Risk Analysis, Vol. 4, No. 3, 1984

Probability of Causation and the Attributable Proportion of Risk

Louis Anthony Cox, Jr.[1]

Received October 17, 1983; revised March 5, 1984

A uranium miner who smokes develops lung cancer: what is the probability that radiation, rather than tobacco, caused it? This paper briefly explains the principles and limits of probability models for which this question makes sense, and then shows how principles of risk accounting can be applied to obtain a solution to the general problem of attributing risk in the presence of joint, possibly interacting, causes. A procedure for calculating each factor's "share" in a jointly caused risk is proposed, and shown to be a generalization of the "probability of causation" concept. Problems of implementation and interpretation for the proposed attribution procedure are discussed, and illustrative error bounds are derived for a simple decision rule, in which probability of causation or attributable risk share calculations are made using aggregate data as a proxy for unknown individual data.

KEY WORDS: Probability of causation; attributable risk; risk attribution; risk accounting; joint causes.

1. INTRODUCTION

Some of the most perplexing difficulties in the theory of tort-law liability arise when the cause of a plaintiff's injury cannot be uniquely identified. This can happen when more than one factor caused the damage—the problem of *joint causes*—or when there are several potential causes that could have produced the injury, and it is not known which one did so—the problem of *indeterminate cause*.[1-3] Both sorts of indeterminacy can be illustrated by suitably modifying the classic case of *Summers vs. Tice*.[1] Two careless hunters fired in the direction of the plaintiff, who was injured in consequence. Suppose that both bullets had struck him, and that he had died as a result, although (n)either bullet alone would have sufficed to kill him; then determining liability for his death is a problem in joint causes. What actually happened is that only one bullet (actually, a piece of shot) struck him, but it could not be determined which hunter had fired it, thus making the liability problem one of indeterminate cause. (The Court held both hunters liable as joint tort feasors.)

The problems of liability and causation raised by *Summers vs. Tice* occur routinely in health physics, where the source of a health effect may be uncertain, or may be distributed among multiple contributing factors. The paradigmatic example concerns a uranium miner who smokes cigarettes: to what extent should his employer be held liable if this miner develops lung cancer?[2] A similar example now before Congress concerns compensation of cancer victims who may have been exposed to fallout from the nuclear weapons testing program in the Southwest. What share of responsibility for such an individual's cancer should the government accept? In these and similar cases, a factor such as radiation increases the *incidence rate* of damages (e.g., cancer) in the exposed population, adding to the *probability*

[1]Arthur D. Little, Inc., 35-329 Acorn Park, Cambridge, Massachusetts 02140, and Operations Research Center, E-40, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

0272-4332/84/0900-0221$03.50/1

that each exposed individual will be damaged, but for any specific individual it remains unlikely that his particular injury was caused solely or primarily by the factor in question. Thus, it may be virtually certain that someone is damaged by a plant's releases to the environment, but there may be no way of identifying who has been damaged—the case of an "indeterminate plaintiff."[3]

In such cases, establishment of rational policies toward liability and compensation requires the best available tools of risk analysis to help allocate the responsibility for an individual's health effect among the causes that may have contributed to it. One such tool is the Probability of Causation (PC) approach to cases involving indeterminate cause.[4,5] But this approach assumes that the causes are disjoint and must be generalized to deal with cases involving joint causes.[6] A principal goal of this paper is to present such a generalization.

We shall be interested in using formal principles of risk accounting to allocate a risk among the factors that may have produced it. In so doing, we shall explain the principles and limitations of the PC approach, and show how to extend it to account for joint causes. To tie our results to the real-world context of health damages from indeterminate or multiple sources, we will use a simple competing-risk model of health effects. The emphasis is intentionally put on simplicity rather than on biological realism, since our goal here is not to model the mechanisms of carcinogenesis, but to clarify the conceptual foundations of the PC approach and its extensions, and to show when it is appropriate to use this approach and when not.

Section 2 introduces the simplest model of health damage that is able to meet our expository needs, and derives the probability of causation formula associated with this model. The need for a more general approach is also demonstrated, and necessary conditions for the PC approach to be valid are discussed. Section 3 then turns to a different approach, based on three principles or "axioms" of risk accounting, that yields a unique procedure (under certain technical conditions) for allocating risk among multiple causes or suspected causes. It is shown that the PC formula of Section 2 can be derived as a special case of this procedure. Finally, Section 4 discusses limitations and implementation difficulties for both approaches, and examines the consequences of trying to apply them without adequately detailed data.

