Reviews and Commentary

CONCEPTUAL PROBLEMS IN THE DEFINITION AND INTERPRETATION OF ATTRIBUTABLE FRACTIONS

SANDER GREENLAND[1] AND JAMES M. ROBINS[2]

The concept of attributable fraction (1–3) has grown in importance as epidemiologists and epidemiologic data have played a larger role in interventions, regulations, and lawsuits concerning hazardous exposures. For example, in a lawsuit, the court may wish to determine the likelihood that a particular case's illness was caused by the exposure at issue, and the attributable fraction has been interpreted as just this likelihood (e.g., see ref. 4, p. 164).

While the concept is known by many names (including attributable risk (5), etiologic fraction (4, 6, 7), and attributable proportion (8)), we would think this variety would cause no problem as long as the conceptual and algebraic formulations were unambiguous. Unfortunately, at least three distinct concepts have been variously identified as the attributable fraction, although these concepts have usually not been distinguished in the literature. Furthermore, certain equations used to relate attributable

fractions to incidence and relative risk fail to hold in many circumstances. These problems are of some importance because of the recent appearance of attributable fraction concepts in legislation (9, 10). We will show that the conceptual problems appear to arise from a failure of some definitions to take account of time of incidence when evaluating the role of the study exposure in disease etiology. These conceptual problems are distinct from study validity issues (such as misclassification, selection bias, or sampling error) and thus constitute an additional obstacle to valid estimation of exposure effects.

EXCESS VERSUS ETIOLOGIC FRACTIONS

Suppose we are asked to estimate the fraction of leukemia cases attributable to exposure within a cohort of former military personnel who had been exposed to radiation from a nuclear weapons test. It is not clear from this question whether a case "attributable to exposure" is 1) a case for which exposure played an etiologic role, that is, for which exposure was a contributory cause of the outcome (an "etiologic case"), or 2) a case that would not have occurred had exposure not occurred (an "excess case").

All excess cases are etiologic cases, but not vice versa. We will illustrate this point

and show that the distinction of these cases can be of critical importance: From the standpoint of both law and biology it can be essential to measure the fraction of all cases that are etiologic cases. Unfortunately, in traditional definitions, the attributable fraction measures only the fraction of all cases that are excess cases, and this can be much smaller than the fraction of cases that are etiologically attributable to exposure.

To illustrate these points mathematically, suppose that incidence is evaluated over a specified risk period or time interval $(0, t)$ after exposure at time zero (this interval may vary across individuals, although we will treat it as constant in the following development). In the leukemia example, t might be "20 years from the date of the test." Furthermore, suppose that exposure action follows a deterministic model, so that there are only three types of exposed subjects who become cases during the interval:

Type 0: The exposure had no impact whatsoever on the case's incidence time.

Type 1: The exposure made the case's incidence time earlier than it would have been in the absence of exposure (so exposure played a role in the etiology of this case), *but* had exposure never occurred (or had its effect been blocked), this subject would still have become a case by t, although later in the interval.

Type 2: Had exposure never occurred, the subject would *not* have become a case by t because, in the absence of exposure, disease would have occurred after t, or not at all.

Let the number or set of each of these types be denoted A_0, A_1, and A_2, respectively, with $A_+ = A_0 + A_1 + A_2$, and let M equal the total number of cases under study. (M sometimes equals A_+, as in traditional standardized morbidity ratio (SMR) studies in which only an exposed cohort is studied, and M sometimes equals A_+ plus unexposed cases, as in a case-control study.) We think it clear that a case of type 0 is not "attributable" to exposure

and a case of type 2 is. Furthermore, type 2 cases correspond exactly with excess cases, as defined earlier. Type 2 cases are also etiologic cases, as defined earlier.

What about type 1 cases? Like type 2 cases, they are etiologic cases, since exposure played a role in the etiology of their disease. They are *not*, however, excess cases, because they still would have become cases by time t had exposure not occurred. The issue of whether to count these cases as attributable to exposure is important because, as we will show, their number may be large relative to excess (type 2) cases.

Some textbooks can be interpreted to imply that only excess cases contribute to the attributable fraction, so that the latter should be algebraically equivalent to A_2/M (which we will call the excess fraction). Consider the following definitions of attributable fraction (the first two of which assume $M = A_+$): "the proportion of the cases of disease occurring among exposed persons which is in excess in comparison with the nonexposed" (5, p. 74); "[the attributable fraction] conveys a sense of how much of the disease in an exposed population can be prevented by blocking the effect of exposure or eliminating the exposure" (8, pp. 38–9); "the proportion of disease in the target population that would not have occurred had the factor been absent" (6, p. 44); "the proportion of the disease occurrence that would potentially be eliminated if exposure to the risk factor were prevented" (3, pp. 39–40).

Since type 1 cases become cases by the end of the risk period whether or not they are exposed, they cannot be counted among the proportion of disease that would not have occurred had exposure been absent, prevented, or eliminated. Thus, it seems to us that such cases would not be counted by the above definitions. Nevertheless, it is possible that within the risk period, a type 1 case may have suffered a considerable loss of healthy, productive life because of exposure's effect.

