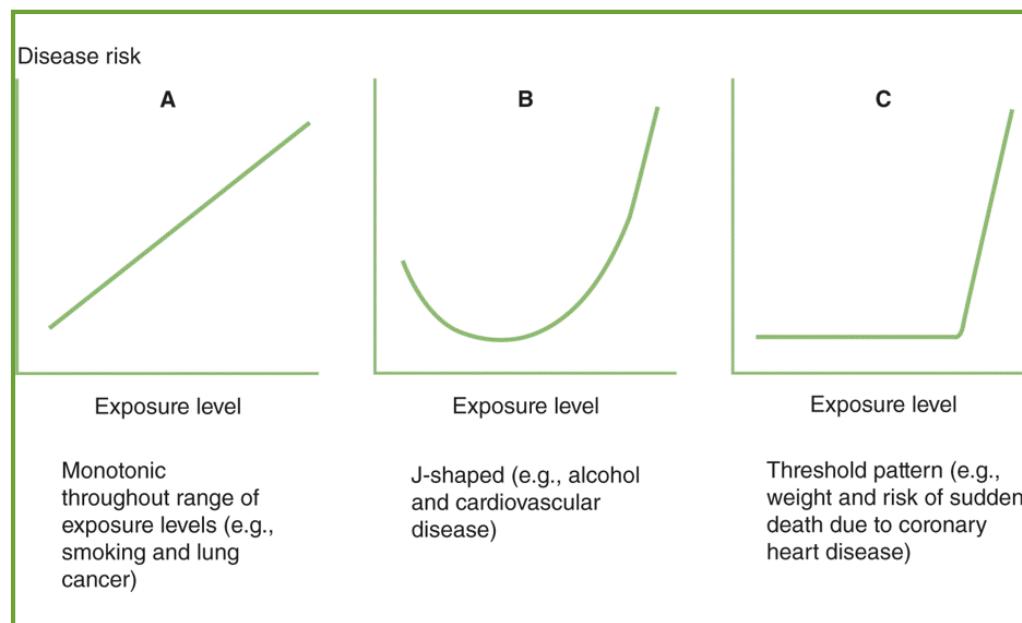
The importance of considering the scale (relative versus absolute excess) when assessing the impact of potentially causal—albeit weak—associations is particularly manifest when assessing interactions. With the widespread use of ratio-based models, assessment of interaction has become virtually synonymous with assessment of multiplicative interaction. Yet as shown in [Chapter 6, Section 6.6](#), evaluation of additive

interactions is crucial for public health purposes and can readily be done in the context of ratio-based models.<sup>43,44</sup>

