Epidemiology

Contributed by: Virginia L. Ernster

Publication year: 2014

The study of the distribution of diseases in populations and of factors that influence the occurrence of disease. Epidemiology examines epidemic (excess) and endemic (always present) diseases; it is based on the observation that most disease does not occur randomly, but is related to environmental and personal characteristics that vary by place, time, and subgroup of the population. The epidemiologist attempts to determine who is prone to a particular disease; where risk of the disease is highest; when the disease is most likely to occur and its trends over time; what exposure its victims have in common; how much the risk is increased through exposure; and how many cases of the disease could be avoided by eliminating the exposure.

In the course of history, the epidemiologic approach has helped to explain the transmission of communicable diseases, such as cholera and measles, by discovering what exposures or host factors were shared by individuals who became sick. Modern epidemiologists have contributed to an understanding of factors that influence the risk of chronic diseases, particularly cardiovascular diseases and cancer, which account for most deaths in developed countries today. Epidemiology has established the causal association of cigarette smoking with heart disease; shown that acquired immune deficiency syndrome (AIDS) is associated with certain sexual practices; linked menopausal estrogen use to increased risk of endometrial cancer but to decreased risk of osteoporosis; and demonstrated the value of mammography in reducing breast cancer mortality. By identifying personal characteristics and environmental exposures that increase the risk of disease, epidemiologists provide crucial input to risk assessments and contribute to the formulation of public health policy.

Epidemiologic studies, based mainly on human subjects, have the advantage of producing results relevant to people, but the disadvantage of not always allowing perfect control of study conditions. For ethical and practical reasons, many questions cannot be addressed by experimental studies in humans and for which observational studies (or experimental studies using laboratory animals or biomedical models) must suffice. Still, there are circumstances in which experimental studies on human subjects are appropriate, for example, when a new drug or surgical procedure appears promising and the potential benefits outweigh known or suspected risks.

Descriptive studies

Descriptive epidemiologic studies provide information about the occurrence of disease in a population or its subgroups and trends in the frequency of disease over time. Data sources include death certificates, special disease registries, surveys, and population censuses; the most common measures of disease occurrence are (1)

mortality (number of deaths yearly per 1000 of population at risk); (2) incidence (number of new cases yearly per 100,000 of population at risk); and (3) prevalence (number of existing cases at a given time per 100 of population at risk).

These measures can be calculated for all causes combined or for specific causes of disease and can pertain to an entire population or to specific subgroups (age or racial). When a rate for a specific subgroup is calculated, the number of people in the subgroup with the disease is the numerator and the population is the denominator. For example, to calculate the incidence rate for endometrial cancer in black women aged 50–54 in 1980, the numerator should include the new cases of endometrial cancer diagnosed in black women in that age group in 1980 and the denominator should give the total number of black women in that age group in the population that year.

The risk of most diseases is related to age. For example, heart disease mortality rates increase markedly with age. Thus, a country with a young population may have low overall rates of heart disease compared to another country whose inhabitants are, on average, much older; still, the risk of heart disease in each specific age group could be similar in the two countries. To compare the risk of disease across populations, differences in the age structure of those populations must be taken into account; similarly, to compare risk in the same population over time, the changing age composition of that population must be taken into account.

The effects of age can be controlled in several ways. For example, with age adjustment by the direct method, the age-specific rates of the populations of interest are applied to the age distribution of a standard population, and the resulting numbers of expected deaths for each population are compared. By the indirect method, the age-specific rates of a standard population are applied to the number of people in each age group in the study population, and the resulting numbers of deaths are compared to the number actually observed. Similar procedures adjust rates for other factors that influence risk of disease, for example, sex or race.

Descriptive measures are useful for identifying populations and subgroups at high and low risk of disease and for monitoring time trends for specific diseases. They provide the leads for analytic studies designed to investigate factors responsible for such disease profiles.

Analytic studies

Analytic epidemiologic studies seek to identify specific factors that increase or decrease the risk of disease and to quantify the associated risk.

Observational studies. In observational studies, the researcher does not alter the behavior or exposure of the study subjects, but observes them to learn whether those exposed to different factors differ in disease rates. Alternatively, the researcher attempts to learn what factors distinguish people who have developed a particular disease from those who have not.

The observational analytic studies most often performed are the cohort study and the case-control study. In the former (also called prospective study), a disease-free group of people, often characterized by a common feature such as occupation, is identified; data on exposures of interest are collected; and the people are observed over time to see if they develop the disease. Occurrence can be compared between the entire cohort (factory workers exposed to certain chemicals) and an external group (the general population) or between subgroups in the study cohort who differ by exposure (workers in the same factory, some exposed and others not exposed to certain chemicals). The relative risk of disease associated with exposure is determined by dividing rates of disease in the exposed group by rates in the nonexposed comparison group. Cohort studies are ideal in the respect that data gathered on exposure predate development of disease. However, such studies require large numbers of people and long follow-up periods to ensure reliable estimates of disease occurrence. Cohort studies are typically expensive and administratively complex.

