

INTERPRETATION AND CHOICE OF EFFECT MEASURES IN EPIDEMIOLOGIC ANALYSES¹

SANDER GREENLAND

The concept of the odds ratio is now well-established in epidemiology, largely because it serves as a link between results obtainable from follow-up studies and those obtainable from case-control studies (1-7). Odds ratios also naturally arise when considering small sample analysis of 2×2 tables and in logistic and log-linear modeling (2-4, 8). This ubiquity, along with certain technical considerations, has led some authors to treat the odds ratio as perhaps a "universal" measure of epidemiologic effect, in that they would estimate odds ratios in follow-up studies as well as case-control studies (6, 9); others have expressed reservations about the utility of the odds ratio as something other than an estimate of an incidence ratio (10, 11).

I believe that such controversy as exists regarding the use of the odds ratio arises from its inherent disadvantages compared with the other measures for biological inference, and its inherent advantages for statistical inference. The purpose of this paper is to compare the interpretations and statistical properties of the common measures of effect in an attempt to delineate clearly the advantages and drawbacks of each measure for epidemiologic inference. I will argue that only incidence differences and ratios possess direct interpretations as measures of impact on average risk or hazard. Consequently, odds ratios are useful only when they serve as incidence-ratio estimates, and logistic and log-linear models are useful only insofar as they pro-

vide improved (smoothed (8)) estimates of incidence differences or ratios.

INTERPRETATIONS UNDER A STOCHASTIC-RISK MODEL

For simplicity, this section will focus on the problem of estimating the effect of a binary exposure factor on the risk of a binary disease outcome over a well-defined time period (the risk period). As discussed later, the conclusions extend to more general cases, including risk considered as a function of time (as in failure-time analysis) and polytomous, continuous, and multiple exposures (as in regression models).

There are two basic conceptual models for viewing individual disease risk: deterministic and stochastic (probabilistic) (7), deterministic being the special case of stochastic in which risks may be zero or one, but not in between. A treatment of the measures in the deterministic case has recently been given elsewhere (11). The arguments of this section generalize that treatment to the stochastic case.

Individual measures

Under the stochastic model, each individual is analogous to a coin to be tossed: for each individual, i , there is a certain unknown probability r_{1i} that disease will occur when the individual is exposed, and a probability r_{0i} that disease will occur when the individual is not exposed. These risks are analogous to the probability that a coin will land heads, and thus may vary between zero and one. One may define the survival probabilities of the individual as $s_{1i} = 1 - r_{1i}$ when exposed and $s_{0i} = 1 - r_{0i}$ when unexposed.

One may also define the odds of disease for the individual, or *risk odds*, by

¹ Division of Epidemiology, University of California, Los Angeles, School of Public Health, Los Angeles, CA 90024.

The author thanks Drs. Hal Morgenstern, Charles Poole, and James Schlesselman for their helpful comments.

$w_{1i} = r_{1i}/s_{1i}$ and $w_{0i} = r_{0i}/s_{0i}$. Unlike risks, the odds are defined only if the survival probabilities s_{1i} and s_{0i} are nonzero.

The effect of exposure on the risk of an individual may be measured in terms of the risk difference $r_{1i} - r_{0i}$, the risk ratio r_{1i}/r_{0i} , the risk-odds difference $w_{1i} - w_{0i}$, or the risk-odds ratio w_{1i}/w_{0i} . Clearly, the risk ratio and risk-odds ratio will be undefined if the risk in absence of exposure r_{0i} is zero. In addition, both the risk-odds difference and the risk-odds ratio will be undefined if either survival probability is zero, i.e., if either risk is one.

Population measures

Epidemiology is largely concerned with inferences about average risks and effects in populations. In a cohort comprising N_1 exposed and N_0 unexposed individuals, the expected number of cases and noncases over the risk period would be as follows:

average risk to the average survival probability; that is,

$$\frac{A}{C} = \frac{\sum_1 r_{1i}/N_1}{\sum_1 s_{1i}/N_1}$$

and

$$\frac{B}{D} = \frac{\sum_0 r_{0i}/N_0}{\sum_0 s_{0i}/N_0}$$

Note, however, that the incidence odds do not equal the simple averages of the risk odds; that is, $A/C \neq \sum_1 w_{1i}/N_1$ and $B/D \neq \sum_0 w_{0i}/N_0$. Although the incidence odds do equal the average risk odds when the risk odds are weighted by the survival probabilities, it will be shown below that the failure of the incidence odds to equal the simple average risk odds seriously handicaps the interpretability of measures based on the incidence odds.