2. COMPETING RISKS AND THE PROBABILITY OF CAUSATION

2.1. Mutual Exclusivity

Suppose that a worker is exposed to N sources of cancer, and let X_i denote his exposure to the ith source (e.g., the magnitude of the dose he receives from Source i). Given the exposure profile $(X_1, X_2, \ldots, X_N)$, and supposing that a cancer has occurred, what is the probability that Source i caused it?

Before answering this question, we must make sure that it makes sense. The question assumes that some one source is the sole cause of the cancer: exactly one of the N sources caused it, and the only uncertainty is, which one? In other words, causes are assumed to be mutually exclusive: If Source i caused the cancer, then Source j did not. If this assumption is false, then the question "what is the probability that Source i caused the cancer?" is misformulated, because it embodies a mistaken presumption about the way the world works.[2]

The question must be re-asked, perhaps as "What is the probability that Source i *helped* cause the cancer?", or as "What *share* of responsibility for the cancer can properly be attributed to Source i?"

For the moment, however, we will assume that causes are disjoint, so that the question makes sense as stated. This assumption of mutual exclusivity can be expressed by using a "competing risk" model of health effects.[7]

2.2. Competing Risks

In the simplest competing risk model, the N sources are thought of as bombarding the exposed individual with invisible bullets or "hits." The intensity of bombardment from Source i is denoted by X_i, and is defined as the average number of hits per second from Source i; for simplicity, we assume that X_i remains constant over time, and that hits from Source i occur according to a purely random (i.e., Poisson) process with average arrival rate X_i. Each hit that contacts individual j has a probability p_j,

[2]Logically, there is a "referential failure," much as in the question "What is the probability that the present King of France will die on February 31?"

depending on the individual, of initiating a cancer. Once initiated by some hit, the cancer has a probability q_j, again depending on the individual, of developing into a full-grown tumor under the influence of the body's endogenous promoters.

Given this "Poisson competing risks" model, and assuming that some individual is observed to have a tumor, what is the probability that Source i caused it? Equivalently, what is the probability that the first hit to successfully initiate a tumor comes from Source i? This question can be answered using the elementary theory of stochastic processes,[8] which shows that the conditional probability of being "successfully" hit first by Source i, given the exposure intensity profile $X_1, \ldots, X_N$ (and *any* exposure duration) is just

$$p_j q_j X_i / (p_j q_j X_1 + \cdots + p_j q_j X_N) = X_i / (X_1 + \cdots + X_N)$$

This quantity may be denoted by PC_i; it is called the *probability of causation* for Source i. Formally, then, we have a definition:

- $PC_i \triangleq Pr$(Source i caused Cancer C|Cancer C has occurred);

and also a theorem:

- Under the assumption of Poisson competing risks,

$$PC_i = X_i / (X_1 + X_2 + \cdots + X_N) \quad (1)$$

Both require the assumption of mutually exclusive causes, to be interpretable.

PC_i does not depend on the individual susceptibility factors, p_j and q_j. If the distribution of these factors in an exposed population remains fixed, then X_i can be reinterpreted as the increase in the incidence rate of cancer caused by Source i, since incidence rate is proportional to hit rate in this simple model. In epidemiological terms, X_i is the *attributable risk*[9] due to Source i (after rescaling from hits per second to cancers per person-year of exposure), and the probability of causation ratio in Eq. (1) can be interpreted as a ratio of attributable risks.

2.3. Need for a More General Approach

The above competing risk model, in which each hit has a chance $p_j q_j$ of causing cancer, is probabilistically equivalent to a Poisson competing risk model in which the hit intensities are reduced from X_i to $X_i' = p_j q_j X_i$, and each hit that occurs is now *certain* to cause a cancer. (PC_i, of course, remains unchanged by such "binomial censoring.") Thus, the probability model on which the PC ratio, Eq. (1), is based is equivalent to a *one-hit* model, in which the first source to "hit" someone gives him cancer. (Indeed, we could simply define a hit as a cancer-causing event.) In this model, Eq. (1) gives the probability of causation for Source i.

Suppose that we generalize to a K-hits model, in which a cancer occurs as soon as K hits have been received, but no sooner.[3] (Conceivably, the susceptibility threshold K could vary from individual to individual, or we could specify that the K hits must all occur within a certain time interval in order to achieve a "hit density" sufficient to provoke a health effect, but these variations would not alter the following points.) How would Eq. (1) change?

