 International Journal of Epidemiology 2002;**31**:422–429

THEORY AND METHODS

Estimating causal effects

George Maldonado[a] and Sander Greenland[b]

Although one goal of aetiologic epidemiology is to estimate 'the true effect' of an exposure on disease occurrence, epidemiologists usually do not precisely specify what 'true effect' they want to estimate. We describe how the counterfactual theory of causation, originally developed in philosophy and statistics, can be adapted to epidemiological studies to provide precise answers to the questions 'What is a cause?', 'How should we measure effects?' and 'What effect measure should epidemiologists estimate in aetiologic studies?' We also show that the theory of counterfactuals (1) provides a general framework for designing and analysing aetiologic studies; (2) shows that we must always depend on a substitution step when estimating effects, and therefore the validity of our estimate will always depend on the validity of the substitution; (3) leads to precise definitions of effect measure, confounding, confounder, and effect-measure modification; and (4) shows why effect measures should be expected to vary across populations whenever the distribution of causal factors varies across the populations.

Introduction

Imagine that the creator of the universe appears to you in a dream and grants you the answer to one public-health question. The conversation might go as follows:

You: What is the true effect of (your exposure here, denoted by E) on the occurrence of (your disease here, denoted by D)?

Creator: What do you mean by 'the true effect'? The true value of what parameter?

You: The true relative risk.

Creator: Epidemiologists use the term relative risk for several different parameters. Which do you mean?

You: The ratio of average risk with and without exposure—what some call the risk ratio[1] and others call the incidence proportion ratio.[2]

Creator: Which incidence proportion ratio?

You: Pardon?

Creator: Do you want a ratio of average disease risk in two different groups of people with different exposure levels?

You: Yes.

Creator: So you want a descriptive incidence proportion ratio?

You: No, not descriptive. Causal. An incidence proportion ratio that isolates the effect of E on D from all other causal factors.

Creator: By 'isolate', you mean a measure that applies to a single population under different possible exposure scenarios?

You: Yes, that's what I mean.

Creator: OK. Which causal incidence proportion ratio?

You: Pardon?

Creator: For what population, and for what time period? The true value of a causal incidence proportion ratio can be different for different groups of people and for different time periods. It's not necessarily a biological constant, you know.

You: Yes, of course. For population (your population here, denoted by P) between the years (your study time period here, denoted by t_0 to t_1).

Creator: By population P, do you mean: (1) everyone in population P, or (2) the people in population P who have a specific set of characteristics?

You: Pardon?

Creator: As I just said, the true value of a causal incidence proportion ratio is not necessarily a biological constant. It can be different for subgroups of a population.

You: Of course. Everyone in population P.

Creator: OK. Comparing what two exposure levels?

You: Exposed and unexposed.

Creator: What do you mean by exposed and unexposed? Exposed for how long, to how much, and during what time period? There are many different ways you could define exposed and unexposed, and each of the corresponding possible ratios can have a different true value, you know.

You: Of course. Ever exposed to any amount of E versus never exposed to E.

Creator: The incidence proportion ratio for the causal effect on D of ever E compared to never E in population P during the study time period t_0 to t_1 is (your causal incidence-proportion-ratio parameter value here).

The point of the above is that, while one goal of etiologic epidemiology is to estimate 'the true effect' of an exposure on disease frequency, we usually do not precisely specify what 'true effect' we want to estimate. We may not be able to do so. For example, before reading this paper would you have required

[a] University of Minnesota School of Public Health, Mayo Mail Code 807, 420 Delaware St. SE, Minneapolis, MN 55455–0392, USA. E-mail: GMPhD@umn.edu

[b] Department of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90095–1772, USA.

less prompting than in the dialog above? How many published papers explicitly state what the authors mean by 'true' relative risk or odds ratio, or whether the estimated measure of association is intended to have a descriptive or causal interpretation? How many papers explicitly define the population or time period of interest? How many etiologic papers overemphasize results that cannot be given a causal interpretation, such as significance tests, *P*-values, correlation coefficients, or proportion of variance 'explained'?

In this paper we discuss the questions 'What is a cause?', 'How should we measure effects?' and 'What effect measure should epidemiologists estimate in etiologic studies?' We begin by adapting the counterfactual approach to causation, originally developed in philosophy and in statistics,[3,4] to epidemiological studies. In the process, we give precise answers to these questions, and we describe how these answers have important implications for etiologic research: (1) Under the counterfactual approach, the measure we term a 'causal contrast' is the only meaningful effect measure for etiologic studies. (2) The counterfactual approach provides a general framework for designing and analysing epidemiological studies. (3) The counterfactual definition of *causal effect* shows why direct measurement of an effect size is impossible: We must always depend on a substitution step when estimating effects, and the validity of our estimate will thus always depend on the validity of the substitution.[3,5–7] (4) The counterfactual approach makes clear that a critical step in study interpretation is the formal quantification of bias in study results. (5) The counterfactual approach leads to precise definitions of effect measure, confounding, confounder, and to precise criteria for effect-measure modification.

In the discussion that follows, we assume that the study outcome is a disease (e.g. lung cancer); this discussion can be readily extended to any outcome (e.g. a health behaviour such as cigarette smoking). We also assume for simplicity that disease occurrence is deterministic; under a stochastic model, the quantities we discuss are probabilities or expected values.[6,7]

The counterfactual approach

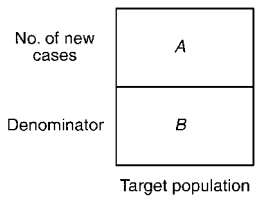
History

In 1748, the renowned Scottish philosopher David Hume wrote 'we may define a cause to be an object followed by another … where, if the first object had not been, the second never had existed.'[3,8] A key innovation of this definition was that it pivoted on a clause of the form 'if C had not occurred, D would not have either', where C and D are what actually occurred. Such a clause, which hypothesizes what would have happened under conditions contrary to actual conditions, is called a *counterfactual conditional*. Despite its early appearance, this counterfactual concept of causation received no formal basis until 1923 when the statistician Jerzy Neyman presented a quantitative conceptual model for causal analysis.[9] This model was originally known as the 'randomization model'[10] and was later called the 'potential-outcomes model' (or, inaccurately, 'Rubin's model') when extended to observational studies.[11] The model has since been widely (though not universally[12]) adopted by statisticians and others seeking a logical foundation for statistical analysis of causation.[4,10,13–16] These developments were paralleled by more extensive analysis of counterfactual reasoning by philosophers.[17–20] A comprehensive review of causality theory is provided by Pearl,[15] who shows how structural-equation models and graphical causal models (causal diagrams) translate directly to counterfactual models, shedding light on all three approaches. A brief review of these connections is given by Greenland,[21] and Greenland *et al*.[22] provide a more extensive review of graphical causal modelling for epidemiological research.

Target

We will use the term *target population* for the group of people about which our scientific or public-health question asks, and therefore for which we want to estimate the causal effect of an exposure. The target population could be composed of one group of people (as in most epidemiological studies), several groups of people (as in an intervention study in several communities), or one person. For simplicity, in the rest of this discussion we assume that the target population is one group of people.

Let the *etiologic time period* be the time period about which our scientific or public-health question asks. The beginning and end of this time period is specified by the study question. For example, in a study of the effectiveness of a back-injury prevention programme in a workplace, the etiologic time period could be any time period after the implementation of the programme. Note that this period may vary among individuals (e.g. the etiologic time period for a study of weight gain during pregnancy and pre-eclampsia spans only pregnancy), and not all of the period need be time at risk. For example, the etiologic time period for a study of intrauterine diethylstilbesterol exposure and subsequent fertility problems could include childhood, a time at no risk of such problems but during which etiologically relevant events (e.g. puberty) occur.



Let A be the number of new cases of the study disease in the target population during the etiologic time period. Let B be the denominator for computing disease frequency in the target population during the etiologic time period. If B is the number of people at risk at the beginning of the period and all individuals are followed throughout the etiologic time period, the disease-frequency parameter

$$R = \frac{A}{B}$$

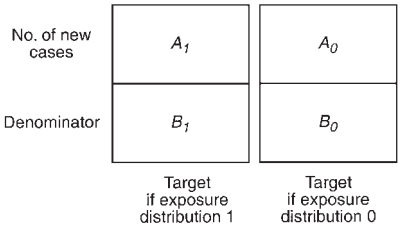
is the proportion getting disease over the period (incidence proportion, average risk). If B is the amount of person-time at risk during the period, R is the person-time incidence rate. If B is the number of people who do not get disease by the end of the period, R is the incidence odds.

Let *target* refer to the target population during the etiologic time period.

Causal effect

Consider one target population during one etiologic time period, but under two different exposure distributions, as illustrated

below. Let the subscript 1 denote one exposure distribution, and let the subscript 0 denote the other. These distributions represent different possible mixtures of individual exposure conditions. With, say, smoking as the exposure, distribution 0 could represent conditions under which 20% of the target population regularly smoked cigarettes during a given time period, whereas distribution 1 could represent conditions under which 40% (instead of 20%) of the population did so during that period.



Then $R_1 = A_1/B_1$ is disease frequency if the target population had experienced exposure distribution 1, and $R_0 = A_0/B_0$ is disease frequency if the same group of people during the same time period had instead experienced exposure distribution 0.

Let a *causal contrast* be a contrast between R_1 and R_0. For example, we define the ratio causal contrast as

$$RR_{causal} = \frac{R_1}{R_0} = \frac{A_1/B_1}{A_0/B_0}$$

where we allow *RR* to denote a risk ratio, rate ratio, or odds ratio. Similarly, we define the difference causal contrast as

$$RD_{causal} = R_1 - R_0.$$

Synonyms for causal contrast are *effect measure* and *causal parameter*.

