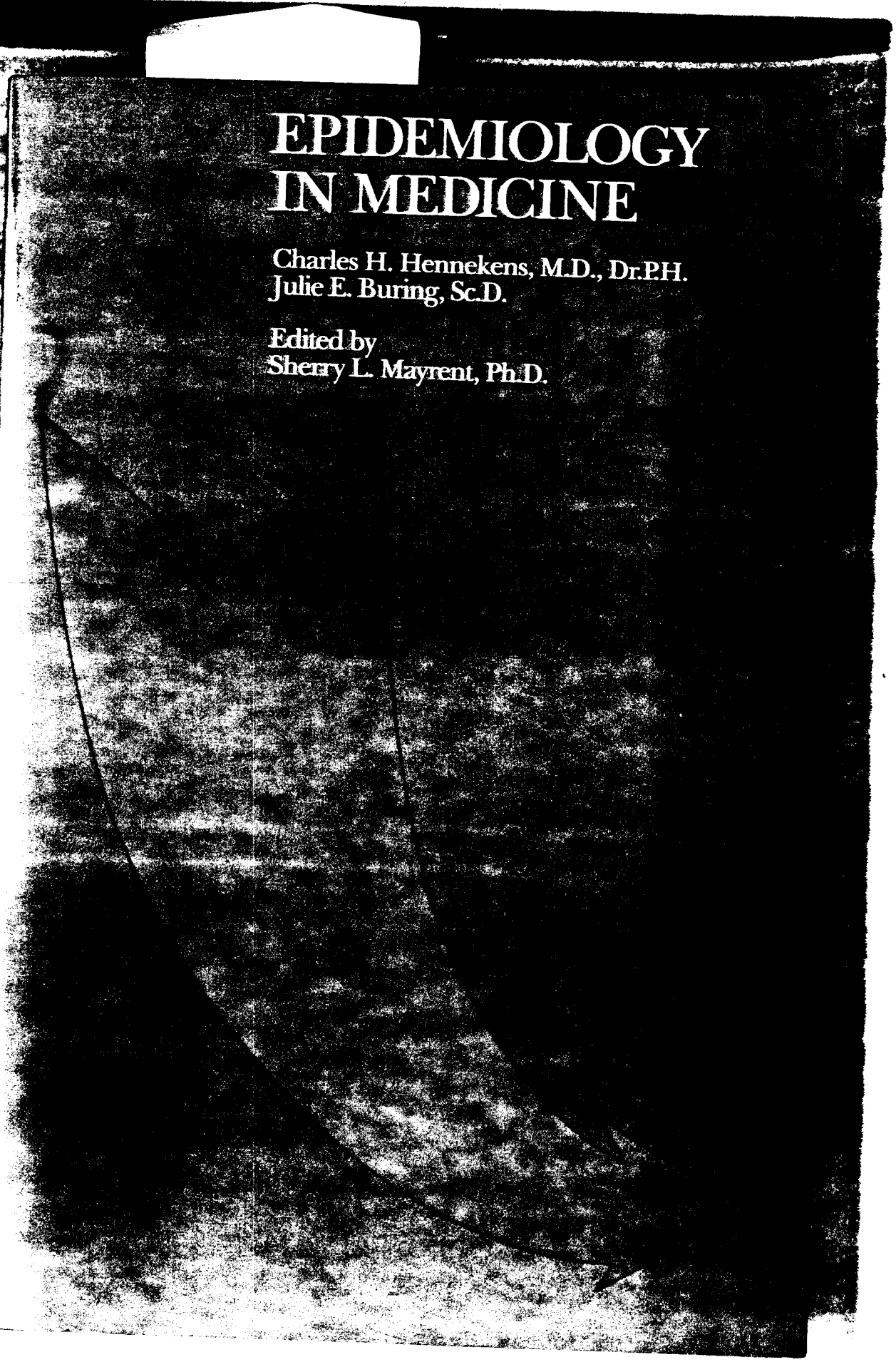
EPIDEMIOLOGY
IN MEDICINE
Charles H. Hennekens, M.D., Dr.P.H.
Julie E. Buring, Sc.D.
Edited by
Sherry L. Mayrent, Ph.D.

