Causal Inference in Public Health

Thomas A. Glass,[1] Steven N. Goodman,[2] Miguel A. Hernán,[3,4] and Jonathan M. Samet[5]

[1]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205; email: tglass@jhsph.edu

[2]Department of Medicine, Stanford University, Palo Alto, California 94305; email: steve.goodman@stanford.edu

[3]Departments of Epidemiology and Biostatistics, School of Public Health, Harvard University, Boston, Massachusetts 02115

[4]Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02115; email: miguel_hernan@post.harvard.edu

[5]Department of Preventive Medicine, Keck School of Medicine, and USC Institute for Global Health, University of Southern California, Los Angeles, California 90089; email: jsamet@med.usc.edu

Keywords

causation, causal modeling, causal framework, epidemiology

Abstract

Causal inference has a central role in public health; the determination that an association is causal indicates the possibility for intervention. We review and comment on the long-used guidelines for interpreting evidence as supporting a causal association and contrast them with the potential outcomes framework that encourages thinking in terms of causes that are interventions. We argue that in public health this framework is more suitable, providing an estimate of an action's consequences rather than the less precise notion of a risk factor's causal effect. A variety of modern statistical methods adopt this approach. When an intervention cannot be specified, causal relations can still exist, but how to intervene to change the outcome will be unclear. In application, the often-complex structure of causal processes needs to be acknowledged and appropriate data collected to study them. These newer approaches need to be brought to bear on the increasingly complex public health challenges of our globalized world.

INTRODUCTION

Overview

The determination that an association is causal can have profound public health consequences, signaling the need or at least the possibility to take an action to reduce exposure to a hazardous agent or to increase exposure to a beneficial one. Consequently, causal inference is implicitly and sometimes explicitly embedded in public health practice and policy formulation. Practitioners decide on interventions on the basis of consequences produced by a presumed causal relationship. Causal inference is embedded in regulatory processes, for example those of the US Environmental Protection Agency (EPA) with regard to major outdoor air pollutants and the hazards of chemicals, and those of the Department of Veterans Affairs, in compensation of US veterans for service-connected conditions and diseases [Agent Orange Act, Pub. L. 102-4 (1991); Clean Air Act 42 U.S.C. § 7401-7671q (2008); 36, 37]. Public health evidence may be prominent in legal proceedings in which judgment about the existence of a causal relationship is pivotal in determining guilt and liability for damages (16, 72). Causal inference is also embedded in many aspects of medical practice through the principles of evidence-based medicine, where decisions about harms or benefits of therapeutic agents are based, in part, on rules for how to measure the strength of evidence for causal connections between interventions and health outcomes (20).

The history of public health and of its quantitative disciplines, epidemiology and biostatistics, can be seen as one long discourse on disease causation, the ultimate targets of which are to find and to mitigate reversible causes (22, 23, 33, 46, 50, 67). Over that history, a variety of "frameworks" for thinking about causation have risen to coincide with the dominant problems of the day and the scientific understanding of their etiology. During the ravages of the cholera epidemics of the nineteenth century, John Snow gathered evidence in support of waterborne transmission, using what Frost later called his ordered "chains of inference" (11, p. IX) (15, 73, 74). With the advent of germ theory, Koch's postulates provided a more systematic and formalized approach that worked well within the specificity of unique germ-disease links (8).

In the 1950s and 1960s, what we call the classic framework for causal thinking was articulated by Sir Austin Bradford Hill, who added to this discourse with his causal criteria against the backdrop of international debate about the causal role of smoking in the epidemic of lung cancer (53, 76). This classic framework was developed to identify the causes of diseases and particularly to determine the role of smoking in lung cancer (33, 71), but its use has been extended to public health decision making, a domain where questions about causal effects relate to the consequences of interventions that have often been motivated by the identification of causal factors. This framework, described below, has proven useful and has driven decision making in public health for decades. However, the framework does not reflect the current, more clearly articulated view of causal processes. Additionally, the guidelines used to evaluate evidence have not changed for decades, even as the causal questions have become more complex, beyond the original intent of this framework.

One important limitation of the classic view of disease causation arising from the Hill criteria has been the lack of a formal basis for evaluating causal hypotheses. Only in the past several decades have investigators explored more formally the foundational mathematical and conceptual issues required for rigorous estimation of causal effects, particularly in circumstances where randomization of treatment assignment that insures exchangeable comparison groups is unfeasible. Since 1970, the frequency and intensity of formal discourse on causation and causal inference have increased, and the field has progressed toward what we term the modern approach, based on the counterfactual or potential outcomes framework (18, 25).

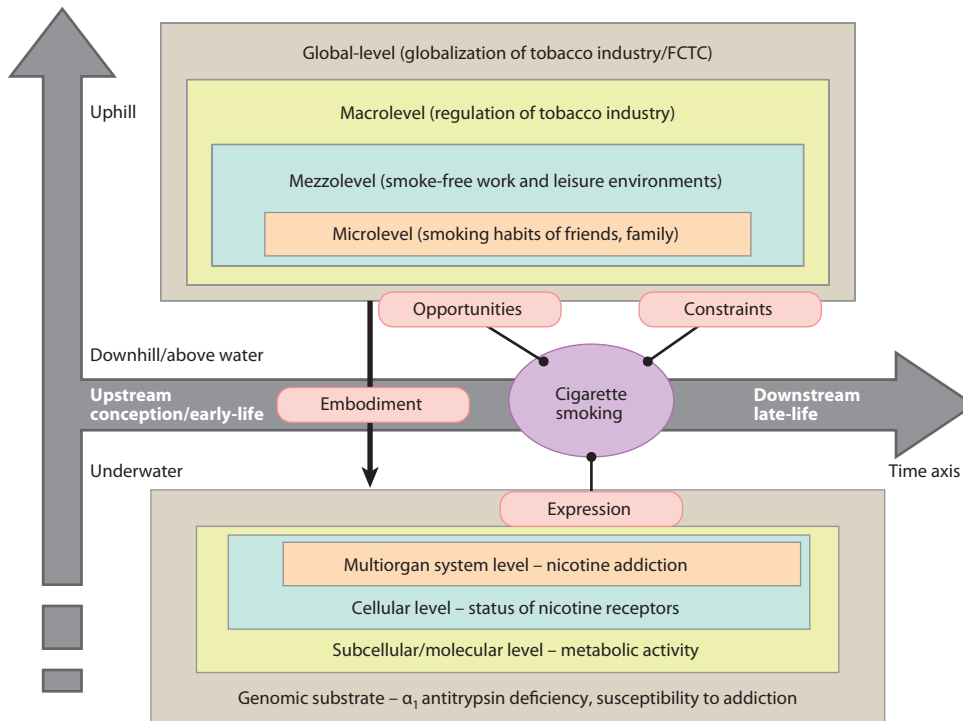In this review, we first describe and comment on the classic framework that is generally