Some textbooks could be interpreted to imply that all etiologic cases—both type 1

and type 2—should contribute to the attributable fraction, so that the latter should be algebraically equivalent $(A_1 + A_2)/M$ (which we will call the etiologic fraction). Consider the following definitions of attributable fraction (all of which assume M is restricted to exposed cases only, i.e., $M = A_+$): "the proportion of exposed cases that are due to the risk factor" (4, pp. 163–4); "the proportion of the actual cases in the index [exposed] domain that are caused by the cause at issue [if the cause is never preventive]" (7, p. 255); "[the attributable fraction] can be interpreted as the proportion of exposed cases for whom the disease is attributable to exposure" (8, p. 38).

Since type 1 cases are caused by the exposure (i.e., exposure is a contributory or component cause of their disease), they would be counted among the proportion of cases "due to" or "attributable to" exposure if "due to" or "attributable to" is given the interpretation of "caused by." Kleinbaum et al. (4) give another definition, of considerable legal interest, to the effect that the attributable fraction " . . . may also be interpreted as the probability that a randomly selected case from the population developed the disease as a result of the risk factor" (4, p. 160). This "probability-of-causation" definition appears to us to correspond to $(A_1 + A_2)/M$, since the latter is the probability that a randomly selected case had exposure as a contributory cause.

Algebraic definitions

Several textbooks also offer algebraic definitions of attributable fractions, and in some situations, the defining formulas are not equivalent to possible interpretations of the verbal definitions. For example, Miettinen (7, pp. 254–5) defines the attributable fraction among the exposed as $(O_1 - E_1)/O_1$, where O_1 is the observed number of exposed cases, that is, $O_1 = A_+$, and E_1 is the number of exposed cases that would have occurred had the exposed population not experienced the exposure effect. Because type 1 (as well as type 0) cases would have become cases by t even if the exposure

effect was absent, it is apparent that $E_1 = A_0 + A_1$ and so $O_1 - E_1 = A_2$, the number of excess cases only. Thus, Miettinen's formula equals the excess fraction. On the other hand, Breslow and Day (5, p. 74) and Rothman (8, p. 38) algebraically define the attributable fraction in terms of incidence densities. As we discuss later, these definitions are not in general equivalent to either the excess fraction A_2/M or the etiologic fraction $(A_1 + A_2)/M$, although under certain biologic models they will be equivalent to the latter quantity (11).

RELATIONS BETWEEN QUANTITIES

We recognize that several of the passages quoted above may have more than one possible interpretation. Nevertheless, it is apparent that there are several different concepts of attributable fraction in use. This observation raises two questions: 1) How far apart will the quantities corresponding to the different concepts be? 2) Which of these quantities are estimated by the attributable fraction estimates offered in the literature?

The answers to both questions hinge on the observation that the quantities A_0 and A_1 are not empirically distinguishable without strong biologic assumptions; only the total $A_0 + A_1$ can be estimated without such assumptions, even if there is no bias in the study. To see this, consider again the leukemia illustration. Let $t = 20$ years, with a total of 24 exposed cases occurring by t, no cases occurring in the first five years after exposure, and six cases occurring in the last five years before t.

Example 1

Suppose the effect of exposure had been to "age" everyone five years with respect to their leukemia risk, that is, the effect of exposure was to make leukemia occur five years sooner among those persons destined to contract leukemia (in the absence of other causes of death). Then the six subjects who became cases in the last five years before t would have remained leukemia-free up to t had exposure not occurred, while

the remaining cases still would have contracted leukemia by t. Hence $A_2 = 6$, and the excess fraction (up to t) among the exposed would be $6/24 = 0.25$. But, under this "uniform aging" biologic model, the exposure was a contributory cause in every one of the 24 cases that occurred, so that $A_1 + A_2 = 24$ and the etiologic fraction among the exposed is $24/24 = 1.0$.

Example 2

Suppose the effect of exposure had been to produce leukemogenic marrow-cell mutations in six of the exposed subjects, with leukemia arising from these mutations within the 20 years, but had no effect on leukemia risk in the remaining subjects. Suppose also that 18 leukemias etiologically unrelated to exposure ("spontaneous cases") occurred in the remainder. Then $A_1 = 0$, $A_2 = 6$, and so the excess and etiologic fractions would be identical, that is, $6/24 = 0.25$.

In both examples, the exposure produced six cases in excess of the 18 that would have occurred had exposure or its effect been absent or prevented, so that the true excess fraction was $6/24 = 0.25$. But the etiologic fraction was four times higher in the first example than in the second, and four times higher than the excess fraction in the first example. Given a perfect unexposed comparison group for this study, we would be able to accurately estimate the number of leukemia cases to expect among the exposed if exposure had been absent. But this information would only allow us to estimate the excess fraction. The residual number $18 = A_0 + A_1$ could not be further partitioned without assumptions about the biologic process leading from exposure to disease.

