

The Probability of Causation Under a Stochastic Model for Individual Risk

James Robins

Occupational Health Program and Department of Biostatistics,
Harvard School of Public Health, 665 Huntington Avenue,
Boston, Massachusetts 02115-9957, U.S.A.

and

Sander Greenland

Department of Epidemiology, UCLA School of Public Health,
Los Angeles, California 90024-1772, U.S.A.

SUMMARY

In this paper we offer a mathematical definition for the probability of causation that formalizes the legal and ordinary-language meaning of the term. We show that, under this definition, even the average probability of causation among exposed cases is not identifiable from epidemiologic data. This is because the probability of causation depends both on the unknown mechanisms by which exposure affects disease risk and competing risks, and on the unknown degree of heterogeneity in the background disease risk of the exposed population. We derive the maximum and minimum values for the probability of causation consistent with the observable population quantities. We also derive the relationship of the "assigned share" (excess incidence rate as a proportion of total incidence rate) to the probability of causation.

1. Introduction

Suppose John Smith, born 1920, worked in an asbestos textile plant for 1 year in 1950 and contracted lung cancer at age 60, and he brings suit against the asbestos supplier. If it were known with certainty that John Smith's lung cancer was due to asbestos from the plant, he would be awarded full compensation. But the court cannot be sure that John Smith's lung cancer was due to asbestos.

It has been suggested that when uncertainty exists as to the cause of a particular individual's lung cancer, award levels in tort suits be based on the concept of "probability of causation" (see Lagakos and Mosteller, 1986; NAS Oversight Committee, 1984; NIH Working Group, 1984). One proposal is that awards be made in direct proportion to the probability of causation. Another proposed payment scheme is (a) if the probability of causation exceeds $\frac{1}{2}$, full compensation is awarded but (b) if the probability of causation is less than $\frac{1}{2}$, no award will be made (see NIH Working Group, 1984). This scheme is based on an interpretation of tort law that holds that no compensation be paid unless it is "more probable than not" that the injury was caused by an action of the defendant.

Implementation of either payment scheme requires the ability to develop estimates of the probability of causation from epidemiologic data. It is commonly supposed that the probability of causation for exposed cases occurring at age t would equal $[R(t) - 1]/R(t)$, where $R(t)$ is the age-specific lung cancer rate in the exposed population divided by the

Key words: Attributable risk; Epidemiologic methods; Probability of causation; Risk assessment.

rate in a comparable unexposed population. In this paper we present a precise mathematical definition for the probability of causation that attempts to formalize the legal and ordinary-language meaning of the term. We will show that, under our definition, $[R(t) - 1]/R(t)$ can be interpreted as the age-specific probability of causation if for each follow-up time, the background hazard of disease (i.e., the disease hazard when unexposed) is the same for all subjects. Nevertheless, because of variation in unmeasured genetic and environmental factors, it is likely that there will be large between-subject heterogeneity in background hazards. As discussed in the above references and in detail below, $[R(t) - 1]/R(t)$ is not, except in unusual circumstances, the age-specific probability of causation in the presence of heterogeneity. It is for this reason that Lagakos and Mosteller (1986) called $[R(t) - 1]/R(t)$ the age-specific "assigned share" rather than the "probability of causation."

In an epidemiologic study, the age-specific disease rates in an exposed population and a comparable unexposed population can be directly estimated from the observed data. Unfortunately, the population disease rates determine only a range of possible values for the probability of causation among all exposed cases. Thus, the probability of causation is nonidentifiable in the sense that, even if we knew the exact values of the underlying disease-rate parameters, a range of values for the probability of causation would be consistent with these parameters. Henceforth, we shall assume that we effectively have an unlimited amount of data, so that additional sources of uncertainty due to sampling error can be ignored.

We will derive maximum and minimum values for the probability of causation consistent with observed epidemiologic data. Within this range, the value of the probability of causation depends on both the degree of heterogeneity in background risk and the type and degree of interaction between exposure and other unmeasured environmental and genetic risk factors. Because of our inevitably limited knowledge of the interaction of exposure with other risk factors and of the degree of background heterogeneity of risk, we cannot estimate the probability of causation from epidemiologic data without making unverifiable assumptions. When (as is almost always the case) competing risks are present, even stronger assumptions are needed.

The chief estimation problems that arise due to competing risks have been recognized in earlier literature on the probability of causation (e.g., Cox, 1987). The present discussion shows that the estimation problems remain even in the absence of competing risks, problems of study methods, or sampling error, and derives bounds for the probability of causation that can be estimated from epidemiologic data. Some of the present results are analogous to the corresponding results for etiologic fractions; this analogy follows directly from the fact that the etiologic fraction may be viewed as the probability of causation under a deterministic model (Greenland and Robins, 1988; Robins and Greenland, 1989). The results given here, however, are more general because they allow for individual probabilities other than 0 or 1.

2. Formalization of the Problem

Throughout, we assume that exposure occurs only at a time $t = 0$ (start of follow-up), so that as of time 0 each individual in the study cohort is either exposed or unexposed. We will base our development on the following formal theory of causation, which allows a precise mathematical characterization of the probability of causation for fixed exposures. We suppose that for each individual i and time t after exposure, there is a probability $s_{1i}(t)$ of remaining free of the study disease up to time t if one is exposed at $t = 0$, and a corresponding probability $s_{0i}(t)$ of remaining free of the study disease up to time t if one is never exposed. This is a stochastic version of Rubin's (1978) causal model. For now, we will assume that competing risks are absent, although we will later drop this assumption. We shall usually assume that the $s_{ki}(t)$ are differentiable. We then define the *individual*

hazards under exposure ($k = 1$) and nonexposure ($k = 0$) as $h_{ki}(t) = -ds_{ki}(t)/s_{ki}(t)dt = f_{ki}(t)/s_{ki}(t)$, where $f_{ki}(t) = -ds_{ki}(t)/dt$ is the probability density of disease at time t for an individual i with exposure status k . Note that, although observing the death time of an exposed individual provides some information about $h_{1i}(t)$, we have no direct observations concerning $h_{0i}(t)$ for exposed individuals. Likewise, we have no direct observations concerning $h_{1i}(t)$ for an unexposed individual.