As soon as we consider the question, it becomes apparent that a new concept is needed. In the one-hit case, knowing which source fired the first successful hit is sufficient to tell us that Source i *caused* the resulting cancer. In a K-hits model, however, knowing exactly how many hits were received from each source would still not allow us to pick out one source as "the" cause of cancer (unless all K hits happened to come from the same source—an event whose probability is $[X_i/(X_1 + \cdots + X_N)]^K$ for Source i). Since even *complete* factual information would not allow us to give a yes–no answer to the question "Did Source i cause the cancer?" if Source i contributed some of the K-hits while other sources contributed the rest, asking for the *probability* that Source i caused the cancer, in the absence of complete information, would be bootless. It is the concept of "cause" itself that requires revision.

3. A THEORY OF ATTRIBUTION

To deal with cases where more than one source, cause, or factor contributes to the occurrence of an undesirable event, we need an idea of *partial* causation. This has nothing to do with uncertainty about facts, and must not be confused with the probability of causation idea introduced above for disjoint causes. Instead, it refers to the *proportion* or share of a total

[3] This "cumulative risk" model may not be appropriate for cancers, but is plausible for many other risks.[8]

risk that is properly attributable to each of the factors producing it. The broader concept of attributable proportion of risk must replace the notion of probability of causation when there are joint causes.

But how can this construct of "attributable proportion of risk" be defined, measured, or interpreted for practical use without indulging in purely subjective value judgments? One approach is to examine the consistency properties, or risk accounting axioms, that we would expect attributable proportions to obey, based on intuitive notions of what makes a risk (at least partially) attributable to a factor, and then examine the restricted set of measures, if any, that are logically consistent with these desired properties.

Suppose that the total risk produced by exposure profile $X = (X_1, X_2, \ldots, X_N)$ is some number $R(X)$, where X denotes the N-vector of factor levels that an individual is exposed to (i.e., the exposure profile), and $R(X)$ is the resulting level of risk. For example, $R(X)$ might represent the probability that an individual exposed to X will get cancer; other interpretations are possible for other contexts. For simplicity, we make the convenient mathematical assumptions that:

(1) $R(X)$ and the X_i are measured on numerical scales (i.e., they are numbers rather than, say, distributions) and both X and $R(X)$ are nonnegative.
(2) $R(X)$ is a smooth nondecreasing function of its arguments (having continuous first partial derivatives). Thus, increasing the level of exposure to any factor never decreases—but may increase—the resulting risk.
(3) $R(X)$ has been normalized so that $R(0) = 0$, perhaps after subtracting out a constant "background" level of risk not attributable to any of the components of X.

Given $R(X)$ and X, we want to find the proportion of the total risk, $R(X)$, that is attributable to Source i. This "attributable proportion" will be denoted by p_i, and we require that $p_1 + p_2 + \cdots p_N = 1$. In general, p_i may depend both on the response function $R(X)$ and on X itself, and so may be written as $p_i(R, X)$ to emphasize this dependency.

What properties should the attributable proportions of risk assigned to different factors obey? Two obvious ones are

- $p_i \geq 0$ for $i = 1, \ldots, N$ (no "negative" shares); and
- $p_1 + p_2 + \cdots + p_N = 1$ (completeness of attribution).

In addition, we impose three more substantive conditions, as follows. Let $p_i(R, X)$ be written as $a_i(R, X) \cdot X_i / R(X)$, where $a_i = a_i(R, X)$ is the *per-unit* risk rate attributed to Source i.

- *Axiom 1* (Independence of Accounting): If $R(X)$ is the sum of two component functions, $R_a(X)$ and $R_b(X)$, then $a_i(R, X) = a_i(R_a, X) + a_i(R_b, X)$.
- *Axiom 2* (Scale Invariance): The proportion p_i of risk attributed to Source i should not depend on the units used to measure the exposure levels $X_1, X_2, \ldots, X_N$.
- *Axiom 3* (Aggregation Invariance): If $R(X)$ depends on X_i and X_j only through their sum, $X_i + X_j$, (so that X_i and X_j act as if they were part of the "combined" factor $X_i + X_j$), then $p_i(R, X) + p_j(R, X)$ also depends on X_i and X_j only through their sum, and $a_i = a_j$.

These conditions are discussed and interpreted in terms of the per-unit risk rates a_i in Ref. (10), and their intuitive motivation is explained in Refs. (10–12).

These three "risk accounting axioms" can be shown mathematically to imply a host of other intuitively desirable properties, such as that no portion of the risk $R(X)$ is attributed to Factor i unless the level of risk depends on X_i. Indeed, we have the following strong result.