Figure 1

Axis of nested hierarchies for tobacco control. Reprinted with permission from Samet & Wipfli (65). FCTC, Framework Convention on Tobacco Control.

attributed to Sir Austin Bradford Hill and the advisory committee that prepared the 1964 US Surgeon General's report on smoking (33, 71). We follow with a brief review of the modern framework based on the counterfactual, or potential, outcomes model for estimating causal effects. The latter approaches are unified by an analytic effort to approximate the experimental paradigm that balances treated (exposed) and untreated (unexposed) groups on other factors. We next carry this counterfactual approach to the broad and multilevel nature of causal questions, as formulated over the past several decades, and consider causal inference in the context of such questions and their implications for public health actions (14). We end with consideration of how these new approaches—broader frameworks for formulating causal questions and developing analytical tools to answer them—can be used to reduce uncertainty associated with causal determinations. The

interplay between strength of evidence and remaining uncertainties typically figures prominently in decision making. More pragmatically grounded and transparent approaches are needed as we face such challenges as the rise of obesity throughout the world—an example that necessitates a multilevel framing of underlying causal processes, with structures extending from the genes of individuals to the foods sold worldwide by multinational corporations, as the basis for formulating interventions (35). This type of framework has already proven valuable in approaching tobacco control (**Figure 1**). The upstream drivers of the epidemic are clear at this point in its course: a large and powerful global industry led by a handful of powerful multinational corporations. The role of factors at other levels has also been characterized: cultural acceptance of smoking, laws, peers, and the family. Now, we are probing the genetic basis of susceptibility to

nicotine addiction and tobacco-caused diseases. Within the modern framework, such structure leads to questions and counterfactuals at multiple levels: At the highest level, what would be the disease burden, absent the upstream factor (e.g., the tobacco industry), and at the lowest level, what would be the disease risk for genetically susceptible individuals, absent the environmental factor (e.g., smoking)? The structure also raises the possibility of interventions at multiple levels, reflecting how interventions might be carried out in practice.

A Brief Detour into Philosophy

Although public health scientists and practitioners have disagreed fiercely at times about what is required of causal explanations, the idea that causal relationships can be proven has rarely been seriously questioned. But in the long and contentious discourse on causation in philosophy (3), one can discern two distinct classes of causation theory. On one side are the descendants of Locke and John Stuart Mill, who argued that causation can be verified through the careful implementation of the scientific method and the power of experimentation. On the other side is a parallel line of discourse that extends from David Hume, who argued that even though nature may contain real causal "connexions" between phenomena, causation cannot be empirically verified (36). This skeptical tradition had no better spokesman than Bertrand Russell (64), who in a famous essay delivered to the Aristotelian Society of 1912, wrote, "The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm" (p. 1).

Although the science of epidemiology and the practice of public health fall clearly into the pragmatic tradition of Locke and Mill, evidence of the influence of Hume and Russell can be found in the early skepticism of R.A. Fisher (9) and Karl Pearson, the father of modern statistics, who argued that the correlation between two variables, once known, is all there is to know, a view that persists with some epidemiologists. In their review of causal inference in epidemiology, Lipton & Ødegaard (47) ask what is really added to the statement that smokers are at X-fold increased risk of lung cancer by the statement that smoking is a cause. From a policy point of view, the use of causal language has obvious advantages, and it has been widely embraced not only by researchers but by policy makers. The legacy of Hume and Russell urges us to be cautious because assigning causal significance to some phenomena also provides an easy target for skeptics and, potentially, affected stakeholders to derail reasonable interventions on the basis of an absence of proof. Public health practitioners and researchers are interested primarily in effecting change and not in engaging in philosophical debates, but the ghost of Russell reminds us that the invocation of causal language has powerful consequences, both good and bad.

The challenge of determining causation in public health has always been shaped by the limitations of the available data, the understanding of the underlying biological or sociological processes, and our ability to intervene in the real world. Faced with sometimes limited data and an often poor understanding of a network of connected factors in a complex world, we revert to pragmatism. Public health science seeks the certainty of the experiment as its organizing principle. Holland (34) says it succinctly in a famous paper, "Put as bluntly and as contentiously as possible, in this article I take the position that causes are only those things that could, in principle, be treatments in experiments" (p. 954).

This statement is formalized in the potential outcomes framework, which compares what is observed to what might have been observed, all other things being equal, under a counterfactual scenario. The potential outcomes framework is a powerful tool that has implications for how we see the world and to determine what types of questions can be answered in a useful way for public health purposes and what kinds of questions are beyond our capacity to answer (25, 55, 61–63).