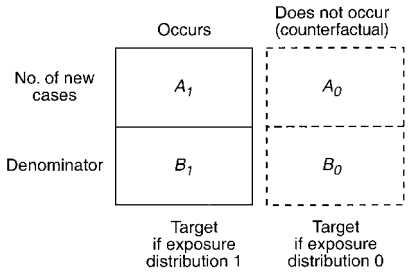
A causal contrast compares disease frequency under *two* exposure distributions, but in *one* target population during *one* etiologic time period. This type of contrast has two important consequences. First, the only possible reason for a difference between R_1 and R_0 is the exposure difference. A causal contrast, therefore, measures the *causal effect* of the difference between exposure distributions 1 and 0 in the target population during the etiologic time period.[2,4–7,23]

Second, a causal contrast cannot be observed directly, as we explain below, and therefore a different type of measure must be used as a substitute for it.

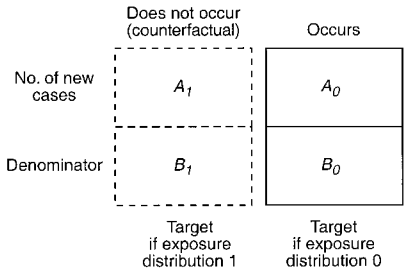
Counterfactuals

Why is it not possible to directly observe a causal contrast? Because at least one of the disease-frequency parameters needed for a causal contrast, R_1 and R_0, must be *counterfactual* and therefore unobservable. A parameter (such as a disease frequency) that describes events under *actual conditions* is said to be *actual* (or factual); in contrast, a parameter that describes events under *a hypothetical alternative to actual conditions* is said to be *counterfactual*.[2,3,7,17] Counterfactual parameters cannot be observed because, *by their very definition*, they describe consequences of conditions that did not exist—they describe events following hypothetical alternatives to actual conditions, not actual conditions. The entire collection of outcome parameters for the target, actual and counterfactual (here, R_1, R_0 and all R_i under all other exposure conditions), is sometimes called the set of *potential outcomes*, to note that each is a possibility before the exposure distribution becomes fixed.[11,23]

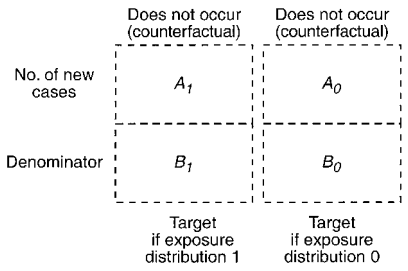
For a given R_1 and R_0, one and only one of the following three scenarios may occur. (1) R_1 is an actual disease frequency; it occurs, and therefore it can be observed. R_0, then, must be counterfactual; as a hypothetical alternative to R_1 it does not occur, and therefore it cannot be observed (as illustrated below).



(2) R_0 is an actual disease frequency; it occurs, and therefore it can be observed. R_1, then, must be counterfactual; as a hypothetical alternative to R_0 it does not occur, and therefore it cannot be observed.

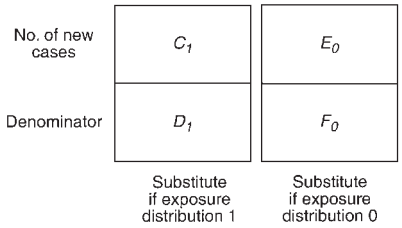


(3) Both R_1 and R_0 are counterfactual disease frequencies—both are hypothetical alternatives to the actual disease frequency that occurs under the actual exposure distribution (which is neither exposure distribution 1 nor 0), and therefore neither R_1 nor R_0 can occur and be observed.



Substitutes

A causal contrast requires two quantities, at least one of which must be counterfactual and therefore unobservable. How, then, can we estimate a causal contrast? By using *substitutes* (illustrated below) for the counterfactual disease frequencies in the target. As before, the subscript indicates the exposure distribution.



In a substitute under exposure distribution 1, let C_1 be the numerator of a disease-frequency measure, and let D_1 be the denominator (number of people or amount of person-time at risk). Likewise, in a substitute under exposure distribution 0, let E_0 be the numerator, and let F_0 be the denominator.

In epidemiological practice, a substitute will usually be a population other than the target population during the etiologic time period. It may be the target population observed at a time other than the etiologic time period, or a population other than the target population. In theory, however, a substitute can be *any* source of information about a counterfactual parameter.

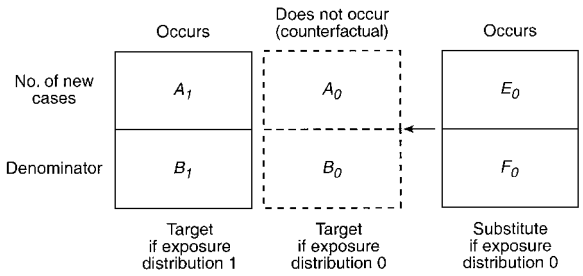
Below we describe how these quantities are used to predict[5–7] or impute[11] the counterfactual quantities in the causal contrast of interest.

Target experiences exposure distribution 1

If the target experiences exposure distribution 1, $R_1 = A_1/B_1$ occurs and therefore can be observed directly, but we must substitute E_0/F_0 for the counterfactual disease-frequency parameter $R_0 = A_0/B_0$; hence, we must substitute the association measure

$$RR_{association} = \frac{R_1}{\text{Substitute for } R_0} = \frac{R_1}{E_0/F_0} = \frac{A_1/B_1}{E_0/F_0}$$

for the causal contrast RR_{causal}. That is, we substitute what we can observe (a contrast in two populations or two time periods) for what we would like to observe directly, but cannot (a contrast in one population and one time period). In the diagrams below, an arrow indicates a substitution of an actual frequency for a counterfactual one.

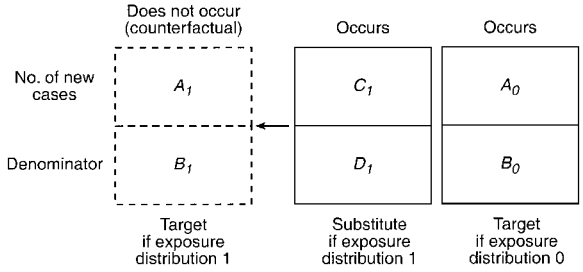


Target experiences exposure distribution 0

If the target experiences exposure distribution 0, $R_0 = A_0/B_0$ occurs and therefore can be observed directly, but we must substitute C_1/D_1 for the counterfactual disease-frequency parameter $R_1 = A_1/B_1$; hence, we must substitute the association measure

$$RR_{association} = \frac{\text{Substitute for } R_1}{R_0} = \frac{C_1/D_1}{R_0} = \frac{C_1/D_1}{A_0/B_0}$$

for the causal contrast RR_{causal}.

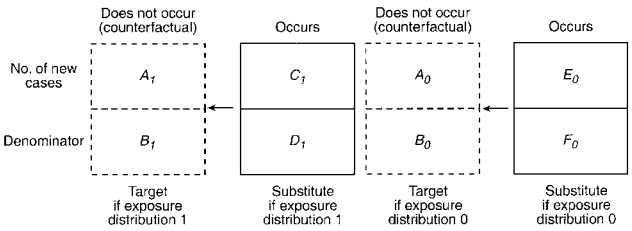


Target experiences neither exposure distribution 1 nor 0

If the target experiences neither exposure distribution 1 nor 0, we must substitute C_1/D_1 for the counterfactual $R_1 = A_1/B_1$, and E_0/F_0 for the counterfactual $R_0 = A_0/B_0$; hence, we substitute the association measure

$$RR_{association} = \frac{\text{Substitute for } R_1}{\text{Substitute for } R_0} = \frac{C_1/D_1}{E_0/F_0}$$

for the causal contrast RR_{causal}.



In this scenario, the exposed substitute is often the exposed subset of the target population, and the unexposed substitute is often the unexposed subset of the target population. That is, C_1, D_1, E_0, and F_0 are often subsets of A_1, B_1, A_0, and B_0, respectively. They may, however, have no overlap at all with the target population; this is the case whenever we generalize from a study to an external target population (see below).

Implications for aetiologic studies

The counterfactual approach and the concept of a causal contrast have important implications for designing, analysing, and interpreting aetiologic studies.

Choice and interpretation of effect measure

Under the counterfactual approach, causal contrasts are the only meaningful effect measures for aetiologic studies. Note that many measures are not causal contrasts; for example, the following are not, because they cannot be expressed as contrasts of a target under two exposure distributions of intrinsic interest: correlation coefficients, percent of variance explained (R^2), *P*-values, χ^2 statistics, and standardized regression coefficients.[24]

Causal contrasts can be given precise interpretations. RR_{causal} can be interpreted as the net proportionate change in disease frequency caused by the difference in exposure distributions 1 and 0 in the target population during the aetiologic time period.

RD_{causal} can be similarly interpreted as the net absolute change in disease frequency.[2,5–7] Because a causal contrast is a contrast in a target—not in a substitute or a 'study base'—it should be interpreted as a measure of causal effect in the target.

We caution that not all population causal contrasts can be interpreted as averages of *individual* causal effects of exposure, or as averages of effects on subpopulations. This limitation arises when the denominators (B_i) of the disease-frequency measures are affected by exposure, as when the B_i represent non-cases or person-years, so that the R_i represent odds or incidence rates.[2,6,25] For example, if the B_i represent person-years, so that RR_{causal} and RD_{causal} are the causal rate ratio and rate difference, and exposure, age, and sex all affect the rate, then RR_{causal} will not equal the average rate ratio across age and sex groups and RD_{causal} will not equal the average rate difference across these groups. For explanations of such problems, see refs[2,6,7,25].