The $h_{ki}(t)$ may be interpreted as follows. If individual i is at exposure level k and still disease-free at time t , then the probability that this individual will get disease in the next short period of time Δt is $h_{ki}(t)\Delta t$. Thus, our model is a stochastic model for individual outcomes. That is, the incidence time of individual i when exposed and when unexposed is random, rather than fixed. Some authors feel more comfortable with a deterministic model, i.e., a model in which for each individual i there is a time d_{1i} at which the disease would occur when exposed and a time d_{0i} at which disease would occur when unexposed. This deterministic model is actually a special case of the stochastic model—the special case in which $s_{ki}(t) = 1$ if $t < d_{ki}$ and $s_{ki}(t) = 0$ if $t \geq d_{ki}$.

We formally define a cohort study to be unconfounded if the distribution of the functions $h_{0i}(\cdot)$ in the unexposed population and in the exposed population are identical. [We write $h_{ki}(\cdot)$ when we are referring to the entire function rather than to the particular value $h_{ki}(t)$ that the function takes at time t .] An epidemiologic cohort study will be unconfounded if, at the time of exposure, the exposed and unexposed cohorts do not differ on unmeasured risk factors for disease. In an unconfounded cohort study the expected survival curve of the unexposed cohort will equal the expected survival curve that would have been observed in the exposed cohort if that cohort had been unexposed. Until the discussion section, we consider only unconfounded studies.

We will say there is *heterogeneity of background risks* if (as is almost always the case) the functions $s_{0i}(\cdot)$ vary across individuals. A point worthy of emphasis is that in most applications one should expect such heterogeneity to be quite severe, as there are likely to be unmeasured genetic (as well as environmental) factors that vary across individuals and strongly affect individual risk [as measured by $1 - s_{0i}(t)$]. Note that heterogeneity of background hazards [i.e., $h_{0i}(\cdot)$ varying across individuals] is equivalent to heterogeneity of background risks.

Define $r_i(t) = h_{1i}(t)/h_{0i}(t)$ and $p_i(t) = [h_{1i}(t) - h_{0i}(t)]/h_{1i}(t) = [r_i(t) - 1]/r_i(t)$. $r_i(t)$ is the individual hazard ratio at t for individual i . In the following sections we will present a number of results on estimating averages of the $p_i(t)$. To derive these results, we will assume that $h_{1i}(t) \geq h_{0i}(t)$ for each individual i and time t , which we will call Assumption (a). We will give a number of theorems on the estimation of population averages of the $p_i(t)$. Our interest in $p_i(t)$ arises from the following consideration, which we label Assumption (b): Suppose that, for any exposed subject surviving to t , the probability of contracting disease in the interval $[t, t + \Delta t)$ due to a stochastic mechanism not involving exposure approaches $h_{0i}(t)\Delta t$ as Δt goes to 0; that is, the probability that the individual contracts disease from a mechanism involving exposure approaches $[h_{1i}(t) - h_{0i}(t)]\Delta t$. Then, if Assumption (a) holds and the individual contracts the disease at t , $p_i(t)$ corresponds to the ordinary-language and legal notions of the probability that the individual's disease was caused by exposure.

Informally, our basic premise is that $p_i(t)$ represents the probability of causation, provided that $h_{1i}(\cdot)$ and $h_{0i}(\cdot)$ never cross, and exposure never blocks a causal mechanism leading to disease. These assumptions are satisfied by many but not all biological models. See Section 8 for further discussion. Throughout, we will suppose that Assumptions (a) and (b) are correct. We wish to stress that all of the mathematical results derived in this paper concern bounds on population averages of the $p_i(t)$ under Assumption (a). Our only use of Assumption (b) is to allow us to identify $p_i(t)$ with the ordinary-language and legal notions of the probability of causation.

Let $P(t)$ denote the average of the $p_i(t)$ over all exposed subjects in the population who contract disease at t . Under Assumptions (a) and (b), $P(t)$ is the probability that the disease of a randomly sampled exposed person contracting disease at t was caused by exposure. We thus have *two* types of time-specific "probability of causation": one, $p_i(t)$, applies to *individuals*; the other, $P(t)$, applies only to *populations*. We will show that, even in the absence of confounding, $P(t)$ is not identifiable from epidemiologic data. We can, however, estimate the population (cohort) survival probabilities $S_k(t) = E_k[S_{ki}(t)]$ and the incidence (population hazard) rates $H_k(t) = -dS_k(t)/S_k(t)dt$. E_k denotes an average over subjects in cohort k for $k \in \{0, 1\}$. Note that $S_k(t) = \exp[-\int_0^t H_k(u) du]$ is the proportion of population k surviving to time t .

For the moment, assume that $r_i(t) = r(t)$ is the same for all individuals i at all times t . Then the individual probability of causation $p_i(t) = p(t) = [r(t) - 1]/r(t)$ is the same for all exposed subjects contracting disease at t , and $P(t) = p(t)$. Now define the (population) rate ratio as $R(t) = H_1(t)/H_0(t)$ and the (population) rate fraction as $[H_1(t) - H_0(t)]/H_0(t) = [R(t) - 1]/R(t)$. If there is no confounding and no heterogeneity of background risks [i.e., $s_{0i}(\cdot)$ is the same for all individuals i], then, as shown below, $R(t) = r(t)$ and thus $p(t) = [R(t) - 1]/R(t)$. $[R(t) - 1]/R(t)$ is sometimes called the time-specific assigned share.

As we will discuss later, an average of $P(t)$ over time can be of interest in compensation issues. Let t_i be the time of disease onset for subject i . We define the *average probability of causation* P to be the average of the $p_i(t_i)$ over all exposed subjects who develop the disease over the follow-up period. Mathematically,

$$P = \frac{\int_0^x P(u)H_1(u)S_1(u) du}{1 - S_1(x)} = \frac{\int_0^x [H_1(u) - H_0(u)]S_1(u) du}{1 - S_1(x)},$$

where x is the time at which follow-up ends. Note that if $P(t)$ is a constant C over t , this constant C must equal P , since $\int_0^x H_1(u)S_1(u) du = 1 - S_1(x)$. As we will show below, P (like the time-specific probabilities) is not identifiable from epidemiologic data.