Theorem 1: There is exactly one measure of "attributable proportion of risk" that satisfies Axioms 1–3 for any risk function $R(X)$ meeting our assumptions.

It is given by

$$p_i(R, X) = \frac{a_i^* X_i}{a_1^* X_1 + a_2^* X_2 + \cdots + a_N^* X_N} \qquad (2)$$

where a_i^* is the average marginal risk rate for Factor i as exposure increases uniformly from 0 to X. Sym-

bolically, a_i^* may be written as

$$a_i^* = \int_0^1 R^i(tX)\,dt \qquad (3)$$

where $R^i(\cdot)$ denotes the marginal risk of Factor i (i.e., the partial derivative of $R(\cdot)$ with respect to its ith argument). Note that

$$a_1^* X_1 + a_2^* X_2 + \cdots + a_N^* X_N = R(X).$$

The proof of this result is discussed in Refs. (10–12). Its significance for our purposes is that Eq. (2) provides a generalization of the PC formula that holds even for interacting joint causes. Of course, the attributable risk proportion p_i cannot be interpreted as a probability except in special cases, but it may nonetheless be used as a measure of the share of responsibility for a jointly caused risk that is attributable to Source i.

To show that the "attributable proportion" concept, p_i, is in fact a generalization of the "probability of causation," we will show how Eq. (1) can be derived as a special case of Eq. (2).

In the one-hit Poisson competing risk model of Section 2, the probability that an individual exposed to "effective" hit rates $X_1, X_2, \ldots, X_N$ from Sources $S_1, S_2, \ldots, S_N$, respectively will *not* have been hit (i.e., will not have received a cancer) by time T is $\exp-(X_1 + \cdots + X_N)T$.[8] Therefore, the probability that he *will* have received a cancer by T is described by the risk function $R(X) = 1 - \exp-(X_1 + \cdots + X_N)T$. Performing the differentiation and integration in Eq. (3), we find that $a_i^* = R(X)/[X_1 + \cdots + X_N]$, and substituting this into Equation (2) gives the final result, $p_i(R, X) = X_i/[X_1 + \cdots + X_N]$. Thus, we have:

> *Theorem 2*[4]: In the case of a one-hit Poisson competing risk model, the proportion of risk attributable to Source i [according to the definition of attributable proportion in Eq. (2)] is just the conditional probability that Source i caused the damage, given that damage has occurred.

What is perhaps more interesting is:

> *Theorem 3*[4]: In any K-hits Poisson competing risk model, Eqs. (2) and (3) still give $p_i(R, X) = X_i/[X_1 + \cdots + X_N]$ as the *proportion* of risk attributable to Source i.

For $K > 1$, however, this quantity can no longer be interpreted as the probability that Factor i caused the cancer.

Similarly, if risks are described by the logistic dose-response function $R(X) = [1 + \exp-(\beta_1 X_1 + \cdots + \beta_N X_N)]^{-1} - .5$, then it can be shown that $p_i(R, X) = \beta_i X_i/[\beta_1 X_1 + \cdots + \beta_N X_N]$. (These and similar results follow directly from Axiom 3.)[10] Note that in the case of the logistic dose-response curve, however, *there is no simple relation between the attributable proportion of risk, $p_i(R, X)$, and the increase in incidence rate* (the epidemiologist's "attributable risk") *due to Source i.* The increase in incidence rate induced by X_i is sensitive to what other factors are present, and Eq. (2) holds these other factors partially accountable for the increase. This is characteristic of Eq. (2): it charges factors for their "negative externalities" when there is synergism among them, rather than simply holding each factor accountable for the portion of risk that would disappear if it were removed.[10] This is in keeping with our *a priori* consistency condition that 100% of the total risk, and no more, be apportioned among the factors.

Theorem 1 suggests an integrated approach to the attribution of risk in the presence of joint and indeterminate causes. If $R(X)$ denotes the probability of a health effect, given exposure X, and if a joint probability density function $f(X)$ can be assessed, then a corresponding joint probability distribution for the $p_i(R, X)$ can be deduced by the usual probabilistic technique of changing variables. A discrete-case example of this approach is given in Section 4.3. If X is known, as in the K-hits model, then the distribution of $p_i(R, X)$ collapses to a single number: both the multiplicity and the indeterminacy of causes are accounted for in the process [Eqs. (2) and (3)] of going from $R(\cdot)$ and X to $p_i(R, X)$, assuming that $R(\cdot)$ and X are known.