General framework for design and analysis of aetiologic studies

The counterfactual approach leads to a general framework for designing and analysing aetiologic studies. Because all aetiologic designs should estimate causal contrasts, different designs can be viewed simply as different ways of (1) choosing a target that corresponds to the study question, and (2) choosing substitutes and sampling subjects from target and substitutes into the study to balance tradeoffs among bias, variance, study costs and study time. This approach works for all etiologic studies.[26] Beginning with Fisher and Neyman's work on permutation tests,[16] careful study of counterfactual models has also led to the invention of new analysis methods and new study designs.[11,27–29]

The above framework applies to randomized trials as well as observational studies. In fact, it was *invented* for the analysis of randomized trials, and only later extended to non-experimental studies.[3,4,16,17] A typical randomized trial is an example of the scenario discussed above in which the target experiences neither exposure distribution 1 nor 0. Here the treatment arms are substitutes for the target under different treatments. For example, when a drug is approved for treatment of a particular disease, a generalization is being made from the clinical trials on which the approval was based to some external (target) population of patients with that disease. In effect, the treatment and placebo arms in those trials serve as substitutes for the target under different treatment scenarios.

Definition of confounding and confounder

The concept of a causal contrast facilitates precise and general definitions of confounding and confounder. *Confounding* is present if our substitute imperfectly represents what our target would have been like under the counterfactual condition. An association measure is *confounded* (or biased due to confounding) for a causal contrast if it does not equal that causal contrast because of such an imperfect substitution.[2,5–7,30]

Under scenario 1, in which the target experiences exposure distribution 1, confounding occurs if $E_0/F_0 \neq A_0/B_0$. The bias due to confounding in the ratio and difference associations may be measured by

$$\frac{RR_{association}}{RR_{causal}} \text{ and } RD_{association} - RD_{causal},$$

which for this scenario equal

$$\frac{\dfrac{A_1/B_1}{E_0/F_0}}{\dfrac{A_1/B_1}{A_0/B_0}} = \frac{A_0/B_0}{E_0/F_0} \text{ and } \left(\frac{A_1}{B_1} - \frac{E_0}{F_0}\right) - \left(\frac{A_1}{B_1} - \frac{A_0}{B_0}\right) = \frac{A_0}{B_0} - \frac{E_0}{F_0}.$$

Under scenario 2, in which the target experiences exposure distribution 0, confounding occurs if $C_1/D_1 \neq A_1/B_1$. The bias due to confounding in the ratio and difference associations may be measured by

$$\frac{RR_{association}}{RR_{causal}} = \frac{\dfrac{C_1/D_1}{A_0/B_0}}{\dfrac{A_1/B_1}{A_0/B_0}} = \frac{C_1/D_1}{A_1/B_1} \text{ and}$$

$$RR_{association} - RD_{causal} = \left(\frac{C_1}{D_1} - \frac{A_0}{B_0}\right) - \left(\frac{A_1}{B_1} - \frac{A_0}{B_0}\right) = \frac{C_1}{D_1} - \frac{A_1}{B_1}.$$

Finally, under scenario 3, confounding may occur if $E_0/F_0 \neq A_0/B_0$ *or* $C_1/D_1 \neq A_1/B_1$. The bias due to confounding in the ratio and difference associations are

$$\frac{RR_{association}}{RR_{causal}} = \frac{\dfrac{C_1/D_1}{E_0/F_0}}{\dfrac{A_1/B_1}{A_0/B_0}} = \frac{A_0/B_0}{E_0/F_0} \bullet \frac{C_1/D_1}{A_1/B_1} \text{ and}$$

$$\begin{aligned} RR_{association} - RD_{causal} &= \left(\frac{C_1}{D_1} - \frac{E_0}{F_0}\right) - \left(\frac{A_1}{B_1} - \frac{A_0}{B_0}\right) \\ &= \left(\frac{C_1}{D_1} - \frac{A_1}{B_1}\right) + \left(\frac{A_0}{B_0} - \frac{E_0}{F_0}\right), \end{aligned}$$

which is just the product or sum of the confounding factors under scenarios 1 and 2. Thus, $RR_{association}$ will be biased for RR_{causal} unless the product of its two confounding factors is 1, and $RD_{association}$ will be biased for RD_{causal} unless the sum of its two bias factors is zero. Note that, if confounding is present, at least one (and usually both) of the measures will be biased.

Roughly speaking, a *confounder* is a variable that at least partly explains why confounding is present. Many authors attempt to define a confounder more precisely as a variable that is a risk factor for disease and is associated with exposure but not affected by exposure. This definition has several limitations. One is that it applies only to the classical condition in which there is just one variable to consider. That variable may be a compound of several variables, such as an age-sex-race stratification used for standardization or Mantel-Haenszel analysis. Often, however, we must consider several variables at once while keeping them distinct, as when some have been measured and others have not. In that case, the status of a variable as a confounder, as well as the degree and direction of confounding, can change drastically according to which variables are controlled.[5,7,22,31,32] One consequence is that control of a variable that meets the above definition can at times introduce more confounding than it removes.[5,7,22,32] This happens, for

example, when there is little or no confounding to explain; in that case we may still find many variables that satisfy the above definition, but whose confounding effects have balanced out. When this happens, control of one but not the others can increase confounding; see ref.[5] for an illustration.

More generally, the fundamental equalities that must be met to control confounding are $E_0/F_0 = A_0/B_0$ in scenario 1 and $C_1/D_1 = A_1/B_1$ in scenario 2; in scenario 3, both equalities are needed except in the special case discussed above. Both these 'no-confounding' equalities, however, represent summary relations, and place no constraints on particular covariates or their effects.[5,7] Control of confounding thus depends on creating strata within which these equalities are satisfied, rather than on the particular variables used to create the strata.[2,5,7,15,22,32] Methods to aid in identifying sufficient sets of variables for control have been developed using counterfactual and graphical causal models.[7,15,22,32]

Properties of effect-measure modifiers

The size of a causal effect for a given pair of exposure distributions can be different for different targets. To see this, let P_{doomed} be the proportion of individuals in the target population who would get disease during the etiologic time period regardless of their exposure status ('doomed' with respect to the study exposure), $P_{causative}$ the proportion in the target population who would get disease during the etiologic time period if and only if exposed, and $P_{preventive}$ the proportion in the target population who would get disease during the etiologic time period if and only if not exposed. The proportion of individuals in the target who would get disease if *exposed* is $P_{doomed} + P_{causative}$; the proportion who would get disease if *not exposed* is $P_{doomed} + P_{preventive}$. Then, we can write a causal risk ratio as the ratio of these proportions:[2,5]

$$RR_{causal} = \frac{P_{doomed} + P_{causative}}{P_{doomed} + P_{preventive}}.$$

This formula shows that the size of a causal risk ratio not only tends to vary with the proportion of individuals in the target population whose outcome is altered by exposure (who are counted in $P_{causative}$ and $P_{preventive}$), but also tends to vary with the proportion of individuals in the target population for whom disease is inevitable by the end of the etiologic time period (who are counted in P_{doomed}).[2]

The causal risk difference does not depend on P_{doomed}.[2,5] To see this, we can write a causal risk difference as follows:

$$\begin{aligned} RD_{causal} &= (P_{doomed} + P_{causative}) - (P_{doomed} + P_{preventive}) \\ &= P_{causative} - P_{preventive}. \end{aligned}$$

This formula shows that the size of the causal risk difference will tend to vary only with the proportion of individuals in the target population whose outcome is altered by exposure.

It follows that a factor that affects $P_{causative}$ or $P_{preventive}$ can modify the size of a ratio or difference effect measure, and can modify the size of a ratio effect measure even if it affects only P_{doomed}.[2,5] Thus, one should not be surprised if an effect measure varies from one population to another or from one time period to another unless one expects other causal factors to have similar distributions across the populations or periods.

Implications for consistency criteria and meta-analysis

Because variations in the distribution of other factors can easily produce variations in effect measures, the consistency of an association measure across populations should not be viewed as a necessary causal 'criterion'. Conversely, if one expects other causal factors to have similar distributions across a set of populations, one should in particular expect consistency in the distribution of uncontrolled confounders across the populations and hence similar amounts of confounding in the association measures; thus, consistency (homogeneity) of an association measure does not in itself provide logical support for causality, even if the distribution of all other factors is consistent across the populations.

These deductions from the counterfactual formulation provide a logical basis for earlier reservations about the consistency criterion:

> A pertinent question is on what grounds consistency is to be decided. To ask for the same risk ratios to recur under many diverse circumstances is to ask for homogeneity, which is certainly to ask too much.[33]

In other words, the consistency criterion has general applicability only as a qualitative criterion rather than a quantitative one, and then only on the (often reasonable) assumption that either $P_{causative}$ or $P_{preventive}$ is negligible. The same deduction adds force to arguments that meta-analyses are better conducted as a search for sources of systematic variation among study results, rather than as an exercise in estimating a fictional common effect.[34,35]

The amount of bias in effect measures should be quantified

We must always use measures of association as surrogates for causal measures. This gives rise to the question, 'How different are measures of association from causal measures?' In other words, how much bias is inherent in the measures of association that we estimate? In practice, these questions are usually answered informally—that is, it is a matter of 'judgement'. The magnitude of bias, however, is a complicated function of many parameters, and informal evaluation may be inadequate.[27,29,36–38] Many authors hence recommend that formal methods, such as sensitivity analysis[27,29,36–38] and validation substudies,[36] be used to quantify the magnitude of bias.

Formal evaluation of bias requires formulas that describe the magnitude of bias as a function of relevant parameters. The counterfactual approach can help here. For example, it can be used to show that in special cases the approximate expected value of a relative risk estimate equals the causal relative risk times a bias factor for confounding, times a bias factor for losses to follow-up, times a bias factor for subject sampling, times a bias factor for subject non-response, times a bias factor for subjects excluded from analysis, times a bias factor for information bias.[39] This result can be used in a sensitivity analysis to evaluate bias under different plausible scenarios.