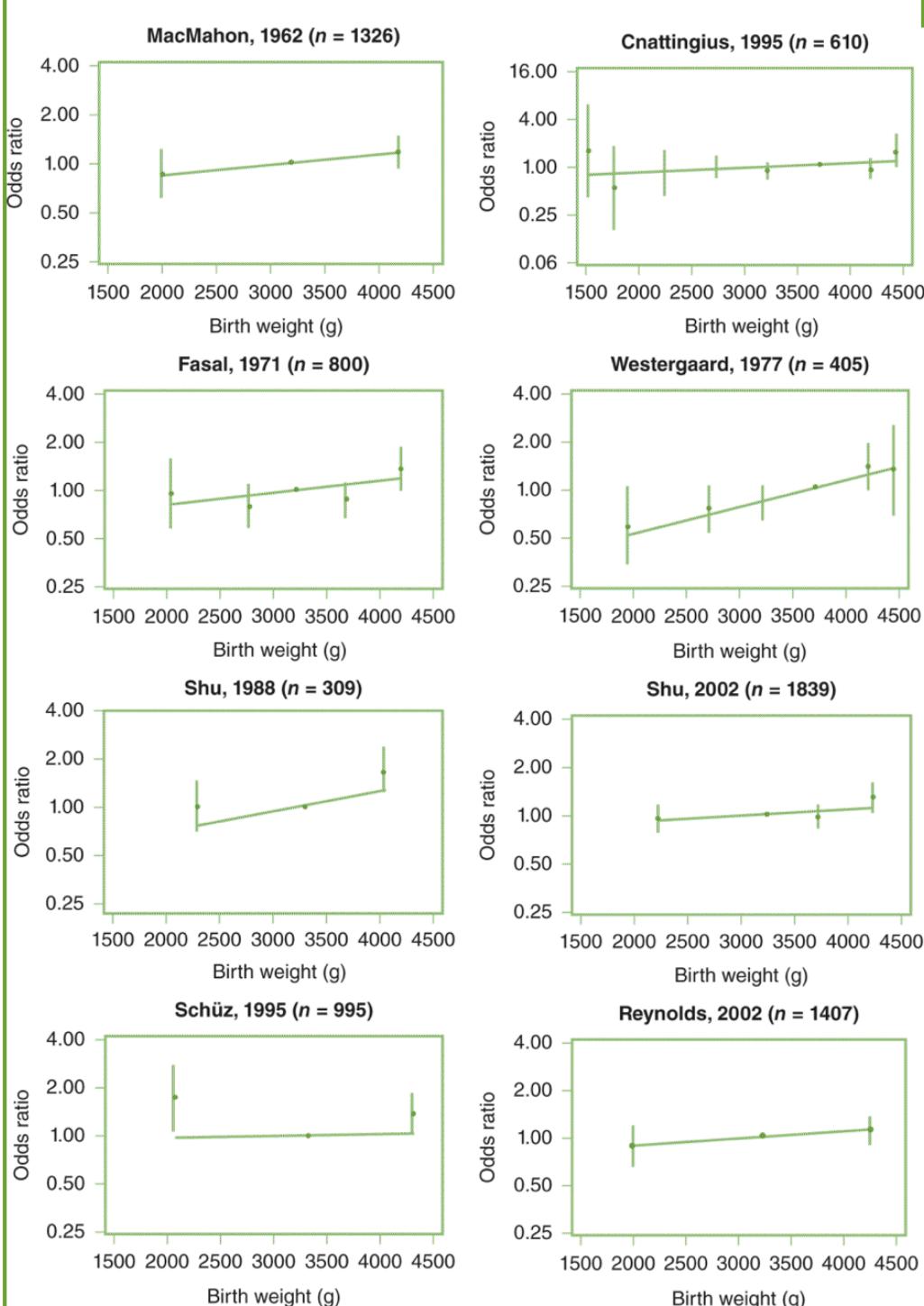
4. *Dose-response (graded pattern)*: The observation of a straightforward monotonic relationship between exposure dose and risk of an outcome is regarded as strong evidence that a cause–effect relationship exists (**FIGURE 10-5A**). There are numerous examples of this type of pattern in the epidemiologic literature (e.g., smoking and lung cancer, blood pressure and stroke). In a meta-analysis of the relationship between birth weight and leukemia (discussed later), the authors used the observation of consistent linear dose-response associations across studies as further evidence that the association was likely causal (**FIGURE 10-6**).<sup>45</sup> Meta-analysis (see [Section 10.4](#)) has been increasingly used to evaluate dose-response patterns. For example, in a meta-analysis done by Zhang et al., the authors pooled 45 cohort studies and detected a linear relationship between resting heart rate and both cardiovascular and noncardiovascular diseases.<sup>46</sup> Causal relationships, however, may also be characterized by other types of patterns reflecting the biological mechanisms underpinning these relationships. Thus, for example, the association between alcohol intake and cardiovascular mortality seems to follow a J-shaped relationship (**FIGURE 10-5B**) probably because, at low intake levels, alcohol may be protective against atherosclerotic disease through an increase in serum high-density lipoprotein concentration, platelet activation inhibition (and thus coagulation inhibition), and antioxidant activity; however, at higher levels, its harmful effects on blood pressure may predominate.<sup>47</sup> Another type of pattern is that in which the excess risk appears only above a certain exposure level (i.e., a certain threshold) (**FIGURE 10-5C**). As an example, in early analyses of the Framingham Heart Study data, relative weight seemed to be related to an increased incidence of atherothrombotic brain infarction in men aged 50–59 years only

at high levels (FIGURE 10-7).<sup>48</sup> An exposure–outcome association pattern may, in addition, be dose independent—for example, that seen in allergic disorders to certain environmental exposures, such as medications, pollen, and others.

Although confounding or bias is regarded as having less explanatory value when there is a linear dose-response pattern of the type shown in Figure 10-5A, it must be emphasized that this may not be the case if there is a correlation between the level of the exposure of interest and the level of a confounder (or the level of information bias). An example is the relationship of excessive alcohol intake to lung cancer, which may show a graded pattern because of the graded relationship between alcohol (the putative risk factor of interest) and the confounder (smoking). In addition, as previously discussed (Chapter 4, Section 4.3.3), both nondifferential misclassification when there are more than two exposure categories and differential misclassification may produce a spuriously graded relationship between exposure and outcome.