The following example, while somewhat absurd in its extremity, shows a rare situation in which biologic knowledge is so strong that both fractions can be precisely computed. It also illustrates how, for inevitable outcomes, the excess fraction will approach zero as follow-up time t becomes large, while in general the etiologic fraction will not do so.

Example 3

Consider the 1860 United States birth cohort, with overt (nervous-system) rabies as the exposure, death as the outcome, and $t = 120$ years from infection. Then the etiologic fraction among the exposed is one, since (as far as is known) overt rabies caused death in all its victims before the advent of modern life-support systems. Nevertheless, all these victims would have died within 120 years anyway—so that the excess mortality produced by rabies (or anything) by 120 years of follow-up is zero. Thus, the excess fraction is zero.

Examples 1–3 demonstrate that the excess fraction A_2/M and the etiologic fraction $(A_1 + A_2)/M$ may be arbitrarily far from one another. More generally, the excess fraction, A_2/M, can never exceed the etiologic fraction, $(A_1 + A_2)/M$, and must be strictly less than the latter if $A_1 > 0$ (as will be the case if exposure is not a necessary cause and t is large enough). It follows that an unbiased estimate of the excess fraction will often be a null-biased estimate of the etiologic fraction. As is apparent from example 1, this bias can be dramatic in realistic cases, and increases with follow-up time t. Parallel results can be obtained under a stochastic model for individual effects (11).

As noted before, strong biologic assumptions may be needed to determine the etiologic fraction. If the exposure is never preventive, the necessary and sufficient condition for the etiologic fraction to equal the excess fraction is that $A_1 = 0$, that is, that no cases caused by exposure would become cases in the absence of exposure. There are several biologic conditions under which this will be so. For example, in a situation in which an exposure is a necessary cause of the outcome (as in many foodborne disease outbreaks), $A_0 = A_1 = 0$ for that exposure, and so both fractions will be one.

Relations to incidence time

Whether an etiologic case is an excess case depends on how much exposure advanced the time of disease incidence. For example, an etiologic case occurring at follow-up year 10 will not be an excess case by year 25 unless exposure advances incidence time by more than 15 years. Thus, the excess fraction directly depends on the amount by which exposure advances incidence time, albeit in a crude fashion.

In contrast, an etiologic case remains an etiologic case regardless of the degree to which exposure advances incidence time. To take an extreme example, suppose in a study of the first battle of the Somme (in 1915), we wish to determine the fraction of deaths caused by machine-gun hits (exposure is thus being hit by a machine-gun bullet). Consider a soldier hit in the head and killed instantly by a machine gun just 10 seconds before an artillery shell exploded in his trench. The cause of the soldier's death was a machine-gun hit, and so the soldier is an etiologic case. This is so regardless of whether, had the machine gun missed, the soldier would have been killed by the artillery burst 10 seconds later or the soldier would have survived the artillery burst and died 70 years later.

The preceding example shows that the etiologic fraction (which is the fraction of cases for whom exposure advanced the time of incidence) is insensitive to how much exposure advances the time of incidence. This remains so even if one inteprets the etiologic fraction as the "probability of causation": in the preceding example, the probability that the soldier was killed by a machine-gun hit is one, regardless of how long the soldier would have survived had the machine gun missed. As will be discussed later, such insensitivity renders the etiologic fraction and probability of causation inappropriate for certain applications.

INCIDENCE FRACTIONS

One often sees the attributable fraction defined as the fraction of the incidence rate "attributable" to exposure, that is, the excess incidence rate in the exposed expressed as a proportion of the total incidence rate in the exposed. As is apparent from the literature (2–8), there are several different ways to define "incidence rate." Each possible definition of incidence rate leads to a different quantity for the attributable fraction. For some definitions, the quantity is not equivalent to either the excess or the etiologic fraction as defined above.

Incidence-proportion fractions

Consider first the definition in which the "incidence rate" is the proportion of a closed (i.e., uncensored) cohort that contracts a disease over a specified time interval, that is, "incidence rate" is taken to be the incidence proportion (7) (i.e., average risk (4) or cumulative incidence (4, 8)). Given an exposed cohort of size N_1, the incidence proportion (IP) over the interval is $IP_1 = A_+/N_1$, whereas the proportion that would have contracted the disease had exposure been absent is $IP_0 = (A_0 + A_1)/N_1$. It follows that the incidence-proportion difference expressed as a fraction of the exposed incidence proportion is

$$\frac{IP_1 - IP_0}{IP_1} = \frac{A_+/N_1 - (A_0 + A_1)/N_1}{A_+/N_1}$$

$$= \frac{A_+ - A_0 - A_1}{A_+} = \frac{A_2}{A_+}.$$

The latter term is simply the excess fraction among the exposed. Thus, defining the attributable fraction as the fraction of the incidence proportion "attributable" to exposure is algebraically equivalent to the excess fraction definition given earlier. Note, however, that the definition in terms of incidence proportions is restricted to closed cohorts (since the incidence proportion must be defined in reference to a closed cohort), whereas the excess fraction is defined for any population.