Assume there is no confounding, in the sense that had exposure been completely

	Subcohort 1 (exposed)	Subcohort 0 (unexposed)	Total
Disease occurs	$A = \sum_1 r_{1i}$	$B = \sum_0 r_{0i}$	M_1
Disease does not occur	$C = \sum_1 s_{1i}$	$D = \sum_0 s_{0i}$	M_0
Total	N_1	N_0	T

where \sum_1 and \sum_0 mean summation over all individuals in subcohort 1 (exposed) and subcohort 0 (unexposed), respectively. The proportion expected to contract the disease in a group is the cumulative incidence (1) or *incidence proportion* (12). The incidence proportions A/N_1 and B/N_0 are interpretable as average risks in their respective groups, i.e., $A/N_1 = \sum_1 r_{1i}/N_1$ and $B/N_0 = \sum_0 r_{0i}/N_0$.

One can also construct another measure of disease occurrence in the above population. The ratio of the number expected to contract disease to the number expected to not contract disease in a group is the disease odds or *incidence odds*, which takes on the values A/C for the exposed and B/D for the unexposed. The incidence odds A/C and B/D are interpretable as ratios of the av-

absent, the average risk would have been the same among the subcohorts that were in fact the exposed and the unexposed (12); that is, $\sum_1 r_{0i}/N_1 = \sum_0 r_{0i}/N_0$. The incidence-proportion difference is then given by

$$\begin{aligned} \frac{A}{N_1} - \frac{B}{N_0} &= \sum_1 r_{1i}/N_1 - \sum_0 r_{0i}/N_0 \\ &= \sum_1 r_{1i}/N_1 - \sum_1 r_{0i}/N_1 \quad (1) \end{aligned}$$

$$= \sum_1 (r_{1i} - r_{0i})/N_1. \quad (2)$$

Thus, this incidence difference is interpretable as both the absolute change in the average risk of the exposed subcohort produced by exposure (expression 1, the average-risk difference) and the average absolute change in risk produced by exposure among exposed individuals (expression 2, the average risk-difference).

The incidence-proportion ratio is given by

$$\begin{aligned} \frac{A/N_1}{B/N_0} &= \frac{\sum_1 r_{1i}/N_1}{\sum_0 r_{0i}/N_0} \\ &= \frac{\sum_1 r_{1i}/N_1}{\sum_1 r_{0i}/N_1}. \end{aligned} \quad (3)$$

Thus, the incidence-proportion ratio is interpretable as the proportionate change in the average risk of the exposed subcohort produced by exposure (expression 3, the average-risk ratio). Nevertheless, it is not interpretable as the average proportionate change in risk produced by exposure among exposed individuals, i.e., the average risk-ratio

$$\sum_1 (r_{1i}/r_{0i})/N_1, \quad (4)$$

which is undefined if unexposed risks of zero occur. If, however, the individual risk ratios r_{1i}/r_{0i} are all equal to a constant value, expressions 3 and 4 and the incidence proportion ratio will all equal that value.

The incidence-odds ratio is given by

$$\begin{aligned} \frac{A/C}{B/D} &= \frac{\sum_1 r_{1i}/\sum_1 s_{1i}}{\sum_0 r_{0i}/\sum_0 s_{0i}} \\ &= \frac{\sum_1 r_{1i}/\sum_1 s_{1i}}{\sum_1 r_{0i}/\sum_1 s_{0i}}. \end{aligned} \quad (5)$$

This expression represents the proportionate change in the incidence odds in the exposed produced by exposure. Nevertheless, it is *not* equivalent to the proportionate change in the average odds in the exposed produced by exposure,

$$\frac{\sum_1 w_{1i}/N_1}{\sum_1 w_{0i}/N_1}. \quad (6)$$

Neither of the last two expressions is equivalent to the average of the individual odds ratios among the exposed,

$$\sum_1 (w_{1i}/w_{0i})/N_1 \quad (7)$$

which is undefined if any risks of one or unexposed risks of zero occur. Thus, the incidence-odds ratio lacks any simple interpretation in terms of exposure effect on

average risk or odds, or average exposure effect on individual risk or odds. Parallel arguments show that the same is true of the incidence-odds difference.