Discussion

By their very definition, counterfactuals cannot be observed. Some people find this property disconcerting and reject

counterfactuals as a foundation for causal inference, even though they may use statistical methods that require hypothetical (and hence unobserved) study repetitions for proper interpretation. One reason for their discomfort is that the counterfactual definition of effect seems to contradict the common-sense notion that we can observe effects. This seeming contradiction arises because of the unfortunate tendency to use the word 'effect' for different concepts. Sometimes 'effect' refers to an observed (actual) outcome event, such as 'John Smith's lung cancer was an effect of his smoking'. Often, however, 'effect' refers to an effect *measure* such as RR_{causal}, which has at least one counterfactual (and hence unobservable) component. Although we observe the effects of a cause, we can only *infer* the cause of an effect, because our inferences will always depend on substitutions that may be called into question.

Causal inference is possible because we can make logically sound *conditional* inferences about counterfactuals, despite the fact that we do not observe them. Indeed, following earlier writings[2–7,11,16,17,23,27–29] we have shown how basic problems of causal inference can be made logically precise (and hence subject to logical analysis) by translating them into problems of inference about counterfactuals. Two other well-developed systems of reasoning about cause and effect, structural-equations models and causal diagrams, turn out to yield results equivalent to those obtained using counterfactuals.[15,22,32] This equivalence points to a basic unity among logically sound methods for causal inference.

The physicist Richard Feynman considered science to be 'confusion and doubt, … a march through fog'.[40,p.380] As it does in physics,[41,42] counterfactual analysis can cut through some of the 'fog' in epidemiology, for it leads to a general framework for designing, analysing, and interpreting etiologic studies. It has already led to a number of analysis innovations,[11,16,27–29,32] and we have found it an excellent teaching tool. We hope that this paper will prove useful in enabling epidemiologists to view problems from the counterfactual perspective.

Acknowledgements

This publication was made possible by support from grant number NIH/1R29-ES07986 from the National Institute of Environmental Health Sciences (NIEHS), NIH. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIEHS, NIH. We are grateful to Timothy Church, Aaron Cohen, Bud Gerstman, Jay Kaufman, Stephan Lanes, Malcolm Maclure, Wendy McKelvey, Mark Parascandola, Judea Pearl, Carl Phillips, Charles Poole, Eyal Shahar, and the referees for their helpful comments on earlier drafts of this manuscript.

References

[1] Kelsey JL, Whittemore AS, Evans AS, Thompson WD. *Methods in Observational Epidemiology*. New York: Oxford University Press, 1996.

[2] Greenland S, Rothman KJ. Chapter 4: Measures of effect and measures of association. In: Rothman KJ, Greenland S (eds). *Modern Epidemiology. 2nd Edn*. Philadelphia: Lippincott-Raven, 1998.

[3] Lewis DK. Causation. *J Philos* 1973;**70:**556–67.

[4] Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psych* 1974;**66:**688–701.

[5] Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* 1986;**15:**413–19.

[6] Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol* 1987;**125:**761–68.

[7] Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999;**14:**29–46.

[8] Hume D. *An Enquiry Concerning Human Understanding*. LaSalle: Open Court Press, 1748, p.115.

[9] Neyman J. (1923) Sur les applications de la thar des probabilities aux experiences Agaricales: Essay des principle. [English translation of excerpts by D. Dabrowska and T. Speed]. *Statist Sci* 1990;**5:**463–72.

[10] Copas JB. Randomization models for matched and unmatched 2 × 2 tables. *Biometrika* 1978;**60:**467–76.

[11] Rubin D. Bayesian inference for causal effects: the role of randomization. *Ann Stat* 1978;**6:**34–58.

[12] Dawid AP. Causal inference without counterfactuals (with discussion). *J Am Statist Assoc* 2000;**95:**407–48.

[13] Fisher RA. *The Design of Experiments*. Edinburgh: Oliver and Boyd, 1935.

[14] Sobel ME. Causal inference in the social and behavioral sciences. In: Arminger G, Clogg CC, Sobel ME (eds). *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum Press, 1995.

[15] Pearl J. *Causality*. New York: Springer, 2000.

[16] Rubin DB. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statist Sci* 1990;**5:**472–80.

[17] Simon HA, Rescher N. Cause and counterfactual. *Philos Science* 1966; **33:**323–40.

[18] Stalnaker RC. A theory of conditionals. In: Rescher N (ed.). *Studies in Logical Theory*. Oxford: Blackwell, 1968.

[19] Lewis D. *Counterfactuals*. Oxford: Blackwell, 1973.

[20] Harper WL, Stalnaker RC, Pearce G. *Ifs*. Dordrecht: Reidel, 1981.

[21] Greenland S. Causal analysis in the health sciences. *J Am Statist Assoc* 2000;**95:**286–89.

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

[23] Rubin DB. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics* 1991;**47:**1213–34.

[24] Greenland S, Maclure M, Schlesselman JJ, Poole C, Morgenstern H. Standardized regression coefficients: a further critique and a review of alternatives. *Epidemiology* 1991;**2:**387–92.

[25] Greenland S. Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology* 1996;**7:**498–501.

[26] Maldonado G, Greenland S. The causal-contrast study design (abstract). *Am J Epidemiol* 2000;**151:**S39.

[27] Rosenbaum PR. *Observational Studies*. New York: Springer-Verlag, 1995.

[28] Robins JM. Causal inference from complex longitudinal data. In: Berkane M (ed.). *Latent Variable Modeling with Applications to Causality*. New York: Springer, 1997, pp.69–117.

[29] Robins JM, Rotnitzky A, Scharfstein DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran E (ed.). *Statistical Models in Epidemiology*. New York: Springer, 1999.

[30] Greenland S, Morgenstern H. Confounding in health research. *Ann Rev Public Health* 2001;**22:**189–212.

[31] Fisher L, Patil K. Matching and unrelatedness. *Am J Epidemiol* 1974; **100:**347–49.

[32] Pearl J. Causal diagrams for empirical research (with discussion). *Biometrika* 1995;**82:**669–710.

[33] Susser M. Falsification, verification and causal inference in epidemiology. In: Rothman KJ (ed.). *Causal Inference*. Chestnut Hill, MA: Epidemiology Resources, 1988, pp.46.

[34] Greenland S. A critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994;**140:**290–96.

[35] Poole C, Greenland S. Random effects meta-analyses are not always conservative. *Am J Epidemiol* 1999;**150:**469–75.

[36] Greenland S. Basic methods for sensitivity analysis and external adjustment. In: Rothman KJ, Greenland S (eds). *Modern Epidemiology, 2nd Edn*. Philadelphia: Lippincott-Raven, 1998, pp.343–57.

[37] Maldonado G. Informal evaluation of bias may be inadequate (abstract). *Am J Epidemiol* 1998;**147:**S82.

[38] Leamer EE. Sensitivity analyses would help. *Am Econ Rev* 1985;**75:** 308–13.

[39] Maclure M, Schneeweiss S. The confounding product (abstract). *Am J Epidemiol* 1997;**145:**S55.

[40] Gleick J. *Genius. The Life and Science of Richard Feynman*. New York: Vintage Books, 1992.

[41] Penrose R. *Shadows of the Mind*. New York: Oxford, 1994, Chapters 5 and 6.

[42] Price H. *Time's Arrow and Archimedes' Point*. New York: Oxford, 1996, Chapters 6 and 7.

International Journal of Epidemiology 2002;**31**:429–430

Commentary: Counterfactuals: help or hindrance?

AP Dawid

I welcome this attempt to clarify some of the often perplexing issues, both definitional and philosophical, underlying the formulation and estimation of causal quantities of interest. At the same time I am a little disappointed that the authors' case[1] has not been made with deeper analysis and greater clarity. In particular, I believe that their emphasis on a counterfactual understanding of causality is mostly superfluous and, at some points, misleading. See Dawid[2]—henceforth CIWC—for a detailed account of this position, as well as some dissenting views. In the terminology of their paper, Maldonado and Greenland are considering, as their *experimental unit u*, a specified population, in given circumstances, studied over a given etiologic time period (§1 of CIWC echoes the authors' valuable emphasis on the need for absolute clarity in the definitions and external referents of the theoretical terms employed). Their *treatment t* is the 'exposure distribution' applied to the population. Although Maldonado and Greenland insist on a clear definition of exposure at the individual level, at the desired population level this is less precisely specified (e.g. 20% of the population smoke); this could, and ideally should, be described in greater detail (exactly who smoked, and for how long). However, it appears implicit in the authors' account that populations may be regarded as sufficiently large and homogeneous that such individual level detail can be 'averaged out' over the population, so that we can neglect the effect, on the observed overall proportion, of sampling variability and other such phenomena. Such a 'large population' assumption must also underlie their working assumption that 'response', as measured by population proportion affected, is 'deterministic'. It is not clear to me exactly what else is intended by this description. In particular, does it imply that, were we to study two different populations, we would expect to observe identical responses to the same exposure distribution?—the property termed 'uniformity' in CIWC. This property is a very strong one that I would not normally expect to hold, but it is at least empirically testable. When it does hold, we can find a *perfect* substitute population u_0 for a given population u_1. On applying, say, exposure distribution 1 to u_1 and exposure distribution 0 to u_0, we could then observe, in effect, *both* R_1 and R_0 (where, as in the paper, R_i, or more fully $R_i(u_1)$, denotes the disease frequency, if the target population u_1 had experienced exposure distribution i)—and thus directly measure any causal contrast. So, when the above uniformity property can be taken to hold, 'counterfactuals' become observable and unproblematic.

A weaker form of uniformity, which we may term 'conditional uniformity', might apply when there are covariates that affect response. Conditional uniformity asserts that, if two different populations have identical values for the covariates, and identical exposure distributions, then they will deliver identical responses—still a strong assumption, and again testable (for a specified set of covariates). When this holds, it is once again, in principle, possible to find a perfect substitute, by matching on all the relevant covariates. The authors appear to be mainly concerned with the practical difficulty that the chosen substitute u_0 might *not* be perfectly matched, leading to $R_1(u_0) \neq R_1(u_1)$, and so typically 'biasing' the substitute causal contrast. This important point is well made and helpful, but its connection with 'confounding' as usually understood is far from clear.

University College London, Department of Statistical Science, Gower Street, London WC1E 6BT, UK.