4. LIMITATIONS AND IMPLEMENTATION DIFFICULTIES

We have shown that the "ratio of attributable risks" formula for probability of causation given by

[4]Assuming that $R(X) = Pr(\text{Hit by time } T | X)$ for some T. This assumption is discussed further in Section 4.2.

Eq. (1) (with the X_i terms interpreted as being proportional to attributable risks) has a much broader justification than the probability of causation idea alone supports; but that this justification fails in non-additive risk models, like the logistic dose-response model, where the X_i exposure terms are not proportional to resulting risks (increased incidence rates). In such cases, the "attributable proportion of risk" formula, Eq. (2), may offer a satisfactory alternative to the *PC* formula. There are several difficulties standing between this theoretical formula and practical applications, however, that the would-be practitioner must be prepared to deal with. These difficulties face the *PC* formula as well.

4.1. Not a Theory of Compensation

Both Eq. (1) and Eq. (2) deal with the attribution of risk to *factors* (e.g., radiation, cigarette smoking, and so forth) appearing in the exposure profile. Neither goes on to consider who should be held socially responsible for these factors. Should an individual who chooses to smoke cigarettes (or work in a nuclear power plant) with full knowledge of the associated health risks be able to recover from the cigarette manufacturer (or reactor operator) if he develops cancer? Or should he bear the full consequences of his decision? Eqs. (1) and (2) provide no guidance for answering such questions: they only tell what portion of the cancer risk is attributable to each factor. To determine who should pay compensation for a cancer (if anyone) and how much should be paid, it is necessary to look beyond just the attribution of risk, to nonbiological factors such as intent, knowledge of consequences, voluntary and informed assumption of risk, relative cost of risk avoidance, and incentive effects of compensation rules.[1, 2, 6] A theory of risk attribution is necessary but not sufficient for developing wise compensation policies.

4.2. Ambiguity of Probability

Given an exposure profile X, and a risk function $R(\cdot)$ relating X to the resulting probability of cancer, Eq. (2) gives a way of determining the proportion of the risk $R(X)$ that is attributable to each factor X_i, $i = 1, 2, \ldots, N$. But how is the risk function or dose-response relation to be defined and determined for practical applications? $R(\cdot)$ may be thought of as a *biological model* relating exposure levels to cancer probabilities, and the proportions or probabilities defined by Eqs. (1) and (2) are interpretable only in terms of such models; they have no independent validity. But the validity of a probability model $R(\cdot)$ can never be unambiguously determined.

The difficulty lies with the concept of probability, and springs from the fact that no unambiguous "objective" probability of an event can ever be determined, if by "probability" we mean an additive measure satisfying the Kolmogorov axioms on a field of events.[13] The reason is that an infinity of measures (i.e., assignments of probability numbers to events) satisfy the axioms of probability on any field of events, and no number of observations suffices to single out a unique "correct" measure, except in degenerate cases. In Bayesian terms there is no "correct" way of assigning priors. Thus, probabilities are underdetermined by observations. This is in contrast to the naive view, sometimes taken in risk analysis, that regards probability as a property of events (rather than of fields of events), and which seeks to measure "the" probabilities of events in isolation, much as a physicist might try to measure the weights of objects. When we recognize that probability is not a property, but simply any assignment of numbers consistent with the axioms of probability theory, the futility of this quest for "objective" probability becomes apparent.

In some cases, of course, there are enough observations so that the posterior probabilities assigned to events are virtually independent of the priors, thus making the indeterminacy of the initial probability measure unimportant.[14] But in the case of a biological model, $R(\cdot)$, "the" probability assigned to getting cancer, given an exposure X, is apt to keep changing as more information is gained about the exposed individual and the details of his exposure. Given only that the victim is a randomly selected member of the population, one choice of $R(\cdot)$ might seem appropriate; given that she is female, a more refined model $R'(\cdot)$ might be used; information about her age, race, and blood type could lead to still more appropriate models, $R''(\cdot)$, $R'''(\cdot)$, and $R''''(\cdot)$; and still the list of relevant facts influencing "the" probability $R(X)$ of cancer for this individual, given exposure X, goes on. There is no reason to expect that this sequence of models will in any sense converge in its predictions, and both time and resources for information collection would surely be exhausted before all relevant facts for determining $R(\cdot)$ could be assembled.