Approaches to Causal Inference in Public Health

The classic approach to causal inference in public health, described quite similarly across textbooks and widely used in practice, has its roots in the seminal debate around smoking as a cause of lung cancer in the 1950s and 1960s (33, 71). At that time, the results of epidemiological studies had shown associations of smoking with increased risk for lung cancer and other cancers, for coronary heart disease, and for "emphysema" and "bronchitis." The most relevant data came from case-control and cohort studies and findings from animal models and lab studies characterizing the components of tobacco smoke. Rising mortality rates from lung cancer and coronary heart disease provided a strong imperative for taking action to reduce cigarette smoking. However, taking action required that smoking be established as the cause of the increases in mortality. Even as the epidemiological evidence mounted, the tobacco industry implemented a wide-ranging strategy to question the credibility of epidemiological evidence generally and of the most pivotal studies specifically (54). This tactic of creating doubt about the evidence heightened tension around the challenge of interpreting the findings of epidemiological research, and its use attests to the societal importance of causal determinations. The manufacture and dissemination of doubt remain strategies today, widely used by stakeholders whose interests are potentially threatened by a causal finding (49).

The framework that was put forth for causal inference in the 1960s involved expert judgment grounded in a set of guidelines or criteria (**Table 1**). The long-standing discussion among philosophers was acknowledged as these guidelines were elaborated, but the need for a pragmatic and timely approach foreshortened debate. The framework was effective for smoking and lung cancer, one of its first applications. Smoking is a potent cause, increasing the risk of lung cancer about 20-fold and leading to most cases of lung cancer; consequently, the evidence from observational studies was consistent and strong, and temporality

was clear. As described by their originators and as used in practice, these criteria (or what Hill calls "viewpoints") are not absolute nor does inference of a causal relationship require that all criteria be met. In fact, only temporality is requisite. Some features of evidence, most notably specificity, have proven to have little applicability for noncommunicable diseases that have multiple causes. The classic approach is vulnerable to subjectivity in the evaluation of evidence and to manipulation of the evidence, and stakeholders potentially affected by the finding that an association is or is not causal may take opposing positions on evidence interpretation. Additionally, as constructed and applied, the framework assumes a simplistic direct relationship between cause and putative effect without explicit consideration of the structure of the underlying causal processes. For example, tobacco smoking is an indisputable cause of lung cancer, but more distally in the causal process, a small number of multinational tobacco companies produce most of the cigarettes sold and smoked worldwide (**Figure 1**). The inference about cause became the rationale for intervention, but the causal conclusions were not couched in the consequences of specific actions to reduce or eliminate cigarette smoking. And later, public health action was aimed at the individual smoker, rather than at the upstream system of cigarette manufacture, advertising, and distribution. This limited focus is a key characteristic of the traditional approach; causal determinations were made by epidemiologists and others in public health about various risk factors without considering the effect of a specific way of changing them.

Today, public health practice can be seen to be influenced by both the classic and modern frameworks, as exemplified in the following case studies. In setting outdoor air quality standards in the United States, causal inference and associated counterfactuals figure in the decision process. Two sections of the US Clean Air Act (108 and 109) address the major outdoor air pollutants, requiring the Administrator of the EPA to set National Ambient Air Quality

Table 1   Guidelines for causal inference. Data from the 1964 *Smoking and Health: Report of the Advisory Committee to the Surgeon General* **(71) and from Hill 1965 (33)**

US Surgeon General Report's criteria	Hill's criteria
Consistency of association	Strength
Strength of association	Consistency
Specificity of association	Specificity
Temporal relationship of association	Temporality
Coherence of association	Biological gradient
	Coherence
	Experiment

Standards (NAAQS) such that "the attainment and maintenance of which in the judgment of the Administrator, based on such criteria and allowing an adequate margin of safety, are requisite to protect the public health" (p. 5697). The phrase "such criteria" refers to the accumulated evidence on harm, giving emphasis to that reported since the last review of the NAAQS. The present process for a pollutant, e.g., ozone, begins with a review of the evidence, assembled in the Integrative Science Assessment (**Figure 2**). The process for causal inference draws on the long-standing classic approach and classifies the strength of evidence in a five-level scheme ("not likely," "inadequate," "suggestive," "likely," and "causal"). The classification, in part, determines the effects that are subsequently considered in the risk analysis, which estimates the pollutant-related burden of disease and the consequences of potential changes to the NAAQS. Those effects for which the evidence reaches the level of "likely" or "causal" are generally advanced for consideration in the risk analysis and consequently figure in the policy judgment made by the Administrator on revising the NAAQS for a pollutant. The risk analysis models the counterfactual distribution of health outcomes under different scenarios of pollution reduction and under no intervention. The risk analysis has the modern approach as its conceptual underpinning.

The International Agency for Research on Cancer (IARC) of the World Health Organization operates its Monograph Program, which conducts systematic reviews to classify agents by their carcinogenicity (39). The general approach involves a meeting of a multidisciplinary working group that reviews evidence relevant to a particular agent in four broad categories: (*a*) exposure, (*b*) studies of cancer in humans, (*c*) studies of cancer in experimental animals, and (*d*) mechanistic and other relevant data. The human and animal evidence is separately considered, and for each category, the strength of evidence for causation is classified in a four-level hierarchical schema: sufficient, limited, inadequate, or suggesting lack of carcinogenicity. The evidence is evaluated with an approach based in the Hill or classic criteria. Evidence for the role of particular mechanisms is evaluated as "weak," "moderate," or "strong," and investigators consider the relevance of the mechanism to cancer in humans. The overall classification is based primarily on the animal and human findings (**Figure 3**), but the mechanistic evidence can figure in the classification as well. This approach, for example, resulted in the 2011 classification of radiofrequency electromagnetic radiation, the type emitted by mobile phones, as a possible human carcinogen, Group 2B in the IARC schema (2).