Things becomes much murkier if no uniformity assumption can be made, since then no perfect substitute exists. Even if two populations u_1 and u_0 could be regarded as *a priori* exchangeable, so that $R_0(u_0)$ and $R_0(u_1)$ initially have the same distribution, $R_0(u_0)$ may no longer be an appropriate (unbiased) substitute for $R_0(u_1)$ *after* exposing population 1 to exposure distribution 1 and observing $R_1(u_1)$, since that observation might carry some information about the level of immunity in population 1, so changing the distribution of $R_0(u_1)$ but not that of $R_0(u_0)$. As pointed out in §11 of CIWC, such effects are highly sensitive to untestable assumptions made about the joint distribution of $[R_1(u_1), R_0(u_1)]$ (although bounds are available, which become tighter as the situation approaches that of uniformity).

We can eliminate some of these difficulties by using data of the form $R_1(u_1)$ and $R_0(u_0)$ (for a number of distinct but exchangeable populations) to estimate the marginal probability distributions of each of $R_1(u^*)$ and $R_0(u^*)$ for some *new*, as yet unexposed, 'test population' u^*, exchangeable with those studied. (Analogues of the authors' cautions against bias will continue to apply if the populations u_1, u_0 and u^* are not perfectly exchangeable; but the analysis can then be modified accordingly, so as to take into account differences in observed [CIWC, §8] and/or unobserved [CIWC, §6] concomitant variables.) I argued in CIWC that comparison of these estimated *marginal distributions* for $R_1(u^*)$ and $R_0(u^*)$ is all that is required for causal inference about the effects of switching between treatments (exposure distributions) on the test population u^*. Note particularly that, in contrast to inference about a causal contrast such as $R_1(u_1)/R_0(u_1)$, such a comparison is *not* affected by untestable assumptions about the *joint* distribution of $[R_1(u_1), R_0(u_1)]$. So in this setting the assumption of coexisting potential responses is unnecessary, and can indeed be positively harmful. (The issue is not exactly that either response is 'counterfactual'—before the exposure decision for u^*, each of $R_1(u^*)$ and $R_0(u^*)$ can in principle still be observed, and so is 'hypothetical' rather than counterfactual; rather, the point is that, for any population u, we can never, even in principle, observe *both* $R_1(u)$ and $R_0(u)$ together—they are 'complementary' —and so we can never learn about their dependence structure.)

I do not find the authors' treatment of 'effect-measure modifier' helpful, since it is phrased in terms of quantities I find I cannot meaningfully relate to. Their P_{doomed} is very much a feature of the empirically unknowable joint distribution of $[R_1(u_1), R_0(u_1)]$, and as such I regard it as pointlessly metaphysical. On purely commonsense grounds, at the individual level, response to either exposure will normally be dependent on a host of further stochastic factors, as well as on exposure, so that I find it difficult to accept that this individual response somehow already existed prior to its realization (an attitude I dubbed 'fatalism' in §7 of CIWC). How then can I compare the actual realized response with a counterfactual response under a different exposure, which, even if allowed as a proper subject of discourse, should still be regarded as stochastic? But if there is no predetermined value of this comparison, there can be no such thing as a 'doomed' patient—any more than there can be a penny that, when tossed, will land tails up.

It is significant that P_{doomed} disappears in the expression for RD_{causal}, but not in that for RR_{causal}. This is a reflection of the fact that, in the terminology of §9 of CIWC, RD_{causal} is a 'sheep', having also a perfectly good non-counterfactual interpretation; while RR_{causal} is a 'goat', and simply not an appropriate subject of discourse.

In CIWC I emphasized the importance of the distinction between inference about 'Effects of Causes', referring to predictions about a new population u^* under various hypothetical exposure distribution; and 'Causes of Effects', referring to a comparison of an observed $R_1(u_1)$, for a population u_1 already subjected to exposure distribution 1, with $R_0(u_1)$, the purely counterfactual response it would have displayed had it actually been subjected to exposure distribution 0. An application of inference about Causes of Effects might arise in a legal liability suit, in which an ex-soldier sues the army for having caused his leukaemia through exposing him to depleted uranium contained in anti-tank shells. Epidemiological evidence about the expected consequences of such exposure, and about the natural incidence of leukaemia, would clearly be of relevance; but since such evidence can only directly address the question of Effects of Causes, its correct incorporation and analysis in this context raises some very subtle issues—for example, how to allow for the fact that some individuals might be more susceptible than others, irrespective of exposure?

I argued in CIWC that inference about Effects of Causes is reasonably straightforward, and does not require any recourse to counterfactuals; while inference about Causes of Effects is beset by ambiguities that are compounded, rather than being resolved, by being set in a counterfactual framework. Although it is not always clear from the way they are phrased, the problems considered in the paper currently under discussion are largely concerned with the simpler problem of Effects of Causes —where I do not see a counterfactual analysis contributing much beyond unnecessary complication of concepts and notation. Moreover, there is a danger that readers may be misled into thinking that the paper supplies tools for valid analysis of problems involving Causes of Effects. *Caveat emptor!*

In summary, while I value the authors' emphasis on clarity of definition and their discussion of the problem of bias, I do not consider that they have proved their case that 'counterfactual analysis can cut through some of the fog in epidemiology'. In my own view, such analysis is more likely to obscure the clarity of the view.

References

[1] Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol* 2001;**30:**1035–42.

[2] Dawid AP. Causal inference without counterfactuals (with Discussion). *J Am Statist Assoc* 2000;**95:**407–48.

 International Journal of Epidemiology 2002;31:431–432

Commentary: Estimating causal effects

Jay S Kaufman[a,b] and Sol Kaufman[c]

Maldonado and Greenland have provided a great service to our field in crafting this broadly accessible and eminently readable review of causal principles in epidemiological research.[1] Attention to these issues yields substantial benefits in study design, analysis, and interpretation, and this new elucidation promises to raise the quality of epidemiological thought and practice widely by introducing the concepts to a new generation of researchers, and clarifying them further for the rest of us. Indeed, it is fitting that this review should appear in this journal, as Greenland and Robins' seminal article on this topic appeared in 1986 in these very pages.[2] The authors have achieved an admirable level of clarity and simplicity in their presentation. Some of the devices for obtaining this conceptual simplicity, however, succeed at the risk of obscuring other important issues, and we comment on a few of these below. This is not to suggest that an alternative presentation may have been preferred, but rather merely to briefly explore a few of the many questions that are understandably avoided in the paper.

The authors organize their presentation around an aggregate model, rather than the individual causal model that dominates elsewhere.[2,3] While this choice leads most directly to the comprehension of epidemiological contrasts, it also circumvents several considerations. To begin with, because causation ultimately operates at the individual level, an elucidation at that level, via potential response variables, helps to demystify the 'black box' causal behaviour of a total population. Potential-response variables indicate, for each individual and for each exposure level under consideration, the disease response of that individual had it received that exposure. With this framework in hand, one can attach a clearer meaning to the stated assumption in the paper that 'disease occurrence is deterministic',[1,p.1036] to wit, individual potential responses are fixed rather than random quantities. In other words, the characteristics identifying an individual are sufficient to uniquely determine that individual's response to any given level of exposure.

Following the authors, we consider the simplified case of only two levels of exposure, 'exposed' and 'not exposed', and two levels of response, 'disease' and 'no disease'. If one were to assume, further, that exposure distributions 1 and 0 are 'everyone exposed' and 'everyone not exposed', respectively, then the 'numbers of new cases', $\mathbf{A_1}$ and $\mathbf{A_0}$, are easily seen to be the numbers of individuals in the target population having potential response 'disease', if 'exposed' and if 'not exposed', respectively. The authors actually consider more general exposure distributions characterized by 'per cent (or proportion) of population subjected to each exposure level' (allowing for possibly more than two levels). This generalization may be problematic in that the proportions, alone, are insufficient to determine the aggregate numbers $\mathbf{A_1}$ and $\mathbf{A_0}$ unless one makes the highly unrealistic assumption that all individuals have the same potential responses. There is, however, another way out of this impasse, which is to assume that the different exposure levels in a mixed exposure distribution are assigned by random partition of the target population into subsets of the appropriate size. One might consider such randomization to arise by design, as in experimental studies,[4] or by nature, as in observational studies.[5] A consequence is that $\mathbf{A_1}$ and $\mathbf{A_0}$ are also random, and the quantities of interest would then become their expected values.[1,p.1036]

Another important source of variability is the sampling of the study population from the target population. The authors clearly subsume this under the rubric of confounding, inasmuch as bias in the estimation of the causal effect due to sampling variability arises from use of a substitute population that does not precisely correspond to the outcome experience (actual and counterfactual) of the target population. While this approach has many advantages, it also risks some confusion. Later in the paper the authors state that various epidemiological study designs represent different ways of 'choosing substitutes and sampling subjects from target and substitutes into the study…',[1,p.1039] re-establishing a distinction between the two concepts that they had just wed. Furthermore, many readers may understandably be uncomfortable with the resulting definition of a confounder as a variable that 'partly explains why confounding is present',[1,p.1039] since they may attribute 'explanation' to *causal* confounders only. If we subsume sampling variability under the general category of confounding, then we find that we may indeed reduce confounding through conditioning on some covariates even when these covariates 'explain' nothing, in that they are causally irrelevant to the etiologic process linking exposure of interest to disease.[6] This discomfort may be heightened by the apparent inconsistency of taking expected values to deal with a stochastic potential response model[1,p.1036] or with random assignments in mixed exposure distributions (as was noted above to be a necessary aspect of mixed distributions), but not taking expected values when sampling variability is involved. Finally, we must accept that many other authors distinguish confounding from sampling variability. Stone, for example, asserts that confounding pertains to distributions in the total population from which the sample was taken, and that confounding is present only if there exist unmeasured covariates which affect outcome, and are not independent of exposure, conditional on the measured covariates.[7] For the

[a] Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, NC 27599–7400, USA.