The limiting case of this argument can be illustrated for the Poisson competing risk K-hit model with $K = 2$. Assume that there are only two sources, with intensities X_1 and X_2, respectively, so that the prior probability of getting cancer by time T is $R(X_1, X_2) = 1 - \exp-(X_1 + X_2)$. Then Eqs. (2) and (3) give $X_i/(X_1 + X_2)$ as the proportion of risk attributable to Source i, $i = 1,2$. In an adversarial setting, however, in which Source i was being sued for damages based on this formula, a strong rebuttal could be made along the lines sketched above. It could be argued that the risk function $R(X)$ just described is not appropriate for use after the fact; that instead, X_i should be defined as the *number* of hits (out of $K = 2$) actually received from Source i (instead of Source i's intensity), and that $R(X_1, X_2)$ should be defined as 1 if $X_1 + X_2 = 2$ and as 0 for $X_1 + X_2 = 0$ or 1. This $R(X)$ corresponds to a "perfect information" model.

A theory of risk attribution for such discontinuous risk functions, analogous to the one presented in Section 3, has been developed,[10] and shows that $X_i/(X_1 + X_2) = X_i/2$; is still the proportion of risk attributable to Source i with these new definitions. The salient point, however, is that X_i is *unknown and, in practice, unknowable*: if defense insists on attributing risk in response to the number of hits actually contributed by Source i, instead of on the basis of Source i's average intensity, then the data required for calculating Source i's attributable proportion will be unavailable. Thus, defense can argue that Source 2's share in causing the victim's cancer is unknown: it is either 0%, 50%, or 100% (with prior probabilities of $[X_1/(X_1 + X_2)]^2$, $[2X_1X_2/(X_1 + X_2)^2]$, and $[X_2/(X_1 + X_2)]^2$, respectively, where these X_i's are rates), and there is no way of knowing which. If we agree with this formulation, then the burden of proof is insurmountable, and we are stuck, with the problem of an "indeterminate risk share," characterized by a probability distribution over a source's attributable proportion of risk. How compensation decisions should take such distributions into account is unclear.

4.3. Use of Aggregate Data for Individuals

Ignoring the unique factors determining an individual's response to exposure may lead to use of an inappropriate biological model, $R(X)$, for him; on the other hand, it is hopelessly impractical to attempt to incorporate all such factors, or to try to discover exactly the details of his exposure history. Suppose then that we use average response data from a group of "similar" individuals to compute "proxy" attributable risk proportions for the factors that may have caused an individual's cancer. What types and frequencies of errors might result from such applications of group data to individual cases?

To answer this question, it is necessary to know or assume something about the distribution of response characteristics among the exposed individuals in the group for which aggregate response data are available. Suppose, for example, that individuals are exposed to two Poisson competing risk sources, as in the model of 2.2, and that the "effective" hit rate for an individual (= true hit rate × conditional probability of developing cancer, given a hit) from Source i depends on the individual and on the source. Thus, different sources represent different factors. Specifically, assume that for each true hit from Source i, individual j has a small chance P_{ij}, independently and uniformly distributed between 0 and some small α_i $(0 < \alpha_i < 1)$ in the population of interest, of getting cancer. The different values of P_{ij} reflect individual differences in susceptibility to different factors, and have a random component. The "true" hit rate for an individual from Source i will be denoted by X_i and is assumed to be known. Now let the probability of causation estimate, $\overline{PC_i}$, be calculated from the observed aggregate attributable risk rate ($= \alpha_i X_i/2$) in the population; and let PC_i denote the "true" PC ratio for a randomly selected individual drawn from that population (i.e., the PC ratio that would be calculated for that individual using his own true effective hit rate, if it were known). Thus,

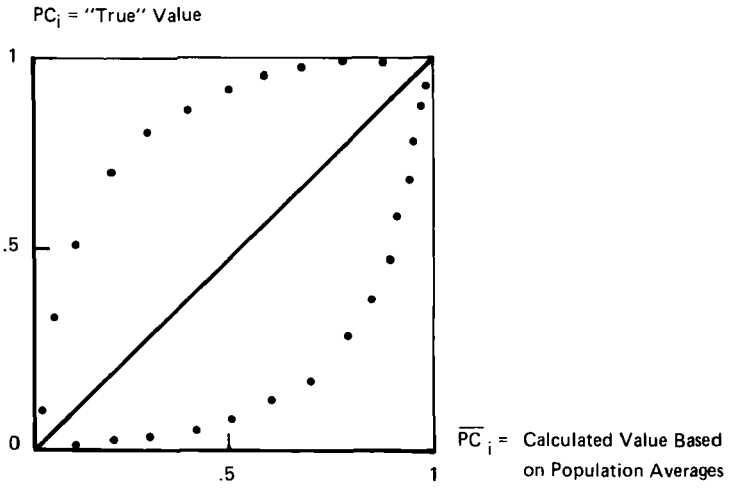
$$\overline{PC_i} = (\alpha_i X_i/2)/[(\alpha_1 X_1/2) + (\alpha_2 X_2/2)]$$
$$= \alpha_i X_i/[\alpha_1 X_1 + \alpha_2 X_2],$$

