Relation of Probability of Causation to Relative Risk and Doubling Dose: A Methodologic Error That Has Become a Social Problem

Sander Greenland, DrPH

ABSTRACT

Epidemiologists, biostatisticians, and health physicists frequently serve as expert consultants to lawyers, courts, and administrators. One of the most common errors committed by experts is to equate, without qualification, the attributable fraction estimated from epidemiologic data to the probability of causation requested by courts and administrators. This error has become so pervasive that it has been incorporated into judicial precedents and legislation. This commentary provides a brief overview of the error and the context in which it arises. (*Am J Public Health*. 1999;89:1166–1169)

Many misconceptions are common in the literature relating epidemiology to compensation decisions. I here focus on 2 related problems that are routinely unrecognized by epidemiologists, administrators, and courts. The first problem is that the probability of causation cannot be computed solely from the relative risk. In particular, when exposure accelerates the time of disease occurrence, the standard epidemiologic estimates of probability of causation will tend to underestimate that probability. The second problem is that the exposure dose at which the probability of causation exceeds 50% (the point at which exposure causation is more likely than not) may fall well below the "doubling dose" (the dose at which the incidence of disease is doubled).

These problems have been explained in many articles.[1–9] The remarks that follow are an attempt to summarize, in an elementary fashion, some of the major points of the analyses given in those articles, in hope that the public health community will become aware of these points and their importance to judicial and legislative decisions. They are adapted from court declarations I submitted in support of plaintiff groups filing suit against operators of facilities that processed nuclear weapons material: Apollo-Parks (Pennsylvania) and Hanford (Washington). These declarations argued against acceptance of epidemiologic estimates of the probability of causation. It is my belief that failure to recognize the problems discussed here subsequently has resulted in logically unsound and possibly unjust compensation rules, such as those incorporated into radiogenic cancer legislation.[10] My arguments are confined to outcomes (such as cancer) in which time of incidence is crucial; different considerations apply to "all-or-none" outcomes (such as birth defects),[3–5] for which risks but not rates are important.

Preliminary Concepts

To resolve questions of causation, courts and compensation boards often have to determine the relation of an observed level of exposure (which I call "the exposure") to the development of a disease in a given plaintiff. In particular, as discussed in a number of judicial comments (including those concerning Agent Orange[11] and silicone-implant litigation[12,13]), courts have proposed and sometimes imposed that litigation should proceed only if the probability of causation for exposed individuals exceeds 50% (the "more-likely-than-not" rule). Here, the *probability of causation* (PC) is the probability that the exposure was a contributory cause of the plaintiff's disease; the exposure is a *contributory cause* of the plaintiff's disease if, but for exposure, that disease would have occurred later in life or not at all.

Suppose we consider persons with the disease who are similar to the plaintiff with respect to their exposure history and risk factors for disease. The *etiologic fraction* is then the fraction of these individuals for whom exposure was a contributory cause of the disease.[3,7] A common and relatively uncontroversial step in deriving the probability of causation for the plaintiff is to equate this probability with the etiologic fraction.

Consider next the concept of relative risk. In studies of outcomes for which timing is important, the term "relative risk" usually refers to the *incidence rate ratio* (IR), which is the ratio of the incidence rates with and without the exposure in question:

$$\text{IR} = \frac{\text{Incidence rate if exposed}}{\text{Incidence rate if not exposed}}$$

The magnitude of this rate ratio depends on the exposure level under consideration: If exposure is causal, we should expect the rate in the numerator of the incidence rate ratio to increase as the exposure level is increased, and hence we should expect the rate ratio to increase as the exposure level is increased. In legal actions, the exposure level at which the incidence rate ratio equals 2 (a doubling of

Correspondence should be sent to Sander Greenland, DrPH, Department of Epidemiology, School of Public Health, University of California, Los Angeles, 22333 Swenson Dr, Topanga, CA 90290-3434.

This paper was accepted April 19, 1999.

the incidence rate among those exposed) is sometimes called the "doubling dose" of exposure. As discussed later, a common mathematical error among experts is to assume that this doubling dose is the dose at which the probability of causation is 50%.