If the individual risk-odds ratios w_{1i}/w_{0i} are all equal to a constant value (as assumed, for example, by a logistic-risk model), expressions 6 and 7 will equal that value, yet the incidence-odds ratio need *not* equal that value. For example, in a population in which 10 per cent of individuals had $r_{1i} = 0.60$ and $r_{0i} = 0.20$, and the remainder had $r_{1i} = 0.035$ and $r_{0i} = 0.006$, all the risk-odds ratios (and, thus, expressions 6 and 7) would equal 6.0. However, the incidence proportion would be $0.10(0.60) + 0.90(0.035) = 0.0915$ under exposure and $0.10(0.20) + 0.90(0.006) = 0.0254$ under nonexposure, yielding an incidence-odds ratio of $0.0915(1 - 0.0254)/0.0254(1 - 0.0915) = 3.9$.

Connections to confounding criteria

The failure of the incidence-odds ratio to equal expressions 6 or 7, even when w_{1i}/w_{0i} is constant, can be seen as an analog of the "paradoxical" behavior of the odds ratio noted by Miettinen and Cook (10) in their example 3. That example showed that the elevation in the crude incidence odds produced by exposure can fall short of the elevation in odds produced in any subgroup, even if confounding (as defined above) is entirely absent. Such paradoxical behavior cannot occur with the incidence-proportion difference or ratio (10). Boivin and Wacholder (9) failed to note that the crude odds ratio in example 3 of Miettinen and Cook (10) is an unbiased estimate of expression 5 (the true exposure effect on the incidence odds in the exposed); as a result, they postulated that odds ratio non-confounding corresponds to the crude odds ratio being equal to a weighted average of stratum-specific odds ratios. This postulate overlooks the "defect" in the incidence-odds ratio demonstrated above, i.e., the crude incidence-odds ratio may equal neither an average effect nor an effect on

average odds, and yet still unbiasedly represent the effect of exposure on the incidence odds.

Density measures

The above arguments deal only with comparisons of proportions getting disease (or not) in a simple closed cohort, in which everyone is observed throughout their risk period. Some populations reasonably approximate this model, especially in the fields of perinatal epidemiology and technology assessment. Even when loss to follow-up occurs, various methods still allow one to estimate the incidence proportions for the original cohort (4, 6). Nevertheless, most chronic disease studies are based on open (dynamic) populations, in which the incidence proportion cannot be directly measured or even simply defined. This has led to the development and use of concepts of person-time rates and incidence density (1, 2, 4-7) and comparisons based on these "density" measures. Somewhat lengthy development is required to connect incidence-density comparisons to exposure effects on incidence proportion (1, 4, 7), and the resulting connection is fairly abstract (much like the concept of incidence density itself). Nevertheless, if the disease is "rare" and censoring is unrelated to risk, the incidence-density ratio will approximate the incidence-proportion ratio and do so more closely than the incidence-odds ratio (5).

Incidence-density measures may also be directly linked to individual failure-time (incidence) distributions: Suppose at time t the hazard (13) for an individual i in a population is $h_i(t)$, the instantaneous incidence density in the population is $ID(t)$, and the size of the population is $N(t)$; then $ID(t) = \sum h_i(t)/N(t)$ (a proof is given in the Appendix). Thus, like incidence proportion and risk, but unlike incidence odds and risk odds, the population measure (incidence density) is a simple average of the individual parameters (here, hazards). As a consequence, incidence-density differences and ratios may be interpreted as differences and ratios of average hazards. As described

in Implications for Modeling, these interpretations generalize to link density measures and failure-time models.

IMPLICATIONS FOR MODELING

The preceding observations have important implications for inferences about parameters in biologic models for individual disease risks or hazards. Under general risk-difference or risk-ratio models and certain mixtures of these, the form of covariate effects at the individual level will (in the absence of uncontrolled confounding) be followed at the population level by the incidence proportions. To see this, consider first a model stating that the risk of an individual i with covariate level x is given by $\alpha_i + d(x; \beta)$, where α_i is a random effect independent of x , and $d(x; \beta)$ is a general risk-difference function, e.g., βx (β and x may be vectors). If the size of the observed population at level x is $N(x)$, the incidence proportion at level x will be $\sum(\alpha_i + d(x; \beta))/N(x) = \alpha(x) + d(x; \beta)$, where $\alpha(x) = \sum \alpha_i/N(x)$, the mean of the α_i at level x , and the sums are over individuals observed at level x (the no-confounding assumption given earlier translates into assuming the $\alpha(x)$ are constant across x ; the assumption of independence of the α_i and x is, however, sufficient for inference on β). Assume next a model in which risk is given by $r(x; \gamma)\lambda_i$, where $r(x; \gamma)$ is a general risk-ratio function, e.g., $\exp(\gamma x)$. The incidence proportion at level x will then be $\sum r(x; \gamma)\lambda_i/N(x) = r(x; \gamma)\lambda(x)$, where $\lambda(x)$ is the mean of the λ_i at level x . Finally, assume an additive mixture model $\alpha_i + d(x; \beta) + r(x; \gamma)\lambda_i$; the incidence proportion will then be $\alpha(x) + d(x; \beta) + r(x; \gamma)\lambda(x)$, where $\alpha(x)$ and $\lambda(x)$ are as before.