EPIDEMIOLOGY IN MEDICINE

CHARLES H. HENNEKENS, M.D., Dr.P.H.
Professor of Medicine and Preventive Medicine, Department of Medicine and Preventive Medicine, Harvard Medical School, Brigham and Women's Hospital; Adjunct Professor of Epidemiology and Biostatistics, Boston University School of Public Health, Boston

JULIE E. BURING, Sc.D.
Associate Professor, Department of Preventive Medicine, Harvard Medical School, Brigham and Women's Hospital; Adjunct Associate Professor of Epidemiology and Biostatistics, Boston University School of Public Health, Boston

EDITED BY
SHERRY L. MAYRENT, Ph.D.
Associate in Medicine (Epidemiology), Harvard Medical School, Brigham and Women's Hospital, Boston

FOREWORD BY
SIR RICHARD DOLL, F.R.S.
Regis Professor Emeritus of Medicine, Oxford University, Oxford, England

LITTLE, BROWN AND COMPANY
BOSTON/TORONTO

First Edition
Tenth Printing

Library of Congress Catalog Card No. 87-45377

ISBN 0-316-35636-0

Printed in the United States of America

RRD VA

TO THE ME

Attributable Risk

As has been discussed, the relative risk represents the likelihood of disease in exposed individuals relative to those who are nonexposed. The risk difference (RD) or attributable risk (AR) is a measure of association that provides information about the absolute effect of the exposure or the excess risk of disease in those exposed compared with those nonexposed. This measure is defined as the difference between the incidence rates in the exposed and nonexposed groups and can be calculated as follows:

$$AR = I_e - I_0$$

In a cohort study, the attributable risk is calculated as the difference of cumulative incidences (risk difference) or incidence densities (rate difference) depending on the study design. For example, in the study of OC use and bacteriuria (Table 4-10), the attributable risk would be calculated as follows:

$$AR = CI_e - CI_0 = \frac{a}{(a+b)} - \frac{c}{(c+d)}$$

$$= \frac{27}{482} - \frac{77}{1908}$$

$$= 0.01566 = 1566/10^5$$

Thus, the excess occurrence of bacteriuria among OC users attributable to their OC use is 1566 per 100,000. The attributable risk is used to quantify the risk of disease in the exposed group that can be considered attributable to the exposure by removing the risk of disease that would have occurred anyway due to other causes (the risk in the nonexposed). Thus, the interpretation of the attributable risk is dependent on the assumption that a cause-effect relationship exists between exposure and disease. If there is no association between exposure and disease, there will be no difference between the incidence rates in the exposed and nonexposed groups, so that AR = 0. If, however, there is a causal association between the exposure and disease and the attributable risk is greater than 0, its value indicates the number of cases of the disease among the exposed that can be attributed to the exposure itself, or alternatively, the number of cases of the disease among the exposed that could be eliminated if the exposure were eliminated. Thus, the attributable risk can be useful as a measure of the public health impact of a particular exposure.

To estimate the proportion of the disease among the exposed that is

attributable to the exposure, or the proportion of the disease in that group that could be prevented by eliminating the exposure, the attributable risk is often expressed as a percentage (AR%). The attributable-risk percent, also referred to as the attributable-rate percent, attributable proportion, or etiologic fraction, is calculated as the attributable risk divided by the rate of disease among the exposed or:

$$AR\% = \frac{AR}{I_e} \times 100$$

$$= \frac{(I_e - I_0)}{I_e} \times 100$$

In the cohort study of OC use and bacteriuria, the attributable-risk percent would be calculated as follows:

$$AR\% = \frac{(AR)}{I_e} \times 100$$

$$= \frac{1566/10^5}{27/482} \times 100$$

$$= 27.96\%$$

Thus, if OC use does cause bacteriuria, about 28 percent of bacteriuria among women who use OCs can be attributed to their OC use and could therefore be eliminated if they did not use OCs.

For most case-control studies, the attributable risk cannot be calculated because the incidence rates of disease among the exposed and nonexposed groups are not available. It is, however, possible to calculate the attributable-risk percent using the following formula [4]:

$$AR\% = \frac{(RR - 1)}{RR} \times 100$$

From the data on OC use and MI presented in Table 4-9, the relative risk of MI associated with current OC use was 1.6, yielding an attributable-risk percent of:

$$AR\% = \frac{(1.6 - 1)}{1.6} \times 100$$

$$= 37.5\%$$

This study therefore suggests that if OC use causes MI, nearly 38 percent of MIs among young women who used OCs could be attributed to that exposure or could be eliminated if they were to stop using OCs.