CAUSAL INFERENCE AS A COMPARISON OF OUTCOMES UNDER DIFFERENT PUBLIC HEALTH INTERVENTIONS

As described above, a key role for causal inference in public health is the comparison of the distribution of health outcomes after
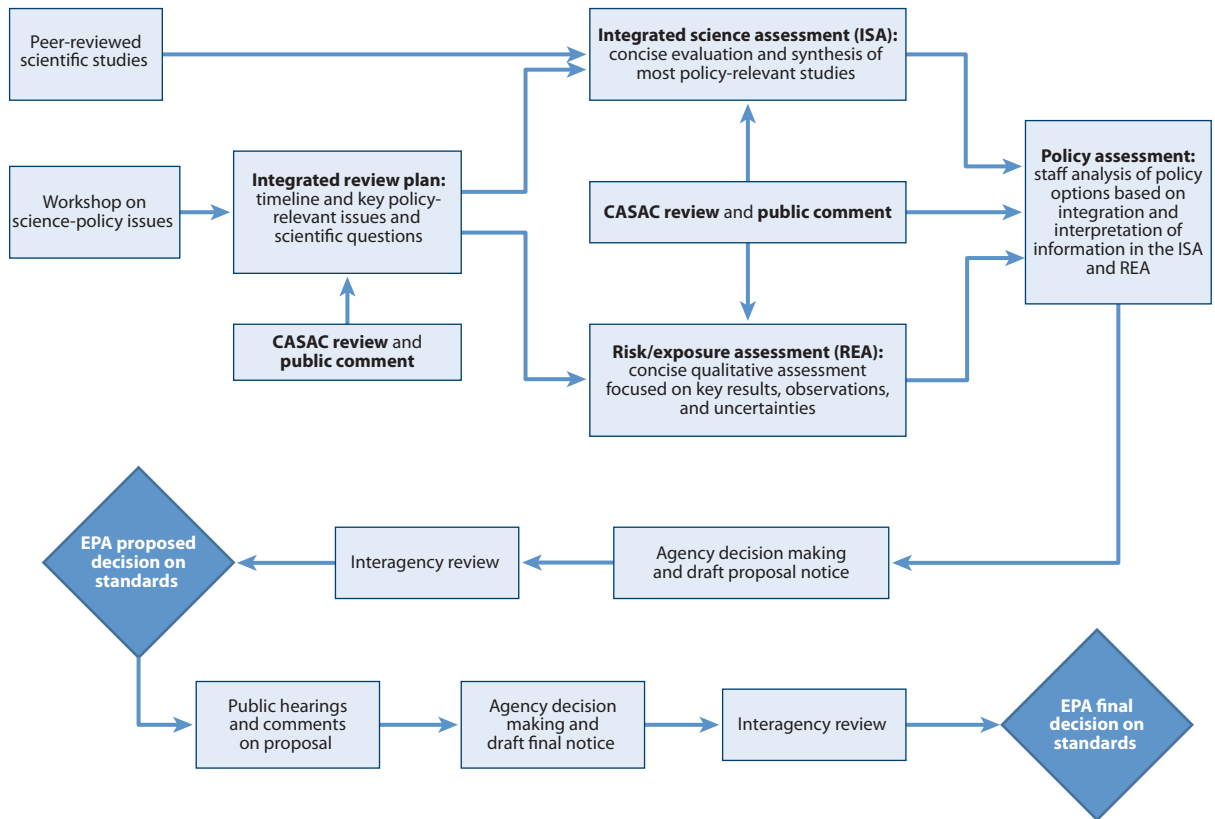
Figure 2

The National Ambient Air Quality Standards (NAAQS) review process. Source: Memorandum from US Environmental Protection Agency (EPA) Administrator Jackson addressing the revisit of the National Ambient Air Quality Standards (NAAQS) review process, May 21, 2009 (70). CASAC, Clear Air Scientific Advisory Committee.

different interventions. In an ideal world, these comparisons would be conducted via randomized experiments, and all public health decisions would be based on the findings of those experiments. For example, the integration of smoking-cessation programs into the health care system would ideally rely on the findings from long-term randomized studies comparing the efficacy of the intervention in large groups of people from the target population that adhered to the intervention with control groups. Similarly, the decision to increase taxation or regulation of tobacco products would be based in studies that randomly allocated these policies across communities or counties. Unfortunately, such randomized

experiments are often unethical, impractical, or simply too lengthy for timely decision making. As a result, causal inferences for public health are usually derived from observational studies, buttressed by other lines of evidence if available.

The use of observational, rather than experimental, data for causal inference in public health raises several concerns. One particularly relevant concern for public health is that the interventions under consideration may be vaguely defined, if at all, limiting the relevance of the findings for public health decision making. For example, the comparison of observed mortality rates between obese and lean people suggests a possible causal relation between
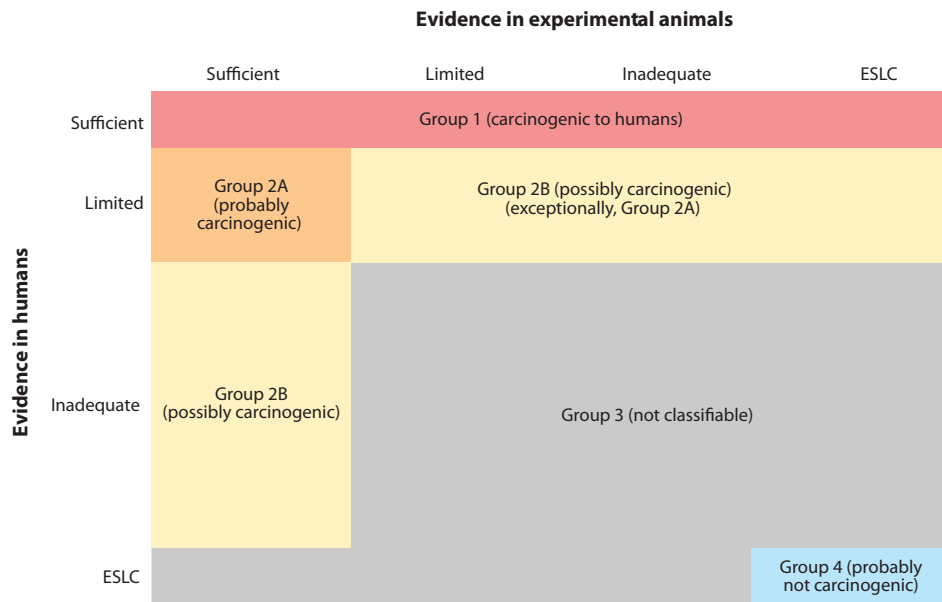
Evidence in experimental animals

