

Nursing Research: Generating and Assessing Evidence for Nursing Practice

Study of health and disease within a population

Epidemiology is the study and analysis of the distribution (who, when, and where), patterns and determinants of health and disease conditions in a defined population.

It is a cornerstone of public health, and shapes policy decisions and evidence-based practice by identifying risk factors for disease and targets for preventive healthcare. Epidemiologists help with study design, collection, and statistical analysis of data, amend interpretation and dissemination of results (including peer review and occasional systematic review). Epidemiology has helped develop methodology used in clinical research, public health studies, and, to a lesser extent, basic research in the biological sciences.[1]

Major areas of epidemiological study include disease causation, transmission, outbreak investigation, disease surveillance, environmental epidemiology, forensic epidemiology, occupational epidemiology, screening, biomonitoring, and comparisons of treatment effects such as in clinical trials. Epidemiologists rely on other scientific disciplines like biology to better understand disease processes, statistics to make efficient use of the data and draw appropriate conclusions, social sciences to better understand proximate and distal causes, and engineering for exposure assessment.

Epidemiology, literally meaning "the study of what is upon the people", is derived from Greek *epi* 'upon, among', *demos* 'people, district', and *logos* 'study, word, discourse', suggesting that it applies only to human populations. However, the term is widely used in studies of zoological populations (veterinary epidemiology), although the term "epizootology" is available, and it has also been applied to studies of plant populations (botanical or plant disease epidemiology).[2]

The distinction between "epidemic" and "endemic" was first drawn by Hippocrates,[3] to distinguish between diseases that are "visited upon" a population (epidemic) from those that "reside within" a population (endemic).[4] The term "epidemiology" appears to have first been used to describe the study of epidemics in 1802 by the Spanish physician Villalba in *Epidemiología de España*. [4] Epidemiologists also study the interaction of diseases in a population, a condition known as a syndemic.

The term epidemiology is now widely applied to cover the description and causation of not only epidemic, infectious disease, but of disease in general, including related conditions. Some examples of topics examined through epidemiology include as high blood pressure, mental illness and obesity. Therefore, this epidemiology is based upon how the pattern of the disease causes change in the function of human beings.

History [edit]

The Greek physician Hippocrates, known as the father of medicine,[5][6] sought a

P

logic to sickness; he is the first person known to have examined the relationships between the occurrence of disease and environmental influences.[7] Hippocrates believed sickness of the human body to be caused by an imbalance of the four humors (black bile, yellow bile, blood, and phlegm). The cure to the sickness was to remove or add the humor in question to balance the body. This belief led to the application of bloodletting and dieting in medicine.[8] He coined the terms endemic (for diseases usually found in some places but not in others) and epidemic (for diseases that are seen at some times but not others).[9]

Modern era [edit]

In the middle of the 16th century, a doctor from Verona named Girolamo Fracastoro was the first to propose a theory that these very small, unseeable, particles that cause disease were alive. They were considered to be able to spread by air, multiply by themselves and to be destroyable by fire. In this way he refuted Galen's miasma theory (poison gas in sick people). In 1543 he wrote a book *De contagione et contagiosis morbis*, in which he was the first to promote personal and environmental hygiene to prevent disease. The development of a sufficiently powerful microscope by Antonie van Leeuwenhoek in 1675 provided visual evidence of living particles consistent with a germ theory of disease.[citation needed]

During the Ming Dynasty, Wu Youke (1582–1652) developed the idea that some diseases were caused by transmissible agents, which he called *Li Qi* (疫气 or pestilential factors) when he observed various epidemics rage around him between 1641 and 1644.[10] His book *Wen Yi Lun* (瘟疫论/Treatise on Pestilence/Treatise of Epidemic Diseases) can be regarded as the main etiological work that brought forward the concept.[11] His concepts were still being considered in analysing SARS outbreak by WHO in 2004 in the context of traditional Chinese medicine.[12]

Another pioneer, Thomas Sydenham (1624–1689), was the first to distinguish the fevers of Londoners in the later 1600s. His theories on cures of fevers met with much resistance from traditional physicians at the time. He was not able to find the initial cause of the smallpox fever he researched and treated.[8]

John Graunt, a haberdasher and amateur statistician, published *Natural and Political Observations ... upon the Bills*

of Mortality in 1662. In it, he analysed the mortality rolls in London before the Great Plague, presented one of the first life tables, and reported time trends for many diseases, new and old. He provided statistical evidence for many theories on disease, and also refuted some widespread ideas on them.[citation needed]

John Snow is famous for his investigations into the causes of the 19th-century cholera epidemics, and is also known as the father of (modern) epidemiology.[13][14] He began with noticing the significantly higher death rates in two areas supplied by Southwark Company. His identification of the Broad Street pump as the cause of the Soho epidemic is considered the classic example of epidemiology. Snow used chlorine in an attempt to clean the water and removed the handle; this ended the outbreak. This has been perceived as a major event in the history of public health and regarded as the founding event of the science of epidemiology, having helped shape public health policies around the world.[15][16] However, Snow's research and preventive measures to avoid further outbreaks were not fully accepted or put into practice until after his death due to the prevailing Miasma Theory of the time, a model of disease in which poor air quality was blamed for illness. This was used to rationalize high rates of infection in impoverished areas instead of addressing the underlying issues of poor nutrition and sanitation, and was proven false by his work.[17]

Other pioneers include Danish physician Peter Anton Schleisner, who in 1849 related his work on the prevention of the epidemic of neonatal tetanus on the Vestmanna Islands in Iceland.[18][19] Another important pioneer was Hungarian physician Ignaz Semmelweis, who in 1847 brought down infant mortality at a Vienna hospital by instituting a disinfection procedure. His findings were published in 1850, but his work was ill-received by his colleagues, who discontinued the procedure. Disinfection did not become widely practiced until British surgeon Joseph Lister 'discovered' antiseptics in 1865 in light of the work of Louis Pasteur.[citation needed]

