International Journal of Epidemiology

Vol. 9 No. 1
Printed in Great Britain

Measures of Disease Incidence Used in Epidemiologic Research

HAL MORGENSTERN[1,2] DAVID G KLEINBAUM[3,4] and LAWRENCE L KUPPER[4]

Morgenstern H [Department of Epidemiology and Public Health, School of Medicine, Yale University, 60 College Street, New Haven, CT 06510], Kleinbaum D G and Kupper L L. Measures of disease incidence used in epidemiologic research. *International Journal of Epidemiology* 1980, 9: 97–104.

This paper distinguishes between 2 concepts for measuring the incidence of disease: risk and rate. Alternative procedures for estimating these measures from epidemiologic data are reviewed and illustrated. An attempt is made to integrate statistical principles with epidemiologic methods while minimizing the use of higher mathematics. Several theoretical and practical criteria are discussed for choosing the appropriate incidence measure in the planning of a study and for selecting the best method of estimation in the analysis.

In a follow-up study of human populations, the fundamental concerns to the investigator are the measurement, analysis, and interpretation of an observed set of new events: incident cases of disease, deaths from one or more causes, or new occurrences of any health-related event. Despite the long history of this type of research, dating back well over a century, much of the current literature on epidemiologic methods (including textbooks) regards disease incidence as a single concept that is readily operationalized from longitudinal data. The purpose of this paper is to distinguish 2 basic concepts for quantifying incidence and to compare alternative methods for estimating each of them. While most of the ideas that follow are not new, it is felt that the theoretical work of statisticians has not been adequately applied to the understanding of empirical research. Perhaps, one explanation for this situation is the increasing lack of mutual understanding between health researchers who are primarily concerned with findings and statisticians who are more concerned with analytic methods. An attempt is made in this paper to merge statistical principles with epidemiologic methods while minimizing the use of higher mathematics.

[1] Department of Epidemiology and Public Health, School of Medicine, Yale University, 60 College Street, New Haven, CT 06510, USA.
[2] Center for Health Studies, Institution for Social and Policy Studies, Yale University, 89 Trumbull Street, New Haven, CT 06520, USA.
[3] Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, NC 27514, USA.
[4] Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, NC 27514, USA.

The work on this paper was partly funded by two grants from: the National Cancer Institute (#1-PO1-CA-16359-05) and the Henry J. Kaiser Family Foundation.

BASIC THEORETICAL CONCEPTS

Two distinct concepts may be employed to assess the expected occurrence of new (incident) events: risk and rate. Risk is the probability of an individual developing a given disease or health status change over a specified period, conditional on not dying from any other causes during that period. Thus, risk can vary between zero and one and is dimensionless. It is important to recognize that the concept of risk requires a specific period referent – e.g., the *5-year* risk of developing lung cancer. Defining risk as a probability is required not because we necessarily believe that disease occurrence is a random event, but rather to express our ignorance of the causal process and how to observe it. Of course, it is recognized that the term 'risk' or its derivatives (e.g., risky, at-risk, risk factor, etc.) is commonly used in a much broader framework than suggested by the above definition. Nevertheless, in order to give the concept scientific meaning, it is important to treat risk as a conditional (a priori) probability referring to an individual.

The concept of rate involves a more complex formulation than does risk and therefore is somewhat more difficult to understand. In general, a rate is an instantaneous potential for change in one quantity per unit change in another quantity – where, typically, the second quantity is time (e.g., velocity). The (instantaneous) rate of development of a particular disease in a population might be expressed as the number of new cases per year. The

