Annu. Rev. Public Health. 2000. 21:101–19

LOOKING BACK ON "CAUSAL THINKING IN THE HEALTH SCIENCES"

J. S. Kaufman[1,2] and C. Poole[1]
[1]*Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, North Carolina 27599-7400; e-mail: Jay_Kaufman@unc.edu*
[2]*Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27516-3997*

Key Words causation, inference, confounding, observational studies, epidemiologic methods

■ **Abstract** It has now been over a quarter of a century since the publication of Mervyn Susser's *Causal Thinking in the Health Sciences* (1973, Oxford University Press), the first book-length treatment of causal reasoning and inference in our field. Major contributions of this work were its holistic focus on the origins of health outcomes in the context of ecologic systems and its invigoration of the literature on causal criteria in epidemiology. Although a recent resurgence of interest in social context has revivified many points made by Susser, a formal basis for causal analysis consistent with this ecologic perspective has failed to emerge in public health research. Susser's discussion of causal criteria, on the other hand, helped spur a vigorous dialogue that has persisted unabated to the present day. Although the basic outline of the criteria has evolved little, their applications, interrelations, and relative contributions to causal judgments have been the subject of continued and sometimes contentious debate.

INTRODUCTION

It has now been over a quarter of a century since the publication of Mervyn Susser's *Causal Thinking in the Health Sciences* (51). This work has profoundly influenced both theory and practice for analysis and interpretation of public health data, particularly through its holistic focus on the origins of health outcomes in the context of ecologic systems and its invigoration of the literature on causal criteria in epidemiology. The focus on ecologic systems is a theme that M Susser & E Susser (59) and M Susser (58) have rearticulated and elaborated in recent years, and, as often occurs in scientific discourse, it has rebounded in popularity after a period of relative neglect (4, 28). Although this resurgence of interest in social context has revivified many of the points made by Susser in 1973, the formalization of this ecologic perspective has unfortunately advanced little in the subsequent quarter-century. The progress toward more refined and systematic articulations

of causal logic that have appeared in the epidemiologic and statistical literature in recent decades has been characterized by an explicit conceptual foundation in atomistic interventions. The emergent properties of causal systems, as distinct from the consideration of multiple discrete actions, remain largely undescribed in any formal sense in the epidemiologic literature, with the possible exception of the population dynamics of infectious disease (18).

Susser's discussion of causal criteria occupies only a brief 22 pages in the original text, but it helped spur a vigorous discussion of the use of such criteria, which has persisted unabated to the present day, including substantial refinements by Susser himself (52, 56, 57). Although the basic outline of the modern set of criteria has evolved little since formulation by a Surgeon General's Advisory Committee (60) and Hill (15), their philosophical justification, relative merits, and specific interpretation have been the subject of continued and sometimes contentious debate. Despite the relatively minor role afforded criteria for judgment in the 1973 text, they have become a central tool for the epidemiologic community in grappling with the broader issues of causal reasoning.

WHAT IS A CAUSE?

History

The notions of cause and effect are so ingrained in our everyday cognition that most researchers would find it difficult to define the words in any but a circular fashion; causes are conditions and events that produce effects, and effects are conditions and events produced by causes (43). These fundamental concepts have been the object of considerable philosophical inquiry, including exposition and debate by Hume in the eighteenth century, Mill in the nineteenth century, and Russell in the twentieth century (2). Nineteenth-century application of these notions to health outcomes ranged from the well-known contributions of Henle and Koch regarding microbial causes of disease (5) to the articulation of social causation of disease by Virchow (1). Attribution of causality in biomedical research was influenced greatly by the development of statistical inference in the early twentieth century, notably by Fisher and Neyman (47). After World War II, alongside the evolution of the randomized trial as the methodologic standard for accumulating evidence of causal attribution in clinical medicine, there evolved a literature attempting to systematize and justify causal inference from passive observation, stemming from the debate over an association between cigarette smoking and lung cancer (14, 15).

Counterfactual Foundation

Throughout the philosophical debate on causality, a consistent thread that reappeared and endured was the notion that a cause-and-effect relation could be understood in terms of similar but slightly varying versions of reality. Hume, for example, suggested a definition of "cause" based on whether the second object in

a temporal sequence of two objects would ever have existed had it not been for the prior object (23). This counterfactual definition of causality, which has achieved preeminence, defines a cause in relation to an effect as a contrast in the state of the latter between scenarios in which the former is (hypothetically) perturbed by some local intervention (31). In recent years, such potential-outcomes conceptualizations of causality have generally adopted Lewis's notion of similar alternate worlds in attempting to envision a variation of reality that is as close as possible to the real world, except for the perturbation of a single factor of interest (34).

This counterfactual definition can be formally expressed for the average causal effect in a population as the conditional probability distribution of the outcome (effect), given the hypothetical local action of setting variable X, which would distinguish between the alternate but closest worlds under consideration. That is,

$$Pr[Y = y \mid \text{Set}(X = x)] \qquad 1.$$

for discrete Y and X, where Y is the putative effect and X is the putative cause, and the hypothetical action under consideration is the physical manipulation of X (Set) to hold one of any number of specific values x. The implicit contrasts between the conditional distributions of Y in the expression above {i.e. $Pr[Y = y \mid \text{Set}(X = x_1)]$ vs $Pr[Y = y \mid \text{Set}(X = x_2)]$, etc} provide a necessary formal definition for exactly what we mean by a cause X and an average effect Y in a population (31, 32). The statistical literature has often used an equivalent notational convention, which would represent the quantity in expression 1 as simply:

$$Pr(Y_x = y) \qquad 2.$$

where Y_x is the value that Y would take under the (hypothetical) action $\text{Set}(X = x)$ and the contrast is between various regimens $\text{Set}(X = x_1)$, $\text{Set}(X = x_2)$, etc (16, 33, 45, 49). As discussed below, however, an average causal effect of zero may still allow for causality if the exposure can both instigate and prevent the outcome; an average effect of zero over the population implies either that there is no causality or that induced and prevented cases are balanced (10; 11:60–62).