and

$$PC_i = \tilde{X}_i/(\tilde{X}_1 + \tilde{X}_2)$$

where $\tilde{X}_i$ is drawn at random from the interval $\alpha_i X_i/2 \pm \alpha_i X_i/2$. What is the relationship between the "proxy" PC ratio, $\overline{PC_i}$, and the "true" one, PC_i?

Figure 1 presents the answer graphically in terms of 90% confidence bands. Given that a $\overline{PC_i}$ value of 10% is calculated from aggregate health statistics for "similar" individuals, there is a 90% confidence that



Method of Calculation:

$$(1-\alpha)\,\%\text{ Confidence Interval: } \frac{\alpha\overline{PC_i}}{1+(\alpha-1)\,\overline{PC_i}} \leqslant PC_i \leqslant \frac{\overline{PC_i}}{\alpha+(1-\alpha)\,\overline{PC_i}}$$

Assumptions:

- p_1 and p_2 for an individual are uniformly distributed from 0 to α_i
- "1-hit" model with binomial censoring
- $P_i = \Pr(\text{Cancer} \mid \text{Hit by Source } i),\ i = 1, 2$

$$\overline{PC_i} = \frac{\bar{X}i}{\bar{X}_1 + \bar{X}_2} \qquad PC_i = \frac{\tilde{X}_i}{\bar{X}_1 + \bar{X}_2} \qquad \tilde{X}_i \sim U[0, 2\bar{X}_i]$$

Fig. 1. 90% confidence bands for "true" *PC* given calculated *PC*.

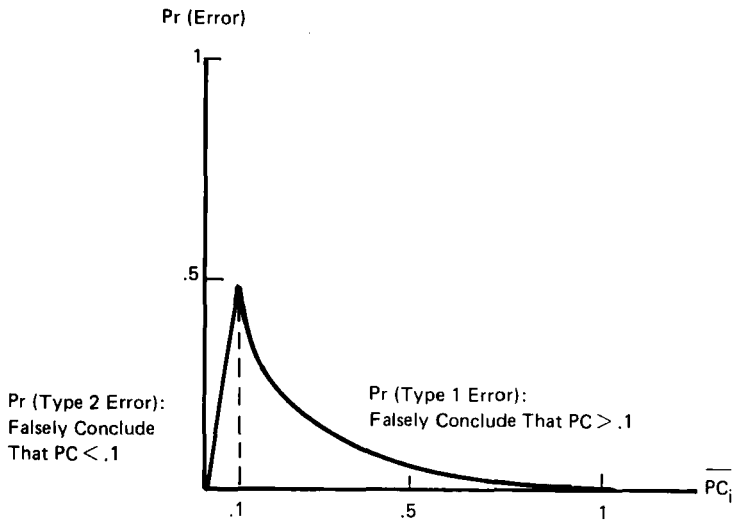
the chosen individual's "true" PC_i value will be between 1% and 53%. Given a computed PC_i value of 51% based on aggregate statistics, the individual's $\overline{PC_i}$ value will plausibly lie somewhere between 9% and 92%.

Choosing a nonuniform distribution for the "effectiveness" coefficient that converts actual to "effective" hits would alter the confidence bands in Fig. 1. But the qualitative message is clear: *the relation between "true" and "proxy" PC values may be very weak*, depending on distributional assumptions.

Figure 2 shows the implications of this observation for a simple decision rule. The rule is: make Source i compensate the victim if and only if $\overline{PC_i} \geqslant 10\%$. Figure 2 shows the probability that this rule will falsely compensate (or falsely fail to compensate) a victim as a function of the estimated probability of causation $\overline{PC_i}$. For example, if PC_i is estimated to be $\overline{PC_i} = 30\%$ for a certain individual, based on his gender, age, exposure history, etc., and on the average response rates for this group, then there is a 13% chance that the true value of PC_i for that individual is less than 10%. Thus, the chance of falsely awarding compensation to such an individual is 13%. Similarly, if PC_i is estimated to be $\overline{PC_i} = 3\%$, then there is 14% chance that the true PC_i value is over 10%, and that compensation is incorrectly denied. If the prior distribution of $\overline{PC_i}$ were known, then these Type 1 and Type 2 error probabilities for a randomly selected individual could be adjusted by changing the proposed compensation threshold of $\overline{PC_i} = 10\%$.