Figure 3

The International Agency for Research on Cancer (IARC) classifications based on evidence from human and experimental evaluations. From IARC. For further information, see the Preamble to the IARC monographs on the evaluation of carcinogenic risk to humans, 2006 (40). ESLC, evidence suggesting lack of carcinogenicity.

obesity and death but offers little guidance for action: Should solutions be found in exercise programs in the workplace, reduction of sizes of sugared sodas available in retail stores, liposuction (26, 32)? Although obesity may meet criteria for a causal factor in the classic framework, the association between obesity and mortality offers little insight for preventive action. One alternative is to focus on the contrast between individuals randomly assigned to dietary modification versus those who are not or a contrast between communities randomized to taxation of sugary drinks versus those who are not. The findings from such experiments would provide direct, actionable information about the effects of interventions against obesity. The observational study that compares obese and lean people provides only indirect evidence and lacks a formally testable causal relation in the absence of further specification.

One way to address this concern and bridge the gap between the observational data and public health decision making is to design observational analyses in such a way that the observational data emulate those from hypothetical randomized experiments with relatively well-defined interventions. For example, observational data could be used to mimic a hypothetical randomized experiment involving dietary interventions by comparing the observed outcomes of individuals who change versus those who do not change their diet during the study period; or data could be used to mimic a hypothetical randomized experiment of food policy by comparing health outcomes between schools that did and did not restrict access to sugary drinks. This approach is built into the counterfactual or potential outcomes framework proposed by Neyman (51), expanded by Rubin (61, 62), and generalized to time-varying exposures by Robins (55, 56). A counterfactual approach to causal inference in public health requires that the causal effects are defined in terms of contrasts between the distributions of

the health outcomes under different (hypothetical) well-defined interventions.

Comparing relatively well-defined public health interventions is only the first problem for causal inference from observational data, however. Even well-defined intervention groups will not usually be directly comparable because the key characteristics of individuals in each group are likely to differ. For example, individuals who change their diet may also adopt a healthier lifestyle than those who do not, and schools that change their food policies may serve populations with less economic inequality than do those schools whose policies remain unchanged. This noncomparability problem, commonly referred to as confounding, is a fundamental problem for causal inference using observational data.

The most common approach to mitigate confounding is to measure as many variables as possible that are responsible for the noncomparability and to adjust for them in the statistical analysis. The available methods to adjust for measured confounders are stratification, matching, standardization, inverse probability weighting, and g-estimation. In practical applications with sparse or high-dimensional data, these adjustment methods are implemented with the help of statistical models. For example, adjustment via stratification is often carried out using conventional regression models.

Sometimes the measured confounders are used to estimate each study participant's probability of receiving the exposure of interest. For binary exposures (e.g., yes/no), this probability is referred to as the propensity score (60). If the propensity score is available for adjustment, then the individual variables are not necessary. Inverse probability weighting and g-estimation are methods based on propensity scores. Propensity scores can also be used to adjust for confounding via stratification (e.g., by adding the propensity score as a covariate in the regression model), matching, and standardization.

For the above methods to provide valid causal inferences, all the confounders must have been identified and appropriately measured, a condition that is not empirically testable. One alternative method to eliminate confounding from the effect estimate is instrumental variable estimation (17, 31). Unlike the other methods, instrumental variable estimation does not require investigators to measure any confounders. Rather, it requires them to identify and appropriately measure an instrument, which is roughly defined as a variable that has an effect on the exposure and that is unassociated with the outcome except through its effect on the exposure. Unfortunately, it is impossible to verify empirically that a particular variable is an appropriate instrument. Furthermore, valid instruments can provide only lower and upper bounds for the magnitude of the causal effect of interest. Typically, these bounds are not helpful for decision making because they range from beneficial to harmful effects. As a result, most applications of instrumental variables make additional untestable assumptions to obtain point estimates for the effect of interest.

When exposures are time-varying, a new potential problem arises: Perhaps the confounders (also time-varying) are themselves affected by prior exposure levels. In the presence of this exposure-confounder feedback process, some of the above methods—stratification and matching—cannot be generally used for valid causal inference. Valid adjustment for measured confounding requires the use of the parametric g-formula (a generalization of standardization) (55, 68), inverse probability of marginal structural models (27, 58), or g-estimation of nested structural models (which include some forms of instrumental variable estimation for time-varying exposures as a particular case) (28, 57). These methods, developed by Robins and collaborators since 1986, are often referred to as causal methods because they can be applied to obtain valid causal inferences, even in complex settings with time-varying confounders affected by prior exposure (29, 55).

Another recent addition to causal inference methodology is the use of causal diagrams (directed acyclic graphs, or DAGs). Although not

a data-analysis method themselves, causal diagrams are used to represent the structure of the causal networks linking exposure, outcome, confounders, and other variables, requiring an explicit formulation of the relationships among these factors. Thus, causal diagrams are a helpful tool to detect, graphically, possible sources of bias and to guide investigators in the design of their data analysis (19, 30, 52).