In case-control st
lation of interest is
the distribution of
resentative of the w
estimate incidence
the overall incidenc
of as a weighted ave
egories, with the w
each category, I_T ca
posed group (I_e) tim
lation who have the
nonexposed (I_0) tim
is expressed mathen

$$I_T = (I_e)(P_e) + (I_0)(P_0$$

Since the relative ris
posed and nonexpos
bers of a population
rate in the nonexpo
risk can be estimate
for I_e in the formula

$$I_T = (I_0)(OR)(P_e) +$$

$$= I_0[(OR)(P_e) + (1$$

To determine the inc
ply solved for I_0, as f

$$I_0 = \frac{I_T}{(OR)(P_e) + P_0}$$

Once the incidence r
multiplied by the o
among the exposed.
tributable risk can th

For example, in D
lung cancer, 1350 of
without lung cancer
years, yielding a rel
incidence of 480/mil
proportion of smok
dence rates of lung
500 per million per y

In case-control studies in which the incidence rate in the total population of interest is known or can be estimated from other sources and the distribution of exposure among the controls is assumed to be representative of the whole population, these parameters can be used to estimate incidence rates in the exposed and nonexposed groups. Since the overall incidence rate of disease in a population (I_T) may be thought of as a weighted average of the incidence rates in various exposure categories, with the weights related to the proportions of individuals in each category, I_T can be calculated as the incidence rate among the exposed group (I_e) times the proportion of individuals in the total population who have the exposure (P_e), plus the incidence rate among the nonexposed (I_0) times the proportion of nonexposed persons (P_0). This is expressed mathematically as follows:

$$I_T = (I_e)\,(P_e) + (I_0)\,(P_0)$$

Since the relative risk is the ratio of the incidence rates among the exposed and nonexposed, the incidence rate among the exposed (I_e) members of a population is equal to the relative risk times the comparable rate in the nonexposed ($RR \times I_0$). In a case-control study, the relative risk can be estimated by the odds ratio (OR), and thus we can substitute for I_e in the formula:

$$\begin{aligned} I_T &= (I_0)\,(OR)\,(P_e) + (I_0)\,(P_0) \\ &= I_0\,[(OR)\,(P_e) + (P_0)] \end{aligned}$$

To determine the incidence rate in the nonexposed, this equation is simply solved for I_0, as follows:

$$I_0 = \frac{I_T}{(OR)\,(P_e) + P_0}$$

Once the incidence rate among the nonexposed is determined, it can be multiplied by the odds ratio to provide an estimate of the incidence among the exposed. Given these two incidence rates (I_e and I_0), the attributable risk can then be calculated.

For example, in Doll and Hill's case-control study [9] of smoking and lung cancer, 1350 of 1357 men with the disease and 1296 of 1357 men without lung cancer had smoked cigarettes regularly for the previous 10 years, yielding a relative risk of 9.1. Using an estimate of lung cancer incidence of 480/million/year derived from other sources [7] and the proportion of smokers and nonsmokers among the controls, the incidence rates of lung cancer can be estimated, as shown in Table 4-17, as 500 per million per year among the exposed and 55 per million per year

Table 4-17. Case-control study of cigarette smoking and lung cancer, with calculation of the relative risk, estimates of the incidence rates in the exposed and nonexposed, and the attributable risk

	Lung cancer		
	Cases	Controls	Total
Cigarette smoking			
Yes	1350	1296	2646
No	7	61	68
Total	1357	1357	2714

$$RR = \frac{(1350)(61)}{(1296)(7)} = 9.1$$

$$I_0 = \frac{I_T}{(RR)(P_e) + (P_0)} = \frac{480/10^6}{(9.1)(0.955) + (0.045)} = 55/10^6$$

$$I_e = (RR)(I_0) = (9.1)(55/10^6) = 500/10^6/\text{year}$$

$$AR = I_e - I_0 = 500 - 55 = 445/10^6/\text{year}$$

Data from R. Doll and A. B. Hill, A study of the aetiology of carcinoma of the lung. *Br. Med. J.* 2:1271, 1952; and R. Doll, Bronchial carcinoma: Incidence and aetiology. *Br. Med. J.* 2:521, 1953.