4.4. Conclusions

The foregoing comments suggest that the theory of attributable risk proportions presented in Section 3—and including the "Probability of Causation" for-



Method of Calculation:

$$\Pr(\text{Type 1 Error} \mid \overline{PC_i} > .1) = \frac{1 - \overline{PC_i}}{18\,\overline{PC_i}}$$

$$\Pr(\text{Type 2 Error} \mid \overline{PC_i} < .1) = \frac{9\,\overline{PC_i}}{2 - 2\,\overline{PC_i}}$$

Assumptions:

- p_1 and p_2 for an individual are uniformly distributed from 0 to α_i
- "1-hit" model with binomial censoring
- Pr (Cancer | Hit by Source i) $= p_i$, i = 1, 2

Fig. 2. Analysis of error rates for 10% compensation threshold.

mula as a special case—cannot easily be put into practice unless the practitioner is willing to accept some loose ends and substantial error probabilities. From a theoretical perspective, there is some question about what model $R(X)$ should be used *after* the fact of carcinogenisis, when an infinite amount of detail about the individual affected is available. In the competing risk "1-hit" model, an ultimately "correct" model of what happened to the individual would necessarily give a PC_i value of either 0 or 1 to each factor—but we can never in practice tell which.

If a probability distribution for PC_i is assessed, based on aggregate data for similar individuals, and conditioned on the "average" $\overline{PC_i}$ that would be computed for such individuals under the exposure conditions experienced by the individual in question, then the fact must be faced that aggregate data do not necessarily provide very much information about a single individual. In fact, the distribution of PC_i for an individual may be widely dispersed over most of the range from 0 to 1 for nearly all values of the group-based estimate $\overline{PC_i}$. Consequently, attempts to use aggregate data to compute "proxy" PC values, and then to use these estimates as a basis for compensation decisions, are apt to lead to decisions with substantial levels of both Type 1 and Type 2 errors.

REFERENCES

1. W. M. Landes and R. A. Posner, "Causation in Tort Law: An Economic Approach," *Journal of Legal Studies* **XII**, 124 (1983).
2. S. Shavell, "An Analysis of Causation and the Scope of Liability in the Law of Torts," *Journal of Legal Studies* **IX**, 499 (1980).
3. R. Delgado, "Beyond *Sindell*: Relaxation of Cause-In-Fact Rules for Indeterminate Plaintiffs," *California Law Review* **70**, 881 (1982).

4. V. P. Bond, "The Cancer Risk Attributable to Radiation Exposure: Some Practical Problems," *Health Physics* **40**, 108–111 (January, 1981).
5. R. J. Catlin and N. Parmentier, "Implications of Radiation Compensation Criteria for the Nuclear Industry," (Nuclear Safety Analysis Center, Electric Power Research Institute, October 1982).
6. L. A. Cox and J. R. Fiksel, "A Critical Review of the Probability of Causation Method," (Arthur D. Little, Inc. report to American Nuclear Insurers and Mutual Atomic Energy Liability Underwriters, 1984).
7. C. L. Chiang, "On the Probability of Death from Specific Causes in the Presence of Competing Risks," in J. Neyman (ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* Vol. 4, (University of California Press, 1961) p. 169.
8. R. E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing: Probability Models* (Holt, Rinehart, and Winston, New York, 1975) esp. Exercise 1, p. 134, and p. 84.
9. J. Schlesseman, *Case Control Studies: Design, Conduct, Analysis* (Oxford University Press, New York, 1982) pp. 28 and 41.
10. L. A. Cox, "A New Measure of Attributable Risk for Public Health Applications," (unpublished manuscript, Arthur D. Little, Inc.)
11. L. J. Billera and D. C. Heath, "Allocation of Shared Costs: A Set of Axioms Yielding a Unique Procedure," *Mathematics of Operations Research* **7**, 121–135 (1982).
12. L. J. Mirman, D. Samet, and Y. Tauman, "An Axiomatic Approach to the Allocation of a Fixed Cost Through Prices," *Bell Journal of Economics* **14**, 32–39 (1983).
13. T. L. Fine, *Theories of Probability* (Academic Press, New York, 1973).
14. M. H. DeGroot, *Optimal Statistical Decisions* (McGraw–Hill, New York, 1970) Chapter 10.