3. The Impact of Heterogeneity

In this section we will show that the population rate ratio $R(t)$ need not equal any weighted average of the individual hazard ratios $r_i(t)$. In fact, even if $R(t)$ is a constant R and the $r_i(t)$ equal a constant r , we will show that R does not in general equal r . These results have some important consequences for the estimation of the probability of causation. For example, suppose $r_i(t)$ is a constant r across individuals and time; then each of the probabilities of causation $p_i(t)$, $P(t)$, and P defined above are equal to the constant $(r - 1)/r$. In this situation, if $R(t)$ were also equal to a constant R , one might still hope that $(R - 1)/R$ would serve as a reasonable estimate of the probability of causation $(r - 1)/r$. We will however show that, when $R > 1$, any value of r greater than R is consistent with the observed data. For example, even when $R = 1.01$, r could be arbitrarily large and thus the true probability of causation $(r - 1)/r$ could be arbitrarily close to 1! It follows that the true probability of causation could be close to 1, even if the rate fraction, $(R - 1)/R$, were equal to $(1.01 - 1)/1.01 = .01$. This result has a close connection with the results of Hougaard (1986) on frailty distributions. Specifically, if $r_i(t)$ is a constant r over i and t , and for fixed j the ratios $h_{0i}(t)/h_{0j}(t)$ follow a stable distribution (Hougaard, 1986) and do not depend on t , then $R(t)$ will equal a constant R that is strictly less than r .

With $R = 1.01$, $(r - 1)/r$ could be close to 1 only in the presence of severe heterogeneity in the background risks. One class of biological models that is consistent with $R = 1.01$ and that gives such extreme heterogeneity is the following deterministic model: An

individual i contracts disease before an individual j when both are exposed (i.e., $d_{1i} < d_{1j}$) if and only if individual i contracts disease before individual j when both are unexposed ($d_{0i} < d_{0j}$). We call such a deterministic model *rank-preserving*, since an individual's ranking by incidence time when unexposed is preserved under exposure. Under such a model, $h_{0i}(d_{1i}) = 0$ for all individuals, and thus the probability of causation is 1 for each exposed case of the disease [since $h_{0i}(t) = 0$ means no mechanisms not involving exposure act at t].

From the above discussion, it might appear that, in an unconfounded epidemiologic study with $R > 1$, the estimable parameter $(R - 1)/R$ is a lower bound for the probability of causation. In fact, as we show next, this is not necessarily the case.

4. Bounds for the Probability of Causation

In this section we characterize the nonidentifiability of $P(t)$ and P by deriving a range for the possible values that are consistent with the distribution of observed data. The next section will discuss the implications of these results for compensation.

In a cohort study (with exposure occurring at $t = 0$), the probability distribution for the observed data can be equivalently characterized by (a) the time-specific rates $H_1(t)$ and $H_0(t)$ in the exposed and unexposed cohorts, (b) the survival curves $S_1(t)$ and $S_0(t)$ for the exposed and unexposed cohorts, or (c) the unconditional densities $f_1(t)$ and $f_0(t)$ for the exposed and unexposed cohorts, where $f_k(t) = H_k(t)S_k(t)$. If our cohort study is unconfounded and $h_{1i}(t) \geq h_{0i}(t)$ for all subjects at all times [i.e., Assumption (a) holds], then $S_1(t) < S_0(t)$ for all t such that $S_0(t) > 0$ and $t > t_b$, where t_b is the greatest lower bound of times such that $h_{1i}(t) > h_{0i}(t)$ for some subjects (t_b is sometimes called the minimum biologic induction period or minimum latent period). Nevertheless, Assumption (a) does *not* imply that $R(t) \geq 1$ for all t . Without loss of generality, in what follows we can assume that $t_b = 0$, since the time axis can always be right-translated by t_b (Robins and Greenland, Technical Report No. 36, Department of Biostatistics, Harvard School of Public Health, 1988).

Given that $S_1(t)$ and $S_0(t)$ (and functions of them) are the only statistically identifiable quantities, what bounds can we place on $P(t)$ and P ? The following two theorems provide these bounds, under the assumptions that the study is unconfounded and $h_{1i}(t) \geq h_{0i}(t)$ for all subjects at all times.

Theorem 1 For each t such that $f_1(t) > f_0(t)$, $P(t)$ can be anywhere in the closed interval $[(f_1(t) - f_0(t))/f_1(t), 1]$. At each time t such that $f_0(t) > f_1(t)$, $P(t)$ can be anywhere in the closed interval $[0, 1]$. If $R(t) > 1$, then the rate fraction $[R(t) - 1]/R(t)$ will always represent a possible value for $P(t)$.

Proof See the Appendix.

By Theorem 1, the lower bound for $P(t)$ is

$$\max\left[0, \frac{f_1(t) - f_0(t)}{f_1(t)}\right] = \frac{f_1(t) - \min[f_1(t), f_0(t)]}{f_1(t)}.$$

The bounds given in Theorem 1 still hold if deterministic models are excluded, but with open intervals replacing closed intervals. Furthermore, if $r_i(t)$ is a constant r over t and i and $R(t)$ is a constant R , then the rate fraction $(R - 1)/R$ will be a lower bound for the probability of causation $(r - 1)/r$: Under the stated assumptions, we have $(R - 1)/R = [h_1(0) - h_0(0)]/h_1(0) = [f_1(0) - f_0(0)]/f_1(0)$ and $(r - 1)/r = P(0)$, from which the result follows by Theorem 1.

Theorem 2 If the outcome of interest is death (from any cause), and follow-up continues until all exposed cohort members have died, then P can be anywhere in the closed interval $[\int_0^\infty \max[f_1(t) - f_0(t), 0] dt, 1]$. If $R(t) = R$ is constant over time, then P can be anywhere in the interval $[(R - 1)/[R^{R/(R-1)}], 1]$ and $(R - 1)/R$ will always lie within that interval.

Proof See Robins and Greenland (Technical Report No. 36, Department of Biostatistics, Harvard School of Public Health, 1988).

P attains its upper bound (of 1) under our rank-preserving model. P attains its lower bound (given in the preceding theorem) under a deterministic model in which exposure affects only those individuals whose survival time when exposed is less than the first time $t > 0$ for which $f_1(t) = f_0(t)$.