The Scope of Causality in Epidemiology

The expansive sociological perspective in Susser's 1973 text sought to reach beyond the point-exposure paradigm that characterized previous applications of causal logic to epidemiologic problems, such as the relation between cigarette smoking and lung cancer. The smoking literature typically considered smoking behavior as an isolated entity, and implicitly posited the state of health that would occur in individuals or groups under alternative, potentially counterfactual states of this behavior. Susser, however, argued that this perspective was unnecessarily restrictive: "States of health do not exist in a vacuum apart from people. People form societies, and any study of the attributes of people is also a study of the manifestations of the form, the structure, and the processes of social forces" (51:6). This admonition has been characterized more recently as an objection to myopic

overemphasis on the proximate causes of disease and a concomitant neglect of the upstream or distal factors that culminate in the final product of a causal chain (28, 48).

To facilitate recognition of this broader set of causal relations, Susser defined a potential cause as ". . . any factor, whether event, characteristic, or other definable entity, so long as it brings about change for better or worse in a health condition" (51:3). Later, he rearticulated the definition of a cause as simply ". . . something that makes a difference" (57:638). These informal definitions, which Susser defended as pragmatic, were decidedly incompatible with the counterfactual definition of causality that achieved preeminence after Rubin's work in the 1970s (45). Indeed, much of the statistical literature is explicitly hostile to these more diffuse conceptualizations of causality, basing formalizations on principles of experimental design that evolved earlier in the century (16, 46, 47). Susser's defense of a broader set of potential causes was that it "enjoins a model of multiple causes" and, furthermore, that the restriction of potential causes to "active agents of change" would imply that ". . . a large part of the epidemiologic pursuit of causes [would be] lost or reduced to insignificance" (57:636).

This tension between pragmatic (sociologic) and manipulative (counterfactual) bases for causality continues to echo throughout the literature in disciplines characterized by nonexperimental research (49). Proponents of a broader set of causal factors tend to decry the restriction to manipulable quantities that follows implicitly from analogies to experimental control of variables. Proponents of the counterfactual definition of causality, on the other hand, argue that application of this less restricted definition of causality, specifically the inclusion of immutable attributes of individuals (e.g. age, sex, and race) as potential causes, leads to logical quandaries that remain unresolved (17, 29).

The price to be paid in return for the pragmatism of a broader definition of putative causes that includes "static determinants" is the lack of any formal definition of causal effect (as in expressions 1 and 2 above). When causal definitions are tied to human action, by analogy with experimental manipulation, there is no ambiguity about the casual criterion of making a difference; the difference referred to is for the outcome distributions under various, potentially hypothetical, manipulative regimens. When such manipulation is not tenable, even hypothetically, then making a difference can correspond only to contrasts between conditional distributions such as $Pr(Y = y \mid X = x_1)$ and $Pr(Y = y \mid X = x_2)$, where x_1 and x_2 are observed levels of X. This contrast provides no distinction between association through causation and association through a common antecedent cause (e.g. confounding), a philosophical objection to probabilistic causation that has a long history (36).

Although described as "pragmatic," the more inclusive definition for causal agents has no necessary relation to the results of any potential public health interventions. For example, in response to the question, "What caused the recent airline crash?" one might respond, "Gravity." Indeed, a crash clearly could not have occurred in the absence of gravity, but it exists as one of a nearly infinite

number of circumstances that are necessary for the incident to occur and yet have no practical significance because they are not subject to variation or amenable to intervention. Furthermore, the informal and broad definition of putative causes does not distinguish necessary conditions from those that Susser has referred to as causal imposters. If we observe that lung cancer tends to occur in individuals with nicotine-stained fingers, we may be able to predict the outcome event well in a data set that records lung cancer and finger stains, but we would be disappointed to discover that our intervention effort of distributing latex gloves to smokers is ineffective in preventing incident cancer. Making this sort of distinction, conceptually and practically, is clearly a central goal of Susser's 1973 book, and yet the broader definition of putative causes provides no logical foundation for such a distinction.

The scope of epidemiologic inquiry has therefore been clarified in recent decades by the formal consideration of causation as counterfactual contrast. Immutable traits such as race and sex are never causes for individuals in either the logical or pragmatic sense, but rather these are additional examples of Susser's causal imposters. To say that Black race is a cause of sickle cell anemia is to assert only that Blacks as a group have a higher prevalence of disease than Whites as a group, because no individual can plausibly be subject to alternative geographic ancestry in the same way that we can conceive, for example, of individuals smoking or not smoking cigarettes (3, 29). Likewise, to say that a specific individual would not have Down syndrome had she been conceived from a different germ cell from the one that, in actual fact, was fertilized (i.e. one without a mitotic error) is simply to assert the tautology that, were she a different person, she would be a different person. The sweep of causality is therefore contingent on the limits of human action, leading the manipulative definition of causation to be necessarily anthropocentric and fickle, depending on the current reach of technology (25).

We may, of course, fruitfully engage in hypothetical contrasts at the population level that have no interpretability at the individual level. Ratio measures of effect (e.g. risk ratio or odds ratio) may be undefined for individuals who would not have achieved the outcome in the absence of exposure, even though these measures may be interpretable in aggregate as averages across individuals (7). Likewise, an age-adjusted rate is the rate that would pertain if the index population had the age distribution of the standard population. Populations may plausibly vary in age structure, even if for individuals there is no meaningful interpretation to a hypothetical alternative birth cohort. Therefore, we may contrast, for example, the prevalence of Down syndrome in a population with the prevalence that would pertain if, counter to fact, the maternal age distribution were different. It is apparent from this line of thinking that adjustment for factors such as age, race, and sex may be sensible when seeking the unbiased causal effect estimate for an exposure, even without viewing these immutable traits as causal agents, but that the effect estimates for such trait variables have no sensible causal interpretation themselves (17).