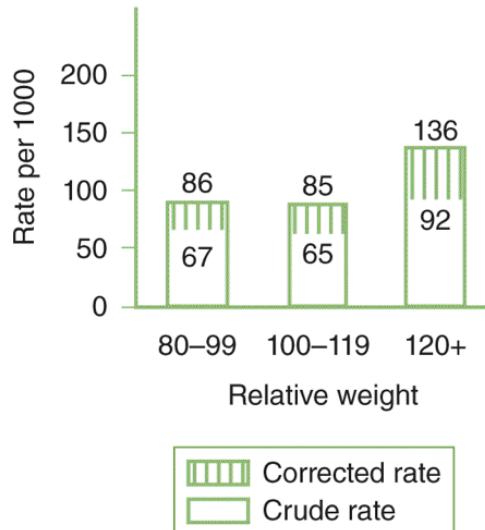


**FIGURE 10-5** Some patterns of cause–effect relationships.



**FIGURE 10-6** Dose-response in 8 studies of birth weight and leukemia included in a meta-analysis.

Data from Hjalgrim LL, Westergaard T, Rostgaard K, et al. Birth weight as a risk factor for childhood leukemia: a meta-analysis of 18 epidemiologic studies. *Am J Epidemiol.* 2003;158:724-735.<sup>45</sup>



**FIGURE 10-7** Risk of atherothrombotic brain infarction in relation to relative weight,\* men aged 50–59 years.

\*Relative weight was determined by comparing the weight of the individual to the median for the corresponding age and sex group.

Courtesy of the Harvard University Press. Dawber TR. *The Framingham Study: The Epidemiology of Atherosclerotic Disease*. Cambridge, MA: Harvard University Press; 1980.<sup>48</sup>

5. *Biological plausibility*: For an association to be causal, it has to be plausible (i.e., consistent with the laws of biology). Biological plausibility may well be one of the most problematic guidelines supporting causality, as it is based on *a priori* evidence that may not stand the test of time. Thus, for example, vis-à-vis the state-of-the-art scientific knowledge of his time, Snow's hypothesis that cholera was produced by a live organism lacked biological plausibility altogether. Weed and Hursting<sup>49</sup> go as far as to suggest the dispensability of the biological plausibility criterion and cite Schlesselman's contention that it "may occasionally impede acceptance of new facts."<sup>50</sup>(p201)

Notwithstanding these limitations, biological plausibility is a useful guideline when it is consistent with the epidemiologic patterns of the exposure–outcome associations. Consider, for example, the J-shaped relationship of alcohol to cardiovascular

mortality mentioned previously. Its biological plausibility is based on the known dose-dependent relationships of alcohol to serum high-density lipoprotein, coagulation factors, and blood pressure as well as on the knowledge of the roles of these factors in the causal pathways resulting in atherosclerosis.

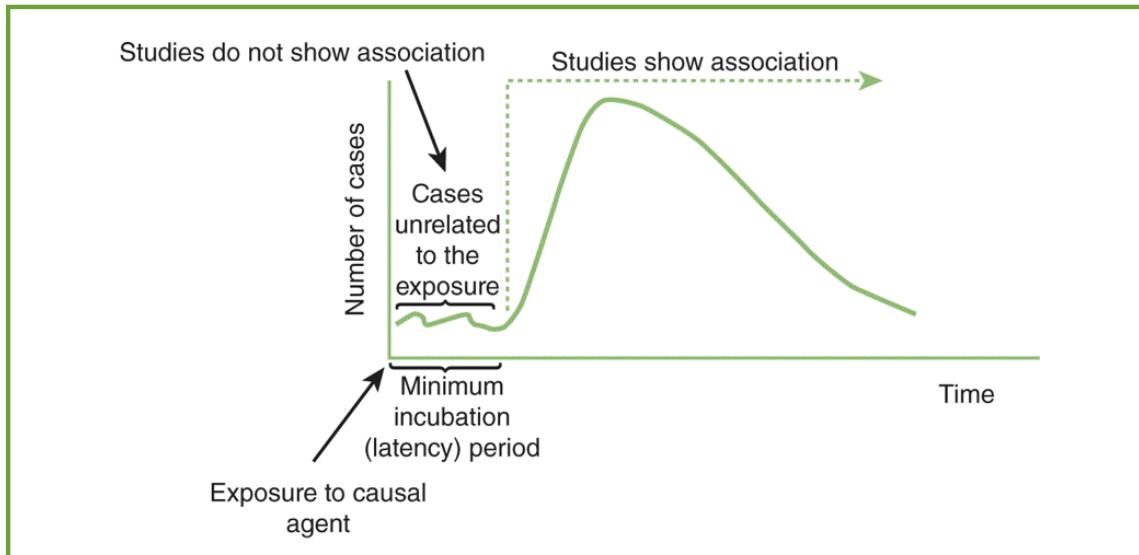
6. *Consistency*: Consistency of results across epidemiologic studies gets at the heart of inductive reasoning and is highly esteemed by epidemiologists as a guideline to infer causality in observational studies, particularly when ratio-based measures indicate weak associations. Observation of consistent, albeit weak, associations provides the main rationale for the use of meta-analytic techniques for policy decision making (see the next section).