Theorem 1 shows that when $R(t) > 1$, the time-specific rate fraction $[R(t) - 1]/R(t)$ may overestimate or underestimate the time-specific probability of causation $P(t)$. In particular, $P(t)$ may be 1, whatever the value of $R(t)$. Since (as probability densities) $f_1(t)$ and $f_0(t)$ must integrate to 1, it must be true that, for some values of t , $f_1(t) < f_0(t)$ if $f_1(\cdot) \neq f_0(\cdot)$ and both densities are smooth. At any such value of t , $R(t)$ may be arbitrarily large (e.g., 100), making the time-specific rate fraction arbitrarily close to 1, and yet $p_i(t) = 0$ for all subjects dying at such t [so that $P(t) = 0$ as well].

We now show how models for the interaction of exposure with other risk factors can help predict whether $[R(t) - 1]/R(t)$ is an overestimate or an underestimate of $P(t)$. To do so we shall need the following definitions.

Definition 1 We say that the effect of exposure on hazards is *additive* if, at each time t , the difference between the exposed and unexposed hazards at t is the same for all individuals. That is, for all individuals i and j , and times t ,

$$h_{1i}(t) - h_{0i}(t) = h_{1j}(t) - h_{0j}(t). \quad (1)$$

Definition 2 We say that the effect of exposure is *superadditive* if

$$h_{1i}(t) - h_{0i}(t) \geq h_{1j}(t) - h_{0j}(t) \quad (2)$$

whenever $h_{0i}(t) > h_{0j}(t)$, with a strict inequality in expression (2) for some i, j , and t . That is, the amount exposure adds to the hazard *increases* with increasing background hazard.

Definition 3 We say that the effect of exposure is *subadditive* if inequality (2) holds whenever $h_{0i}(t) < h_{0j}(t)$, with strict inequality for some i, j , and t . That is, the amount exposure adds to the hazard *decreases* with increasing background hazard.

Definition 4 We say that a population is *well ordered* by hazard if, for all individuals i and j , either $h_{0i}(t) \geq h_{0j}(t)$ or $h_{0i}(t) \leq h_{0j}(t)$ for all t .

Note that none of the above four conditions (additivity, superadditivity, subadditivity, or well-ordering by hazard) is itself identifiable from epidemiologic data. Nevertheless, as shown in the following two theorems (proved in the Appendix), the above conditions have important implications for the identifiability of the probability of causation $P(t)$.

Theorem 3 If in the study population, the functions $\Delta_i(\cdot) = h_{1i}(\cdot) - h_{0i}(\cdot)$ are distributed independently of the functions $h_{0i}(\cdot)$, then $[R(t) - 1]/R(t) = P(t)$.

One situation in which the functions $\Delta_i(\cdot)$ and $h_{0i}(\cdot)$ would be independent is if the mechanisms by which exposure causes disease act independently of the mechanisms by which disease is caused in unexposed individuals.

Corollary 3.1 If the effect of exposure is additive, then (even when there is heterogeneity in the background hazards), $[R(t) - 1]/R(t) = P(t)$.

Corollary 3.2 If there is no heterogeneity in the background hazards, then $[R(t) - 1]/R(t) = P(t)$.

Theorem 4 If the effect of exposure is superadditive (subadditive) and the population is well ordered by hazard, then $[R(t) - 1]/R(t)$ is strictly less (greater) than the probability of causation $P(t)$. Nevertheless, even if the effect of exposure is known to be superadditive and well ordered by hazard, $P(t)$ may still be arbitrarily close to 1.

5. Potential Implications for Compensation

If awards in tort suits are to be made in proportion to the probability of causation, it appears reasonable that, in the absence of further covariate information, fair compensation for an exposed individual i observed to die at time t would be in proportion to the probability of causation $P(t)$ [since the individual probability $p_i(t)$ is unknowable]. Under such a payment scheme, the average award to exposed cases (by the time all the exposed individuals have died) will be the average probability of causation $P = \int_0^\infty P(t) f_1(t) dt$. Therefore, the total liability assumed by the defendant will be proportional to P . We have shown that such a fair award scheme cannot be constructed from epidemiologic data without making nonidentifiable biologic assumptions.

Specifically, it follows from Theorem 3 that, if the effect of exposure on disease is additive, the optimal payment scheme (that is, proportional to the probability of causation) is payment in proportion to $[R(t) - 1]/R(t)$, regardless of the degree of background heterogeneity. It follows from Theorem 4, however, that if the effect of exposure on disease is superadditive or subadditive, this scheme will not be optimal. It will also not be optimal under a rank-preserving deterministic model, which has $p_i(t) = 1$ for all exposed cases of disease occurring at t ; under this model, the optimal payment scheme with payments proportional to the probability of causation is full compensation to all exposed individuals, regardless of the value of the rate fraction $[R(t) - 1]/R(t)$.

To illustrate the potential conflicts produced by assuming different biological models, suppose $R(t) = R = 1.01$. If the effect of exposure is additive or there is no heterogeneity in the background rates, we would pay only $(1.01 - 1.00)/1.01 = 1\%$ compensation to all exposed cases; but if in fact a rank-preserving deterministic model holds, then the total payments will be 100 times too low (if payments are intended to be proportional to the probability of causation). On the other hand, if we assume a rank-preserving model holds, we would pay 100% compensation to all exposed cases; but if in fact either the effect of exposure is additive or there is no heterogeneity in the background rates, then the total payments will be 100 times too high. Thus, a payment scheme with awards proportional to the probability of causation is sensitive to both misspecification of exposure interaction with other risk factors and heterogeneity of the background rates. As we discuss elsewhere (Robins and Greenland, Technical Report No. 7, Occupational Health Program, Harvard School of Public Health, 1989), a more robust alternative payment scheme can be based on expected years of life lost.

6. Multiple Levels of Exposure

Suppose that exposure has K ordered levels indexed by k , with $k = 0$ corresponding to no exposure. Our previous notation for individual and population hazards and survival distributions extends in a straightforward manner to this situation; for example, $h_{ki}(t)$ is the hazard of individual i at time t when individual i is exposed to exposure level k , and

$H_k(t)$ is the rate in the subpopulation actually exposed to exposure level k . We continue to assume that any exposure $k > 0$ occurred by the time follow-up began.