Challenges to Implementing the Potential Outcomes Framework

Although the potential outcomes approach is robust in the context of a range of causal questions of high value to public health, its use raises some questions. For example, should we consider causal questions about inherent features of the individual (such as sex, race/ethnicity, or age) that cannot be reasonably translated into hypothetical interventions (4, 41, 42, 47, 66)? And how should investigators address individual (e.g., body weight) or social (e.g., neighborhood income level) factors that can be translated into hypothetical interventions but for which many possible interventions exist? The potential outcomes approach highlights that when we estimate associations of health outcomes with factors not amenable to change, the question of how to change the outcomes caused by those factors remains open. Consequently, investigations into the association between nonmanipulable factors and health outcomes can be seen as a prelude to other studies on hypothetical interventions (1). For example, if observational studies tell us that individuals living in poor neighborhoods experience higher cancer rates than do those living in more affluent neighborhoods, then the next suite of investigations might consider potentially manipulable carcinogen exposures or diets that differ between the communities under study. The initial finding of a higher cancer rate in poor communities is critical in motivating studies to find causes that can be manipulated. Absent such further research, epidemiology becomes more of a descriptive tool for sociologic analysis and

less of an instrument for providing evidence leading to interventions to improve health.

The potential outcomes framework can also be combined with a multilevel framework to bring context back to epidemiology and public health (5, 6, 12, 48, 59). The causal role of higher-level contextual factors can be evaluated as long as they can be defined as comparisons between alternative interventions or policies. However, even when hypothetical interventions on national or regional policies can be imagined (though often impossible to implement), many of these contextual exposures are uniform within a society, which makes it difficult to gather the data needed to conduct an evaluation. As a result, in practice, epidemiologists and public health practitioners can be induced to prioritize the study of proximal, downstream interventions at the individual level. For example, it is easier to conduct, or emulate using observational data, randomized trials of smoking-cessation programs that target individuals than to conduct trials about the behavior of well-funded corporate entities with vested interests and political connections.

The potential outcomes framework has been extended in several directions to accommodate multilevel causal processes (75). Formal modeling approaches have arisen in infectious disease to handle endogeneity and interference (21, 43–45, 69). Complex systems approaches have begun to offer new frameworks for causal processes across multiple geographic and time scales (7, 13, 14, 24). They call for a mapping of the agents and processes involved in producing outcomes and, consequently, are useful for framing many of the most pressing public health challenges that result from processes at levels ranging from local to global. They need to be brought to bear on public health problems as appropriate. They point to the data that should be collected, how the data should be organized, and how the data should be analyzed in the potential outcomes framework. Complex systems approaches may also provide insights into the consequences of outcomes carried out by different actors and at different levels.

THE FUTURE OF CAUSAL INFERENCE IN PUBLIC HEALTH

In this article, we have offered a brief overview of how causal inference can evolve to enhance public health decision making; we have given insight into how this goal could be accomplished. Current causal inference methods are relevant and useful because they are directed not at identifying causes, but at identifying effects of interventions. The classic criteria for causal inference do not clearly separate these two goals, leading to debates about the attribution of cause that are, in fact, implicitly about the appropriate intervention. Even if we understood a causal chain perfectly, i.e., knew every factor that could be considered a cause, we still might not know how best to change the outcome. Newer causal inference methods move us away from the philosophical exercise of identifying causes and force us to consider more profoundly how to improve health through specific interventions.

Returning to the example of tobacco, for public health purposes, estimating the impact of a given reduction in individual smoking is less important than estimating the consequences for health of smoking-cessation programs versus cigarette taxes. The latter exercise provides a guide to action. But what is different about these two interventions, aside from their estimated effects, is how we evaluate them. The effect of smoking cessation is amenable to randomized evaluation, but characterizing the consequences of raising taxes and other forms of social intervention may not be. In such situations, we must use observational data to emulate the experiment that cannot be conducted. The greater the departure from the randomized experiment for evaluation, the greater the reliance on modeling and subject-matter knowledge, including sociologic and other theories. This need to turn to observational data poses a potential dilemma for public health; if we succumb to focusing on interventions that are easy to evaluate, we may ignore upstream interventions for which the randomized experiment cannot be conducted or emulated but which may have the greatest potential to effect change. Causal inference frameworks and methods help us to identify intervention options and to determine how best to assess their effects, but they do not necessarily inform on the relevant levels of intervention to consider and which interventions should be attempted.

The preceding discussion shows us that causal inference methods cannot be ignored by those who endeavor to improve the public's health. A focus on the effects of interventions rather than causes brings the science of public health in closer alignment with its practice. New causal inference methods force us to confront, as previous methods did not, how interventions will affect public health. However, a number of steps must be taken to move these methods from academia to practice. First, teaching in public health, particularly in MPH (master of public health) programs, often emphasizes the classic framework. This limited focus needs to be changed so we can birth a new cohort of public health professionals who have a better understanding of causation and the relevance of the potential outcomes framework for their work. Second, accessible, high-profile examples of the utility of the modern framework need to be developed and disseminated through publication and presentations at professional meetings that public health professionals attend. A very useful case study could be developed, for example, around the multicomponent strategy used to address cigarette smoking in New York City (10). During the period 2002–2003, cigarette smoking dropped steeply in New York City following the implementation of an aggressive strategy with components including increased taxes, an indoor smoking ban covering most workplaces, increased cessation services, and education. These methods could be used to address the public health impact of the ban on sales of large-sized sugar-containing beverages, which came into force in 2013. This widely publicized intervention offers a valuable test case.

The potential outcomes framework should be embraced as appropriate to gauge the potential effectiveness of public health actions.

We have touched on new analytical tools developed to sharpen analyses of observational data within this framework, recognizing that true randomized trials are not possible for many issues. Other approaches capture the complexity of causal processes with formality sufficient to be useful as a framework for data collection and analysis and to identify targets for intervention. As public health data are collected, they need to have enough richness for this purpose. Public health professionals need not shy from causal inference using these newer approaches because of perceived complexities.

As the origins of questions confronting public health professionals become more complex and global, we are increasingly challenged to understand the world sufficiently and to capture its complexity in our models and interventions to identify areas for change. From obesity to climate change, how we should measure the effects of causes and where investments are best directed become questions with enormous health and social consequences. With so much at stake, and with quantities of linked information from multiple levels—from gene to environment—which past generations never had, our quantitative and conceptual tools must keep pace. The utility of long-used, familiar approaches for statistical analysis and causal inference to interpret the broad sweep of evidence on the causal determinants of human health is diminishing. Public health practitioners and researchers must understand the limitations of those methods and commit to learning what new approaches offer if they are to be reliable scientific guides for the health of future generations.