Incidence-density fractions

Consider next definitions in which the "incidence rate" is the instantaneous inci-

dence density (ID) (hazard rate or person-time rate), so that the attributable fraction (at time u) is defined as $(ID_1 - ID_0)/ID_1$, where ID_1 and ID_0 are the incidence densities (at time u) when exposure is present and absent (cf. references 4, 5, and 8).

As has been noted elsewhere (12–16), it is possible for an exposure which only causes and never prevents disease to have $ID_1 < ID_0$ over certain time intervals following exposure (the "crossing hazards" phenomenon). The instantaneous incidence-density ratio ID_1/ID_0 will be less than one and the quantity $(ID_1 - ID_0)/ID_1$ will be negative over such intervals. Since neither the excess nor the etiologic fractions can be negative for purely causal exposures, such examples show that $(ID_1 - ID_0)/ID_1$ is not equivalent to either fraction. The following example shows that, even if there are no competing risks and ID_1, ID_0, and their ratio are constant over time, the quantity $(ID_1 - ID_0)/ID_1$ may still be far from either fraction.

Example 4

Suppose at time zero, we randomly sample a large number of exposed persons (indexed by i) and proceed to follow them. Assume that each person i would have had a death time D_{0i} if unexposed, but would die instead at time $D_{1i} = D_{0i}/2$ when exposed. Finally, assume that the D_{0i} are exponentially distributed with expectation T; as a consequence, the D_{1i} will be exponentially distributed with expectation $T/2$ (this is a simple special case of the accelerated-life model given by Cox and Oakes (17, equation 5.4)). Since exposure cuts everyone's lifetime in half, the etiologic fraction is 1. However, the expected death rates ID_1 and ID_0 in the presence and absence of exposure will be $2/T$ and $1/T$, respectively, so that $ID_1/ID_0 = 2$ and $(ID_1 - ID_0)/ID_1 = 0.5$, much less than the true etiologic fraction. Furthermore, the incidence proportions under exposure and nonexposure at time T will be

$$1 - \exp[-(2/T)/T] = 1 - e^{-2} \text{ and}$$
$$1 - \exp[-(1/T)/T] = 1 - e^{-1},$$

so that at time T the excess fraction will be

$$[(1 - e^{-2}) - (1 - e^{-1})]/(1 - e^{-2}) = 0.27,$$

much less than $(ID_1 - ID_0)/ID_1$. Note that these results hold if ID_1 and ID_0 are interpreted as either instantaneous or average (interval) incidence densities. (In this example, $(ID_1 - ID_0)/ID_1$ does equal the proportionate reduction in life expectancy due to exposure. This relation is, however, a consequence of the constancy of the death rates, and does not hold in general.)

The quantity $(ID_1 - ID_0)/ID_1$ has been termed the "assigned shares" in the risk assessment literature (15, 16); because it can take on negative values, we propose to instead call it the incidence-density fraction. This fraction has no general relation to excess and etiologic fractions in that it may fall above, between, or below the other fractions. Nevertheless, it does have systematic relations to the other fractions under certain biologic models. For example, under certain models, $(ID_1 - ID_0)/ID_1$ will equal the etiologic fraction (16, 18), and, under a broader class of models, it is better than the excess fraction as a lower bound for the etiologic fraction (18). Thus, the incidence-density fraction may be useful in the estimation of the etiologic fraction, provided one does not lose sight of the assumptions required for such use.

If the incidence-density fraction is computed using average instead of instantaneous densities, it can, in special circumstances, approximate the excess fraction. Consider again a closed cohort of initial size N_1 with average incidence density ID_1 if exposed and ID_0 if unexposed, and incidence proportion IP_1 if exposed and IP_0 if unexposed. Let IDR be the incidence-density ratio ID_1/ID_0, and let IPR be the incidence-proportion ratio IP_1/IP_0. If the disease is rare over the study interval, IDR will approximate IPR (3, 4), so that $(ID_1 - ID_0)/ID_1 = (IDR - 1)/IDR$ will approxi-

mate the excess fraction $(IP_1 - IP_0)/IP_1 = (IPR - 1)/IPR$ for the cohort. Furthermore, since the ratio of the average densities (IDR) will exceed IPR in a closed cohort if IPR > 1, $(IDR - 1)/IDR$ can serve as an upper bound for the excess fraction in the cohort, even if the disease is not rare.

Attributable fractions and relative risks

One often sees expressions for computing an attributable fraction (AF) from some form of relative risk (RR) (i.e., a risk, rate, or odds ratio), for example $AF_e = (RR - 1)/RR$, where AF_e is the attributable fraction among the exposed, or $AF_p = P_c(RR - 1)/RR = P_c AF_e$, where AF_p is the population attributable fraction and P_c is the exposure rate among cases (2–8). The results given earlier show that if RR is interpreted as the incidence-density ratio, these computing formulas are not always valid for estimating either the excess or the etiologic fraction, whereas if RR is as the incidence-proportion ratio, the computing formulas will be valid for estimating the excess fraction in a closed cohort. It follows that if RR is replaced by an odds ratio, the computing formulas will validly approximate the excess fraction only insofar as the odds ratio approximates the incidence-proportion ratio.