The *rate fraction* (RF)[4–7] is defined by

$$RF = \frac{\text{Incidence rate if exposed} - \text{Incidence rate if not exposed}}{\text{Incidence rate if exposed}}$$

This quantity is the excess incidence rate produced by exposure, expressed relative to the incidence rate if exposed. If we define I_e as incidence rate if exposed and I_u as incidence rate if not exposed, then we can see that $IR = I_e/I_u$ and that

$$RF = \frac{I_e - I_u}{I_e} = \frac{I_e/I_u - I_u/I_u}{I_e/I_u} = \frac{IR - 1}{IR}$$

The rate fraction has been given many other names, including "attributable fraction," "attributable risk," and "assigned share," although the first 2 terms have been used to refer to many other concepts as well (see Greenland and Robins[3]).

Logical Errors in Estimates of Probability of Causation

The key fallacy in much of the literature and testimony regarding the probability of causation is the use of the following generally incorrect equations: Etiologic Fraction = Rate Fraction and Probability of Causation = Rate Fraction, which may be succinctly expressed as PC = RF (although it is more commonly and ambiguously expressed as PC = "Attributable Risk"[14–18]). To understand why these equations are *not* correct in general, it is essential to clarify the meaning of "contributory cause." The discussion to follow focuses on the situation perhaps most often faced by the courts, that in which the cause of disease in an individual of a specific age, sex, and exposure level is at issue. Further issues that arise when considering population causation (e.g., summarizing over age) have been analyzed elsewhere.[3–5]

Consider a case of thyroid cancer occurring at 50 years of age in a woman who had an excess radiation exposure (above natural background) that was only 20% of the putative doubling dose. Ordinary language and common sense tell us that the exposure contributed causally to this disease occurrence if, *but for exposure*, the plaintiff's disease either (1) would have occurred at a later time or (2) would not have occurred at all. I label this occurrence an *accelerated occurrence* if exposure causally contributed to the disease in the first sense (i.e., without exposure, the disease would have occurred at a later time). I label this occurrence an *all-or-none occurrence* if exposure causally contributed to disease in the second sense (i.e., without exposure, the disease would not have occurred at all). Finally, I label this occurrence an *unaffected occurrence* if exposure made no difference in the timing of the disease (i.e., exposure failed both "but for" conditions and so was not a contributory cause of this occurrence of disease). In the case of either an accelerated or an all-or-none occurrence, exposure harmed the individual because exposure reduced the amount of time that the individual was able to live without the disease.

As an illustration, consider a hypothetical cohort of 100 000 women born in 1942 who experienced excess radiation exposure that was 20% of the doubling dose because they lived downwind of a facility that had released radioactive emissions for many years. Suppose that the incidence rate of thyroid cancer in this cohort in 1992 was $I_e = 12$ occurrences per 100 000 woman-years but would have been $I_u = 10$ occurrences per 100 000 woman-years in the absence of the excess exposure. Here, the rate ratio is $IR = I_e/I_u = 12/10 = 1.2$, and the rate fraction is $RF = (12-10)/12 = (1.2-1)/1.2 = 2/12 = 17\%$.

Using the formula PC = RF, many experts, judges, and administrators would incorrectly conclude from these figures that 17% of exposed occurrences arising in 1992 were affected by exposure (e.g., see references 12–20). Such a conclusion is fallacious because one cannot determine which of the exposed occurrences would have taken place at a later time had individuals not been exposed. For example, it is possible that all 12 of the thyroid cancers occurring in 1992 would have occurred in 1993 or later if no one had been exposed; that is, all 12 of the women contracting thyroid cancer could have suffered accelerated occurrences of cancer. Even if this were the case, none of these women would be entitled to compensation under various schemes that equate probability of causation to rate fraction,[10,17,20] because the rate fraction makes it appear that only 17% of the women were affected by the exposure.

One may well ask: If all 12 of the women contracting thyroid cancer in 1992 suffered at least a year's acceleration, who are the 10 women who would have contracted thyroid cancer in 1992 if there had been no excess radiation exposure? They would be women in the cohort (born in 1942) who contracted thyroid cancer before 1992 because of the exposure. These 10 women would have thyroid cancers that would have not occurred until 1992 but for the exposure. Furthermore, these 10 women would not overlap at all with the 12 women who contracted thyroid cancer in 1992 because of the exposure.