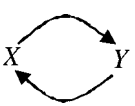
MULTIFACTORAL ETIOLOGY

The Logic of Multiple Causes

One of the major contributions of Susser's 1973 text and subsequent work is that social context is essential to understanding the complex causal pathway to achieved human health status. "The examination of an ecological complex . . . obliges us to replace simple schemes of single causes with a scheme of multiple causes" (51:42). Using various analogies, including webs (27), fractals (19), and Susser's Chinese boxes (51, 59), several authors have sought to clarify this point that no cause has an independent effect (i.e. no cause acts alone), but rather is always conditioned on contextual circumstances. For example, phenylketonuria is a cause of severe mental retardation but only occurs in susceptible individuals in the context of dietary exposure to phenylalanine (44:14). The holistic conceptualization of Susser's multilevel eco-epidemiology involves two distinct underlying phenomena. The first is merely an appreciation for the ubiquity of effect modification or synergy. The second is a recognition of the role of variable structure in the behavior of complex causal networks. Neither of these two issues is inconsistent with the contemporary formalization of causality based on counterfactual contrasts. Indeed, it is precisely the importance of context and structure that forces us to consider the counterfactual state in the closest possible world to that in which it actually occurred, save for hypothetical manipulation of the factor of interest and all changes wrought by that manipulation (23). The closest possible world is presumably the one in which all contextual factors not affected by the intervention are constrained to be identical, allowing for assessment of the effect in its natural structure and circumstances. This is precisely why many have questioned whether defining attributes of individuals can constitute valid causes, because it would be impossible to imagine having altered a factor such as sex and presume that life-long social context for the individual could remain constant.

Systems and Levels of Organization

Although the dominant formalization of causal reasoning involving counterfactual contrasts is rooted in hypothetical atomistic interventions, there is no need to consider these in isolation. Rather, several atomistic interventions can be considered as potentially co-occurring, yielding a system for which a defined intervention regime (i.e. a set of hypothetical actions) has a potentially identifiable effect (35). The difficulty is not in the conceptualization of multifactorial causal action, but rather in deciding exactly which is the salient effect of interest (39) and in the technical problems associated with unbiased effect estimation (32, 37). The effect of a putative causal factor may be unidentifiable, even in a very simple system. For example, in a system represented by the graph



3.

which Susser referred to as a symmetrical relationship (51:113), the effect of *X* on *Y* cannot be uniquely specified from observational data. If we re-express this relation longitudinally in discrete time order, however, a commensurate but now identifiable causal structure emerges. Susser referred to this new structure as a reciprocal relationship (51:68), in which the subscript is the time step at which the variable is observed,

$$X_1 \rightarrow Y_2 \rightarrow X_3 \rightarrow Y_4, \ldots, X_{t-1} \rightarrow Y_t \qquad 4.$$

In this fashion, many causal structures can be reinterpreted as directed and acyclic, allowing for the possibility of identifying the causal effect of a hypothetical intervention applied to one or more of the elements in the system, even when the functional form of the relations represented as directed arcs is left unspecified (8, 33).

Nonetheless, the logical impossibility of uniquely identifying the causal effect from passive observation of a well-defined system remains a commonly encountered problem. This is more likely to occur when some factors are latent (i.e. unmeasured) or when multifactorial causality is considered (i.e. compound hypothetical interventions). Even when the total effect of a factor may be identifiable, direct and indirect effects will often not be separately identifiable (38). For example, investigators frequently refer to the estimation of the independent contribution of one variable while "holding the others constant" through multivariate regression (e.g. 51:71). If the term "independent" here is taken to mean "direct," then this interpretation is at odds with a focus on structured relations that is wisely emphasized by Susser and others. For example, the attempt to identify a direct effect of social class on mortality by controlling for other known factors that influence mortality rates (51:50) would generally be interpretable as such only if these other known factors were not also affected in some way by social class (41). Given that social class is an upstream or distal variable that is reliably predictive of a large variety of other more proximal factors [in fact, some authors refer to social class as a fundamental or ultimate cause for exactly this reason (24)], it seems unlikely that conditioning on covariates would tend to provide a valid estimate for the direct (independent) effect of this or any other upstream quantity.

PROCEDURES FOR ESTABLISHING CAUSAL ASSOCIATIONS

Screening Causal Models for Extraneous Factors

The path to causal inference is one of a continual process of elimination. As Susser advised, "A good research design eliminates as many as possible of the extraneous variables that may intrude on the relationship we hope to establish between hypothesized cause and effect" (51:90). Maclure has expressed this same idea somewhat more succinctly: "We should always ask 'Why else?' and then try

to show 'Why not' " (26:138). The alternative explanation for causation in the case of an epidemiologic association that occurs in the real world (i.e. one that is not due to some error in observation such as selection bias or misclassification) is confounding. The nature and the detection of confounding have occupied countless pages of journals and texts. Nonetheless, using the definition of causal effect in terms of counterfactual contrasts that was described in expression 1, a succinct and unambiguous definition of confounding becomes readily apparent. The relation between X and Y is confounded by an extraneous variable when

$$Pr(Y = y \mid X = x) \neq Pr[Y = y \mid \text{Set}(X = x)] \qquad 5.$$

(e.g. 12:364). That is, there is net confounding between X and Y when the probability distribution of Y, given that X is observed to take a specific value x, is not equal to the distribution of Y, given that X is set by atomistic intervention to take the specific value x, for all observed values of x. The use of causal criteria is intended to provide hints as to whether a validly observed association can be taken to be causal, in that it represents what would occur under the hypothetical intervention represented by the Set statement in equation 5. If conditioning on some set of covariates Z would yield an equality sign in the expression in place of the inequality sign (i.e. conditioned on Z, passive observation is equivalent to atomistic intervention), then confounding by Z has been controlled (8).