RELEVANCE OF THE MEASURES

In the preceding sections, we have argued for the need to distinguish three concepts of attributable fraction: the excess fraction, the etiologic fraction, and the incidence-density fraction. In this section, we would like to examine the relevant domain of application of these quantities in public health.

If disease status as of time t is the only relevant aspect of an application, the excess fraction is the relevant measure. Consider, for example, the issue of the effect of oxytocin use on intrapartum death rates: From a public health perspective, the outcome of interest would be *whether* a death occurred by time t (end of delivery), not when the death occurred, and so the excess fraction (or its preventive analogue) would be the relevant measure. The instances in which the treatment delayed or accelerated an inevitable death would be of interest in studying the mechanism of treatment action, but would not count for or against the effectiveness of the treatment in preventing or causing intrapartum deaths. In many other planning and policy questions, the excess caseload that exposure would produce over an interval must be estimated, and here again the excess fraction is the relevant parameter.

In many situations, *when* the disease occurs is (or should be) of as much or more public health (and legal) concern than whether it occurs by some time t. For a disease inevitable by t (as in example 3, in which the disease is death and $t = 120$ years), time of occurrence is the only relevant issue. The excess fraction does not capture this beyond a simple dichotomy, and it is an inadequate measure if time of occurrence in the interval $(0, t)$ is important.

Unfortunately, even if we know exactly what the etiologic fraction is, it is not necessarily a useful measure of the effect of exposure on disease occurrence. To see this, compare the impact of the genetic conditions that produce Tay-Sachs disease and Huntington's chorea. Both conditions lead to premature death, and both may be considered to have etiologic fractions for death (among the exposed) that approach one. Nevertheless, persons who develop Tay-Sachs disease die in early childhood, whereas persons with the gene for Huntington's chorea usually survive well into adulthood and can lead rich, if shortened, lives. The etiologic fraction is not sensitive to this distinction.

Interestingly, in the preceding example, the excess fraction at age 20 years would clearly distinguish between the two conditions (since it would be near one for Tay-Sachs and near zero for Huntington's chorea), as would the incidence-density frac-

tion in early childhood. More generally, however, we would suggest turning attention to direct measures of exposure effect on incidence time whenever the latter is important. For example, one could examine expected years of life lost (mean reduction in life expectancy).

Consider again examples 1 and 2: One can estimate the average number of years of leukemia-free life lost by exposed cases, provided one can construct a reasonable estimate of the leukemia-free survival curve in the absence of exposure (19). The latter construction reduces to the common methodological problem of finding a good comparison group for the exposed population, and thus (unlike the etiologic fraction) is approachable by standard epidemiologic methods.

For public health purposes, impact measures such as reduction in life expectancy make use of incidence-time information ignored by the excess fraction, while avoiding the strong biologic assumptions usually required to estimate the etiologic fraction and some of the problems (such as crossing hazards) that can occur with incidence-density fractions. The greater emphasis on attributable fractions in the epidemiologic literature may be in part due to their simplicity, and in part due to the fact that of the possible measures, only the excess and incidence-density fractions are directly estimable from case-control data without restrictive assumptions (18, 19). These are not, however, sufficient reasons for neglecting other measures of impact.

Although in the previous example years of life lost is a more relevant measure of exposure impact than the etiologic fraction, this will not always be so, since relevance will often strongly depend on social and ethical issues. For example, a large etiologic fraction for homelessness as a risk factor for death would be of social concern, even if removing the exposure (homelessness) would result in only slight additional survival time for some persons (e.g., providing dormitories would prevent deaths due to freezing, even though some rescued persons

might soon die of effects of chronic alcoholism). Consideration of other examples, in which years of life lost or the excess fraction would be considered more relevant, shows that no single measure can be regarded as universally preferable.

IMPLICATIONS FOR INDIVIDUAL COMPENSATION

Although no single measure is universally preferable, the etiologic fraction has become established in current legal thinking regarding compensation for harmful exposure, usually under the heading of "probability of causation" (15, 16). Unfortunately, as we have shown, one cannot estimate the etiologic fraction without resorting to very strong biologic assumptions; this fact can have dramatic implications for personal-injury suits.

Because of the inability to identify exposure-induced cases, Hatch (9) has proposed that monetary awards for personal-injury suits be made in the following manner: First, the dollar amount V appropriate to compensate a single exposure-induced case is determined; then, each exposed case is awarded the (exposed) attributable fraction of this amount, $AF_e \cdot V$. If only excess (A_2) cases are considered relevant (as in the perinatal example), one could simply substitute an estimate of the excess fraction for AF_e. But if all persons who contract exposure-induced disease are considered exposure victims (as in the leukemia example), the exposed etiologic fraction should be used for AF_e, for only the etiologic fraction is interpretable as the proportion of exposed cases with exposure-induced disease. Thus, the dilemma is not resolved: Assuming the model in example 1, it would be reasonable to claim that exposure harmed all the exposed cases; after all, if not for exposure, all the cases would have had more years of healthy life than they did. Nevertheless, the very same data (24 exposed cases observed when 18 should be expected under nonexposure) are compatible with the model in example 2, in which exposure harmed only one fourth of

the exposed cases. Note that larger numbers would do nothing to resolve this dilemma.