It is always logically possible that every exposed person with the disease was harmed by the exposure, so that PC = 100% no matter how close to 1 the true relative risk. This point was demonstrated long ago.[3–5] I believe that many experts do not see this fact because they fail to take into account the possibility that some or all exposed occurrences may be accelerated occurrences. Furthermore, even when confronted by this possibility, they fail to recognize that epidemiologic data cannot distinguish accelerated occurrences from unaffected occurrences. The only way one can estimate the relative proportion of accelerated and unaffected occurrences—and hence estimate the probability of causation—is by positing a specific biologic model for the disease process.[3–7] Under some biologic models, the rate fraction will equal the probability of causation; under many more models, however, the rate fraction will underestimate the probability of causation.[3–6]

I am aware of no real example in which enough is known of cancer biology to justify a claim that the rate fraction approximates the probability of causation. Nonetheless, many experts claim in court (if not in journals) that PC = RF is a good approximation without supplying any evidence to support this claim. In an even more unscientific manner, some experts will stand by the PC = RF assertion on the grounds that society (or the court) needs a probability of causation formula immediately. This type of rationale ("there is a dire need and we must fill it immediately") could be used to justify any of the thousands of quack cures for cancer that have appeared in the 20th century.

Algebraic Analysis

To explain the problem in algebraic terms, suppose that A_T exposed persons contracted the disease during the time period in question and that, of these individuals, A_0 are unaffected, A_1 were accelerated by exposure, and A_2 represented all-or-none occurrences of disease. By definition, exposure played a role in the etiology (development) z; put more colloquially, exposure harmed persons whose disease was either accelerated or all or none. Hence, the fraction of exposed persons with the disease who were harmed by the exposure is $(A_1 + A_2)/A_T$. This quantity is the etiologic fraction.[3–7] Furthermore, if we randomly select an exposed person with disease from the total A_T, the chance that

exposure harmed that person (i.e., the chance that the person had an accelerated or an all-or-none occurrence) is also $(A_1+A_2)/A_T$. The latter quantity is thus also the probability of causation.[3–7]

Consider next the proportion of exposed disease occurrences that would not have taken place in the absence of exposure. This quantity is the proportion of all-or-none occurrences among all exposed occurrences, and it is equal to A_2/A_T. It has been called the *excess fraction*,[3,4,7] because it corresponds to the excess caseload produced by exposure, expressed as a fraction of the total caseload.

The failure to distinguish the etiologic fraction $(A_1+A_2)/A_T$ from the excess fraction A_2/A_T is a major problem in most of the literature.[3] Writers often use terms such as attributable risk, attributable proportion, attributable fraction, etiologic fraction, and probability of causation in an interchangeable fashion. The term attributable risk is particularly misleading, because neither the etiologic fraction nor the excess fraction equals a disease risk; yet, despite its misleading nature, the term dominates the American literature (especially the biostatistics literature, which recognizes the points discussed here no more often than does the epidemiology literature). Often, a writer or expert will describe in words the probability of causation $(A_1 + A_2)/A_T$ but then proceed to incorrectly estimate this probability by using a formula that approximates the excess fraction A_2/A_T (e.g., see references 14–18, 20); the result is an underestimate of the probability of causation.

To illustrate the problem, suppose effects of exposure on the exposed population are infrequent and there is no bias present. Then the quantity approximated by the usual relative risk estimates is the ratio of cases that would occur with and without exposure, $A_T/(A_1 + A_0)$; this quantity is traditionally called the *standardized morbidity ratio* (SMR).[7] In this situation, the rate ratio I_e/I_u will approximate the standardized morbidity ratio,[3–5] and so the rate fraction will approximate the excess fraction A_2/A_T:

$$\frac{IR-1}{IR} \approx \frac{SMR-1}{SMR} = \frac{A_T/(A_1+A_0)-1}{A_T/(A_1+A_0)}$$

$$= \frac{A_T-(A_1+A_0)}{A_T} = \frac{A_2}{A_T}$$

It follows that the quantity approximated by the usual estimates of the attributable fraction or attributable risk among the exposed population is the excess fraction.[3,4,7] This quantity will not be close to the etiologic fraction $(A_1+A_2)/A_T$ unless the number of accelerated cases, A_1, is small relative to the number of all-or-none cases, A_2.