The practical steps taken to ascertain that confounding is absent or has been controlled are the subject of innumerable treatments throughout the social sciences (10). The decision to condition in some way on a set of covariates to control for confounding rests on the criterion of ignorability (40, 42), which in turn is a function of the counterfactual contrasts implied by the right-hand side of equation 5. The probability distribution of Y under counterfactual states of X is unobservable, however, and thus the achievement of the ignorability (conditional independence) criterion is not a direct guide for analytic strategy decisions (33). Under the assumption of a specified set of structural relations between study variables (measured and unmeasured), it is often possible to identify a sufficient covariate set for adjustment, but the covariate structure is derived from subject matter knowledge rather than from the data and remains prone to misspecification (8). The lists of causal criteria have therefore emerged as rules of thumb to guide causal judgments—essentially as informal tests of whether alternative explanations (e.g. confounding) are likely to exist for the hypothesis of causality.

Association

Although the basic outline of criteria for making causal judgments has evolved little since the elaboration of the Surgeon General's list of five items into Hill's list of nine, these guidelines have generated a talmudic literature on their nature, logic, and application (e.g. 5, 61). Citing precedent in the sociological literature, Susser elevated three criteria to the status of absolute requirements: association, time order, and direction (57:638). Indeed, the first of these was not even discussed in the 1973

text, because it was presumed given: The criteria are provided to infer causation from association, and so association is presumably a prerequisite for even posing the question. The perspective achieved by basing causal logic on counterfactual contrasts, however, reveals that association is by no means necessary for causation, even if observed association is an essential starting point for causal judgment. For example, consider the bivariate counterfactual set defined in Greenland & Robins (9). If half of all individuals in a population would die owing to the exposure (i.e. they would live only if not exposed) and the other half of the individuals would die due to lack of exposure (i.e. they would live only if exposed) and if exposure is randomly assigned to the population, then the average causal effect is zero in expectation, there is no observed association between exposure and death in infinitely large samples, and yet the observed outcome for every single individual is due to his or her obtained exposure status (10). More general statements of this scenario have been expressed by several authors (e.g. 50:462–63). Given that many common exposures have the capacity to both cause and prevent outcomes (e.g. automobile airbags), this is neither a fanciful nor merely academic consideration.

Criteria of Judgment

In the final chapter of *Causal Thinking in the Health Sciences* (51:140–73), Susser discussed the five criteria used in a well-known Surgeon General's Advisory Committee report (60) to help judge whether the epidemiologic association between cigarette smoking and lung cancer reflected causality: time order, consistency, strength, specificity, and coherence. Susser's elaboration and expansion of this list over the ensuing years (52–54, 57) forms the most detailed and prolonged attempt to develop criteria for causality in the field of epidemiology. Several sustained themes have been evident in this work from the start.

One Strategy Among Several One often neglected theme is that the use of causal criteria formed but one of five strategies Susser originally described for coming to judgments about causality from epidemiologic results (51:73–170). The other four were to simplify the conditions of observation in study design and execution, to screen for confounders analytically, to elaborate associations analytically, and to use significance tests and power analyses to address the role of chance. Perhaps in response to the strong, widespread, and mostly favorable interest in the criteria-based strategy, Susser has devoted much of his attention to that one over the years. The main developmental themes have been to lengthen the list of criteria, to create hierarchies within the list, to show how the criteria may overlap and mutually reinforce each other, to distinguish between their uses in affirming and in refuting causality, and to grade their relative degrees of support and detraction on a semiquantitative scale.

To some extent, the criteria-based strategy has grown merely by shifting elements of causal inference from other strategies to that one. Specifically, the entire probability strategy (i.e. significance testing and power analysis) became a causal

criterion in 1986 (54), and the most general consideration in the strategy of simplifying conditions of observation (i.e. a hierarchical classification of basic structures of study design) was moved to one part of a bifurcated consistency criterion in 1991 (57). Nevertheless, the many important considerations of study design and implementation that remain, as well as the strategies of screening for confounders and elaborating associations in their entirety, survive as crucial elements of Susser's approach to causal inference that are not (yet) subsumed by the list of causal criteria. This vital feature of Susser's formulation, that there is more to causal inference than the application of causal criteria, deserves wider recognition.

Subjectivity of Judgment A second theme that has permeated the literature on causal criteria is the subjective nature of the judgments the criteria serve. Susser stated from the outset that judgments about causality "are reached by weighing the available evidence; there are no absolute rules, and different workers often come to conflicting conclusions" (51:140–41), and he has seldom, if ever, failed to repeat this observation in subsequent writings. Attitudes toward this inherent subjectivity, however, have been hard to discern. Is it a good thing to be encouraged or a bad thing to be minimized without hope of eliminating it entirely? In his first extended case study, Susser described how Pearson favored the criteria of probability and strength although his antagonist, Wright, favored the criteria of consistency and coherence in their debate over the efficacy of typhoid inoculation (52:3–9). Although Wright ultimately turned out to be correct, Susser concluded neither that consistency and coherence are more important than probability and strength nor that the preferential emphasis of different criteria by the two disputants was a regrettable feature of their debate.

At the time Susser presented that case study, there were only a few causal criteria on his list, and no attempt had yet been made to rank or quantify them. Also at that time, Kuhn was explaining his own unworried reaction to disagreement among scientists in other fields on the relative merits of competing theories, despite the scientists' complete agreement on the criteria that make one theory preferable to another (22). Kuhn considered five standard criteria for evaluating the adequacy of a theory, on which he believed nearly all scientists placed positive valuation: accuracy, consistency, scope, simplicity, and fruitfulness. He described, with historical reference to choices between geocentric and heliocentric theories in astronomy and between the phlogiston and oxygen theories of combustion in chemistry, how competing theories can trade superiority with regard to a given criterion (e.g. accuracy) in different areas of applicability and how one theory can be superior to another on one criterion yet inferior on another. His conclusion about science in general seems consonant with conclusions that might have been reached by readers of Susser's early writings on causal inference in epidemiology:

> When scientists must choose between competing theories, two men fully committed to the same list of criteria for choice may nevertheless reach different conclusions. Perhaps they interpret simplicity differently or have

> different convictions about the range of fields within which the consistency criterion must be met. Or perhaps they agree about these matters but differ about the relative weights to be accorded to these or to other criteria when several are deployed together. With respect to divergences of this sort, no set of choice criteria yet proposed is of any use. One can explain . . . why particular men made particular choices at particular times. But for that purpose one must go beyond the list of shared criteria to characteristics of the individuals who make the choice. One must . . . deal with characteristics which vary from one scientist to another without thereby in the least jeopardizing their adherence to the canons that make science scientific (22:324).