limitation of the latter quantity is that it does not reflect the size or experience of the candidate population (at-risk) from which the cases develop. Consequently, epidemiologists are more concerned with the 'relative rate' (1), which is defined as the instantaneous potential for change in disease status (i.e., the occurrence of new cases) per unit of time, relative to the size of the candidate population at time t. More descriptively, the (relative) incidence rate reflects the 'force of morbidity' (or mortality) on the population and therefore has no direct interpretation on the individual level. Also, in contrast to the concept of risk, the incidence rate refers to a 'point in time' and has no period referent. Furthermore, the rate is not dimensionless but is expressed in units of 1/time (e.g., years^{-1}). As a result, the incidence rate, unlike risk, does not have an upper limit of one but, in fact, theoretically can approach infinity. This feature can be seen by simply altering the unit of time. For example, an incidence rate of .05 weeks^{-1} is equivalent to 2.6 years^{-1} , which is greater than one. Suppose an earthquake killed every member of a village at exactly the same moment; the death rate at that moment would be infinite. Yet, the risk of a village member being killed by the earthquake on that day, month, year, or any period of time that includes the disaster cannot exceed one.

ESTIMATION OF RISK AND RATE

The most commonly used method for estimating risk from follow-up data is to calculate the proportion of candidate subjects at the onset, who develop the disease during the subsequent follow-up period. The numerator is the number of newly detected cases (I), and the denominator is the number of disease-free subjects at the beginning of follow-up (N_o). While this fraction is generally called an incidence 'rate' by researchers, it should be noted that the term is a misnomer since the proportion is an estimate of a probability (1). We will adopt the term 'cumulative incidence' (2) to express the proportion of a 'fixed' cohort for which the disease is detected during a subsequent follow-up period (Δt); thus, the risk (R) is estimated by calculating the cumulative incidence (CI),

$$\hat{R}_{\Delta t} = CI_{\Delta t} = I/N_o. \quad (1)$$

The cohort is 'fixed' in the sense that no entries are allowed during the follow-up period. Yet, the size of the cohort is likely to be reduced during the follow-up period as a result of deaths and other sources of attrition. This presents a problem, however, since risk was defined as a *conditional* probability; and strictly speaking, a subject should not be included in the denominator (N_o) unless he/she is followed for the entire duration of Δt or is known to develop the disease of interest. Consequently, even assuming no unnecessary loss to follow-up and complete autopsies done on all deaths, we do not know whether a subject, who died of another cause, would have developed the disease of interest during the remaining follow-up period. This problem is often referred to as one of 'competing risks' or 'competing causes of death' and becomes increasingly a more relevant issue as the follow-up period gets longer.

Another difficulty in using the CI to estimate risk is that subjects may be followed for different durations as a result of the study design. For example, a large clinical trial typically enrols subjects into the study throughout the data collection period, which may last for years. Thus, regardless of the attrition and mortality rates, the length of follow-up will vary considerably among subjects. Therefore, any attempt to estimate risk by applying equation (1) will not effectively make use of information from all subjects (3).

Closely associated with the above issues is the distinction between a 'fixed' cohort and a 'dynamic' population. While no subjects may be added to a fixed cohort after the onset of follow-up, the composition of a dynamic population is more fluid, allowing for additions during the follow-up period. Very often, a fixed cohort is selected from a dynamic, source population to which causal inferences are made. If the source population is a geographically defined region (e.g., a county), the study cohort is likely to become increasingly dissimilar to or unrepresentative of the source population as the follow-up proceeds. For instance, after 10 years, the remaining cohort will have aged 10 years while the mean age or the distribution of ages in the dynamic source population may not have changed. Consequently, the cases that develop in the study cohort will become increasingly unrepresentative of the cases that develop in the source population. The reason for this difference is that while subjects are not added to the study cohort after the onset of follow-up, the composition of the source population is dynamic and may therefore remain in equilibrium.