DISCLOSURE STATEMENT

LITERATURE CITED

1. Adler NE, Rehkopf DH. 2008. U.S. disparities in health: descriptions, causes, and mechanisms. *Annu. Rev. Public Health* 29:235–52

2. Baan R, Grosse Y, Lauby-Secretan B, El Ghissassi F, Bouvard V, et al. 2011. Carcinogenicity of radiofrequency electromagnetic fields. *Lancet Oncol.* 12:624–26

3. Bunge M. 2009. *Causality and Modern Science*. New Brunswick, NJ: Transaction

4. Dawid AP. 2000. Causal inference without counterfactuals. *J. Am. Stat. Assoc.* 95:407–24

5. Diez-Roux AV. 1998. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am. J. Public Health* 88:216–22

6. Diez Roux AV. 2004. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc. Sci. Med.* 58:1953–60

7. Diez Roux AV. 2011. Complex systems thinking and current impasses in health disparities research. *Am. J. Public Health* 101:1627–34

8. Evans AS. 1993. *Causation and Disease: A Chronological Journey*. New York: Plenum

9. Fisher RA. 1958. Cancer and smoking. *Nature* 182:596

10. Frieden TR, Mostashari F, Kerker BD, Miller N, Hajat A, Frankel M. 2005. Adult tobacco use levels after intensive tobacco control measures: New York City, 2002–2003. *Am. J. Public Health* 95:1016–23

11. Frost WH. 1936. Introduction. In *Snow on Cholera: Being a Reprint of Two Papers*, ed. J Snow, pp. IX-XXI. New York: Commonwealth Fund

12. Frumkin H. 2003. Healthy places: exploring the evidence. *Am. J. Public Health* 93:1451–56

13. Galea S, Riddle M, Kaplan GA. 2010. Causal thinking and complex system approaches in epidemiology. *Int. J. Epidemiol.* 39:97–106

14. Glass TA, McAtee MJ. 2006. Behavioral science at the crossroads in public health: extending horizons, envisioning the future. *Soc. Sci. Med.* 62:1650–71

15. Goldstein M, Goldstein IF. 1978. Snow on cholera. In *How Do We Know: An Exploration of the Scientific Process*, ed. M Goldstein, IF Goldstein, pp. 25–62. New York: Plenum

16. Green MD, Freedman DM, Gordis L. 2011. Reference guide on epidemiology. In *Reference Manual on Scientific Evidence*, ed. Fed. Judic. Cent., Natl. Res. Coun., pp. 549–632. Washington, DC: Natl. Acad. Press

17. Greenland S. 2000. An introduction to instrumental variables for epidemiologists. *Int. J. Epidemiol.* 29:722–29

18. Greenland S, Brumback B. 2002. An overview of relations among causal modelling methods. *Int. J. Epidemiol.* 31:1030–37

19. Greenland S, Pearl J, Robins JM. 1999. Causal diagrams for epidemiologic research. *Epidemiology* 10:37–48

20. Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, et al. 2000. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. *JAMA* 284:1290–96

21. Halloran ME, Struchiner CJ. 1995. Causal inference in infectious diseases. *Epidemiology* 6:142–51

22. Hamlin C. 1995. Could you starve to death in England in 1839? The Chadwick-Farr controversy and the loss of the "social" in public health. *Am. J. Public Health* 85:856–66

23. Hamlin CS. 2006. The history of methods of social epidemiology to 1965. In *Methods in Social Epidemiology*, ed. JM Oakes & J Kaufman, pp. 21–41. San Francisco: Wiley. 1st ed.

24. Hammond RA. 2009. Complex systems modeling for obesity research. *Prev. Chronic. Dis.* 6:A97

25. Hernán MA. 2004. A definition of causal effect for epidemiological research. *J. Epidemiol. Community Health* 58:265–71

26. Hernán MA. 2005. Invited commentary: hypothetical interventions to define causal effects—afterthought or prerequisite? *Am. J. Epidemiol.* 162:618–20; discussion 21–22

27. Hernán MA, Brumback B, Robins JM. 2001. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J. Am. Stat. Assoc.* 96:440–48

28. Hernán MA, Cole SR, Margolick J, Cohen M, Robins JM. 2005. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidem. Drug Saf.* 14:477–91

29. Hernán MA, Hernández-Diaz S, Robins JM. 2004. A structural approach to selection bias. *Epidemiology* 15:615–25

30. Hernán MA, Hernández-Diaz S, Werler MM, Mitchell AA. 2002. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am. J. Epidemiol.* 155:176–84

31. Hernán MA, Robins JM. 2006. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 17:360–72

32. Hernán MA, Taubman SL. 2008. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int. J. Obes. (Lond.)* 32(Suppl. 3):S8–14

33. Hill AB. 1965. The environment and disease: association or causation? *Proc. R. Soc. Med.* 58:295–300

34. Holland PW. 1986. Statistics and causal inference. *J. Am. Stat. Assoc.* 81:945–60

35. Huang TT, Drewnosksi A, Kumanyika S, Glass TA. 2009. A systems-oriented multilevel framework for addressing obesity in the 21st century. *Prev. Chronic Dis.* 6:A82

36. Hume D. 1896 [1739]. *A Treatise of Human Nature*, ed. LA Selby-Bigge. Oxford: Clarendon. Repr. Ed.

37. Inst. Med. (IOM). 2008. *Improving the Presumptive Disability Decision-Making Process for Veterans*, ed. JM Samet, CC Bodurow. Washington, DC: Natl. Acad. Press. 440 pp.

38. Inst. Med. (IOM), Div. Health Promot. Dis. Prev. 1994. *Veterans and Agent Orange. Health Effects of Herbicides Used in Vietnam*. Washington, DC: Natl. Acad. Press. 791 pp.