To Kuhn, the idiosyncratic factors dependent on individual biography and personality that inevitably inject the subjective component into theory choice (which, in the context of causal inference in epidemiology, we might idealize as the choice between causal and null hypotheses) are far from "eliminable imperfections in [the] rules of choice." To the contrary, they are "responses to the essential nature of science" (22:330). Specifically, Kuhn saw that criteria for theory choice in science overall functioned much as he almost certainly would have seen causal criteria functioning in epidemiology: more as "values" than as "rules." Weed and colleagues, diligent sociologic observers of causal inference practices in epidemiology, have ably documented the multitudinously idiosyncratic ways in which practicing epidemiologists pick and choose among causal criteria and interpret and weight them differently (62–65).

Kuhn saw a number of striking advantages to the recognition that criteria of choice can function as values when incomplete as rules. One obvious advantage is that this recognition "accounts in detail for aspects of scientific behavior which the tradition [in the philosophy of science] has seen as anomalous or even irrational" (22:331). An even more important advantage to Kuhn, however, is that viewing criteria for scientific judgment or choice as values and not as rules "allows the standard criteria to function fully in the earliest stages of theory choice, the period when they are most needed but when, on the traditional view, they function badly or not at all" (22:331). One cannot help but draw a parallel to the special role that causal criteria are deemed to play in epidemiology at the earliest stages of research on a hypothetically causal association.

The fact that most newly suggested theories do not survive, which is certainly true about most newly suggested epidemiologic hypotheses about causes and preventives of disease, meant to Kuhn that a period of uncertainty and debate is required in which some scientists work under the old theory while others explore the possibilities of the new one.

> Such a mode of development *requires* a decision process which permits rational men to disagree, and such disagreement would be barred by the shared algorithm which philosophers have generally sought. If it were at

> hand, all conforming scientists would make the same decision at the same time. . . . I doubt that science would survive the change. What from one viewpoint may seem the looseness and imperfection of choice criteria conceived as rules may, when the same criteria are seen as values, appear an indispensable means of spreading the risk which the introduction or support of novelty always entails (22:332).

What other than idiosyncrasies of individual biography and personality could account for the suspension of disbelief that has kept epidemiologic research on electric and magnetic fields and childhood cancers alive for two decades despite its violent clash with the criterion of biologic coherence. What other than characteristics that vary from one scientist to another could explain the almost giddy manner in which a kaleidoscopically varying array of hypothetically beneficial nutrients is promoted by its epidemiologic patrons long before they have had a chance to be confronted with the consistency criterion? The kinds of idiosyncrasies that cause individual epidemiologists to use their differing algorithms are easy to catalog. They include sociopolitical ideology, religious conviction, pride of discovery, desire for career advancement, reluctance to change stands once taken, undue reliance on one's own results, and, of course, personal financial gain.

Kuhn clearly viewed the subjective nature of criteria for theory choice, or the function of such criteria, as values rather than as rules, as a good and essential feature of science in general. It is clear in the specific context of causal inference in epidemiology that many epidemiologists share Kuhn's view of the inevitability of the subjective nature of the enterprise. But do we consider this subjectivity not only unavoidable, but advantageous, as Kuhn did? Here the answer is not so clear, especially as systems of causal criteria come to be elaborated in ways that make them more encompassing, detailed, and quantitative. Susser, for example, once called causal criteria "rules," but he quickly disclaimed this appellation in favor of "guidelines" (53). As his own system of criteria becomes increasingly elaborate, however, it increasingly resembles a book of rules.

Some years after the publication of Kuhn's *The Structure of Scientific Revolutions* (20), Feyerabend complained of what he called an "ambiguity of presentation."

> [A]re we here presented with *methodological prescriptions* which tell the scientist how to proceed; or are we given a *description*, void of any evaluative element, of those activities which are generally called 'scientific'? Kuhn's writings . . . are *ambiguous* in the sense that they are compatible with, and lend support to, both interpretations. Now this ambiguity . . . is not at all a side issue. . . . More than one social scientist has pointed out to me that now at last he has learned how to turn his field into a 'science'—by which of course he meant that he had learned how to *improve* it. The recipe . . . is to restrict criticism, to reduce the number of comprehensive theories to one, and to create a normal science that has this one theory as its paradigm.

> Students must be prevented from speculating along different lines and the more restless colleagues must be made to conform and 'to do serious work.' *Is this what Kuhn wants to achieve?* (6:198)

Kuhn's reply began with an appeal to the personal origins of his theory of science:

> [U]nlike most philosophers of science, I began as an historian of science, examining closely the facts of scientific life. Having discovered in that process that much scientific behaviour . . . persistently violated accepted methodological canons, I had to ask why those failures to conform did not seem at all to inhibit the success of the enterprise. When I later discovered that an altered view of the nature of science transformed what had previously seemed aberrant behaviour into an essential part of an explanation for science's success, the discovery was a source of confidence in that new explanation. My criterion for emphasizing any particular aspect of scientific behaviour is therefore not simply that it occurs, nor merely that it occurs frequently, but rather that it fits a theory of scientific knowledge. . . . That my criteria for discriminating between the essential and non-essential elements of observed scientific behaviour are to a significant extent theoretical provides also an answer to what Feyerabend calls the ambiguity of my presentation. Are Kuhn's remarks . . . he asks, to be read as descriptions or prescriptions? The answer, of course, is that they should be read in both ways at once. If I have a theory of how and why science works, it must necessarily have implications for the way in which scientists should behave if their enterprise is to flourish (21:236–37).