As a hypothetical illustration of the inherent theoretical difficulties in using the CI to estimate risk, consider a fixed cohort of 5 healthy subjects who are followed for 5 years (or less) for detection

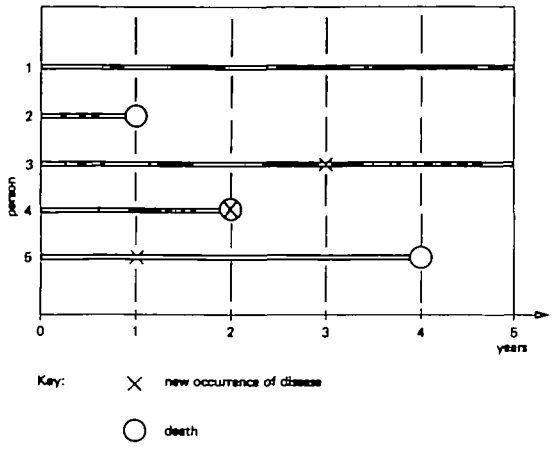

FIGURE 1 *Hypothetical 5-year follow-up experiences of 5 persons for detection of disease X, used for illustrating the calculation of incidence measures (see text)*

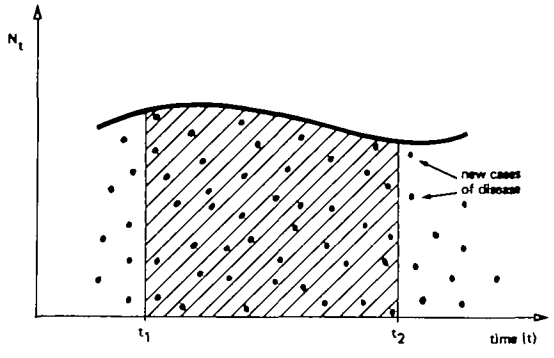

FIGURE 2 *Graphical illustration of the occurrence of new (incident) cases over time in a candidate population (of size N_t at time t)*

of an incurable disease X. Figure 1 is a diagrammatic representation of the follow-up experience of these 5 persons, none of whom are lost to follow-up for any reason except death. Calculation of the 5-year CI presents a problem because the size of the denominator is ambiguous. If the total sample size is used, it must be acknowledged that one noncase (no. 2) died before the end of the 5-year period and therefore might have eventually developed the disease. On the other hand, excluding this subject from the denominator could result in an overestimate of the risk. The dilemma is further complicated by an individual follow-up experience of less than 5 years for 2 of the cases (nos. 4 and 5).

An estimate of the incidence rate is obtained by calculating an average rate for the age/time duration of follow-up, analogously to the use of 'speed' as an average velocity. This average incidence rate will be called the 'incidence density' (ID), as suggested by Miettinen (2). An intuitive appreciation of ID can be gained by representing the new cases of disease as 'points' in a 'sea of population-time', as illustrated in Figure 2. A candidate population is followed between time t_1 and t_2 for detection of new cases as indicated by dots on the graph. The numerator of the ID is the number of dots under the curve between t_1 and t_2 — i.e., the number of new cases (I). The denominator is the amount of shaded area under the curve between t_1 and t_2 — i.e., the product of population x time (PT), which may be expressed in person-years, person-days, or the like. Thus, the estimate of the average incidence rate, or ID, between t_1 and t_2 in the candidate population of size N_t at time t is,

$$ID = I/PT \qquad (2)$$

Figure 2 suggests that ID may be viewed as the concentration or 'density' of new case occurrences in a sea of population time. The more dots per unit area under the curve, the greater is the ID. In this sense, ID reflects the average force of morbidity in a given population and may be computed directly using expression (2). While the (instantaneous) incidence rate — also called a 'hazard function,' 'probability density function', or 'failure rate' — may be mathematically predicted, a priori, using certain theoretical models of pathogenesis (e.g., 4, 5), this paper will be restricted to the calculation of average rates or ID's.