39. Int. Agency for Res. on Cancer (IARC). 2006. Preamble. In *IARC Monographs on the Evaluation of Carcinogenic Risk to Humans*, Vol. 86: *Cobalt in Hard Metals and Cobalt Sulfate, Gallium Arsenide, Indium Phosphide and Vanadium Pentoxide*. Lyon, Fr.: World Health Organ.

40. Int. Agency for Res. on Cancer (IARC). 2012. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Lyon, Fr.: IARC. **http://monographs.iarc.fr/**

41. Kaufman JS, Cooper RS. 1999. Seeking causal explanations in social epidemiology. *Am. J. Epidemiol.* 150:113–20

42. Kaufman JS, Kaufman S. 2001. Assessment of structured socioeconomic effects on health. *Epidemiology* 12:157–67

43. Koopman J. 2002. Epidemiology. Controlling smallpox. *Science* 298:1342–44

44. Koopman JS. 2002. Modeling infection transmission—the pursuit of complexities that matter. *Epidemiology* 13:622–24

45. Koopman JS. 2005. Infection transmission science and models. *Jpn. J. Infect. Dis.* 58:S3–8

46. Krieger N. 1994. Epidemiology and the web of causation: Has anyone seen the spider? *Soc. Sci. Med.* 39:887–903

47. Lipton R, Ødegaard T. 2005. Causal thinking and causal language in epidemiology: It's in the details. *Epidemiol. Perspect. Innov.* 2:8

48. Macintyre S, Ellaway A, Cummins S. 2002. Place effects on health: How can we conceptualise, operationalise and measure them? *Soc. Sci. Med.* 55:125–39

49. Michaels D. 2008. *Doubt Is Their Product: How Industry's Assault on Science Threatens Your Health*. New York: Oxford Univ. Press. xii, 372 pp.

50. Morabia A, ed. 2004. *A History of Epidemiologic Methods and Concepts*. Basel: Birkhäuser

51. Splawa-Neyman J, Dabrowska DM, Speed TP. 1990. On the application of probability theory to agricultural experiments. Essay on principles (1923). *Stat. Sci.* 5:465–572

52. Pearl J. 1995. Causal diagrams for empirical research. *Biometrika* 82:669–710

53. Phillips CV, Goodman KJ. 2004. The missed lessons of Sir Austin Bradford Hill. *Epidemiol. Perspect. Innov.* 1:3

54. Proctor RN. 2011. *Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition*. Berkeley: Univ. Calif. Press

55. Robins J. 1986. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* 7:1393–512

56. Robins JM. 1987. Errata to "a new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect" *Mathl. Modelling* 7(9–12), 1393–1512 (1986). *Comput. Math. Appl.* 14:917–21

57. Robins JM. 1993. Analytic methods for estimating HIV-treatment and cofactor effects. In *Methodological Issues of AIDS Behavioral Research*, ed. DG Ostrow, RC Kessler, pp. 213–90. New York: Plenum

58. Robins JM, Hernán MA, Brumback B. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11:550–60

59. Rockhill B. 2005. Theorizing about causes at the individual level while estimating effects at the population level: implications for prevention. *Epidemiology* 16:124–29

60. Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55

61. Rubin DB. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66:688–701

62. Rubin DB. 1978. Bayesian inference for causal effects: the role of randomization. *Ann. Stat.* 6:34–58

63. Rubin DB. 1997. Estimating causal effects from large data sets using propensity scores. *Ann. Intern. Med.* 127:757–63

64. Russell B. 1912. On the notion of cause. *Proc. Aristot. Soc.* 13:1–26

65. Samet JM, Wipfli HL. 2013. Ending the tobacco epidemic: from the genetic to the global level. In *Structural Approaches in Public Health*, ed. M Sommer, R Parker. New York: Routledge. In press

66. Susser M. 1997. Steps toward discovering causes: divergence and convergence of epidemiology and clinical medicine. *Epidemiol. Prev.* 21:160–68

67. Susser M, Stein Z. 2009. *Eras in Epidemiology: The Evolution of Ideas*. New York: Oxford Univ. Press

68. Taubman SL, Robins JM, Mittleman MA, Hernán MA. 2009. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *Int. J. Epidemiol.* 38:1599–611

69. Tchetgen EJ, VanderWeele TJ. 2012. On causal inference in the presence of interference. *Stat. Methods Med. Res.* 21:55–75

70. US Environ. Prot. Agency. 2009. *Memorandum: Process for Reviewing National Ambient Air Quality Standards*, L Jackson, May 21. Washington, DC: USEPA. **http://www.epa.gov/ttn/naaqs/pdfs/NAAQSReviewProcessMemo52109.pdf**

71. US Dep. Health Educ. Welf. (DHEW). 1964. *Smoking and Health*. Report of the Advisory Committee to the Surgeon General. Rep. DHEW Publ. No. [PHS] 1103. Washington, DC: US Gov. Print. Off.

72. US Court Fed. Claims. 2010. *Cedillo v. Secretary of Health and Human Services.* Case No. 2010–5004 (98-916V). Washington, DC: US Court Fed. Claims. **http://www.uscfc.uscourts.gov/sites/default/files/cedillo.fedcir.pdf**

73. Vandenbroucke JP. 2001. Changing images of John Snow in the history of epidemiology. *Soz. Praventivmed.* 46:288–93

74. Vandenbroucke JP, Eelkman Rooda HM, Beukers H. 1991. Who made John Snow a hero? *Am. J. Epidemiol.* 133:967–73

75. VanderWeele TJ. 2010. Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociol. Methods Res.* 38:515–44

76. Vineis P. 1991. Causality assessment in epidemiology. *Theor. Med.* 12:171–81

Contents

Environmental and Occupational Health

Public Health Practice

Social Environment and Behavior

Health Services

Indexes

Errata

An online log of corrections to *Annual Review of Public Health* articles may be found at
http://publhealth.annualreviews.org/