among the nonexposed. Thus, the excess rate of lung cancer among smokers that is attributable to smoking is

$$AR = 500/10^6/\text{year} - 55/10^6/\text{year}$$
$$= 445/10^6/\text{year}$$

If the exposure is preventive, so that I_e is less than I_0, the attributable risk is meaningless. However, an analogous measure, the preventive fraction (PF), can be defined [24]

$$PF = \frac{I_0 - I_e}{I_0}$$

Population Attributable Risk

While it is useful to estimate the proportion of cases for whom the disease is attributable to their exposure, it is also of interest to estimate the excess rate of disease in the total study population of exposed and nonexposed individuals that is attributable to the exposure. This measure, referred to as the population attributable rate or risk (PAR), helps determine which exposures have the most relevance to the health of a com-

munity [20]. Th
disease in the p
(I_0):

$$PAR = I_T - I_0$$

Alternatively, th
utable risk by th
(P_e):

$$PAR = (AR)(P_e)$$

The population
(Table 4-10) can

$$PAR = I_T - I_0$$
$$= 104/2390$$
$$= 316/10^5/\text{ye}$$

or alternatively a

$$PAR = (AR)(P_e)$$
$$= 1566/10^5 \times$$
$$= 316/10^5/\text{yea}$$

Thus, if OC use v
teriuria that could
100,000. The attri
than the populati
exposure on the r
those with the exp
nation of exposed
lation attributable
either the prevalen
be assumed to be a
sure in the total po
an outside source.
bitrary distribution
exposed and 200 r
risk computed for
estimate of the tru
source outside the s
be calculated.

munity [20]. The population attributable risk is calculated as the rate of disease in the population (I_T) minus the rate in the unexposed group (I_0):

$$PAR = I_T - I_0$$

Alternatively, this measure can be calculated by multiplying the attributable risk by the proportion of exposed individuals in the population (P_e):

$$PAR = (AR)\,(P_e)$$

The population attributable risk of bacteriuria associated with OC use (Table 4-10) can therefore be calculated as:

$$\begin{aligned} PAR &= I_T - I_0 \\ &= 104/2390 - 77/1908 \\ &= 316/10^5/\text{year} \end{aligned}$$

or alternatively as:

$$\begin{aligned} PAR &= (AR)\,(P_e) \\ &= 1566/10^5 \times (482/2390) \\ &= 316/10^5/\text{year} \end{aligned}$$

Thus, if OC use were stopped, the excess annual incidence rate of bacteriuria that could be eliminated among women in this study is 316 per 100,000. The attributable risk among the exposed will always be greater than the population attributable risk since the impact of removing the exposure on the number of cases of disease will always be greater for those with the exposure than for a total population, which is a combination of exposed and nonexposed individuals. To calculate the population attributable risk for a group broader than the study population, either the prevalence of exposure observed in the study population must be assumed to be an adequate reflection of the prevalence of the exposure in the total population, or another estimate must be available from an outside source. If the cohort for study has been chosen with an arbitrary distribution of exposed and nonexposed individuals (such as 200 exposed and 200 nonexposed subjects), then a population attributable risk computed for the study population would be meaningless. If an estimate of the true prevalence of the exposure were available from a source outside the study, however, the population attributable risk could be calculated.

Analogous to the attributable-risk percent among exposed individuals, the population attributable-risk percent (PAR%) expresses the proportion of disease in the study population that is attributable to the exposure and thus could be eliminated if the exposure were eliminated. The population attributable-risk percent is calculated by dividing the population attributable-risk by the rate of the disease in the population or

$$\text{PAR\%} = \frac{\text{PAR}}{I_r} \times 100$$

For example, in the cohort study of OC use and bacteriuria (Table 4-10), the incidence rate of bacteriuria in the total study population was 104 per 2390 or 4351.5 per 100,000, and the population attributable risk was 316 per 100,000. The population attributable-risk percent can then be calculated as follows:

$$\text{PAR\%} = \frac{316}{4351.5} \times 100$$
$$= 7.3\%$$

Thus, if OC use causes bacteriuria, about 7 percent of all the bacteriuria in the study population (and 28 percent of bacteriuria among women taking OCs) could be prevented if OC use were eliminated.