Readers of the earliest writings on causal criteria by Susser (51) and others (15, 60) may be forgiven for sensing the same ambiguity Feyerabend understandably saw in Kuhn, who had not made it clear that he believed he had discerned patterns for success in the actual behavior of scientists. But as Susser's system of causal criteria in particular has become more elaborate, it has become increasingly clear that it is a prescriptive system, not a descriptive one. When he wrote in 1973, "Specificity enhances the plausibility of causal inference, but lack of specificity does not negate it" (51:153), he could have been a sociologist, describing how most epidemiologists interpret specificity and nonspecificity, or a normative philosopher, advising epidemiologists on how they should interpret specificity and its opposite. But in his 1977 case study of the debate on smoking and lung cancer (52), Susser made quite clear his view that Berkson was wrong, not merely in a minority, when he invoked lack of specificity in the apparent effects of a cause as evidence against causality. Here the intent is unmistakably normative.

How, in light of a criterion that nonlinear or nonmonotonic dose-response curves neither affirm causality nor detract from it (54–57), are we to interpret observations of large numbers of epidemiologists who, themselves, interpret such curves as evidence against causality? Do these observations count as evidence against the

validity of a sociologic claim about the behavior of epidemiologists? Or do they count as evidence that many epidemiologists are either unaware of this tenet or are aware of it but are willfully disobedient? The more detailed and quantitative a system of causal criteria becomes, the more forced we become to draw the latter conclusion.

In 1986, Susser wrote, "Specificity in the causes of an effect is persuasive; specificity in the effects of a cause is much less so" (53, 54). The distinction was reflected in his semiquantitative system by a plus-or-minus sign (±) for the degree of support conferred by specificity of cause and two plus signs (++) for specificity of effect. The following year, however, he wrote, "Specificity in the causes of a given effect is persuasive; specificity in the effects of a given cause usually less so," and the difference in degree of affirmation provided by the two kinds of specificity was now too small to be measured by the semiquantitative system; both received a single plus sign (+) (55). What accounted for this shift in just a few months? Was it a more comprehensive or thorough survey, of the kind Weed might undertake, of the interpretations actually offered by epidemiologists? No evidence is at hand to support this hypothesis. More likely it was a shift in the personal, subjective algorithm of Susser himself. Perhaps a future development might be to refine the plus sign metric to reflect the smaller difference in the degree of affirmation he now derives from the two kinds of specificity. The system as a whole seems on its way to becoming a proposal for the kind of "shared algorithm which philosophers have generally sought," which Kuhn decried not only as impossible but as counterproductive to progress in science.

An Antidote to Skepticism A final theme evident throughout Susser's writings on causal criteria reveals quite explicitly one very strong element in his own subjective makeup. It is the view that epidemiologists' judgments are too predisposed in favor of the null hypothesis and against causality. In the 1973 book, he cast the strategy of using criteria of judgment as a kind of causality-friendly antidote against "a bias toward skepticism in most of the strategies so far outlined" (51:141). In many of his more fully developed case studies—typhoid vaccine, poliomyelitis vaccine, the effect of smoking on lung cancer, the effect of social milieu on intellectual development—causality turns out in retrospect to be the right answer. Historical examples of precipitous judgments in favor of causality and stories in which skepticism about causality turn out to be warranted are not prominently featured among his case studies.

Susser has defended epidemiology against a charge that it does not adhere to "the scientific method." He has even dared to dispute the claim that epidemiological studies have demonstrated an association, but do not and cannot demonstrate causality with ". . . a radical counter-assertion. Epidemiology provides a unique instrument for establishing environmental causes of disease in human beings. Indeed, I shall go so far as to say it is the sole decisive instrument" (56:65). Thus, Susser's still-evolving system of causal criteria is not a sociologist-of-epidemiology's description to epidemiologists of their own collective behavior. Instead, it constitutes

the personal algorithm for incorporating subjective judgment into causal inference of one who has been among our field's leading figures for much of this century and one who believes that causal relationships should be established more frequently and quickly than they have been in the past. Its influence has undoubtedly been and will continue to be toward increasing the frequency and certainty of judgments in favor of causality and toward decreasing the frequency and certainty of judgments in acceptance of null hypotheses.

Conflicts Among Criteria Finally, it may be worthwhile to note a theme that has not been addressed by discerning areas of overlap and mutual reinforcement among causal criteria [e.g. that increasing the specificity of a cause or effect may result in an increase in strength of association (53)]. This neglected theme is the possibility for causality to be affirmed by the violation of a causal criterion or for the fulfillment of a criterion to detract from causality. Nonspecificity in the effects of a cause, for instance, can be and has been used as an argument in favor of causality. Two examples may serve to illustrate. One comes from Susser's case study of the specificity of maternal exposure to polychlorinated biphenyls as a cause of cola coloring in newborn babies (53). He argues that this specificity of cause "does strengthen causal inference, especially when those mothers exhibit such other signs of [polychlorinated biphenyl] exposure as chloracne." Thus, in this example at least, specificity of cause is strengthened by nonspecificity of effect. As another example, we conjecture that the nonspecificity of cigarette smoking's accepted causative effect on cancers at several anatomic sites weighs in on the affirmative side of the subjective judgments of many epidemiologists when they are considering the less plentiful and more equivocal evidence of effects on additional cancer sites, such as the stomach (30).

An example in which fulfillment of a causal criterion provides evidence against a causal interpretation is offered by a recent study of alcohol consumption and mortality (13). The investigators, noting that the dose-response curve for this association "is usually reported as 'U-' or 'J-shaped,' with moderate consumers having the lowest risk of mortality and high consumers the highest," interpret their finding of a graded association as evidence against the hypothesis that alcohol consumption reduces mortality for light and moderate drinkers. Thus, when existing evidence creates an expectation of a nonlinear, nonmonotonic dose-response curve, the observation of a linear or monotonic curve can count as evidence against causality or its mirror image, prevention (cf 66). Thorough elaboration of a comprehensive system of causal criteria would include an erstwhile search for other examples in which causal criteria may cut against the usual grain or perhaps even conflict with each other.