In case-control studies, people who have already developed the disease of interest (cases) and people who are free of the disease (controls) are studied. Data are collected from individuals in both groups on personal characteristics or previous exposure to factors that are suspected to be determinants of the disease, and the two groups are compared. Although the relative risk of disease associated with exposure cannot be calculated directly in case-control studies, it can be closely estimated by a quantity known as the odds ratio, that is, the odds of exposure in the cases divided by the odds of exposure in the controls. Because it is not necessary to observe study subjects over time for the occurrence of disease, case-control studies take less time than cohort studies and require fewer subjects to provide statistically meaningful results; they are therefore generally less expensive to conduct. However, it is not always possible to be certain that the factor suspected of causing the disease preceded the development of the disease; the time sequence is important for establishing causality. Furthermore, the accuracy of exposure information often depends on the recall of past events (dietary practices in early life, dosages of past medications, or occupational exposures to particular chemicals); and this recall may be poor.

Experimental studies. In experimental studies, the investigator alters the behavior, exposure, or treatment of people to determine the impact of the intervention on the disease. Usually two groups are studied, one that experiences the intervention (the experimental group) and one that does not (the control group). Ideally, subjects are randomly assigned to the groups and, where possible, those responsible for assessing outcomes are unaware of (blind to) the assignments. Intervention studies include risk-reduction trials of attempts to alter risk factors (such as cessation of smoking by middle-aged men at risk of heart disease); screening trials of the effectiveness of procedures that detect early subclinical cases (such as the Papanicolaou test to screen for cervical cancer); or clinical trials of the effectiveness of specific treatments (such as studies comparing survival after coronary bypass surgery to survival on medication alone).

Outcome measures include incidence, mortality, and survival rates in both the intervention and control groups. Some studies extend only to rates of compliance (the percentage of individuals assigned to the smoking cessation program who attend the program compared to those not assigned who may seek help on their own) or to risk-factor changes (the percentage of smokers assigned to the smoking cessation program who quit smoking

compared to those not assigned) rather than to measuring disease occurrence per se (heart disease rates in those assigned to the smoking cessation program compared to those not assigned).

Association versus causation

An association of a factor with a disease does not necessarily mean that the two are causally related. For example, people who carry matches may have higher rates of cancer, but matches do not cause the disease. Criteria for establishing causality between a factor and a disease include the strength of the observed relationship, supported by any biological evidence, consistency with results of other epidemiologic studies, and the correct time sequence between exposure to the factor and development of the disease.

Observed associations may be chance findings. Biostatistical methods, which are essential in modern epidemiology, are used to test the possibility that observed results are due to chance and consider the effects of other factors on the associations that are found. The association of a factor with a disease may arise because both are related to another factor (known as a confounding factor). For example, smoking is independently related to both alcohol consumption and oral cancer; thus, individuals who drink alcohol are at increased risk of oral cancer in part because of their smoking. The relative risk of oral cancer attributed to alcohol consumption would be artificially high if smoking (a confounding factor) were not taken into account. Design and analysis techniques deal with the effects of potential confounding factors. *See also:* DISEASE; EPIDEMIC.

Virginia L. Ernster

Bibliography

G. D. Friedman, *Primer of Epidemiology*, 4th ed., 1994

C. H. Hennekens and J. E. Buring, *Epidemiology in Medicine*, 1987

S. B. Hulley, *Designing Clinical Research*, 1988

J. L. Kelsey, W. D. Thompson, and A. S. Evans, *Methods in Observational Epidemiology*, 2d ed., 1996

Additional Readings

C. Baglioni et al., Insomnia as a predictor of depression: A meta-analytic evaluation of longitudinal epidemiological studies, *J. Affect. Disord.*, 135(1):10–19, 2011 DOI: http://doi.org/10.1016/j.jad.2011.01.011

P. J. Fos, *Epidemiology Foundations: The Science of Public Health*, Jossey-Bass, San Francisco, CA, 2011

K. Krickeberg, T. M. H. Pham, and V. T. Pham, *Statistics for Biology and Health: Epidemiology: The Key to Prevention*, Springer, New York, 2012

K. J. Rothman, *Epidemiology: An Introduction*, 2d ed., Oxford University Press, New York, 2012

P. Webb and C. Bain, *Essential Epidemiology: An Introduction for Students and Health Professionals*, 2d ed., Cambridge University Press, Cambridge, UK, 2011