The same problem arises for legislation mandating full compensation in individual damage suits if and only if the probability that the plaintiff's disease was induced by exposure exceeds 50 per cent (10). For a randomly sampled case, this probability is $(A_1 + A_2)/A_+$, the etiologic fraction. Use of an excess fraction to estimate this probability would yield an estimate biased against the plaintiff; use of the incidence-density fraction would also yield a biased estimate except in special cases (16, 18).

Multiple factors

The existence of biologic interactions raises difficult issues for the use of attributable fractions in compensation (20). For example, if some persons develop lung cancer solely because of their exposure to asbestos and smoking, such persons will contribute to the excess and etiologic fractions for both asbestos and smoking. As is well known (8), in such situations the attributable fractions for asbestos and smoking as causes of lung cancer among the jointly exposed can (and in fact do) sum to more than one. If a jointly exposed case receives compensation from each party responsible for each exposure, and the compensation from each is determined as an attributable fraction times the total loss incurred by the case, the total of all awards could exceed the total loss.

One theoretical solution to this problem is as follows: Suppose that two factors x and y are at issue, and let AF_x, AF_y, and AF_{xy} be the proportion of cases exposed to both factors for whom disease was attributable to x but not y, to y but not x, and to both x and y, respectively. Here, "disease attributable to" a factor or factors can mean either an excess case (i.e., disease would not have occurred without the factor(s)) or a case with an etiology involving the factor(s). Next, suppose it is decided that, for cases with disease attributable to both x and y, the party responsible for x should

pay a proportion P_x of the total compensation and the party responsible for y should pay the remainder. Finally, let V be the valuation for the total loss one case incurs from the disease. If cases with disease attributable to x and y cannot be distinguished from the other jointly exposed cases, a jointly exposed case could receive $(AF_x + P_x AF_{xy}) V$ from the party responsible for x and $[AF_y + (1 - P_x) AF_{xy}] V$ from the party responsible for y; the total compensation for a jointly exposed case would then be $(AF_x + AF_y + AF_{xy}) V$.

As in the univariate situation, AF_x, AF_y, and AF_{xy} may be estimated using standard epidemiologic methods if they represent excess fractions, whereas their estimation will require strong biologic assumptions if they represent etiologic fractions (18, 20). In particular, expressions for AF_x, AF_y, and AF_{xy} in terms of incidence densities (e.g., as in reference 21) are valid only under certain models.

In contrast to estimation of attributable fractions, determination of P_x is a legal rather than a scientific problem. Note that, among jointly attributable cases, fully 100 per cent of their disease can be causally attributed to either exposure considered singly. In a causal sense, both factors may be viewed as equally responsible for such cases, in that both factors are necessary causes of such cases. This observation does not, however, imply that the two responsible parties should pay an equal share of the compensation for such cases. For example, in current practice, lung cancer victims who are exposed to asbestos and who are smokers usually receive full compensation from the party responsible for asbestos exposure; in effect, then, the courts hold the latter party responsible for all jointly exposed cases.

Years of life lost

The above compensation rules take no account of the amount of healthy life lost by an exposed person. For example, the rules make no explicit distinction between an exposure-induced case that occurs at age

25 years and one that occurs at age 85 years. If expected years of life lost are considered relevant, one could estimate expected years of life lost separately for cases that occur at different ages, using the survival distributions in exposed and unexposed populations, and provide compensation in proportion to expected years lost. Unfortunately, unlike the overall expected years of life lost (but like the etiologic fraction), the expected years of life lost among cases that occur at a particular time is not estimable without strong biologic assumptions (19).

FURTHER ISSUES

Severity of outcome

In the above development, we have assumed that the chief manifestation of exposure playing an etiologic role in an outcome event is that the outcome event occurs earlier than it would have in the absence of exposure. Thus, cases who suffered alteration of severity of outcome without alteration of time of outcome would not be "etiologic cases" in the above sense. Severity of outcome is often not an issue (e.g., in mortality studies), but it can be of crucial importance in some contexts (e.g., compensation for pneumoconiosis). Severity issues can be dealt with in terms of transition times to different degrees of severity. Here, we note only that, if one wished to include as "etiologic" all cases with altered severity, the gap between the etiologic and excess fractions could be larger or smaller than those illustrated here, depending on the situation.