In contrast, the form of covariate effects in a general model for the difference or ratio of the risk odds (e.g., the logistic model, which states the risk odds is given by $\exp(\alpha_i + \beta x)$) will generally not be followed by the incidence odds. In order to force a correspondence between a risk-odds model and the incidence odds, much of the statistical literature implicitly assumes that

the random effects α_i are constant *within* covariate levels. However, such an assumption is unrealistic because it translates into a biologic assumption that all individuals observed at the same covariate level have the same risk, i.e., that the measured covariates are the only important risk factors.

In a manner parallel to that just given for risk models and incidence proportions, one can show that under general hazard difference or general hazard ratio (proportional hazards) models and certain mixtures of these, the form of covariate effects at the individual hazard level will (in the absence of uncontrolled confounding) be followed at the population level by the instantaneous incidence densities.

APPROXIMATE INTERPRETATIONS

While the incidence-proportion difference can be directly interpreted as an effect on average risk, and as an average effect on risk, the incidence-proportion ratio can ordinarily be interpreted in only the first of these ways, and the incidence-odds ratio lacks both interpretations. Under certain circumstances, however, the deficiencies of the ratio measures disappear.

Odds approximation to proportions

If the disease is rare, the incidence odds will approximate the incidence proportions, and consequently the odds and odds ratios can be used as substitutes for the more easily interpreted incidence proportions and their ratios (1-7). Nevertheless, the necessary rarity condition for this substitution should be properly appreciated: The incidence proportion should be low in all exposure and confounder categories of the analysis; it is simply a mistake to require only that the crude incidence be low. This problem arises, for example, in studies of perinatal mortality, in which the crude mortality is usually low, but is high in certain subgroups (e.g., very low birth weight infants). There is a rule of thumb for judging how low incidence must be to allow the rare disease assumption to be invoked: If the odds never exceeds X in any of the

subgroups to be compared, the odds or odds ratio will incorporate no more than $100X$ per cent error in estimating the corresponding proportion or ratio of proportions. For example, if the odds never exceeds 0.10, substitution of odds ratios for proportion ratios will lead to no more than 10 per cent error in estimating the latter. This rule is derived by noting that

$$\frac{A/C}{B/D} = \frac{A/N_1}{B/N_0} \left(\frac{N_1/C}{N_0/D} \right) = \frac{A/N_1}{B/N_0} \left(\frac{1 + A/C}{1 + B/D} \right).$$

The second factor in the last term is the bias in using the odds ratio as an estimate of the proportion ratio; if the odds A/C and B/D are both under X , this bias factor must fall between $1 - X$ and $1 + X$.

Incidence-odds approximation to average risk odds

The rare disease assumption is *not* sufficient to allow one to interpret the incidence odds as the average of the individual risk odds, and thus is not sufficient to allow interpretation of the incidence-odds ratio as the change in average risk odds. For example, in a population in which 2 per cent of individuals had a risk odds of 1.00 and the remainder had a risk odds of 0.01 (which translates to risks of 0.50 and 0.01), the incidence proportion would be $0.02(0.50) + 0.98(0.01) = 0.02$, and so the incidence odds would be $0.02/0.98 = 0.02$, but the average of the risk odds would be $0.02(1.00) + 0.98(0.01) = 0.03$. Examples of this nature are not hard to find in perinatal epidemiology. A sufficient condition for the incidence odds to approximate the average risk odds (if defined) is that all the individual risk odds be low; unlike the rare disease assumption, however, this condition cannot be verified by examining the data.