Unfortunately, A_1 need not be small relative to A_2, and it may often be comparable to or larger than A_2, in which case the rate fraction will be an utterly misleading estimate of causation probability. As an extreme example, suppose the damage done by exposure was that of accelerating the development of disease in all individuals destined to contract disease. Then, when considering the lifetime experience of the exposed cohort, all of the exposed occurrences of disease would be accelerated cases, so A_T would equal A_1, while A_0 and A_2 would both be zero. Consequently, $SMR = A_T/(A_1+A_0) = A_T/A_1 = 1$ and the excess fraction $A_2/A_T = 0$, while $PC = (A_1+A_2)/A_T = A_1/A_T = 1$. In other words, the standardized morbidity ratio would be 1 and the excess fraction would be 0, incorrectly suggesting that there was no exposure effect, and yet the probability of causation would be 100%.

One might object that the extreme structure just described is unrealistic. In reality, however, this extremity is exactly what one should expect if the outcome under study is total mortality in a cohort followed for its entire lifetime, such as the cohort of atomic bomb survivors in Japan. Here, everyone experiences the outcome (death), so there are no "all-or-none" cases, yet everyone may also experience damage and consequent loss of years of life (even if only minor and stress related) owing to the exposure. And, in less extreme situations, there is rarely, if ever, an agreed-upon basis for asserting that A_1 is close enough to 0 to equate the excess and etiologic fractions.

Relation of Doubling Dose and Probability of Causation to Background Risks

Individual and population risks vary with factors other than the exposure in question. Courts often ask whether the doubling dose can vary with these factors and whether the probability of causation can vary with these factors. The answer to both questions is more than yes: As a result of the inevitable complex interactions among risk factors, we should *expect* large variation. Furthermore, it is possible for the variation to be in either direction, in that the doubling dose may either increase or decrease with an increase in background risk; likewise, the probability of causation may either increase or decrease with an increase in background risk. Worse yet, because the numbers of unaffected, accelerated, and all-or-none cases need not vary in a manner proportional to one another, the discrepancy between the rate fraction and causation probability will usually vary with background risk, even when the relative risk does not vary.

Because the probability of causation cannot be estimated from epidemiologic data alone, variation in this probability also cannot be estimated from epidemiologic data alone; a biologic model is again needed. The situation improves only a little if we abandon probability of causation and focus on relative risks as entities in themselves. In theory, variation in relative risks with background risk could be examined with epidemiologic data if the data were so extensive and accurate that one could validly estimate variation in background risk across the myriad subgroups of risk factors (age, sex, occupation, genetic susceptibility, etc.). Unfortunately, epidemiologic data are rarely so extensive and accurate, and, as a consequence, they rarely indicate the potential range of variation in relative risks. Furthermore, epidemiologic data cannot establish the absence of individuals who are exceptionally vulnerable to exposure effects and who constitute a subgroup with an exceptionally high relative risk. At best, epidemiologic data can indicate the maximum frequency of such exceptionally sensitive individuals.

Further Complications

In the present development, I have implicitly assumed that exposure does not affect competing risks and never prevents the study outcome. If either of these assumptions is violated, the rate fraction can be even further from the probability of causation than indicated by the preceding examples and algebraic equations, even if there are no accelerated cases ($A_1 = 0$). For example, suppose exposure prevents 2 of 4 background cases over the period of interest but also causes an additional 3 cases. Then the number of cases occurring under exposure will be $3 + 4 - 2 = 5$, and so the excess fraction due to exposure will be $(5 - 4)/4 = 25\%$. Yet, 3 of the 5 cases under exposure, or 60%, will have been caused by exposure. This discrepancy illustrates a general problem: All epidemiologic measures (such as rate ratios and rate fractions) reflect only the net impact of exposure on a population, rather than the total number of persons affected by exposure.[21]

Conclusions

When an effect of exposure is to accelerate the time at which disease occurs, the rate fraction $RF = (IR - 1)/IR$ will tend to

understate the probability of causation because it does not fully account for the acceleration of disease occurrence. In particular, and contrary to common perceptions, a rate fraction of 50% (or, equivalently, a rate ratio of 2) does not correspond to a 50% probability of causation. This discrepancy between the rate fraction and the probability of causation has been overlooked by various experts in the legal as well as the scientific community, even though it undermines the rationale for a number of current legal standards. Furthermore, we should expect this discrepancy to vary with background risk factors, so that any global assessment of the discrepancy cannot provide assurances about the discrepancies within subgroups.