Consistency among studies, however, should be used as a means to infer causality cautiously, as it may merely reflect consistency of confounding or bias across studies, particularly observational ones.<sup>50</sup> In addition, apparently consistent results across studies may result from publication bias, whereby “positive” results are more likely to be published than null ones (see [Section 10.5](#)).

Conversely, lack of consistency does not necessarily constitute evidence against a causal association. The reasons causal associations may not appear consistent have been described by some authors<sup>51-53</sup> and are summarized as follows:

- *Differences in the specific circumstances of exposure*. Several characteristics of the exposure of interest may cause differences between the results of individual studies, including duration and level. For example, earlier studies of the relationship of estrogen replacement therapy to breast cancer did not take duration into account as accurately as more recent studies<sup>54</sup> and, therefore, were not able to establish firmly the presence of the association.
- *Differences in the timing of the study with regard to the exposure's latency (incubation) period*. When studies are done at different points in time after introduction of a given exposure, they may yield

inconsistent results. Armenian and Lilienfeld have shown that the distribution of neoplasm cases after exposure to a risk factor follows the same log-normal pattern as that seen in infectious diseases. Examples cited by these authors include bladder tumors in dyestuff workers and leukemia after the atom bomb explosion in Hiroshima and Nagasaki.<sup>55</sup> Although their examples were limited to neoplasms, they logically should apply to all chronic diseases. Thus, plotting the distribution of cases by time after exposure initiation (i.e., constructing an epidemic curve<sup>56,57</sup>) may assist in ascertaining at which point in the curve the study was done. When the minimum latency (incubation) period has not yet gone by in a given study population, investigation of a recently introduced agent cannot detect associations between the agent and the outcome of interest (FIGURE 10-8). For example, in the Women's Health Initiative (WHI) trial, no association between initial use of estrogen plus progestin replacement therapy and breast cancer was seen in the first 3 years, which is likely an approximation of the latency between the exposure to hormonal therapy and breast cancer.<sup>58</sup>