In a case-control study, the population attributable-risk percent can be calculated if the proportion of exposed in the control group can be used as an estimate of the proportion exposed in the population (P_e), or if the prevalence of exposure in the population is available from another source. The formula for the population attributable-risk percent can be expressed as:

$$\text{PAR\%} = \frac{(P_e)\,(RR - 1)}{(P_e)\,(RR - 1) + 1} \times 100$$

If the proportion of exposed in the control group can be used as an estimate of P_e, the population attributable-risk percent can be calculated by the equivalent formula [24]:

$$\text{PAR\%} = \text{AR\%} \times (\text{proportion of exposed cases})$$

For example, in the case-control study of OC use and MI presented in Table 4-9, the prevalence of OC use in the population can be estimated as the prevalence of OC use in the controls ($P_e = 304/3120 = 0.0974$). The population attributable-risk percent can be calculated as follows:

$$\text{PAR\%} = \frac{P_e\,(RR$$
$$= \frac{(0.097$$
$$= 5.5\%$$

or

$$\text{PAR\%} = \text{AR\%} \times$$
$$= \frac{1.6 - 1}{1.6}$$
$$= 5.5\%$$

This implies that
menopausal wome
could be preventec

As discussed ear
disease in the gen
among the contro
the attributable ris
is also possible to c
in the case-control
4-17, the populatic

$$\text{PAR} = (\text{AR})\,(P_e)$$
$$= (445/10^6)\,(0.$$
$$= 425/10^6$$

Thus, if cigarette s
are directly attribu

Interpretation of

It is important to
very different type
strength of the ass
vides information
association is likel
vides a measure o
that the associatio

$$\text{PAR\%} = \frac{P_e(RR - 1)}{P_e(RR - 1) + 1} \times 100$$

$$= \frac{(0.0974)(1.6 - 1)}{(0.0974)(1.6 - 1) + 1}$$

$$= 5.5\%$$

or

$$\text{PAR\%} = \text{AR\%} \times \frac{a}{a + c}$$

$$= \frac{1.6 - 1}{1.6} \times \frac{23}{156}$$

$$= 5.5\%$$

This implies that if OC use causes MI, 5.5 percent of MIs among premenopausal women in the study population is attributable to OC use or could be prevented if OC use were discontinued.

As discussed earlier, in case-control studies when the incidence rate of disease in the general population is known and the rate of exposure among the controls is assumed to be representative of the population, the attributable risk can be calculated directly. In such circumstances, it is also possible to calculate the population attributable risk. For example, in the case-control study of smoking and lung cancer presented in Table 4-17, the population attributable risk can be derived as follows:

$$\text{PAR} = (\text{AR})(P_e)$$

$$= (445/10^6)(0.955)$$

$$= 425/10^6$$

Thus, if cigarette smoking causes lung cancer, 425 cases per million men are directly attributable to this habit.

Interpretation of Measures of Association

It is important to remember that relative and attributable risks provide very different types of information. The relative risk is a measure of the strength of the association between an exposure and disease and provides information that can be used to judge whether a valid observed association is likely to be causal. In contrast, the attributable risk provides a measure of the public health impact of an exposure, assuming that the association is one of cause and effect. The magnitude of the

Table 4-18. Relative and attributable risks of mortality from lung cancer and coronary heart disease among cigarette smokers in a cohort study of British male physicians

	Annual mortality rate per 100,000	
	Lung cancer	Coronary heart disease
Cigarette smokers	140	669
Nonsmokers	10	413
Relative risk	14.0	1.6
Attributable risk	$130/10^5$/year	$256/10^5$/year

R. Doll and R. Peto, Mortality in relation to smoking: Twenty years' observations on male British doctors. *Br. Med. J.* 2:1525, 1976.

Table 4-19. Measures of disease frequency

Prevalence (P)	$\frac{\text{Number of persons with disease}}{\text{Number of persons in population}}$ at a point in time
Cumulative incidence (CI)	$\frac{\text{Number of new cases of disease in a given time period}}{\text{Total population at risk}}$
Incidence density (ID)	$\frac{\text{Number of new cases of disease in a given time period}}{\text{Total person-time of observation}}$

relative risk alone does not predict the magnitude of the attributable risk. This can be illustrated by examining the relationship of cigarette smoking with mortality from lung cancer and coronary heart disease in a cohort study of British male physicians [9].

As shown in Table 4-18, this investigation demonstrated a 14-fold increased death rate from lung cancer among smokers of at least one pack of cigarettes daily when compared with nonsmokers. On the other hand, the relative risk of coronary heart disease mortality among current cigarette smokers compared with nonsmokers was 1.6. Thus, cigarette smoking is a much stronger risk factor for mortality from lung cancer than for coronary heart disease. However, if smoking is causally related to both diseases, the elimination of cigarettes would prevent far more deaths among smokers from coronary heart disease than from lung cancer, as shown by the attributable risks of 256 per 100,000 and 130 per 100,000 per year, respectively. The explanation for this is that while death from lung cancer is a relatively rare occurrence, accounting for only 10 deaths per 100,000 population each year among nonsmokers, the annual death rate for coronary heart disease in that same group is 413 per 100,000. Consequently, even a 60-percent increased risk of cor-