While the concept of population-time (e.g., person-years) is not new to epidemiology (6, 7), its implication to the estimation of the incidence rate (versus risk) has often not been appreciated. As an illustration of the calculation of an ID, again consider the example of 5 subjects who are followed for detection of an incurable disease X, as shown in Figure 1. There are 3 incident cases detected during the follow-up period (subjects 3, 4, and 5), and the total number of person-years contributed by this group is the sum of the individual contributions: 5 + 1 + 3 + 2 + 1 = 12. According to equation (2), the ID for disease X is 3/12 = .25 $yr.^{-1}$ Next, suppose the disease is either immediately fatal (no. 4) or cured without subsequent immunity (nos. 3 and 5). In order to calculate the *total* ID of the disease, including multiple occurrences in the same individual (though none were observed), the total number of observed person-years would be: 5 + 1 + 5 + 2 + 4 = 17, resulting in an ID of 3/17 = .176 yr^{-1} . Notice that the candidate population,

whose experience is to be included in the denominator,, is the disease-free population when calculating the ID for first occurrences. On the other hand, when calculating the total ID for all occurrences, the candidate population becomes all living subjects not lost to follow-up. Thus, for example, subject number 3 is 'at risk' for 3 years if the disease is assumed incurable and for 5 years if the disease is assumed immediately fatal without subsequent immunity.

Calculation of the ID from a dynamic population is a more subtle matter, though the method is simplified if we assume the population is *stable* over the follow-up period. Stability implies that the size and the age distribution of the study population remain constant. The total amount of PT is computed by multiplying the size of the candidate population in a certain age category by the actual duration of follow-up. For example, a stable dynamic population of 1000, 50 year-old men followed for 2 years would contribute 2000 person-years. The computation must generally involve small age categories because the underlying assumption of the analysis is that the (instantaneous) rate remains constant over the age/time follow-up period; yet, the incidence rate is known to vary substantially by age for most diseases. The assumption of a stable population and the calculation of person-years as described above is the essence of demographic life table analysis which typically deals with death from all causes as the incident event of interest (1, 6, 8, 9).

Figure 3 is provided as a graphical aid to understanding the calculation of population-time from a stable, dynamic (age-specific) population. For the sake of simplicity, we can imagine that all new entries (A) into the candidate population (N) during the follow-up period (Δt) occur at the midpoint of Δt and that all exits (A) occur at the same time. Thus, the total size of the stable population at any time during the period is N. We can then calculate the total amount of PT by summing the PT within each of the 3 component groups: entries (A), exits (A), and persons who were followed for the entire duration (N-A). This is equivalent to adding the areas of the 3 bands in Figure 3: PT = $A(\Delta t/2) + A(\Delta t/2) + (N\text{-}A)\Delta t = N{\cdot}\Delta t$. The ID is simply computed by dividing the number of new cases detected during the follow-up period among persons within a specific age category by the amount of PT for the sub-population, using the above expression for PT.

The assumption of a constant rate over the follow-up period of age/time is most important to

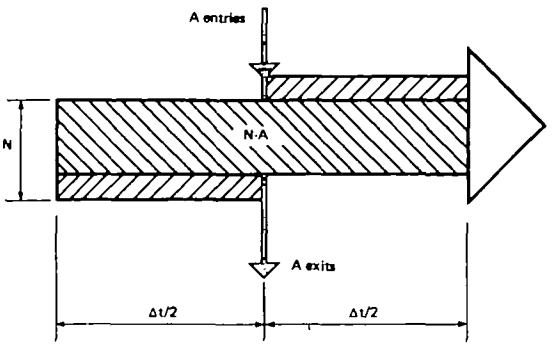

FIGURE 3 *Diagrammatic representation of a stable, dynamic population, used for illustrating the calculation of follow-up experience (see text)*

the analysis of such data and the interpretation of one's findings. To amplify the meaning of this key assumption, consider a fixed cohort of persons (N_o) at time t_o, who are followed until death. Given a constant death rate, estimated as ID, it can be shown that the number of survivors (N_t) at any subsequent time t is an exponential function with a negative slope – similar to a constant rate of radioactive decay. Thus, a constant death rate means that the same proportion of survivors at the beginning of any period of fixed length (Δt) will die during that period. We may express the constant rate (or ID) as,

$$-\mathrm{ID} = N_t'/N_t \qquad (3)$$

where: N_t' = the slope of the curve (N_t) at point t, which is the slope of the tangent to the curve at that point or; simply, the rate of change in the size of N_t. (Note: a negative sign is required because the slope of N_t is always negative.)