Attributable fractions and susceptible proportions

Recently, Khoury et al. (22) have sought to revive interest in the concept of the proportion of persons susceptible to exposure-induced disease. Under a deterministic model for exposure effects, an exposed person is susceptible to exposure-induced disease by time t if, in the absence of competing mechanisms, a mechanism involving exposure would induce disease in

the person by time t. In terms of a sufficient component cause model (8), a person is susceptible to exposure-induced disease if a sufficient cause involving exposure would be completed by time t if exposure is present and competing events do not occur. (This definition of susceptibility should not be confused with the definition used by Miettinen (7) and elsewhere by us (23), in which a susceptible is a person who would be an excess case in the presence of exposure.) Among the exposed, this class of susceptibles includes but is potentially larger than the class of etiologic cases ($A_1 + A_2$), for it includes certain cases whose etiology did not involve exposure (i.e., it includes certain type 0 cases, as well as all type 1 and type 2 cases). Specifically, the class of susceptibles includes those type 0 cases who would (by time t) have contracted disease from a sufficient cause involving exposure if sufficient causes not involving exposure had been absent. As with the etiologic fraction, estimation of the susceptible proportion requires strong biologic assumptions, although broad upper and lower bounds can be estimated for the quantity (22).

Attributable fractions and cofactors

Attributable fractions, like relative risks, are highly dependent on the prevalence of cofactors of exposure. (By "cofactors," we mean factors that enhance (causal cofactors) or reduce (preventive cofactors) exposure effects on risk.) For example, the gene for phenylketonuria can lead to severe mental retardation only when dietary phenylalanine is above a certain level (34). It follows that both the excess and etiologic fractions for the phenylketonuria gene as a cause of mental retardation will depend directly on the distribution of dietary phenylalanine levels in the study population, and will approach zero in a population with uniformly low phenylalanine diets. In a similar fashion, attributable fractions and relative risks depend on the incidence of competing causes (i.e., causal mechanisms that do not involve exposure).

As has been noted elsewhere (8), the dependency of epidemiologic measures on cofactor distributions points out the need to avoid considering such measures as biologic constants. Rather, epidemiologic measures are characteristics of particular populations under particular conditions (analogous to the way in which measures such as relative weight and daily caloric intake are characteristics of particular people under particular conditions, and so are not biologic constants). This should especially be borne in mind if attributable fractions are used to decide compensation, for use of estimates from the literature must assume that the population of potential plaintiffs experienced effects similar to those seen in study populations.

Preventive factors

Results for prevented fractions (2, 4, 8) parallel to those given here for attributable fractions may be obtained by noting that preventive action for a factor is logically equivalent to the factor's absence acting as a cause (24). In particular, two types of prevented fractions should be distinguished when considering a purely preventive exposure. Let A_0 be the number of exposed cases who experienced no preventive action (delay in disease occurrence) from exposure; let A_1 be the number of exposed cases who experienced some preventive action from exposure; let A_2 be the number of exposed persons in the study population who did not become cases but would have if exposure had been absent; let $A_+ = A_0 + A_1 + A_2$. Then A_2/A_+ is the actual caseload reduction produced by exposure, and is the preventive analog of the excess fraction.

On the other hand, $(A_1 + A_2)/A_+$ is the fraction of potential and actual cases who experienced some preventive action from exposure, and is the preventive analog of the etiologic fraction. Like the etiologic fraction, it is not identifiable without strong biologic assumptions; in particular, formulas for estimating this fraction from incidence densities (e.g., reference 4, p. 166) will be valid only under certain models.

Statistical methods

Beginning with the work of Walter (25), there has been extensive development of estimation techniques for attributable fractions, especially adjusted estimators (e.g., see references 26–30). The results presented here show that these estimators should be interpreted with caution. Depending on the sampling design of the study, the estimated parameter may be either an excess fraction or an incidence-density fraction. Furthermore, although adjustment may produce valid estimates of the latter two measures, in general, one should not expect it to produce a valid estimate of the etiologic fraction (18).

Terminology

The number of terms for attributable fractions is perhaps the largest of any concept in epidemiology. Two traditions are extant. One tradition continues to employ Levin's original term "attributable risk" (31), ignoring the fact that this term refers to the risk difference in several widely used textbooks (e.g., references 32 and 33), and that the quantity at issue is not itself a risk. For these reasons, a second tradition arose of amending the term "attributable risk" to "attributable risk proportion" (32), "attributable proportion" (8), "etiologic fraction" (4, 6, 7), or the hybrid term "attributable fraction" (1–3). Neither tradition is adequate to cope with the fact that at least three different quantities should be distinguished: the fractional excess caseload produced by exposure, A_2/M, which we have labelled the "excess fraction"; the fraction of cases for whom exposure played a role in the etiology of their disease, $(A_1 + A_2)/M$, which we have termed the "etiologic fraction"; and the incidence-density difference expressed as a fraction of the exposed incidence density, $(ID_1 - ID_0)/ID_1$, which we have termed the "incidence-density fraction." The incidence-density fraction can be further subdivided into two types, according to whether instantaneous or average densities are used. We have used the

term "attributable fraction" to refer to the family formed by these concepts. We can only hope that our proposed terminology helps resolve the conceptual confusion surrounding attributable fractions.