Assume that our study is unconfounded, i.e., for all exposure levels m and n with $m > n$, at time 0 the distribution of the $h_{mi}(\cdot)$ is the same in the cohorts with exposure levels m and n . Also, assume that (a) $h_{mi}(t) \geq h_{ni}(t)$ if $m > n$ (i.e., at all times t , the individual hazards respond monotonically to exposure); and (b) for any subject exposed by $t = 0$ and surviving to t , the probability that the subject contracts disease by $t + \Delta t$ due to a mechanism not involving exposure approaches $h_{oi}(t)\Delta t$ as Δt approaches 0. Then, for a subject exposed at level k by time 0 and contracting disease at time t , the probability that the subject's disease was caused by exposure is $p_{ki}(t) = [h_{ki}(t) - h_{oi}(t)]/h_{ki}(t)$. Furthermore, the average of the $p_{ki}(t)$ over all individuals at exposure level k who were observed to contract disease at time t , $P_k(t)$, is the probability that the disease of a randomly sampled individual exposed to k and contracting disease at t was caused by exposure.

All our previous results concerning $P(t)$ generalize to $P_k(t)$ when we replace the functions $H_1(\cdot)$, $S_1(\cdot)$, and $f_1(\cdot)$ by the functions $H_k(\cdot)$, $S_k(\cdot)$, and $f_k(\cdot)$, except that the lower bound for $P_k(t)$ is given *not* by Theorem 1, but is instead given by

$$\{f_k(t) - \min\{f_{k'}(t): 0 \leq k' \leq k\}\}/f_k(t) \quad (3)$$

(Robins and Greenland, Technical Report No. 36, Department of Biostatistics, Harvard School of Public Health, 1988).

Example Suppose mortality data were available only on subjects exposed to k asbestos fibers/cc and on unexposed subjects. Then in effect we have but a single exposed and a single unexposed group and our previous results apply. Suppose now data on the mortality experience of subjects exposed at levels $k' < k$ also become available. Then none of our results concerning $P_k(t)$ will be affected by these additional data *except* that the lower bound for $P_k(t)$ will be given by expression (3) rather than by Theorem 1.

In the special case in which $H_k(t)/H_0(t)$ does not depend on t for all k , the lower bound in Theorem 1 [with $f_k(t)$ replacing $f_1(t)$] and expression (3) are identical, as Robins and Greenland show in their unpublished 1988 technical report.

7. Censoring Events, Competing Risks, and Covariates

In most situations there will be censoring events. For example, in a study of lung cancer mortality, deaths from causes other than lung cancer (competing risks) will be censoring events for lung cancer deaths. Even if death from any cause is the outcome of interest, we would still wish to use information on cause of death when estimating the probability of causation. We now examine estimation of the probability of causation for individuals contracting the disease of interest when there are censoring events.

We shall require some additional notation. We let d represent the outcome event of interest (e.g., lung cancer mortality) and c represent any censoring event (e.g., deaths from any other causes). Associated with each individual are four hazards $h_{1di}(t)$, $h_{1ci}(t)$, $h_{0di}(t)$, $h_{0ci}(t)$, where, for example, $h_{0di}(t)$ is individual i 's hazard at t of disease d in the absence of exposure.

For an exposed individual contracting disease d at time t , the probability of causation is $p_{di}(t) = [h_{1di}(t) - h_{0di}(t)]/h_{1di}(t)$ under the assumptions that (a) for each subject i and all times t , $h_{1di}(t) \geq h_{0di}(t)$ and $h_{1ci}(t) \geq h_{0ci}(t)$; and (b) for any subject exposed by $t = 0$ and surviving to time t , the probability that the subject experiences event d and the probability that the subject experiences event c in the interval $[t, t + \Delta t)$ from a stochastic mechanism not involving exposure approach $h_{0di}(t)\Delta t$ and $h_{0ci}(t)\Delta t$, respectively, as Δt goes to 0. These assumptions represent, in a stochastic setting, the informal idea that exposure does not

block any mechanism leading to event d or c , and thus are stronger than Assumptions (a) and (b) given earlier.

We formally define a study to be unconfounded if the joint distributions of the functions $h_{0di}(\cdot)$ and $h_{0ci}(\cdot)$ in the exposed and unexposed cohorts are identical. Henceforth, we suppose that our study is unconfounded and that Assumptions (a) and (b) hold.

For a randomly sampled exposed person observed to contract disease d at time t , the population probability of causation $P_d(t)$ is the average of the $p_{di}(t)$ over individuals observed to contract d at t . If an exposed individual i is observed to die of cause d at t , then, in the absence of further information, payment should be in proportion to $P_d(t)$ if one wishes to provide compensation in proportion to the probability of causation.

In the absence of heterogeneity in the $h_{0di}(t)$ and the $h_{0ci}(t)$, the population probability of causation equals the disease-specific rate fraction (which is identifiable); i.e., under homogeneity,

$$P_d(t) = [H_{1d}(t) - H_{0d}(t)]/H_{1d}(t) = [R_d(t) - 1]/R_d(t),$$

where $H_{1d}(t)$ and $H_{0d}(t)$ are the rates of disease in the exposed and unexposed populations (see Robins and Greenland, Technical Report No. 36, Department of Biostatistics, Harvard School of Public Health, 1988). Nevertheless, in the presence of heterogeneity, $P_d(t)$ will, in general, depend both on the mechanisms by which exposure causes events d and c , and on the degree of heterogeneity in the background hazard functions $h_{0di}(\cdot)$ and $h_{0ci}(\cdot)$.

In their technical report, Robins and Greenland (1988) provide the following generalization of Theorem 1.

Theorem 5 The maximum value of $P_d(t)$ consistent with observed data is 1. The minimum possible value of $P_d(t)$, $\min[P_d(t)]$, equals

$$\frac{H_{1d}(t)S_1(t) - \min[H_{1d}(t)S_1(t), H_{0d}(t)S_0(t)]}{H_{1d}(t)S_1(t)}, \quad (4)$$

where $S_1(t)$ and $S_0(t)$ are the probabilities of surviving to t without experiencing disease or competing risks in the exposed and unexposed populations.

Note that when $H_{1d}(t)S_1(t) > H_{0d}(t)S_0(t)$,

$$\frac{R_d(t) - 1}{R_d(t)} - \min[P_d(t)] = \frac{H_{0d}(t)}{H_{1d}(t)} \left[\frac{S_0(t)}{S_1(t)} - 1 \right]. \quad (5)$$

It follows that if $S_0(t) \approx S_1(t)$, $[R_d(t) - 1]/R_d(t)$ is a reasonable approximation to $\min[P_d(t)]$.