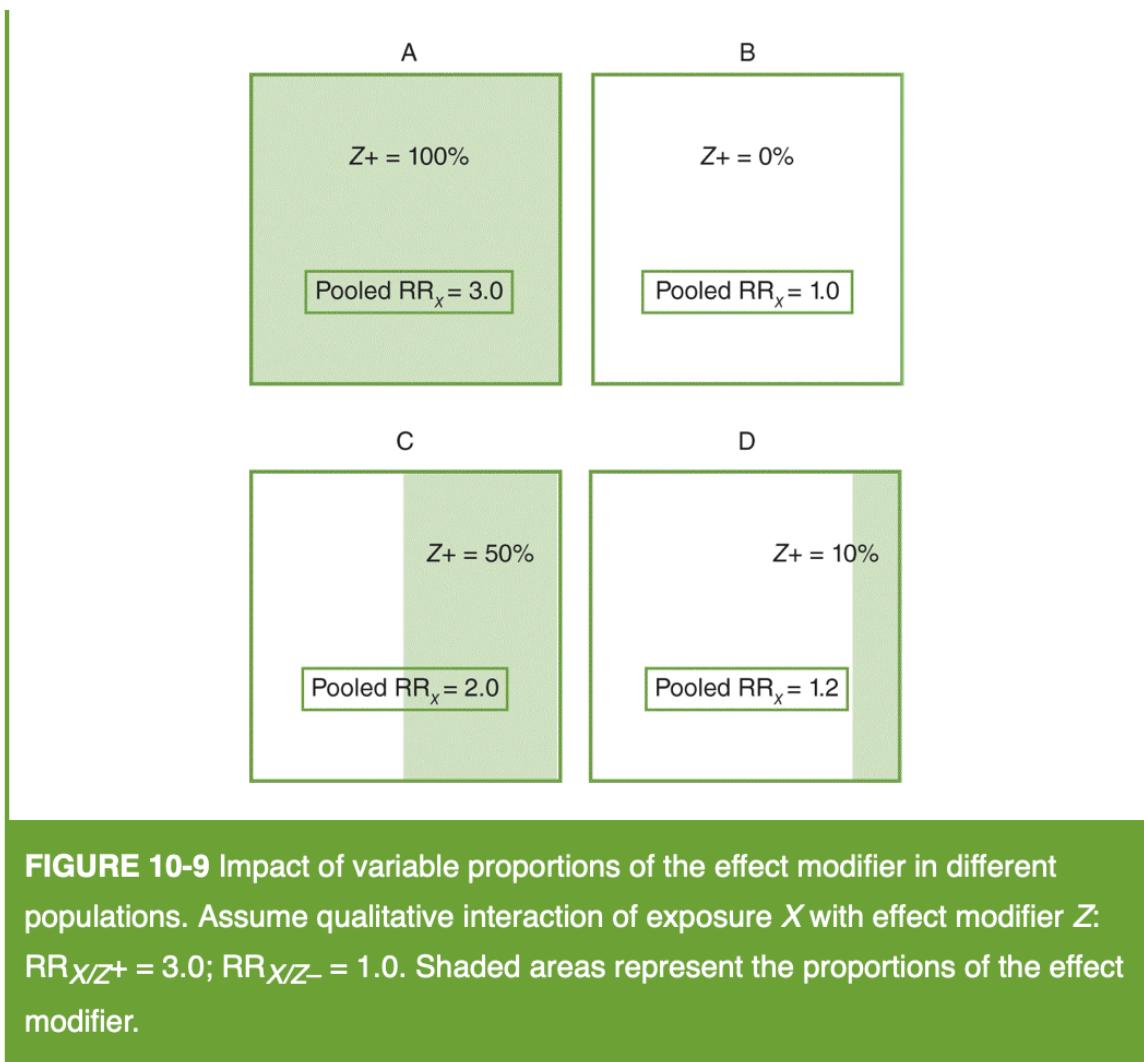
**FIGURE 10-8** Epidemic curve: studies conducted during the minimum incubation (latency) period cannot detect an association between a causal agent and a disease.

Based on Armenian HK, Lilienfeld AM. The distribution of incubation periods of neoplastic diseases. *Am J Epidemiol.* 1974;99:92-100.<sup>55</sup>

- *Differences in design and analytic strategies.* Inconsistent results between studies may also result from differences in the confounders included in the statistical models used in data analyses, the sensitivity and specificity of the definitions of exposure and outcome variables, the power of the study, and the length of follow-up (in cohort studies). Use of broad categories of relevant variables is another problem, as it may hide differences in exposure levels between studies; for example, by using merely the categories “yes” and “no” for smoking, differences may occur in the level of tobacco use from study to study, thus resulting in inconsistent values of the measure of association.
- *Differences in the distribution of a component cause.* Differences in results across studies may also reflect differences in the presence of component causes of a sufficient cause constellation. This notion is best understood in the context of effect modification. For example, if a susceptibility gene for salt-induced hypertension varies from population to population, studies conducted in different populations will detect average (“main”) effects of high salt intake on hypertension of different magnitudes. Assume the extreme example of a qualitative interaction, in which the relative risk is 3.0 when the susceptibility gene  $Z$  (effect modifier) is present but is null (1.0) when the gene is absent ([FIGURE 10-9](#)). In this hypothetical example, for the population in which everyone carries the susceptibility gene ([Figure 10-9A](#)), the relative risk will be 3.0. On the other hand, for populations without the susceptibility gene, the relative risk for hypertension in heavy salt consumers will be 1.0 ([Figure 10-9B](#)). The lower the prevalence of gene carriers, the nearer the average (“main effect”) relative risk will be to 1.0—a phenomenon that has been coined “drowning of susceptibles” (A. Correa, personal communication). Consider, for example, the results of the study by Yu et al.<sup>59</sup> on the relationship between smoking and liver cirrhosis in chronic hepatitis B surface antigen carriers (see [Chapter 6, Section 6.10.2, Table 6-27](#)). In this study, there was marked heterogeneity of the association according to presence of alcohol drinking. Thus, in