[b] Carolina Population Center, University of North Carolina at Chapel Hill, 123 West Franklin Street, Chapel Hill, NC 27516–3997, USA.

[c] Department of Otolaryngology, University at Buffalo, 3435 Main Street, Buffalo, NY 14214, USA.

Correspondence: Jay S Kaufman, Department of Epidemiology (CB#7400), University of North Carolina School of Public Health, McGavran-Greenberg Hall, Pittsboro Road, Chapel Hill, NC 27599–7400, USA. E-mail: Jay_Kaufman@unc.edu

time being at least, these conflicting approaches are sure to generate continued confusion in our field, and in our interactions with statisticians and social scientists.

As a final point, we note that the individual model of causation reinforces an appreciation for implications of the choice of exposures to study and the interpretation of their effect estimates. Causal inference is contingent on the manipulability of the exposure in order to provide some plausible basis for accepting the substitute population as even remotely adequate for the estimation of the counterfactual quantity of interest. This is particularly relevant to social epidemiology, because when the exposure is an individual attribute, such as race or sex, then any choice of substitute population can generally be rejected as grossly inadequate. For example, a team of epidemiologists recently claimed to have found evidence of racial differences in 'tumor virulence' between black and white men with prostate cancer, based on an observational study of mortality in an 'equal access medical care setting'.[8] The choice to locate the study in the 'equal access' setting was motivated by the desire to have the conditional (i.e. covariate-adjusted) mortality experience of white prostate cancer patients serve as a reasonable substitute population for the counterfactual experience of black patients, had they been white. The assertion by the authors that this study reveals some innate biological feature of black race rests on this premise. The discussion by Maldonado and Greenland helps to clarify exactly why we may be left perplexed by such an assertion. It not only requires that we imagine what it means for there to be a counterfactual outcome distribution (i.e. the number of deaths that would have occurred among blacks, had they been white), but also that this quantity is reasonably estimated by the chosen substitute population, a particular group of white men. The approach appears to be quite problematic on both counts.[9]

In closing, we express our congratulations to Maldonado and Greenland for this contribution to the literature. Awareness of the foundations of causal inference in epidemiology has increased in recent years, and this is due in large part to the diligent efforts of Sander Greenland, James Robins, and their students. The present paper serves to provoke further discussion and insight, and to instruct a wider audience of epidemiologists. Through this ongoing process, we benefit our understanding thereby improving our science, and thus, our capacity to intervene upon and improve human health.

References

[1] Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol* 2001;**30:**1035–42.

[2] Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* 1986;**15:**433–39.

[3] Greenland S. Interpretation and choice of effect measures in epidemiologic analysis. *Am J Epidemiol* 1987;**125:**761–68.

[4] Copas JB. Randomization models for the matched and unmatched 2 × 2 tables. *Biometrika* 1973;**60:**467–76.

[5] Robins JM. Confidence intervals for causal parameters. *Stat Med* 1988;**7:**773–85.

[6] Robins JM, Morgenstern H. The foundations of confounding in epidemiology. *Computers and Mathematics with Applications* 1987;**14:** 869–916.

[7] Stone R. The assumptions on which causal inferences rest. *J R Statist Soc (B)* 1993;**55:**455–66.

[8] Robbins AS, Whittemore AS, van den Eeden SK. Race, prostate cancer survival, and membership in a large health maintenance organization. *J Natl Cancer Inst* 1998;**13:**986–90.

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

International Journal of Epidemiology 2002;**31**:432–434

Commentary: Population versus individual level causal effects

Felix Elwert and Christopher Winship

We congratulate Maldonado and Greenland[1] (MG henceforward) on an interesting and provocative paper. Aiming at epidemiological applications, MG identify the causal effect of changing a distribution of exposures to a target population on the population's outcome distribution. Instead of applying a particular treatment to an individual, MG apply a distribution of treatments (the exposure distribution) to a population. By raising the unit of analysis from the individual to the population, MG depart in important respects from the standard model of counterfactual causal inference. Comparing MG's model to the standard model we make two points: First, MG's conceptualization of causal effects on the population level is valuable if the stable unit-treatment assumption (SUTVA) does not hold at lower levels, but the data requirements are steep. Second, as

Department of Sociology, Harvard University, Cambridge, MA 02138, USA.

Correspondence: Felix Elwert, Department of Sociology, Harvard University, Cambridge, MA 02138, USA. E-mail: elwert@fas.harvard.edu

they mention, we emphasize that MG's population level estimates generally cannot be interpreted as estimates of average causal effects (ACE) in the standard individual-level approach.

Individual-level causal effects

We remind the reader of the standard individual-level presentation of the counterfactual model of causal inference, also known as the Rubin Model.[2–4] Here, a particular treatment, t, is applied to a unit of analysis, i, (e.g. a person). The causal effect of t on i, δ_i, is defined as the difference between the outcome of the unit under treatment, $Y_i(t)$, and the outcome of the *same* unit under control, $Y_i(c)$,

$$\text{(I)} \qquad \delta_i = Y_i(t) - Y_i(c).$$

The 'fundamental problem of causal inference'[2] is that $Y_i(t)$ and $Y_i(c)$ cannot be directly observed together, because every unit of analysis is placed either in treatment or in control condition, but not in both at the same time. Therefore direct estimation of causal effects is impossible. As in MG, the solution is to substitute for the counterfactual observation another unit of analysis, j, which resembles i in all causally relevant respects other than treatment status.

Typically, we are not interested in the causal effect for a specific individual, but rather the average causal effect, ACE, in the study population:

$$\text{(II)} \qquad \text{ACE} = \sum_{i=1}^{n} \delta_i / n = \overline{Y_i(t)} - \overline{Y_i(c)} \text{ for } i = 1, \ldots, n.$$

In completely randomized experiments, the standard estimator for this parameter subtracts the mean outcome of the units in the treatment group from the mean outcome of the units in the control group:

$$\text{(III)} \qquad \hat{ACE} = \overline{Y_t(t)} - \overline{Y_c(c)}$$

This approach assumes that there is no interaction between units and that all treated units in the study receive identical treatments. Rubin terms this the 'stable unit-treatment value assumption' (SUTVA).[5,6]

The key virtue of randomization is to create balanced treatment and control groups that resemble each other across all causally relevant variables except treatment status. Techniques such as matching on propensity scores are available to achieve balance even in non-randomized observational studies.[7,8]

Population-level causal effects: utility and data requirements

MG's framework applies exposure distributions to target populations. Consequently, their unit of analysis is the population. This approach has merit, particularly when SUTVA does not hold within the population. Such situations occur frequently, e.g. in educational research where student test scores may be affected by tutoring their classmates received. Here one would want to use classes for units of analysis, rather than students.

Note, however, that the higher the unit of analysis, the more challenging the data requirements due to comparability of units of analysis, and identity of treatments.

The counterfactual model relies on the comparison of units of analysis that resemble each other in all causally relevant aspects except treatment status. To continue our educational example on the population (classroom) level, it would be necessary to find comparable classes, rather than comparable students. If SUTVA does not hold, this would not only involve comparable student populations, but also comparable dependencies between students within classes in order to ensure comparable peer effects.

The standard model further assumes that all units in the treatment group receive identical treatments. (Note that in a population level analogy to the standard individual-level model, a treatment group contains multiple target populations as units of analysis, each of which contains multiple individuals. Comparing a single target population to a single substitute would amount to working with a sample of N = 2.) If the treatment in question is an exposure distribution, as MG stipulate, identity of treatments across units (i.e. target populations) becomes much harder to assert. It depends on two aspects: (1) the exposure distribution's marginal distribution, which records the relative frequency of exposure levels within a target population; and (2) the mapping of distinct exposures from the exposure distribution onto individuals within a target population. If the population is heterogeneous in its members, different mappings of the same exposure distribution will induce different outcomes. Thus, to assure identity of treatments, both the marginal exposure distribution and its mapping have to be held constant across target populations in the treatment group. Due to these challenges, it seems advisable to choose the smallest unit for which SUTVA still holds as unit of analysis.

Dissimilarity of population-level causal contrasts and average causal effects

MG remark that 'not all population causal contrasts can be interpreted as averages of *individual* causal effects of exposure' (p.1039 in their paper). We would like to go further and argue that MG's population-level estimates will hardly ever represent average individual-level causal effects, because their approach generally does not sustain the conditions of a standard individual-level counterfactual analysis.

An example of the causal effect of smoking on lung cancer may convey the guiding intuition. Consider a population of 1000 men. Of these, 40% are highly susceptible to smoking-induced lung cancer and smoke, and 60% are minimally susceptible to cancer and do not smoke. The rate of lung cancer in this population is 40%. We want to estimate the effect of a change in the exposure distribution from 40% to 60% ever-smokers (similar to MG's example on p.1039). We identify a perfect substitute population of 1000 other men, 600 of whom smoke. However, all of these smokers are only minimally susceptible to lung cancer. In this population the cancer rate is 1%. MG's measure of causal contrast would indicate that increasing the exposure to smoking has *decreased* the incidence of lung cancer, even though each individual member would suffer an increased risk of cancer by taking up smoking. The reason is that different individuals smoke in the two populations.

This result makes sense in MG's approach, because it accurately identifies the population-level causal effect of having changed

both the exposure distribution's marginal distribution and its mapping onto the target population. In the individual-level approach this result would be impossible, because the ACE cannot be negative if all δ_i are positive. MG's population level estimates and the standard individual-level ACE are not equivalent.

References

[1] Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol* 2001;**30:**1035–42.

[2] Holland P. Statistics and causal inference. *J Am Statist Assoc* 1986;**81:** 945–70.

[3] Reiter J. Using statistics to determine causal relationships. *American Mathematical Monthly* 2000;**107:**24–32.

[4] Winship C, Morgan C. The estimation of causal effects from observational data. *Annu Rev Sociol* 1999;**25:**659–707.

[5] Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Stat* 1978;**7:**34–58.