CONCLUSION

Mervyn Susser's *Causal Thinking in the Health Sciences* (51) was the first book-length treatment of causal reasoning and casual inference in our field. It pays

substantial dividends on reading and rereading after the many years since its publication. On the topic of causal reasoning, two of the most important developments since the book was published have been the ascension of formal treatments of causation based on counterfactual theory and a concomitant rise in interest in complex systems of causes and effects at levels of organization ranging from the molecular to the societal. Susser has contributed to these developments by criticizing the contention that some static conditions such as race and gender cannot be considered causes and by describing an "eco-epidemiology" with its metaphor of Chinese boxes to depict a multilevel, dynamic-systems approach to guide the next era of epidemiologic research.

On the topic of causal inference, there have been fewer noteworthy developments over the years. Many in epidemiology and other branches of public health continue to rely on the very similar lists of causal criteria that were laid down almost simultaneously by Hill (15), by a Surgeon General's advisory committee (60), and by Susser (51) nearly three decades ago. In this area, Susser has worked essentially alone to lengthen the list of criteria for judging causality, to arrange the criteria into hierarchical categories, to distinguish their roles in affirming and refuting causality, to explore their interrelations, and to begin to quantify their contributions to causal judgments. As his system of causal criteria becomes more elaborate, however, it has raised questions pertaining to Kuhn's distinction between the function of scientific criteria as values or as rules.

In summary, we can see after nearly three decades have passed since the appearance of *Causal Thinking in the Health Sciences* (51) that no final resolution has emerged to the challenges of thinking about causes and their effects or to the formidable task of forming causal judgments about relations between variables. That seminal book propelled a vital discussion of these topics, and its author continues to participate vigorously in the development of these ideas. Our field will be further enriched if he continues to do so for many years to come.

LITERATURE CITED

1. Ackerknecht EH. 1953. *Rudolf Virchow: Doctor, Statesman, Anthropologist*. Madison, WI: Univ. Wisc. Press. 304 pp.
2. Bunge MA. 1979. *Causality in Modern Science*. New York: Dover. 434 pp. 3rd ed.
3. Cooper RS, Kaufman JS. 1998. Race and hypertension: science and nescience. *Hypertension* 32:813–16
4. Diez-Roux AV. 1999. On genes, individuals, society and epidemiology. *Am. J. Epidemiol.* 148:1027–32
5. Evans AS. 1976. Causation and disease: the Henle-Koch postulates revisited. *Yale J. Biol. Med.* 50:175–95
6. Feyerabend P. 1970. Consolations for the specialist. In *Criticism and the Growth of Knowledge*, ed. I Lakatos, A Musgrave, pp. 197–230. Cambridge, UK: Cambridge Univ. Press
7. Greenland S. 1987. Interpretation and choice of effect measures in epidemiologic analysis. *Am. J. Epidemiol.* 125:761–68

8. Greenland S, Pearl J, Robins JM. 1999. Causal diagrams for epidemiologic research. *Epidemiology* 10:38–48
9. Greenland S, Robins JM. 1986. Identifiability, exchangeability, and epidemiological confounding. *Int. J. Epidemiol.* 15:433–39
10. Greenland S, Robins JM, Pearl J. 1999. Confounding and collapsibility in causal inference. *Stat. Sci.* 14:29–47
11. Greenland S, Rothman KJ. 1998. Measures of effect and measures of association. In *Modern Epidemiology*, ed. KJ Rothman, S Greenland, pp. 48–66. Philadelphia, PA: Lippincott-Raven. 2nd ed.
12. Greenland S, Rothman KJ. 1998. Introduction to regression models. See Ref. 11, pp. 359–432
13. Hart CL, Davey Smith G, Hole DJ, Hawthorne VM. 1999. Alcohol consumption and mortality from all causes, coronary heart disease, and stroke: results from a prospective cohort study of Scottish men with 21 years of follow-up. *Br. Med. J.* 318:725–29
14. Hill AB. 1953. Observation and experiment. *N. Engl. J. Med.* 248:995–1001
15. Hill AB. 1965. The environment and disease: association or causation? *Proc. R. Soc. Med.* 58:295–300
16. Holland PW. 1986. Statistics and causal inference. *J. Am. Stat. Assoc.* 81:945–60
17. Kaufman JS, Cooper RS. 1999. Seeking causal explanations in social epidemiology. *Am. J. Epidemiol.* 150:113–20
18. Koopman JS, Lynch JW. 1999. Individual causal models and population system models in epidemiology. *Am. J. Public Health* 89:1170–74
19. Krieger N. 1994. Epidemiology and web of causation: Has anyone seen the spider? *Soc. Sci. Med.* 43:887–903
20. Kuhn TS. 1962. *The Structure of Scientific Revolutions.* Chicago: Univ. Chicago Press. 172 pp.
21. Kuhn TS. 1970. Reflections on my critics. In *Criticism and the Growth of Knowledge*, ed. I Lakatos, A Musgrave, Cambridge, UK: Cambridge Univ. Press
22. Kuhn TS. 1977. Objectivity, value judgment, and theory choice. In *The Essential Tension*, ed. TS Kuhn, pp. 320–43. Chicago: Univ. Chicago Press
23. Lewis D. 1973. Causation. *J. Philos.* 70:556–67
24. Link BG, Phelan J. 1995. Social conditions as fundamental causes of disease. *J. Health Soc. Behav.* 38 (Suppl.):80–94
25. Mackie JL. 1980. *The Cement of the Universe: A Study of Causation.* Oxford, UK: Clarendon. 329 pp.
26. Maclure M. 1988. Refutation in epidemiology: why else not? In *Casual Inference*, ed. KJ Rothman, pp. 131–38. Chestnut Hill, MA: Epidemiology Resour.
27. MacMahon B, Pugh TF, Ipsen J. 1960. *Epidemiologic Methods.* Boston, MA: Little, Brown. 302 pp.
28. McMichael AJ. 1999. Prisoners of the proximate: loosening the constraints on epidemiology in an age of change. *Am. J. Epidemiol.* 150:887–97
29. Morgenstern H. 1997. Defining and explaining race effects. *Epidemiology* 8:609–10
30. Nomura A. 1996. Stomach cancer. In *Cancer Epidemiology and Prevention*, ed. D Schottenfeld, JF Fraumeni Jr, pp. 707–24. New York, NY: Oxford Univ. Press. 2nd ed.
31. Pearl J. 1995. Causation, action, and counterfactuals. In *Computational Learning and Probabilistic Reasoning*, ed. A Gammerman, pp. 235–55. New York: Wiley & Sons
32. Pearl J. 1997. On the identification of nonparametric structural models. In *Latent Variable Modeling with Applications to Causality*, ed. M Berkane, pp. 29–68. New York: Springer-Verlag
33. Pearl J. 1998. Graphs, causality, and structural equations models. *Soc. Methods Res.* 27:226–84
34. Pearl J. 1999. Reasoning with cause and