Incidence-proportion ratio approximation to average risk ratio

The rare disease assumption is also insufficient to allow one to interpret the incidence-proportion ratio as the average of the individual risk ratios (expression 4

above). For example, suppose that in the study population 2 per cent of individuals had $r_{1i} = 0.50$ and $r_{0i} = 0.20$, and the remainder had $r_{1i} = 0.02$ and $r_{0i} = 0.002$. The incidence proportion for this population would be $0.02(0.50) + 0.98(0.02) = 0.03$ under exposure and $0.02(0.20) + 0.98(0.002) = 0.006$ under no exposure, for an incidence-proportion ratio of $0.03/0.006 = 5$; in contrast, the average of the individual risk ratios would be $0.02(0.50/0.20) + 0.98(0.02/0.002) = 10$. A sufficient (but unverifiable) condition for the incidence-proportion ratio to approximate the average risk ratio (if defined) is that all the individual risks be low.

STATISTICAL CONSIDERATIONS

The statistical properties of the measures discussed here have been studied in detail. The literature is vast and highly technical, and I will not attempt a review; the points I wish to consider follow directly from the cited references.

Sparse data efficiency

Epidemiologic studies frequently produce "sparse data," i.e., data that upon stratification by relevant variables (such as matching factors) yield small strata. For example, a twin- or neighborhood-matched pair study that retains the natural pairing will yield data with only two subjects per stratum and thus the data will be sparse. (One should not confuse the term sparse data with "small sample," because sparse data sets may be quite large, as in large matched studies.) For such data, the odds ratio possesses clear if rather technical advantages for formal statistical analysis: Sparse data methods that assume and estimate a constant odds ratio are highly efficient (in the statistical sense) (14), whereas sparse data methods that assume and estimate a constant difference or ratio of proportions can be highly inefficient (15).

Plausibility of homogeneity assumptions

The assumption of constancy (homogeneity) of an effect parameter is statistically

convenient but biologically stringent, and it is good practice to critically examine the assumption before applying a technique based on it. There are no purely logical or general biologic reasons for believing such an assumption, but in certain situations there are purely logical reasons for disbelieving constancy of the difference or ratio of proportions. These reasons arise from the inherent range limitations of these measures. In 2×2 table notation, the difference cannot exceed A/N_1 or fall below $-B/N_0$, and the ratio cannot exceed N_0/B . For example, if the incidence proportion among the unexposed was known to range as high as 0.5 in some strata, the incidence-proportion difference could not exceed 0.5 and the incidence-proportion ratio could not exceed 2.0 in those strata (since incidence proportions cannot exceed 1.0). If the incidence-proportion difference observed in other strata clearly exceed 0.5, one would have to rule out constancy of the difference; similarly, if the incidence-proportion ratio observed in other strata clearly exceed 2.0, one would have to rule out constancy of the proportion ratio.

The odds ratio suffers from no such a priori range limitations, and so for common diseases the constant incidence-odds ratio assumption is logically less vulnerable to objection than are the other constancy assumptions. If, however, the disease is rare, the limit of the size of the incidence-proportion difference and ratio will be so high as to cause no problems.

Modeling considerations

Odds ratios arise naturally as antilogs of simple linear combinations of logistic or log-linear model coefficients (2-4, 8), and thus have been promoted as a link between the results of stratification analyses and modeling (2, 6); similarly, hazard ratios arise naturally as antilogs of linear combination of Cox model coefficients (6, 13). The "naturalness" of these connections, however, reflects only mathematical convenience and should not be taken to impart any special biologic importance to either the measures or the models (7, 16, 17)

(although this convenience may explain the preferences of many statisticians and software developers for the models).

Case-control analysis

If we now consider the earlier 2×2 table as observed case-control data, A/N_1 and B/N_0 cease to be meaningful quantities, and (without external information) analysis must depend on the case-control exposure-odds ratio $(A/B)/(C/D) = AD/BC$. Nevertheless, one should not generally equate this odds ratio to the incidence-odds ratio (as is done in most elementary texts): Depending on the specific methods of case and control selection, the case-control odds ratio may directly estimate the incidence-proportion ratio, incidence-density ratio, or incidence-odds ratio (1, 5, 18, 19). Even when the case-control odds ratio directly estimates the incidence-odds ratio, "disease rarity" will allow the case-control odds ratio to be used as an estimate of the incidence-proportion ratio (1-7), and external information about population rates of disease or exposure will allow direct estimation of the incidence proportions or densities (4, pp. 174-5). Thus, most case-control analyses need not be *interpreted* in terms of odds ratios.

Prevalence data

The situation here somewhat parallels case-control analysis: under certain conditions, the prevalence-odds ratio directly estimates the incidence-density ratio (1, 7). Generally, however, inferences about risk variation from prevalence data will require more restrictive assumptions than will similar inferences from incidence-density case-control studies (1, 4).