Implicit in the above discussion is the idea that the disease rate and risk are mathematically related. Using equation (3) and a little calculus, it can be demonstrated (1) that

$$N_t = N_o \exp\,[-\mathrm{ID}{\cdot}\Delta t].$$

Thus,

$$\hat{R}_{\Delta t} = 1 - N_t/N_o = 1 - \exp\,[-\mathrm{ID}{\cdot}\Delta t]. \qquad (4)$$

When the absolute value of the exponent in (4) – i.e., $\mathrm{ID}{\cdot}\Delta t$ – is small (e.g., $< .10$), the above expression for $R_{\Delta t}$ is approximately equal to ID·t. When Δt is equal to one year, we have $\hat{R}_1 \triangleq \mathrm{ID}$. That is, for rare diseases, the estimated one-year risk is approximately equal to the incidence density (as measured in years^{-1}). It can also be shown from

expression (4) that regardless of the rarity of the disease, the ID is the theoretical limit of the risk for period Δt divided by Δt, as Δt approaches zero.

As previously mentioned, the assumption of a constant incidence rate requires the calculation of ID's for relatively small categories of age – preferably 5-year strata or less, if enough subjects are available for precise estimates. Consequently, Δt must be divided into successive sub-periods (age) of width Δt_j, each with a constant incidence rate (ID_j). The total (or cumulative) Δt-year risk of developing the disease can be expressed as an extension of equation (4) according to the following formula (2):

$$\hat{R}_{\Delta t} = 1 - \exp\left[-\sum_j ID_j \cdot \Delta t_j\right] = 1 - \prod_j (1 - \hat{R}_{\Delta t_j}) \quad (5)$$

where: ID_j = incidence density for the j-th category; Δt_j = the width of the j-th category – i.e., the hypothetical follow-up period; $\hat{R}_{\Delta t_j}$ = the estimated risk for the full period of the j-th age category – i.e., the Δt_j-year risk of developing the disease for a person entering the j-th category, calculated according to equation (4); and Π = the product sign (analogous to Σ). Note that the age intervals used to compute ID's become the hypothetical follow-up periods for calculating risks – e.g., the 20-year risk of developing breast cancer between ages 35 and 55. Therefore, the Δt's in equation (5) are limited by the age range of subjects, not by the actual durations of follow-up which are incorporated into the calculations of ID_j. Combining the risk estimates from several age categories into an overall estimate (i.e., $\hat{R}_{\Delta t}$) assumes the lack of a 'cohort effect'. In other words, it is assumed that age-specific rates remain fixed in the study population over time.

To illustrate the estimation of risk from the follow-up of a dynamic population (assumed stable), consider a population of 250 000 women between the ages of 35 and 55 as counted in a local census. Suppose that for a 3-year period, all newly diagnosed cases of breast cancer are reported to the investigator, as summarized in the first 2 columns of Table 1. For each 5-year age category, we can compute the number of person-years of follow-up and the corresponding age-specific ID's. The estimates in Table 1 assume that the prevalence 'rate' of breast cancer in the population is approximately zero – i.e., the total population (N) is assumed to be disease-free. Using equation (4), the 5-year risks for all age categories have been estimated and are also listed in Table 1. These estimates may be combined into an overall estimate of the 20-year risk for a 35 year-old woman, using equation (5). On the basis of the latter result, we may conclude that a 35 year-old woman in the study population has nearly a 2% chance of developing breast cancer in the next 20 years if she does not first die of another cause during that period.