effect. *Proc. Int. Joint Conf. Artif. Intell.*, pp. 1438–50. San Francisco: Morgan Kaufmann
35. Pearl J, Robins JM. 1995. Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Uncertainty in Artificial Intelligence: Proc. Conf. Artif. Intell., 11th*, ed. P Besnard, S Hanks, pp. 454–53. San Francisco: Morgan-Kaufmann
36. Reichenbach H. 1956. *The Direction of Time.* Berkeley, CA: Univ. Calif. Press. 280 pp.
37. Robins JM. 1997. Causal inference from complex longitudinal data. In *Latent Variable Modeling with Applications to Causality*, ed. M Berkane, pp. 69–117. New York: Springer-Verlag
38. Robins JM, Greenland S. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3:143–155
39. Robins JM, Greenland S, Hu F-C. 1999. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *J. Am. Stat. Assoc.* 94:687–712
40. Rosenbaum PR. 1984. From association to causation in observational studies: the role of tests of strongly ignorable treatment assignment. *J. Am. Stat. Assoc.* 79:43–48
41. Rosenbaum PR. 1984. The consequences of adjustment for a concommitant variable that has been affected by treatment. *J. R. Stat. Soc. A, General* 148:656–66
42. Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrica* 70:43–55
43. Rothman KJ. 1976. Causes. *Am. J. Epidemiol.* 104:587–92
44. Rothman KJ, Greenland S. 1998. Causation and causal inference. In *Modern Epidemiology*, ed. KJ Rothman, S Greenland, pp. 7–28. Philadelphia, PA: Lippincott-Raven. 2nd ed.
45. Rubin DB. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66:688–701
46. Rubin DB. 1978. Bayesian inference for causal effects: the role of randomization. *Ann. Stat.* 6:34–58
47. Rubin DB. 1990. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Stat. Sci.* 5:482–80
48. Shy CM. 1996. The failure of academic epidemiology: witness for the prosecution. *Am. J. Epidemiol.* 145:489–84
49. Sobel ME. 1995. Causal inference in the social and behavioral sciences. In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, ed. G Arminger, C Clogg, ME Sobel, pp. 1–38. New York: Plenum
50. Stone R. 1993. The assumptions on which causal inferences rest. *J. R. Stat. Soc. B* 55:455–66
51. Susser M. 1973. *Causal Thinking in the Health Sciences: Concepts and Strategies of Epidemiology.* New York: Oxford Univ. Press. 181 pp.
52. Susser M. 1977. Judgement and causal inference: criteria in epidemiologic studies. *Am. J. Epidemiol.* 105:1–15
53. Susser M. 1986. Rules of inference in epidemiology. *Regul. Toxicol. Pharmacol.* 6:116–28
54. Susser M. 1986. The logic of Sir Karl Popper and the practice of epidemiology. *Am. J. Epidemiol.* 124:711–18
55. Susser M. 1987. Falsification, verification and causal inference in epidemiology: reconsideration in the light of Sir Karl Popper's philosophy. In *Epidemiology, Health and Society: Selected Papers*, ed. M Susser, pp. 33–58. New York: Oxford Univ. Press
56. Susser M. 1987. *Epidemiology, Health and Society: Selected Papers.* New York: Oxford Univ. Press. 309 pp.
57. Susser M. 1991. What is a cause and how do we know one? A grammar for pragmatic epidemiology. *Am. J. Epidemiol.* 133:635–48

58. Susser M. 1998. Does risk factor epidemiology put epidemiology at risk? Peering into the future. *J. Epidemiol. Community Health* 52:608–11
59. Susser M, Susser E. 1996. Choosing a future for epidemiology. II. From black box to Chinese boxes and eco-epidemiology. *Am. J. Public Health* 86:674–77
60. United States Department of Health, Education and Welfare. 1964. *Smoking and Health: Rep. Advis. Comm. Surg. Gen.* Washington, DC: US Public Health Serv. 387 pp.
61. Weed DL. 1986. On the logic of causal inference. *Am. J. Epidemiol.* 123:965–79
62. Weed DL. 1994. Alcohol, breast cancer, and causal inference: where ethics meets epidemiology. *Contemp. Drug Probl.* 21:185–204
63. Weed DL. 1997. On the use of causal criteria. *Int. J. Epidemiol.* 26:1138–43
64. Weed DL, Gorelic LS. 1996. The practice of causal inference in cancer epidemiology. *Cancer Epidemiol. Biomarkers Prev.* 5:303–11
65. Weed DL, Hursting SD. 1998. Biologic plausibility in causal inference: current method and practice. *Am. J. Epidemiol.* 148:435–25
66. Weiss NS. 1981. Inferring casual relationships: elaboration of the criterion of "dose-response." *Am. J. Epidemiol.* 113:487–90

Annual Review of Public Health
Volume 21, 2000

CONTENTS