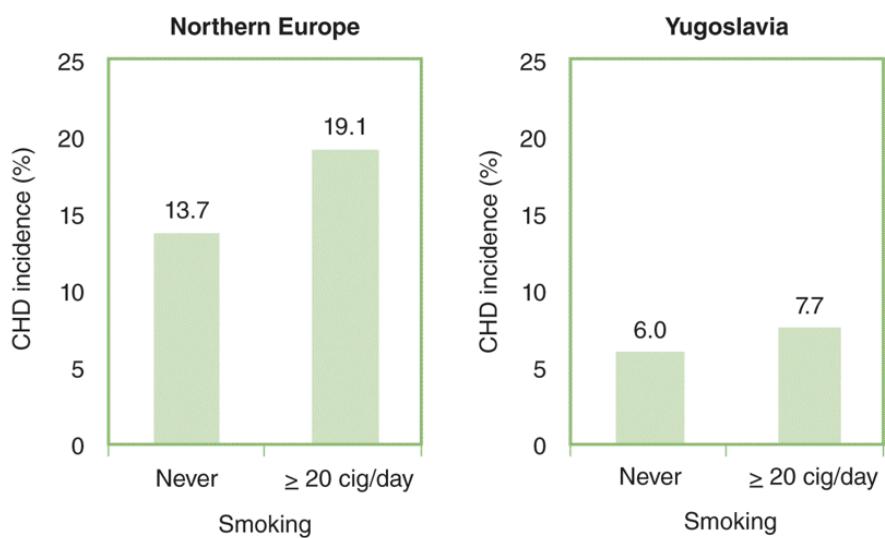
alcohol-drinking individuals, the adjusted relative risk for heavy smoking was about 9.0, whereas in nondrinkers, it was only 1.9. As a result (and assuming that these adjusted point estimates are the true values), it can be inferred that among hepatitis B surface antigen carriers with a high prevalence of drinking compared with those with a low prevalence of drinking, heavy smoking will be a much stronger risk factor for liver cirrhosis. In another study, Fretz et al. found that the association of educational level with the prevalence of subclinical myocardial damage, defined as a concentration of high-sensitivity troponin T (hs-cTnT) of 14 ng/L or more, was found only in black participants. Again, assuming that this study's point estimates are free of random variability, bias, and confounding, studies with mostly white participants would tend to show no association, while those conducted in mostly or exclusively black individuals would yield a prevalence ratio contrasting low (< 12th grade) and high (college or higher) educational levels as high as 1.8.<sup>60</sup> Differences in the proportion of the modifier between a study population and other populations, therefore, affect external validity and require the use of a *transportability* strategy when generalizing results (see [Chapter 6, Section 6.12](#)).<sup>61</sup>

---



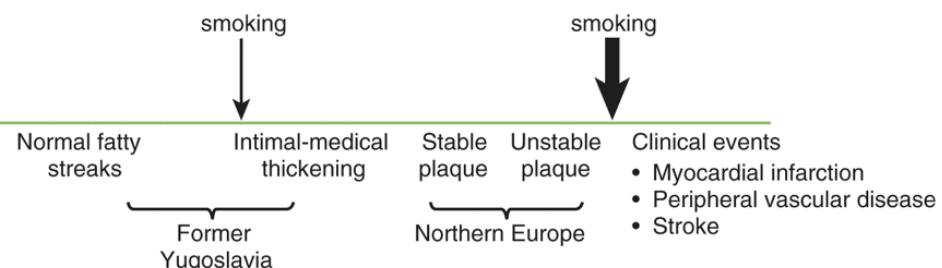
- *Differences in the stage of the natural history of the underlying process.* The natural history of a given disease is often a lengthy process that starts many years before its clinical manifestations. An example is carotid atherosclerosis, which may begin as early as the first or second decade of life.<sup>62,63</sup> Its clinical expression (e.g., myocardial infarction), however, is not common until the sixth and later decades. Traditional risk factors for clinical atherosclerotic disease include high serum cholesterol levels, hypertension, smoking, and diabetes. The role of smoking as a key risk factor had been established early on in the landmark Framingham Heart Study.<sup>64</sup> Its association with CHD, however, was not of uniform magnitude in all locations of the Seven Countries Study.<sup>65</sup> In the latter study, the relationship of heavy smoking ( $\geq 20$  cigarettes/day) to CHD was clearly seen in the Northern European cohort but it was almost