DISCUSSION

The previous sections have demonstrated that the concepts of risk and rate are fundamentally different though they are functionally related. We are therefore left with the 2 related tasks of: (1) specifying which quantity should be estimated in a particular study; and (2) determining whether cumulative incidences (CI) or incidence densities (ID) should be calculated. These questions necessarily involve a number of both theoretical and practical issues. From a theoretical standpoint, the principal basis for selecting the type of incidence

TABLE 1 *Summary of observed data, and computed estimates of average incidence rates (ID) and risks ($\hat{R}$) for the detection of breast cancer in a hypothetical dynamic population (of size N) followed for 3 years, by age category*

	Observed			Computed	
age	N	# new cases	person-years	ID[a] (1/yr.)	$\hat{R}$[b]
35–39	60 000	90	180 000	.00050	.00250
40–44	70 000	168	210 000	.00080	.00399
45–49	65 000	215	195 000	.00110	.00550
50–54	55 000	227	165 000	.00138	.00686
35–54	250 000	700	750 000	–	.01871[c]

a. The incidence density according to equation (2).
b. The estimate of the 5-year risk for a woman at the beginning of each age category, according to equation (4).
c. The estimate of the 20-year risk for a 35 year-old woman, according to equation (5).

measure rests on the objective of the study. For the purpose of making clinical and/or personal decisions about an individual's prognosis, treatment, or the desirability of certain behaviours (e.g., smoking), we would like to know the *risk* of developing a certain condition or health outcome. For example, whether a physician recommends to a patient with a recent myocardial infarction to quit smoking depends on the best assessment of the expected reduction in the probability of experiencing a future coronary event. In contrast to this objective, it has been suggested (2) that the incidence rate is a more fundamental measure for assessing the impact of a disease on a given population and for comparing this impact among 2 or more subpopulations with different distributions of certain characteristics and exposures. Thus, rates may be preferred theoretically when testing aetiologic hypotheses in particular populations. This idea can be appreciated by noting that risk is a function of both the incidence rate during the follow-up period and the length of that period (equation (5)). Therefore, for any fixed rate, risk depends entirely on Δt. As Δt approaches zero, the risk approaches zero; and as Δt approaches infinity the risk approaches one. That is, regardless of the force of morbidity, if we could follow a group of subjects for an infinite period of time, everyone would eventually develop the disease of interest since no one would first die of any other causes (by assumption). Consequently, for relatively long follow-up periods (depending on the incidence rate), the risk (R) becomes less sensitive to differences in the incidence rate. Suppose we wish to compare two populations – an exposed group and an unexposed group – in order to determine whether the exposure is a determinant ('risk factor') of the disease. It follows from the above argument that for a constant ratio in incidence rates comparing exposed and unexposed groups, the risk ratio (RR) moves closer to one (the null value) as the follow-up period gets longer.

Another theoretical issue that influences our choice of incidence measure is the nature of the disease or condition being studied. Three aspects of the disease are relevant in this regard: the rarity of the disease, the length of the latency period, and the period for which a person is a candidate for an occurrence (i.e., 'at risk'). As suggested above, the rarity of the disease is important to the extent that the divergence of the risk ratio from the rate ratio as Δt increases occurs more rapidly for a more frequent disease. Of greater importance, however, is the duration of disease development from first 'exposure' to clinical manifestation – i.e., the latency period. For most chronic diseases with long latency periods, purportedly a matter of years, the actual follow-up period is a small part of the total time for which the population is 'at risk'. Assuming the age-specific incidence rates remain relatively constant over the follow-up period (or portions of it) – i.e., a lack of secular trends, it is of interest to assess how the (average) rate varies by age, using an ID analysis. On the other hand, when the latency (or incubation) period is short, the entire risk period is generally contained within the actual follow-up period. For example, the investigation of a localized outbreak of an infectious disease would usually be extended to include all primary cases. While the incidence rate would vary considerably over the course of the epidemic, we generally would be interested in the proportions of persons in defined groups that develop the disease – i.e., risks, as commonly measured by 'attack rates'.