Define disease d and censoring c to be independent (in the population) if the joint conditional distribution of the individual disease-hazard functions $[h_{1di}(\cdot), h_{0di}(\cdot)]$ does not depend on the censoring functions $[h_{1ci}(\cdot), h_{0ci}(\cdot)]$. If c represents deaths from competing causes and d and c are independent, we shall say that we have independent competing risks. If d and c are known to be independent, then $\min[P_d(t)]$ is no longer given by expression (4) but rather by

$$\{f_{1d}(t) - \min[f_{1d}(t), f_{0d}(t)]\}/f_{1d}(t), \quad (6)$$

where $f_{kd}(t) = H_{kd}(t)S_{kd}(t)$, and $S_{kd}(t) = \exp(-\int_0^t H_{kd}(u) du)$.

If d and c are not independent, then it is possible that $H_{1d}(t) < H_{0d}(t)$ for all t , even if there is no confounding and exposure is never preventive (Slud and Byar, 1988). Whether or not risks are independent, the rate fraction $[R_d(t) - 1]/R_d(t)$ is a possible value for $P_d(t)$ whenever the rate fraction is positive.

Next, define the effect of exposure on the outcome event d or censoring c to be additive, superadditive, or subadditive if Definition 1, 2, or 3 holds with the subscript d or c added. Likewise, we will say that the population is well ordered by disease or censoring hazard if Definition 4 holds with the subscript d or c added. Robins and Greenland (unpublished technical report, 1988) provide the following generalization of Corollary 3.1:

Theorem 6 $P_d(t) = [H_{1d}(t) - H_{0d}(t)]/H_{1d}(t) = [R_d(t) - 1]/R_d(t)$ when the effect of exposure on both d and c is additive.

As the following generalization of Theorem 4 shows, the assumption of independence can have an important impact on bounds for the probability of causation $P_d(t)$.

Theorem 7 If (i) the effect of exposure on d is superadditive (subadditive), (ii) the population is well ordered by disease hazard, and (iii) either d and c are independent, or the effect of exposure on c is additive, then $P_d(t)$ is strictly greater (less) than $[R_d(t) - 1]/R_d(t)$.

For a proof, see Robins and Greenland's 1988 technical report. This theorem gives sufficient but not necessary conditions for $P_d(t)$ to be bounded below by $[R_d(t) - 1]/R_d(t)$.

Example The multistage model of cancer is one of the most commonly used models in risk assessment. Under the multistage model, the hazard of lung cancer in an unexposed person is a constant b times age raised to the k th power. b is a summary measure of the background rates at which individual cells progress through the multiple stages toward cancer. Due to variations in individual genetic factors, one would expect b to vary between individuals. Under the multistage model of cancer, the functions $h_{0i}(t) = b_i t^k$ are well ordered. The exposure effect in a multistage model is not, in general, strictly superadditive; nevertheless, if this model of cancer is correct, then one should expect $[R_d(t) - 1]/R_d(t)$ to underestimate the population probability of causation $P_d(t)$ (Robins and Greenland, technical report, 1988). Awards based on $[R_d(t) - 1]/R_d(t)$ would thus be too small if one desired to make payments in proportion to the probability of causation.

Suppose now that we have data on a covariate Z . Then, in the absence of further information, for a subject with covariate level z observed to contract disease d at time t , awards based solely on probability of causation should be paid in proportion to $P_d(t | z)$. For fixed Z , we can estimate bounds for the probability of causation $P_d(t | z)$ for the subset of individuals with $Z = z$ who contract disease d at time t , using our earlier results applied within covariate levels (under various assumptions about the effect of exposure on d and c within levels of the covariate).

8. Discussion

Throughout this paper we have employed only two assumptions concerning causation, essentially corresponding to the idea that exposure is never preventive. Under these assumptions, $p_i(t)$ formalizes the ordinary-language and legal notions of the probability of causation for an individual i who contracts disease at t . We have shown that, even in the absence of bias and misclassification, the average probability of causation $P(t)$ among exposed individuals developing disease at t cannot be estimated from epidemiologic data without resorting to nonidentifiable biologic assumptions. Furthermore, even in the absence of competing risks, the average probability of causation P among all exposed cases cannot be estimated. This is because these probabilities of causation depend on (i) the unknown mechanisms by which exposure affects disease risk and competing risks, (ii) the unknown degree of heterogeneity in the background risks of disease, and (iii) the unknown degree of

dependence between risk of disease and competing risks. We derived the maximum and minimum values for the probability of causation consistent with observable population distributions.

As discussed below, a major open question is how one should define and measure the probability of causation when one or both of our assumptions fail to hold. For example, how should we treat an exposure that can prevent some cases of disease and cause others? We consider these issues especially problematic, because Assumptions (a) and (b) cannot be tested with epidemiologic data.

Many other risk-assessment issues are raised in attempting to derive the implications of our results. For example:

1. If we suspect that most exposures interact superadditively with unmeasured genetic and environmental risk factors, should we assume that the probability of causation exceeds the time-disease-specific rate fraction (assigned share) $[R_d(t) - 1]/R_d(t)$? If so, how do we decide how much the probability of causation exceeds the time-disease-specific rate fraction?

2. How can measures such as probability of causation be partitioned into a component due solely to exposure and a component due to the joint effects of exposure and some other covariate, such as cigarette smoking? Should we or should we not modify payments based on the size of the respective components?

3. How can the probability of causation be defined and estimated in studies with sustained and time-dependent exposure and covariates? In such situations, even the basic measures of effect are difficult to define and estimate (Robins, Technical Report No. 8, Occupational Health Program, Harvard School of Public Health, 1988). Robins (Technical Report No. 37, Department of Biostatistics, Harvard School of Public Health, 1988) suggests some possible solutions.

4. In realistic epidemiologic studies, bias (e.g., uncontrolled confounding, selection bias), misclassification of exposure (and often disease), and sampling variability are inevitable. Bias and misclassification are especially difficult to quantify. How should the existence of sampling variability, misclassification, and poorly quantified bias affect compensation decisions?

5. In tort suits, award levels are often not based solely on the probability of causation. Award levels may also depend on the extent of damages caused by exposure. If, for a fatal outcome, damage is measured as expected years of life lost and payment is to be in proportion to damages, then the court's interest would be in estimating the expected years of life lost for exposed subjects dying at the age of death of the plaintiff. Elsewhere we consider the identifiability of measures based on years of life lost (Robins and Greenland, Technical Report No. 7, Occupational Health Program, Harvard School of Public Health, 1989).