absent in the cohort from the former Yugoslavia (FIGURE 10-10). This finding may reflect the fact that smoking appears to have a more important role in the development of later (rather than earlier) stages of atherosclerosis (FIGURE 10-11),<sup>66</sup> which were more prevalent in Northern Europe than in the former Yugoslavia. Thus, when assessing differences in association strengths across populations, it is crucial to consider the natural history of the disease and the fact that the role of each risk factor may be not be equally important in all its stages.



**FIGURE 10-10** Age-adjusted cumulative incidence rates of coronary heart disease (CHD) in two cohorts of the Seven Countries Study after a 10-year follow-up.

Data from Keys A. Seven Countries: A Multivariate Analysis of Death and Coronary Heart Disease. Cambridge, MA: Harvard University Press; 1980.<sup>65</sup>



**FIGURE 10-11** Smoking seems to be particularly important in later stages of the

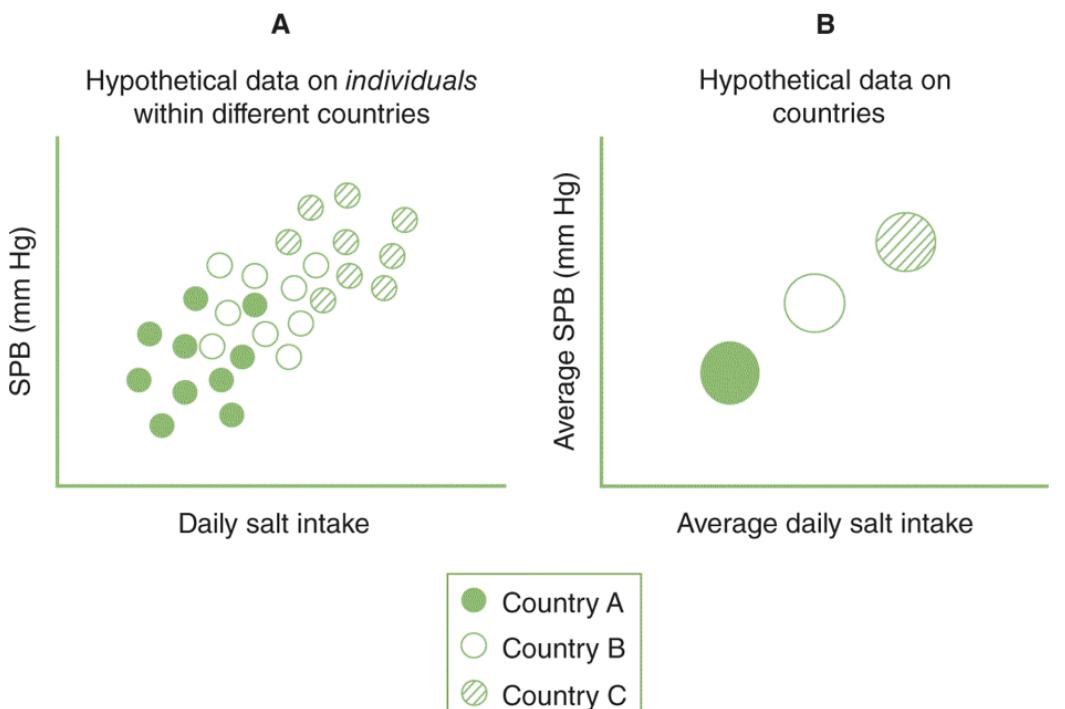
natural history of atherosclerosis. In the former Yugoslavia, earlier atherosclerotic lesions seemed to predominate, so smoking appeared to be a less important risk factor than in Northern Europe, where more advanced lesions likely predominated.