Thirdly, as suggested in a previous illustration, the desired incidence measure depends on whether a case may subsequently be a candidate for future occurrences of the same disease. In the extreme case where the diseased person is rapidly cured without conferring future immunity, the object of the investigation may be to study all occurrences. For example, we might wish to compare the total incidence rate of the common cold for a group which is taking regular supplements of Vitamin C to the rate for a group that is given a placebo. Since each subject may develop multiple occurrences of the condition (assuming a long enough follow-up period), the calculation of cumulative incidences (CI) could not be interpreted as estimates of risk. In fact, such estimates would not be proportions since each person in the denominator could be represented more than once in the numerator. Alternatively, the calculation of ID's would permit an appropriate test of the null hypothesis and an indirect estimate of risk, using equation (5).

Given the specification of the desired incidence parameter, the actual choice of calculating CI's or ID's in a given analysis depends largely on a number of practical issues. One such issue rests on the ability to identify the exact time of disease occurrence or diagnosis when following a fixed cohort. Specifically, the time of clinical manifestation may be ambiguous or the cost of determining the time of the event may be prohibitive. For example, a 10-year follow-up study for detection of new hypertensives would require frequent examinations (say, every year) to make an approximate assess-

ment of disease onset. Without this information, we would not be able to determine how many person-years of follow-up is contributed by each subject, especially 'cases'. In this type of situation, we may be forced to compute CI's. On the other hand, if the amount of PT is known for each subject but varies widely among the noncases (such as in a clinical trial), the direct calculation of the CI's would be meaningless.

A potential hazard of an ID analysis for the unknowing investigator is that the durations of follow-up for the study population are obscured in the computation of PT. For example, 10 persons followed for one year, subsequent to the initiation of a given exposure, amount to the same number of person-years as one person followed for 10 years. Moreover, it may be that one year is not enough time for the disease at issue to develop while 10 years is sufficient for most exposure-related cases to be detected. However, if the study is well designed, this potential difficulty can be handled by stratifying the population into categories of different follow-up periods before calculating ID's.

Very often, especially in areas with relatively complete tumour registries, it is possible to identify nearly all incident cases of a given disease (e.g., a highly fatal cancer) that developed in a well-defined population within a certain period. Rather than fabricating a 'population-at risk' for computation of CI's, it may be preferable to calculate age-specific ID's, assuming a stable population. Furthermore, we might test an aetiologic hypothesis by comparing the set of cases with a sample of noncases from the same population, with respect to exposure histories (2). Essentially, this design combines elements of both the follow-up and case-control studies.

Although the objectives of a study may be to estimate risk, the direct calculation of CI may be ambiguous or even uninterpretable. For example, suppose we follow a fixed cohort of middle-age women for 5 years and wish to assess the putative effect of obesity on mortality between the ages of 40 and 60. Any attempt to estimate risks for obese and non-obese women by calculating CI's will be hopelessly abortive. (What are the denominators?) Yet, the calculation of age-specific ID's is a straightforward procedure if we know the time of death; and these estimates of mortality rates can be used to estimate risks.

Typically, in the test of an aetiologic hypothesis, it is necessary to control in the analysis for a number of factors that may confound the effect of the exposure of interest. Most common procedures for fitting statistical models (e.g., logistic regression) to the data require knowing whether each subject is a case or noncase. Of course, this type of data is not generally incorporated into an ID analysis since a comparison of ID's suppresses individual information on disease status. In recent years, however, modifications of the life table approach using regression to control for covariates have been developed (9). Nevertheless, a more sophisticated understanding of statistics is necessary to utilize these methods effectively.