6. Finally, when Assumption (a) or (b) does not hold, it is not always possible to draw a precise connection between the ordinary-language and legal meaning of causation and a formal definition of "probability of causation," such as ours. Consider a subject who ingests radioactive I^{131} following an unscheduled release of the isotope from a nuclear power plant. Suppose all thyroid cancer is a result of three separate heritable mutations and that I^{131} beta decay can cause only the first of the three mutations. Furthermore, suppose background beta decay is responsible for all other occurrences of the first mutation. If d is thyroid cancer, then $s_{1di}(t)$ and $s_{0di}(t)$ are truly stochastic, provided one accepts the quantum mechanical dictum that whether a particular radioactive atom undergoes beta decay is a truly stochastic event. If the ingested I^{131} does not affect the subject's chance of developing thyroid cancer except through its ability to cause the first mutation and does not affect mortality risk in any other way, $p_{di}(t_i)$ is the conditional probability that the stage-I mutation (and thus the subject's cancer) was due to beta decay of an I^{131} atom released

from the nuclear plant. In this setting, $p_{di}(t_i)$ has the ordinary-language and legal meaning of the probability of causation.

Now suppose that the ingested I^{131} kills a large fraction of the subject's thyroid cells. Suppose further that the risk of thyroid cancer is proportional to the number of thyroid cells, and that the "protective effect" of I^{131} due to cell killing outweighs the adverse effect due to increase in the first mutation rate, so that $h_{1di}(t) \leq h_{0di}(t)$ at all times t after exposure, *but* the subject contracts thyroid cancer due to a stage-1 mutation induced by an I^{131} atom released from the nuclear plant. In ordinary language, the I^{131} caused the subject's cancer, yet $p_{di}(t_i) < 1$, reflecting the fact that Assumptions (a) and (b) of Section 7 no longer hold. We do not know what attitude a judge or jury would take in awarding damages to the subject, even if they knew with certainty that the subject's cancer resulted from a first-stage mutation due to I^{131} from the nuclear power plant.

If the protective benefits of I^{131} outweighed its risks, we could even imagine rationally prescribing treatment with the isotope. Analogous situations in fact arise all the time in cancer chemotherapy and radiotherapy. Should a rationally correct medical decision be considered malpractice? Clearly not, and yet (as the preceding example shows) a change in intent yields a less clear-cut situation. Further examples show that a purely physical description of a hazardous mechanism is insufficient to clearly determine liability. In particular, nonscientific (e.g., ethical) issues regarding intent must also be considered to reach any sensible determinations (Greenland and Robins, 1988).

RÉSUMÉ

Cet article propose une définition mathématique de la probabilité de l'existence d'une relation de cause à effet entre deux événements afin de formaliser le sens que la langue ordinaire attribue généralement à cette notion. Avec cette définition, il montre que la probabilité moyenne de l'existence d'une telle relation de cause à effet chez des sujets exposés à un risque n'est pas pour autant calculable à partir des données épidémiologiques. Ceci provient du fait qu'une telle probabilité dépend à la fois des mécanismes non connus par lesquels l'exposition affecte le risque de maladie, et du degré non connu du risque de maladie lié au passé médical hétérogène de la population exposée. Les valeurs minimales et maximales de cette probabilité sont calculées en liaison avec les effectifs observés de la population; on donne également le lien entre l'excès du taux d'incidence, défini comme une proportion du taux d'incidence totale, et cette probabilité.

ACKNOWLEDGEMENTS

This work was in part supported by grants K04-ES00180, 5-P30-ES00002, and R01-ES03405 from NIEHS.

REFERENCES

- Cox, L. A., Jr. (1987). Statistical issues in the estimation of assigned shares for carcinogenesis liability. *Risk Analysis* 7, 71-80.
- Greenland, S. and Robins, J. M. (1988). Conceptual problems in the definition and interpretation of attributable fractions. *American Journal of Epidemiology* 128, 1185-1197.
- Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* 73, 387-396.
- Lagakos, S. W. and Mosteller, F. (1986). Assigned shares in compensation for radiation-related cancer (with Discussion). *Risk Analysis* 6, 345-380.
- NAS Oversight Committee (1984). *Assigned Share for Radiation as a Cause of Cancer: Review of Radioepidemiologic Tables Assigning Probabilities of Causation*. Report of the Oversight Committee on Radioepidemiologic Tables. Washington, D.C.: National Academy Press.
- NIH Working Group (1984). *Report of the National Institutes of Health Ad Hoc Working Group to Develop Radioepidemiological Tables*. Publication No. 85-2748. Bethesda, Maryland: National Institutes of Health.
- Robins, J. M. and Greenland, S. (1989). Estimability and estimation of excess and etiologic fractions. *Statistics in Medicine* 8, 845-859.

Rubin, D. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6, 34–58.

Slud, E. and Byar, D. (1988). How dependent causes of death can make risk factors appear protective. *Biometrics* 44, 265–269.

Received July 1988; revised March 1989.

APPENDIX

We first prove Theorem 1 in a series of lemmas. We will make use of the following assumptions:

- (1) for each individual i and all times t , $h_{1i}(t) \geq h_{0i}(t)$;
- (2) the epidemiologic study is unconfounded (as defined in Section 2);
- (3) $h_{1i}(t) > h_{0i}(t)$ over some nonzero interval of time for a nonnegligible fraction of the exposed population; and
- (4) there is no minimum induction period, i.e., the first interval over which $h_{1i}(t) > h_{0i}(t)$ begins at $t = 0$. (The last assumption is inessential and is for mathematical convenience only, since we can always redefine time 0 as the first time for which $h_{1i}(t) > h_{0i}(t)$ for some individuals over some nonzero interval.)

Lemma A.1 Assumptions (1)–(4) imply that $S_0(t) > S_1(t)$ if $t > 0$ and $S_0(t) > 0$.

Proof From Assumption (1) we have $s_{1i}(t) \leq s_{0i}(t)$ since $h_{1i}(t) \geq h_{0i}(t)$. Therefore, $S_1(t) \equiv E[s_{1i}(t)] \leq E[s_{0i}(t)] = S_0(t)$ where E , denoting expectation, is in this Appendix a simple average over the subjects i in the *exposed* population. [Since $S_0(t)$ is, by definition, the survival curve of the *unexposed* cohort, $E[s_{0i}(t)] = S_0(t)$ holds only if the study is unconfounded.] Assumptions (3) and (4) imply that the inequality is strict whenever $t > 0$ and $S_0(t) > 0$.