I believe it is the responsibility of health scientists to recognize and acknowledge these limitations of epidemiologic data, rather than continue to offer estimates of causation probabilities that teeter on concealed or unsupported assumptions. These limitations should be distinguished from the less subtle (but equally important) inability of epidemiologic data to reliably identify small relative risks[22]; the latter limitation arises from problems of study design and conduct, whereas the problems discussed here are purely logical. In acknowledging these limits, we may also offer some constructive alternatives to causation probabilities that can be more successfully addressed by epidemiologic research, such as expected years of life lost.[9,23,24] □

Acknowledgments

I wish to thank Alfredo Morabia, Hal Morgenstern, Charles Poole, Daniel Wartenberg, and Noel Weiss for their helpful comments on the manuscript.

References

1. Cox LA. Probability of causation and the attributable risk. *Risk Anal.* 1984;4:221–230.
2. Lagakos SW, Mosteller F. Assigned shares in compensation for radiation-related cancers (with discussion). *Risk Anal.* 1986;6:345–380.
3. Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol.* 1988;128: 1186–1197.
4. Robins JM, Greenland S. Estimability and estimation of excess and etiologic fractions. *Stat Med.* 1989;8:845–859.
5. Robins JM, Greenland S. The probability of causation under a stochastic model for individual risks. *Biometrics.* 1989;45:1125–1138.
6. Beyea J, Greenland S. The importance of specifying the underlying biologic model in estimating the probability of causation. *Health Phys.* 1999;76:269–274.
7. Greenland S, Rothman KJ. Measures of effect and measures of association. In: Rothman KJ, Greenland S, eds. *Modern Epidemiology.* 2nd ed. Philadelphia, Pa: Lippincott-Raven; 1998: 47–64.
8. Parascandola M. What's wrong with the probability of causation? *Jurimetrics.* 1999;39:29–44.
9. Greenland S, Robins JM. Epidemiology, justice and the probability of causation. *Epidemiology.* In press.
10. Hatch OG. Medical/legal aspects of radiation induced cancer. *Health Phys Soc Newsletter.* December 1983:6–8.
11. *In re "Agent Orange,"* 611 F Supp 1262 (ED NY 1985).
12. *Hall v Baxter Healthcare,* 947 F Supp 1387 (D Ore 1996).
13. *Pick v American Medical Systems,* 958 F Supp 1151 (ED La 1997).
14. Bailey LA, Gordis L, Green M. Reference guide on epidemiology. In: *Reference Manual on Scientific Evidence.* Washington, DC: Federal Judicial Center; 1994:168–169.
15. Mettler FA, Upton AC. *Medical Effects of Ionizing Radiation.* 2nd ed. Philadelphia, Pa: WB Saunders Co; 1995.
16. Armstrong B, Thériault G. Compensating lung cancer patients occupationally exposed to coal tar pitch volatiles. *Occup Environ Med.* 1996; 53:160–167.
17. Cole P. Causality in epidemiology, health policy, and law. *ELR News Anal.* 1997;27: 10279–10285.
18. Weiss N. General concepts of epidemiology. In: Faigman DL, Kaye DH, Saks MJ, Sanders J, eds. *Modern Scientific Evidence: The Law and Science of Expert Testimony.* St. Paul, Minn: West Publishing; 1997:321–344.
19. Black B, Lillienfeld DE. Epidemiologic proof in toxic tort litigation. *Fordham Law Rev.* 1984; 52:732–785.
20. Wakeford R, Antell BA, Leigh WJ. A review of the probability of causation and its use in a compensation scheme for nuclear industry workers in the United Kingdom. *Health Phys.* 1998;74:1–9.
21. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol.* 1986;15:413–419.
22. Shapiro S. Meta-analysis, schmeta-analysis. *Am J Epidemiol.* 1994;140:771–778.
23. Robins JM, Greenland S. Estimability and estimation of expected years of life lost due to a hazardous exposure. *Stat Med.* 1991;10:79–93.
24. Boshuizen H, Greenland S. Average age at first occurrence as an alternative occurrence parameter in epidemiology. *Int J Epidemiol.* 1997; 26:867–872.