Table 4-20. Measures of association: cohort studies

Relative risk (RR)	1. Cumulative incidence (risk) ratio = $\frac{\text{Cumulative incidence in exposed } (CI_e)}{\text{Cumulative incidence in nonexposed } (CI_0)}$
	2. Incidence density (rate) ratio = $\frac{\text{Incidence rate among exposed } (I_e)}{\text{Incidence rate among nonexposed } (I_0)}$
Attributable risk (AR)	$I_e - I_0$
Attributable-risk percent (AR%)	$\frac{AR}{I_e} \times 100$
Population attributable risk (PAR)	$I_T - I_0$ or AR × prevalence of exposure (P_e)
Population attributable-risk percent (PAR%)	$\frac{PAR}{\text{Incidence rate of disease in population } (I_T)} \times 100$

onary heart disease mortality associated with cigarette smoking will affect a much larger number of people than a 14-fold increased risk of death from lung cancer. Thus, the potential public health impact of smoking cessation on mortality will be far greater for coronary heart disease than for lung cancer.

In general, the relative risk is the measure used most commonly by those evaluating possible determinants of disease because it represents the magnitude of the association and provides information that can be used in making a judgment of causality. In contrast, once causality is assumed, from the perspective of public health administration and policy, measures of association based on absolute differences in risk between exposed and nonexposed individuals assume far greater importance. These absolute rates express either the actual incidence of a disease that is attributable to an exposure among the exposed (attributable risk) or the number of cases of disease in the total population that could be eliminated by removal of a harmful exposure (population attributable risk).

CONCLUSION

In this chapter we have considered the most commonly used measures of disease frequency (Table 4-19) and association (Tables 4-20 and 4-21)

24. Miettinen, O. S. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am. J. Epidemiol.* 99:325, 1974.
25. Monson, R. R. *Occupational Epidemiology.* Boca Raton, FL: CRC, 1980.
26. Najarian, T., and Colton, T. Mortality from leukemia and cancer in nuclear shipyard workers. *Lancet* 1:42, 1978.
27. Rogentine, G. N., Jr., Yankee, R. A., Gart, J. J., et al. HL-A antigens and disease. *J. Clin. Invest.* 51:2430, 1972.
28. Rogentine, G. N., Trapani, R. J., Yankee, R. A., et al. HL-A antigens and acute lymphocytic leukemia: The nature of the HL-A2 association. *Tissue Antigens* 3:470, 1973.
29. Rogers, M. F., Schonberger, L. B., Hurwitz, E. S., et al. Reye Syndrome surveillance, 1981–1982. *M.M.W.R.* 32:17, 1983.
30. Rosenberg, L., Hennekens, C. H., Rosner, B., et al. Oral contraceptive use in relation to non-fatal myocardial infarction. *Am. J. Epidemiol.* 111:59, 1980.
31. Sattin, R. W., Rubin, G. L., and Hughes, J. M. Hysterectomy among women of reproductive age, United States, update for 1979–80. *M.M.W.R.* 32:1, 1983.
32. Silverberg, E., and Lubera, J. Cancer statistics, 1986. *CA* 36:9, 1986.
33. Stampfer, M. J., Willett, W. C., Colditz, G. A., et al. A prospective study of postmenopausal hormones and coronary heart disease. *N. Engl. J. Med.* 313:1044, 1985.
34. U.S. Bureau of the Census. *Sixteenth Census of the U.S., 1940.* Characteristics of the Population, Part I, United States Summary. Washington, DC: 1942.
35. U.S. Bureau of the Census. *Statistical Abstract of the United States: 1984* (104th ed.). Washington, DC: U.S. Govt. Printing Office, 1983.
36. U.S. D.H.E.W. *Vital Statistics of the United States 1973.* Vol. II, Mortality. Rockville, MD: National Center for Health Statistics, 1975.
37. U.S. D.H.H.S. *Surveillance, Epidemiology, and End Results: Incidence and Mortality Data, 1973–77.* N.I.H. Publication No. 81-2330. Bethesda, MD: National Cancer Institute, 1981.
38. U.S. D.H.H.S. *Vital Statistics of the U.S., 1980.* Vol. II, Mortality, Part B. D.H.H.S. Publication No. (P.H.S.) 85-1102. Washington, DC: National Center for Health Statistics, 1985.