Ultimately, the choice of using CI's or ID's, given the conflicting set of issues listed above, rests on how much difference it makes in practice. This is especially important when evaluating previous studies that are related to our research interest. As already suggested in this paper, there will be an increasing difference between the average rate ratio or incidence density ratio (IDR) and the risk ratio (RR) as: (1) the length of the follow-up period (Δt) increases; (2) the background rate among the unexposed group (ID_0) increases; and (3) the strength of the association between exposure and disease (IDR) increases. However, for rare diseases – e.g., $ID_0 \leqslant .01$ yrs.$^{-1}$ (approximately the rate of CHD among 60 year-old white males) – and follow-up periods less than 10 years, the difference between the 2 ratio measures remains small, especially for modest effects – i.e., IDR $\leqslant 3$. For more frequent diseases, long follow-up periods would not be required or desired, making the difference between the 2 measures of little practical importance. We may conclude, therefore, that in follow-up studies of most chronic diseases, calculation of either CI's or ID's would produce approximately the same estimates of effect. As an example, consider the breast cancer calculations in Table 1. For women between 50 and 54, a direct estimate of the 5-year risk is (227/55 000) x (5/3) = .00688, which is nearly identical to the estimate resulting from the proposed incidence density analysis.

In case-control studies, the odds ratio (OR) is typically used to assess the strength of association between exposure and disease. It was demonstrated many years ago that the OR is an accurate estimate of the prevalence ratio (PR) whenever the prevalence is sufficiently small (10). Consequently, with the assumption of no 'bias' due to selection factors, measurement, or confounding, most epidemiologists regard the OR as an estimate of the 'relative risk' (RR) for 'rare' diseases (11). It should also be recognized, however, that the OR is an estimate of the IDR *without* the rare disease assumption (2). Since the IDR is the theoretical limit

of the RR as Δt approaches zero (2), the OR will also be approximately equal to the RR for short follow-up periods.

In summary, the appropriate choice of incidence measure in a follow-up study depends on several conceptual and practical issues. The investigator must make the selection to satisfy the particular objectives and constraints of each study and analysis.

ACKNOWLEDGEMENT

The authors are grateful to Dr Kenneth Rothman for the visual analogy that led to Figure 2 of this paper.

REFERENCES

(1) Elandt-Johnson R C. Definition of rates: some remarks on their use and misuse. *American Journal of Epidemiology* **102**: 267–271, 1975.

(2) Miettinen O. Estimability and estimation in case-referent studies. *American Journal of Epidemiology* **103**: 226–235, 1976.

(3) Fleiss J L, Dunner D L, Stallone F and Fieve R R. The life table: a method for analyzing longitudinal studies. *Archives of General Psychiatry* **33**: 107–112, 1976.

(4) Whittemore A S. The age distribution of human cancer for carcinogenic exposures of varying intensity. *American Journal of Epidemiology* **106**: 418–432, 1977.

(5) Moolgavkar S H. The multistage theory of carcinogenesis and the age distribution of cancer in man. *Journal of the National Cancer Institute* **61**: 49–52, 1978.

(6) Cutler S J and Ederer F. Maximum utilization of the life table method in analyzing survival. *Journal of Chronic Diseases* **8**: 699–712, 1958.

(7) Sheps M C. On the person years concept in epidemiology and demography. *Milbank Memorial Fund Quarterly* **44**: 69–91, 1966.

(8) Gehan E A. Estimating survival functions from the life table. *Journal of Chronic Diseases* **21**: 629–644, 1969.

(9) Cox D R. Regression models and life-tables. *Journal of the Royal Statistical Society* (Series B) **34**: 187–202, 1972.

(10) Cornfield J. A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast, and cervix. *Journal of the National Cancer Institute* **11**: 1269–1275, 1951.

(11) MacMahon B and Pugh T F. *Epidemiology. Principles and Methods*. Boston, Little, Brown and Company, 1970 Pp 269–273.

(Revised version received 21 September 1979)