Another example is the discrepancy of findings between observational studies and the Women's Health Initiative (WHI), a randomized trial that evaluated the effect of estrogen therapy on cardiovascular disease. Although observational studies have suggested a protective effect for hormone therapy, the WHI trial showed the opposite effect. In addition to residual confounding, it has been suggested that the discrepancy is due to differences between the two types of studies (observational and randomized trial) in time after menopause.<sup>67</sup> Whereas most women included in the observational studies initiated hormone use within 10 years after menopause, in the WHI they did so much later.<sup>68,69</sup> Thus, women included in the WHI were much older than those included in the observational studies. Therefore, underlying atherosclerosis may have been less severe in the latter than in the former, which would result in favorable effects of hormone therapy on lipids and endothelial function in observational studies and harmful prothrombotic and proinflammatory effects of therapy in the older women participating in the WHI trial.<sup>70</sup>

- *Differences in the effectiveness of interventions.* The applied epidemiologist is often interested in studying the effectiveness of a given preventive intervention, such as a smoking cessation program. A crucial issue usually ignored in evaluating consistency of effectiveness values across studies is that this measure is more context specific—and thus less generalizable—than efficacy. The use of vaccines in the field underscores the sharp distinction between efficacy and effectiveness. The efficacy of a vaccine may be high, but if the field conditions are not ideal—due, for example, to deterioration of the vaccine because of lack of refrigeration or poor acceptance by the target population—its effectiveness will be compromised.

Another key consideration is that, as persuasively demonstrated by Comstock,<sup>71</sup> observational studies of interventions yield effectiveness, not efficacy estimates. An example is given by an observational study that suggested a negative effectiveness of a needle exchange program (NEP) for prevention of HIV infection in the Montreal area.<sup>72</sup> The explanation for this paradoxical finding, offered by the authors of this study, was that “because of the availability of clean equipment through pharmacies . . . needle exchange programs may have attracted existing core groups of marginalized, high risk individuals” and that “in view of the high risk population NEPs, the number of needles may have been less than the actual number needed,” which in turn may have led to the use of contaminated needles.<sup>72</sup>(p1001) An obvious conclusion is that a positive effectiveness may have been achieved under different circumstances than those encountered in this particular study population. [Section 10.3](#) discusses the role of the decision tree in estimating the effectiveness of a given intervention.

- *Differences in the variability of the risk factor.* As noted by Wynder and Stellman<sup>73</sup>(p459) with regard to case-control studies, “If cases and controls are drawn from a population in which the range of exposures is narrow, then a study may yield little information about potential health effects.” The issue of little variability in the exposure levels is also applicable to cohort studies. For example, as discussed in [Chapter 1](#), [Section 1.3](#), observational studies using individuals as analytic units have been unable to show consistent relationships between salt intake and hypertension ([FIGURE 10-12A](#)); on the other hand, because of the marked interpopulation variability in average salt intake, ecological studies using country as the unit of analysis have clearly demonstrated a correlation<sup>74</sup> ([FIGURE 10-12B](#)).
-



**FIGURE 10-12** Intracountry variability in salt intake is not sufficient to allow observation of an association between salt intake and systolic blood pressure (SBP). However, given intercountry variability, there is a strong correlation between average salt intake and SBP when “country” is used as the analytic unit. Small circles denote individuals within each country; large circles denote countries.

Thus, the variability of the risk factor level within a population is a key determinant of whether an association can be found in that population. It is recommended, therefore, that presumed risk factor variability be considered by the use of, for example, the coefficient of variability (median standard  $\div$  deviation) or the interquartile range.

Readers might notice the absence of three of Hill’s original guidelines—*coherence*, *analogy*, and *specificity* of an association—from the preceding discussion. We like others,<sup>15,37</sup> believe that these three guidelines are not useful for the following reasons: *Coherence* is hard to distinguish from biological plausibility; *specificity* is inconsistent with state-of-the art knowledge, as it erroneously postulates that a given agent is always associated with only one disease and that the agent can always be found for that disease; and with regard to *analogy*, as pointed

out by Rothman and Greenland,<sup>37</sup> “Whatever insight might be derived from analogy is handicapped by the inventive imagination of scientists who can find analogies everywhere.”

In the next three sections, four topics closely related to the evaluation of the previous guidelines are briefly discussed: *decision tree*, an approach that is useful when estimating overall effectiveness of a program or intervention; *sensitivity analysis*, a technique to evaluate the impact of errors on study results or alternative levels of factors influencing effectiveness; *meta-analysis*, an important tool for the evaluation of strength of a given association and consistency across different studies; and *publication bias*, which may strongly affect results of systematic reviews, including meta-analyses.