SUMMARY

We have argued that the concept of attributable fraction requires separation into the concepts of excess fraction, etiologic fraction, and incidence-density fraction. These quantities do not necessarily approximate one another, and the etiologic fraction is not generally estimable without strong biologic assumptions. For these reasons, care is needed in deciding which (if any) of the concepts is appropriate for a particular application. It appears that the excess fraction (like incidence proportion) will be most relevant in situations that require only consideration of whether disease occurs by a particular time. In situations that require consideration of when disease occurs, direct measures of effect on incidence time may be as relevant as or more relevant than any attributable fraction.

To avoid technical complications, we have not discussed additional problems of causal attribution that can arise when exposure has multiple levels or is sustained over time, and the estimation problems that can arise when considering case-control studies, competing risks, or differential censoring. For more detailed discussions of such problems and proposed solutions, see references 11–20.

REFERENCES

1. Ouellet BL, Romeder J-M, Lance J-M. Premature mortality attributable to smoking and hazardous drinking in Canada. Am J Epidemiol 1979;109: 451–63.
2. Last JM, ed. A dictionary of epidemiology. New York: Oxford University Press, 1983.
3. Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. New York: Oxford University Press, 1986.
4. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. Belmont, CA: Lifetime Learning Publications, 1982.
5. Breslow NE, Day NE. Statistical methods in cancer research. Vol 1. The analysis of case-control studies. (IARC scientific publication no. 32). Lyon: IARC, 1980.
6. Schlesselman JJ. Case-control studies: design, conduct, analysis. New York: Oxford University Press, 1982.
7. Miettinen OS. Theoretical epidemiology. New York: John Wiley and Sons, 1985.
8. Rothman KJ. Modern epidemiology. Boston: Little, Brown, 1986.
9. Hatch OG. Medical/legal aspects of radiation-induced cancer. Health Physics Society Newsletter 1984 (Dec):6–8.
10. Hatch OG. The radiogenic cancer compensation act. Congressional Record 1983;129:Issue 38.
11. Robins JM, Greenland S. The probability of causation under a stochastic model for individual risk. Technical Report no. 4, Occupational Health Program, Harvard School of Public Health, Boston, 1988.
12. Lagakos SW, Mosteller F. A case study of statistics in the regulatory process: the FD&C red no. 40 experiments. JNCI 1981;66:197–212.
13. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods—application to the control of the healthy-worker survivor effect. Mathematical Modelling 1986;7:1393–1512.
14. Robins JM. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. J Chronic Dis 1987;40(suppl 2):139S–62S.
15. Lagakos SW, Mosteller F. Assigned shares in compensation for radiation-related cancers (with comments). Risk Analysis 1986;6:345–80.
16. Cox LA. Statistical issues in the estimation of assigned shares for carcinogenesis liability. Risk Analysis 1987;7:71–80.
17. Cox DR, Oakes D. Analysis of survival data. New York: Chapman and Hall, 1984.
18. Robins JM, Greenland S. Estimability and estimation of excess and etiologic fractions. Stat Med (in press).
19. Robins JM, Greenland S. Estimability and estimation of years of life lost due to hazardous exposure. Technical Report no. 5, Occupational Health Program, Harvard School of Public Health, Boston, 1988.
20. Seiler FA, Scott BR. Mixtures of toxic agents and attributable risk calculations. Risk Analysis 1987;7:81–90.
21. Walker AM. Proportion of disease attributable to the combined effects of two factors. Int J Epidemiol 1981;10:81–5.
22. Khoury MJ, Flanders WD, Greenland S, et al. On the measurement of susceptibility in epidemiologic studies. Am J Epidemiol 1989;129:83–90.
23. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 1986;15:412–18.
24. Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. Scand J Work Environ Health 1988;14:125–9.
25. Walter SD. The estimation and interpretation of attributable risk in health research. Biometrics 1976;32:829–49.
26. Whittemore AS. Statistical methods for estimat-

ing attributable risk from retrospective data. Stat Med 1982;1:229–43.

27. Denman DW, Schlesselman JJ. Interval estimation of the attributable risk for multiple exposure levels in case-control studies. Biometrics 1983; 39:185–92.

28. Bruzzi P, Green SB, Byar DP, et al. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol 1985;122:904–14.

29. Kuritz SJ, Landis JR. Attributable risk ratio estimation from matched-pairs case-control data. Am J Epidemiol 1987;125:324–8.

30. Greenland S. Variance estimators for attributable fraction estimates consistent in both large strata and sparse data. Stat Med 1987;6:701–8.

31. Levin ML. The occurrence of lung cancer in man. Acta Union International Contra Cancrum 1953; 9:531–41.

32. Mausner JS, Bahn AK. Epidemiology. Philadelphia, PA: WB Saunders, 1974.

33. MacMahon B, Pugh TF. Epidemiology: principles and methods. Boston: Little, Brown, 1970.

34. Lloyd JK, Scriver CR (eds.). Genetic and metabolic diseases in pediatrics. London: Butterworths, 1985.