[6] Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health* 2000;**21:**21–45.

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

[8] Rubin DB, Thomas N. Combining propensity score matching with additional adjustments for prognostic covariates. *J Am Statist Assoc* 2000;**95:**573–85.

International Journal of Epidemiology 2002;**31**:434–435

Commentary: Estimating causal effects

Glenn Shafer

This article[1] explains the counterfactual theory of causation, avoiding details and technicalities but providing a clear explanation of most of the terminology that is used when the theory is applied to epidemiology. At the end of the article, the authors mention that some people 'reject counterfactuals as a foundation for casual inference'. The editor has asked me, as one of those people, to explain the difficulties I see with the counterfactual theory. I will try to do so at the same non-technical level at which the article is written.

Although the authors begin their history of the counterfactual approach with a quotation from David Hume,[2] they would probably agree that speculation about 'what might have been' is as old as the human ideas of blame and regret. No doubt the objections to such speculation are equally as old. When your mother tells you that you would have avoided your cold by wearing a jacket, you may object that the result of wearing or not wearing a jacket was not predictable and perhaps not in any sense determined. If you could have acted differently in the matter of the jacket, you and others could have acted differently in other respects, many of which might also have impinged on your health. Who is to say who would have done what had you worn a jacket?

Epidemiologists are usually concerned with the effects of public health risks on whole populations, and we might hope that the average effect of an exposure on a population might be well defined even when the effect on individuals is not, because of the averaging-out of other unpredictable factors. However, as the authors make clear, the counterfactual approach, as it has been developed in the statistical and epidemiological literature in recent decades, insists on the assumption that the effects on individuals are well defined. In this article, for example, they assume that it is determined whether a given individual will fall ill regardless of exposure. So the argument between the advocates of counterfactuals (such as the authors) and the dissenters (such as myself) really does boil down to the ancient argument between those who insist on always giving meaning to a might-have-been and those who demur.

What is the alternative to the counterfactual approach? The obvious alternative is a predictive approach. Using this approach, we say that A causes B in a strong sense if we can predict, using a method of prediction that proves consistently correct, that B will happen if we do A and will not happen if we do not do A. Weaker senses of causation can be expressed using probabilities; we say that the action A is a probabilistic cause of B if it raises the probability of B. This requires an objective concept of probability; it must be verified that B consistently happens more often when A is performed than when it is not, regardless of other factors.

As I explain in my 1996 book, *The Art of Causal Conjecture*,[3] the practical aspects of causal inference (different ways of defining causal effects, ideas of confounding, etc.) can be handled by the predictive approach just as well as by the counterfactual approach—and the predictive approach has a decisive philosophical advantage: it makes clear that the concept of causality has an empirical basis, independent of arbitrarily imagined might-have-beens. I say more about this in my article 'Causality and responsibility',[4] and my recent book with Vladimir Vovk[5] elaborates a foundation for probability theory that can be used to support the predictive approach.

The reader might suspect that the predictive approach and the counterfactual approach say the same thing in different ways.

Graduate School of Management, Rutgers University, 180 University Avenue, Newport, NJ 07102, USA. E-mail: gshafer@andromeda.rutgers.edu

The advocates of the counterfactual approach insist, however, on points that cannot be reconciled with the predictive approach. They begin by insisting on the word *counterfactual*. The very word places us in the situation where A has already been performed and so not(A) is counter to the facts. The counterfactual theory insists that there should be a well-defined answer, in this situation, to the question of what would have happened if A had not been performed. The predictive theory, on the other hand, considers only what can be predicted before the choice between A and not(A) is made. Later, this situation will be in the past, but it will never be in the subjunctive. If no definite prediction is possible about whether B will happen if A is not performed—if only probabilities can be given or not even that—and then A is performed, then there will be no answer as to whether B would have happened had A not been performed.

At the end of their introduction, the authors indicate that they are willing to consider probabilities: 'under a stochastic model, the quantities we discuss are probabilities or expected values'. They then cite two articles by one of the authors, Sander Greenland. They go out of their way, however, to deny causal meaning to the consistency across populations that would be needed to make probabilistic predictions meaningful. I have not been able to understand how the articles by Greenland resolve this contradiction.

Here are some comments that may broaden the picture painted by the authors' citations of literature on counterfactuals outside statistics and epidemiology. David Hume's counterfactual definition was only one of several definitions of cause that he formulated in *An Enquiry Concerning Human Understanding*.[2] David Lewis, a philosopher at Princeton University, is cited as developing the counterfactual definition of causality currently used in the statistics literature, but in conversations with myself and other statisticians Professor Lewis has repeatedly disavowed this interpretation of his work, and during decades following the 1973 book cited, he and his students have published numerous articles devoted to developing an empirical understanding of causality that would be consistent with the predictive approach I have sketched. The authors quite appropriately cite two physicists who favour the counterfactual approach, but their confident assertion that counterfactual analysis cuts through the fog in physics, juxtaposed with the name of Richard Feynman, should not be allowed to obscure the fact that Feynman never advocated the counterfactual approach and that many physicists explicitly oppose it; see for example Layzer.[6]

References

[1] Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol* 2001;**30:**1035–42.

[2] Hume D. *An Enquiry Concerning Human Understanding.* LaSalle: Open Court Press, 1748.

[3] Shafer G. *The Art of Causal Conjecture.* Cambridge, MA: MIT Press, 1996.

[4] Shafer G. *Causality and Responsibility*. Cardozo Law Review 2001; **22(1):**101–23.

[5] Shafer G, Vovk V. *Probability and Finance: It's Only a Game*. New York: Wiley, 2001.

[6] Layzer D. *Cosmogenesis: The Growth of Order in the Universe*. New York: Oxford University Press, 1991.

International Journal of Epidemiology 2002;**31**:435–438

Response: Defining and estimating causal effects

George Maldonado and Sander Greenland

We thank Kaufman and Kaufman (K&K),[1] Dawid,[2] Elwert and Winship,[3] and Shafer[4] for their commentaries on our paper 'Estimating causal effects'.[5] Here we hope to separate misunderstandings from substantial disagreements; we believe the latter arise only in the comments of Dawid[2] and Shafer[4] (and are described in refs. 6–13).

Misunderstandings

According to K&K, 'The authors organize their presentation around an aggregate model, rather than the individual causal model that dominates elsewhere'. This is not entirely true. We organized our presentation around the *target population as specified by the study question*. Our target population could comprise one person, many people, or any collection of interest; our model of effects is therefore aggregate when the study question asks about a population of aggregates (e.g. all counties in California), but it is a model for effects on individuals when the study question asks about a group of individuals.

Kaufman and Kaufman then say 'Following the authors, we consider the simplified case of only two levels of exposure, "exposed" and "not exposed"'. But we did not use this simplification. We wrote that a causal contrast compares outcomes under two exposure distributions that 'represent different

possible mixtures of individual exposure conditions'. We did not imply that only two exposure distributions are possible. On the contrary, our conceptualization allows for any exposure distribution (all possible combinations of exposure timings, exposure metrics, people in the target, and exposure levels).

Kaufman and Kaufman also say 'Another important source of variability is the sampling of the study population from the target population. The authors clearly subsume this under the rubric of confounding.' On the contrary, we do not assume any sampling from the target, nor would we subsume such sampling under confounding. Instead, we wrote '*Confounding* is present if our substitute imperfectly represents what our target would have been like under the counterfactual condition'. Thus, in our conceptualization, confounding results from an imperfect choice of substitute, which could be—but need not be—a result of sampling from the target to form the study population. For example, in an occupational study, workers at plant 1 might be our target population, and plant 2 might be used as a substitute for the counterfactual experience of the workers at plant 1. Here, the study population consists of everyone in the target *plus* workers at plant 2. If the experience of the workers at plant 2 is not a good substitute for the experience of plant 1 workers under the counterfactual exposure distribution, then the resulting effect estimates will be confounded.

The study population need not be sampled from the target population at all. Consider our scenario 3, in which the target experiences neither exposure distribution 1 nor 0. Here the study population would consist of two substitutes, because the target did not experience either of the exposure distributions we want to compare. Neither substitute is *required* to include any members of the target. In theory, the only requirement for a valid causal contrast is that a substitute is a good substitute for the *counterfactual experience* of the target. It need *not* also be a good substitute for the *actual experience* of the target; this is a stronger condition than necessary. Thus, we would strike the word *actual* in K&K's parenthetical statement '(actual and counterfactual)'.

Three of the commentators[1–3] misinterpreted our examples of exposure distributions, in which 20% or 40% of the target population regularly smoked cigarettes during a given time period. We did not intend to imply that one does not know individual exposures. In our conceptualization, *individuals* experience exposures (treatments), and a group of individuals has a distribution that describes each individual's exposure. Therefore, in a study with data on individual exposures, neither Dawid's uniformity assumption nor K&K's assumption of random allocation of exposure is necessary for defining or estimating effects. This point may be seen in definitions of generalized population attributable fractions, in which exposure effects are allowed to vary across covariates (and hence across individuals).[14] The assumptions imposed by K&K and Dawid are indeed made by conventional statistical procedures for effect estimation, but we avoid them because they have no justification in typical observational studies.[15]

Elwert and Winship[3] state 'Instead of applying a particular treatment to an individual, MG apply a distribution of treatments (the exposure distribution) to a population', and from that they incorrectly concluded that our basic unit of analysis is the population rather than the individual. In reality, the ACE of the potential-outcomes model is a special case of our causal-contrast measure. Both measures are derived from the outcomes of *individuals* under different exposures or treatments. In our conceptualization, individuals experience the causal effect of the difference in exposure levels being compared, and the individual outcomes are modelled, even when the average causal effect on a group of individuals is of ultimate interest. Contrary to what Elwert and Winship[3] thought, we do employ 'the mapping of distinct exposures from the exposure distribution onto individuals within a target population'. This mapping is known in most aetiological studies, because exposure information is typically collected on individuals (although not in ecologic studies). We suspect that our simplified notation obscured this important point. Of course, the 'individuals' in our model may be aggregates, such as counties or states, but if so the treatments and outcomes must then be variables defined unambiguously on the aggregate (macro) level (such as laws, expenditures, and mortality rates).