Lemma A.2 Under Assumptions (1)–(4),

$$P(t) = \max\{[f_1(t) - f_0(t)]/f_1(t), 0\}.$$

Proof By definition,

$$P(t) = E\left[\frac{p_i(t)f_{1i}(t)}{f_1(t)}\right] = 1 - \frac{E[h_{0i}(t)s_{1i}(t)]}{f_1(t)}.$$

Therefore, for a given $f_1(t)$ and $f_0(t)$, $P(t)$ is minimized when $E[h_{0i}(t)s_{1i}(t)]$ is maximized. But

$$E[h_{0i}(t)s_{1i}(t)] = E\left[f_{0i}(t)\left(\frac{s_{1i}(t)}{s_{0i}(t)}\right)\right] \leq E[f_{0i}(t)] = f_0(t)$$

by Assumption (1). Also,

$$E[h_{0i}(t)s_{1i}(t)] = E\left[f_{1i}\left(\frac{h_{0i}(t)}{h_{1i}(t)}\right)\right] \leq E[f_{1i}(t)] = f_1(t)$$

by Assumption (1). Therefore, $E[h_{0i}(t)s_{1i}(t)] \leq \min[f_0(t), f_1(t)]$ and so

$$P(t) \geq \max\{[f_1(t) - f_0(t)]/f_1(t), 0\},$$

which proves the lemma. This bound is shown to be sharp in Robins and Greenland (1989).

Lemma A.3 Given $S_0(t)$ and $S_1(t)$, it is possible that $p_i(t) = 1$ for all subjects i .

Proof Given Assumptions (1)–(4), we cannot rule out the rank-preserving deterministic model with $p_i(t) = 1$.

Lemma A.4 If $R(t) > 1$, then $[R(t) - 1]/R(t) > [f_1(t) - f_0(t)]/f_1(t)$.

Proof

$$\frac{R(t) - 1}{R(t)} = \frac{H_1(t) - H_0(t)}{H_1(t)} = \frac{H_1(t)S_1(t) - H_0(t)S_1(t)}{H_1(t)S_1(t)} = \frac{f_1(t) - f_0(t)S_0(t)/S_1(t)}{f_1(t)} > \frac{f_1(t) - f_0(t)}{f_1(t)},$$

when $R(t) > 1$ and $S_0(t) > S_1(t)$. But, by Lemma A.1, $S_0(t) > S_1(t)$.

Taken together, Lemmas A.2, A.3, and A.4 prove Theorem 1.

Proof of Theorems 3 and 4

$$P(t) = 1 - \frac{E[h_{0i}(t)s_{1i}(t)]}{S_1(t)H_1(t)} \quad (\text{A.1})$$

and

$$[R(t) - 1]/R(t) = 1 - H_0(t)/H_1(t). \quad (\text{A.2})$$

Therefore, it follows from equations (A.1) and (A.2) that

$$\begin{aligned} P(t) > \frac{R(t) - 1}{R(t)} &\Leftrightarrow H_0(t) > \frac{E[h_{0i}(t)s_{1i}(t)]}{S_1(t)} \\ &\Leftrightarrow \frac{E[h_{0i}(t)s_{0i}(t)]S_1(t)}{S_0(t)} > E[h_{0i}(t)s_{1i}(t)] \\ &\Leftrightarrow \frac{E[h_{0i}(t)s_{0i}(t)]}{S_0(t)} \frac{E\{[s_{1i}(t)/s_{0i}(t)][s_{0i}(t)]\}}{S_0(t)} > \frac{E[h_{0i}(t)s_{1i}(t)]}{S_0(t)} \\ &\Leftrightarrow E^*[h_{0i}(t)]E^*\left[\frac{s_{1i}(t)}{s_{0i}(t)}\right] > E^*\left[\frac{h_{0i}(t)s_{1i}(t)}{s_{0i}(t)}\right] \end{aligned}$$

where E^* refers to expectation over a distribution in which subject i is chosen with probability p^* proportional to $s_{0i}(t)/S_0(t)$. That is,

$$P(t) > [R(t) - 1]/R(t) \Leftrightarrow \text{cov}^*[h_{0i}(t), s_{1i}(t)/s_{0i}(t)] < 0. \quad (\text{A.3})$$

Now, for fixed t , define $X_i = h_{0i}(t)$. We shall consider the case in which, for any t , $X_i \neq X_j$ if $i \neq j$. When the effect of exposure is superadditive and subjects are well ordered by hazard, we can write $s_{1i}(t)/s_{0i}(t)$ as a monotone decreasing function $g(X_i)$ of X_i , since if $h_{0i}(t) > h_{0j}(t)$, then

$$\frac{s_{1i}(t)}{s_{0i}(t)} = \exp\left[-\int_0^i [h_{1i}(u) - h_{0i}(u)] du\right] < \exp\left[-\int_0^i [h_{1j}(u) - h_{0j}(u)] du\right] = \frac{s_{1j}(t)}{s_{0j}(t)}.$$

But if $g(X_i)$ is a monotone decreasing function of X_i , $\text{cov}[X_i, g(X_i)] < 0$ under any distribution including p^* . This proves Theorem 4 in this case. If X_i can equal X_j for $i \neq j$, the proof is similar but with some technical modifications.

Now by the same argument leading to expression (A.3), it is clear that

$$P(t) = [R(t) - 1]/R(t) \Leftrightarrow \text{cov}^*[h_{0i}(t), s_{1i}(t)/s_{0i}(t)] = 0.$$

But since $s_{1i}(t)/s_{0i}(t)$ is a function of only $\Delta_i(t)$, this will follow from the supposition of Theorem 3.

Corollary 3.1 now follows by noting that if the effect of exposure is additive, $\Delta_i(\cdot)$ is the same for all subjects i , and thus $\Delta_i(\cdot)$ is independent of $h_{0i}(\cdot)$. Similarly, Corollary 3.2 follows by noting that if the $h_{0i}(\cdot)$ are the same for each subject, then the $h_{0i}(\cdot)$ and the $\Delta_i(\cdot)$ are distributed independently and the supposition of Theorem 3 holds.