CONCLUSION

I have argued that, for summarizing exposure impact on risk, the incidence-proportion ("risk") difference and ratio should be the measures of choice, for in the absence of bias only they possess direct interpretations in terms of exposure effect on average risk. The incidence-proportion difference possesses an additional interpre-

tation as an average effect of exposure on risk. The incidence density can be directly interpreted as an average hazard, and thus can be used to estimate parameters of failure-time distributions; if the disease is rare over the risk period, the incidence-density ratio (properly averaged) closely approximates the incidence-proportion ratio and so inherits the latter's interpretation as well. The most common measure in epidemiologic statistics, the odds ratio, is biologically interpretable only insofar as it estimates the incidence-proportion or incidence-density ratio.

Nearly all unbiased etiologic studies can and should provide estimates of exposure effects on incidence proportions or densities. Odds ratios and parameters of multivariate models will often be useful in serving as or in constructing the estimates, but should not be treated as the end product of a statistical analysis of epidemiologic data or as summaries of effect in themselves. It has been argued elsewhere that standardized regression coefficients, correlations, and "variance explained" are also improper summaries of effect (20).

REFERENCES

1. Miettinen OS. Estimability and estimation in case-referent studies. *Am J Epidemiol* 1976; 103:226-35.
2. Breslow NE, Day NE. Statistical methods in cancer research. Vol. 1. The analysis of case-control studies. IARC Scientific Publications no. 32. Lyon: International Agency for Research on Cancer, 1980.
3. Schlesselman JJ. Case-control studies: design, conduct, analysis. New York: Oxford University Press, Inc., 1982.
4. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. Belmont, CA: Lifetime Learning Publications, 1982.
5. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *Am J Epidemiol* 1982;116:547-53.
6. Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. New York: Oxford University Press, Inc., 1986.
7. Rothman KJ. Modern epidemiology. Boston: Little, Brown & Co., 1986.
8. Bishop YMM, Fienberg SE, Holland PW. Discrete multivariate analysis: theory and practice. Cambridge, MA: MIT Press, 1975.
9. Boivin J-F, Wacholder S. Conditions for confounding of the risk ratio and of the odds ratio.

- Am J Epidemiol 1985;121:152-8.
10. Miettinen OS, Cook EF. Confounding: essence and detection. Am J Epidemiol 1981;114:593-603.
 11. Miettinen OS. Theoretical epidemiology. New York: John Wiley & Sons, 1985.
 12. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 1986;15:412-18.
 13. Kalbfleisch JD, Prentice RL. The statistical analysis of failure-time data. New York: John Wiley & Sons, 1980.
 14. Breslow NE. Odds ratio estimators when the data are sparse. Biometrika 1981;68:73-84.
 15. Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. Biometrics 1985;41:55-68.
 16. Greenland S. Limitations of the logistic analysis of epidemiologic data. Am J Epidemiol 1979; 110:693-8.
 17. Semiatycki J, Thomas DC. Biological models and statistical interactions: an example from multi-stage carcinogenesis. Int J Epidemiol 1981;10: 383-7.
 18. Miettinen OS. Design options in epidemiologic research: an update. Scand J Work Environ Health 1982;8(suppl 1):7-14.
 19. Greenland S, Thomas DC, Morgenstern H. The rare-disease assumption revisited: a critique of "Estimators of relative risk in case-control studies." Am J Epidemiol 1986;124:869-76.
 20. Greenland S, Schlesselman JJ, Criqui MH. The fallacy of employing standardized regression coefficients and correlations as measures of effect. Am J Epidemiol 1986;123:203-8.

APPENDIX

Proof that the instantaneous incidence density equals the average hazard

Let $r_i(t' | t)$ be the risk of an individual i up to $t' > t$, given survival to t , and let $IP(t' | t)$ be the proportion of population members at t who would become ill by t' . Then $h_i(t) = \lim_{t' \rightarrow t} r_i(t' | t)/(t' - t)$ (13, p. 6) and $ID(t) = \lim_{t' \rightarrow t} IP(t' | t)/(t' - t)$ (7, p. 31). By the arguments given in the text, $IP(t' | t) = \sum r(t' | t)/N(t)$, where the sum is over the population membership at t ; dividing both sides of this equation by $t' - t$ and letting t' go to t yields $ID(t) = \sum h(t)/N(t)$.