Elwert and Winship[3] also write that 'MG depart in important respects from the standard model of counterfactual causal inference'; this is true, although we do not depart from it in the way that Elwert and Winship describe. Perhaps most importantly, we are interested in the causal effect in a *target population* (the group of individuals about whom our scientific question asks), not in the study population. The two populations are not necessarily the same: The study population may include individuals who are not in the target population but are being used as substitutes for the counterfactual experience of the target (e.g. our scenarios one and two); it may even include no individual from the target population (e.g. our scenario three). The distinction is important because (1) the size of a causal-effect measure is not a biological constant, as it may vary with the composition of the target population, and (2) the target population may not be available for study (e.g. the people enrolled in a randomized trial may not be the target population of public health or medical interest). Thus a large, well-conducted randomized trial would usually not provide an unbiased estimator for a causal-effect measure *unless* the people enrolled in the trial are representative of the target population; this condition is rarely stated in presentations of the potential-outcomes model.

Dawid[4] questions the meaning of deterministic disease occurrence, which we used only for simplicity. Kaufman and Kaufman explain what we mean: 'individual potential responses are fixed rather than random quantities'. P_{doomed}, for example, represents individuals who would always get the study disease if the study were hypothetically repeated, fixing the entire history of that individual except for exposure and factors affected by exposure. Using Dawid's example of a penny toss, a two-tailed penny will always land tails up. Stochastic counterfactuals are discussed in detail elsewhere.[16–18] Dawid also states 'Things become much murkier if no uniformity assumption can be made, since then no perfect substitute exists'. This statement is just wrong: Because of averaging, a substitute may perfectly represent a counterfactual outcome of the target even if there is no uniformity in either group.[18] The practical problem is that we are usually unable to identify a perfect substitute with certainty.

Disagreements

Regarding Dawid's objections to counterfactual causal inference, we recommend readers to sec. 1.4.4 of Pearl[19] and the commentaries

on 'Causal interference without counterfactuals',[6] most of which embrace counterfactual models and address his objections in detail[7–12] (the exception being Shafer[13]). Dawid objects to counterfactual events because (by definition) they do not occur and so cannot be observed (although he concedes a role for them in formulating causal models).[6] We and others[9–12] maintain that this property of counterfactuals leads to insights and reveals assumptions that are hidden by other models. For example, a *P*-value (the probability of observing a statistic as large as observed or *larger*) is defined in terms of counterfactuals (in the 'or larger'), which implies that one should reject use of *P*-values if one rejects inferences based on non-occurring events.

Shafer[4] offers predictive causality as an 'obvious' alternative to counterfactual theories. While predictive theories are interesting, we find them as yet too limiting to supplant counterfactual theories, and quite obscure for teaching purposes. Treatments of predictive causality we have seen have dodged the thorny problem of defining causality by relying on circularities[4,13] (which usually go unnoticed), or on a metaphysical notion of covariate sufficiency[6] (which becomes a derived concept in other theories[18,19]), or on hidden potential outcomes.[13] Shafer[4] indulges in the circularity when he defines weak causation by saying 'A is a probabilistic cause of B if it raises the probability of B'. What does it mean for A to 'raise' a probability if not to cause an increase? Probabilistic counterfactuals[16–18] provide the only non-circular answer we know of, notwithstanding Shafer's inability to understand them[4] or even acknowledge their existence.[13] Shafer's definition of strong causality,[4] 'that A causes B in a strong sense if we can predict, using a method of prediction that proves consistently correct, that B will happen if we do A and will not happen if we do not do A', tacks on a subjective observer ('we' who predict) to an objective definition of causation identical to that based on the 'do' (or 'set') operator of potential-outcome models.[19] This does not strike us as an advantage of Shafer's theory[13] over counterfactual theories.

Shafer[4] nicely sums up another reason why predictive causality has thus far failed to attract the usage that counterfactual theories have: 'The very word [counterfactual] places us in the situation where A has already been performed and so not(A) is counter to the facts. The counterfactual theory insists that there should be a well-defined answer … to the question of what would have happened if A had not been performed. The predictive theory, on the other hand considers only what can be predicted before the choice between A and not(A) is made'. The latter limitation means that Shafer's restrictive version of predictive causality demurs to directly face down the subjunctive causal questions of deep concern to individuals and society. Those questions are explicitly cast in a counterfactual 'but for' form put to American juries, such as 'but for the action of tobacco companies in promoting the use of cigarettes, would the state of Minnesota have had health-care costs as high as it did bear?' The only substance we see in Shafer's criticism of such questions[13] is addressed by prefixing 'health-care costs' with 'expected'.

No one doubts the difficulty of answering such questions, but we dispute Shafer's attempt to address these difficulties by denying meaning to the question.[4,13] The philosophy espoused by Shafer[4] as well as Dawid[6] strikes us and others[8–10] as a form of logical positivism (misattributed to Popper[6,8]) that attempts to hobble science by a fiat of restriction to questions that admit tidy solutions. All else is condemned as 'untestable', 'metaphysical', or 'silly' and therefore not scientific,[6,13] without regard to the importance of the question; witness Dawid's[2] claim that a causal relative risk 'is simply not an appropriate subject for discourse'! In contrast, counterfactual theories allow one to examine such questions logically, and make clear exactly where precise estimates cannot be attained without detailed mechanistic knowledge;[17] they thus show *why* answers to certain causal questions must remain conjectural, and how to shape those conjectures to be consistent with background information (including results from predictive research).[3,17] They also help us shape questions and answers to remove such ambiguity as can be removed,[3,5,17] without introducing the distortions and oversimplifications that seem to attend extreme positivist approaches.[13] If we do not avail ourselves of these advantages, special interests will still exploit them expertly.[20]

While we welcome other coherent theories of causality (such as decision-analytic[6] and graphical theories[19]), they have so far failed to yield a broad set of widely tested statistical methods comparable to that of the potential-outcomes model of counterfactuals (invented by Neyman[21] in the early 1920s, yet often misattributed to Rubin,[3] though not by Rubin himself[22]). Analysts employ this model to good effect whenever they apply a permutation test (such as Fisher's exact test) to randomized-trial data,[15,23] and the vast work by Rubin, Robins, and Rosenbaum has extended the model and methods to observational studies.[11,22,24] Shafer points out correctly that Lewis's counterfactual theory is not equivalent to this model, but fails to point out that Lewis's theory (in which counterfactuals are taken as actual events in 'closest possible worlds') is far more metaphysical than Neyman's model, and lends absolutely no support to Shafer's theory.

Shafer[4] also takes us to task for juxtaposing Feynman's name with citations of physicists who endorse counterfactuals.[5] Indeed, Feynman did not advocate counterfactuals because, to the best of our knowledge, he never discussed them. As for whether 'many' physicists oppose them,[4] the truth will be unknown until there is a more thorough poll of physicists than either we or Shafer (we each cite one) have mustered. Although we doubt whether epidemiologists should base any decision on the poll's outcome, we note that Shafer's cite[4] (like Shafer[13]) fails to even consider probabilistic counterfactuals.

References

[1] Kaufman JS, Kaufman S. Commentary: Estimating causal effects. *Int J Epidemiol* 2002;**31:**431–32.

[2] Dawid AP. Commentary: Counterfactuals: help or hindrance? *Int J Epidemiol* 2002;**31:**429–30.

[3] Elwert F, Winship C. Commentary: Population versus individual level causal effects. *Int J Epidemiol* 2002;**31:**432–34.

[4] Shafer G. Commentary: Estimating causal effects. *Int J Epidemiol* 2002; **31:**434–35.

[5] Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol* 2001;**30:**1–8.

[6] Dawid AP. Causal inference without counterfactuals (with discussion). *J Am Statist Assoc* 2000;**95:**407–48.

[7] Cox DR. Comment. *J Am Statist Assoc* 2000;**95:**424–25.

[8] Casella G, Schwartz SP. Comment. *J Am Statist Assoc* 2000;**95:**425–27.

[9] Pearl J. Comment. *J Am Statist Assoc* 2000;**95:**428–31.

[10] Robins JM, Greenland S. Comment. *J Am Statist Assoc* 2000;**95:**431–35.

[11] Rubin DB. Comment. *J Am Statist Assoc* 2000;**95:**435–38.

[12] Wasserman L. Comment. *J Am Statist Assoc* 2000;**95:**442–43.

[13] Shafer G. Comment. *J Am Statist Assoc* 2000;**95:**438–42.

[14] Greenland S, Drescher K. Maximum likelihood estimation of attributable fractions from logistic models. *Biometrics* 1993;**49:**865–72.

[15] Greenland S. Randomization, statistics, and causal inference. *Epidemiology* 1990;**1:**421–29.

[16] Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol* 1987;**125:**761–68.

[17] Robins JM, Greenland S. The probability of causation under a stochastic model for individual risk. *Biometrics* 1989;**45:**1125–38.

[18] Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Statist Sci* 1999;**14:**29–46.

[19] Pearl J. *Causality*. New York: Springer, 2000.

[20] Rubin DB. Estimating the causal effects of smoking. *Stat Med* 2001;**20:** 1395–414.

[21] Neyman J. Sur les applications de la thar des probabilities aux experiences Agaricales: Essay des principle. Original 1923; English translation of excerpts by Dabrowska D and Speed T. *Statist Sci* 1990;**5:**463–72.

[22] Rubin DB. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Stat Sci* 1990;**5:**472–80.

[23] Greenland S. On the logical justification of conditional tests for two-by-two contingency tables. *Am Stat* 1991;**45:**248–51.

[24] Greenland S. Causal analysis in the health sciences. *J Am Statist Assoc* 2000;**95:**286–89.