In the early 20th century, mathematical methods were introduced into epidemiology by Ronald Ross, Janet Lane-Clayton, Anderson Gray McKendrick, and others.[20][21][22][23] In a parallel development during the 1920s, German-Swiss pathologist Max Askanazy and others founded the International Society for Geographical Pathology to systematically investigate the geographical pathology of cancer and other non-infectious diseases across populations in different regions. After World War II, Richard Doll and other non-pathologists joined the field and advanced methods to study cancer, a disease with patterns and mode of occurrences that could not be suitably studied with the methods developed for epidemics of infectious diseases. Geography pathology eventually combined with infectious disease epidemiology to make the field that is epidemiology today.[24]

Another breakthrough was the 1954 publication of the results of a British Doctors Study, led by Richard Doll and Austin Bradford Hill, which lent very strong statistical support to the link between tobacco smoking and lung cancer.

In the late 20th century, with the advancement of biomedical sciences, a number of molecular markers in blood, other biospecimens and environment were identified as predictors of development or risk of a certain disease. Epidemiology research to examine the relationship between these biomarkers analyzed at the molecular level and disease was broadly named "molecular epidemiology". Specifically, "genetic epidemiology" has been used for epidemiology of germline genetic variation and disease. Genetic variation is typically determined using DNA from peripheral blood leukocytes.[citation needed]

21st century [edit]

Since the 2000s, genome-wide association studies (GWAS) have been commonly performed to identify genetic risk factors for many diseases and health conditions.

While most molecular epidemiology studies are still using conventional disease diagnosis and classification systems, it is increasingly recognized that disease progression represents inherently heterogeneous processes differing from person to person. Conceptually, each individual has a unique disease process different from any other individual ("the unique disease principle"),[25][26] considering uniqueness of the exposome (a totality of endogenous and exogenous / environmental exposures) and its unique influence on molecular pathologic process in each individual. Studies to examine the relationship between an exposure and molecular pathologic signature of disease (particularly cancer) became increasingly common throughout the 2000s. However, the use of molecular pathology in epidemiology posed unique challenges, including lack of research guidelines and standardized statistical methodologies, and paucity of interdisciplinary experts and training programs.[27] Furthermore, the concept of disease heterogeneity appears to conflict with the long-standing premise in epidemiology that individuals with the same disease name have similar etiologies and disease processes. To resolve these issues and advance population health science in the era of molecular precision medicine, "molecular pathology" and "epidemiology" was integrated to create a new interdisciplinary field of "molecular pathological epidemiology" (MPE),[28][29] defined as "epidemiology of molecular pathology and heterogeneity of disease". In MPE, investigators analyze the relationships between (A) environmental, dietary, lifestyle and genetic factors; (B) alterations in cellular or extracellular molecules; and (C) evolution and progression of disease. A better understanding of heterogeneity of disease pathogenesis will further contribute to

elucidate etiologies of disease. The MPE approach can be applied to not only neoplastic diseases but also non-neoplastic diseases.[30] The concept and paradigm of MPE have become widespread in the 2010s.[31][32][33][34][35][36][37]

By 2012, it was recognized that many pathogens' evolution is rapid enough to be highly relevant to epidemiology, and that therefore much could be gained from an interdisciplinary approach to infectious disease integrating epidemiology and molecular evolution to "inform control strategies, or even patient treatment." [38][39]

Modern epidemiological studies can use advanced statistics and machine learning to create predictive models as well as to define treatment effects.[40][41]

Types of studies [edit]

Epidemiologists employ a range of study designs from the observational to experimental and generally categorized as descriptive (involving the assessment of data covering time, place, and person), analytic (aiming to further examine known associations or hypothesized relationships), and experimental (a term often equated with clinical or community trials of treatments and other interventions). In observational studies, nature is allowed to "take its course", as epidemiologists observe from the sidelines. Conversely, in experimental studies, the epidemiologist is the one in control of all of the factors entering a certain case study.[42] Epidemiological studies are aimed, where possible, at revealing unbiased relationships between exposures such as alcohol or smoking, biological agents, stress, or chemicals to mortality or morbidity. The identification of causal relationships between these exposures and outcomes is an important aspect of epidemiology. Modern epidemiologists use informatics as a tool.[citation needed]

Observational studies have two components, descriptive and analytical. Descriptive observations pertain to the "who, what, where and when of health-related state occurrence". However, analytical observations deal more with the 'how' of a health-related event.[42] Experimental epidemiology contains three case types: randomized controlled trials (often used for a new medicine or drug testing), field trials (conducted on those at a high risk of contracting a disease), and community trials (research on social originating diseases).[42]

The term 'epidemiologic triad' is used to describe the intersection of Host, Agent, and Environment in analyzing an outbreak.[citation needed]

Case series [edit]

Case-series may refer to the qualitative study of the experience of a single patient, or small group of patients with a similar diagnosis, or to a statistical factor with the potential to produce illness with periods when they are unexposed.[citation needed]

The former type of study is purely descriptive and cannot be used to make inferences about the general population of patients with that disease. These types of studies, in which an astute clinician identifies an unusual feature of a disease or a patient's history, may lead to a formulation of a new hypothesis. Using the data from the series, analytic studies could be done to investigate possible causal factors. These can include case-control studies or prospective studies. A case-control study would involve matching comparable controls without the disease to the cases in the series. A prospective study would involve following the case series over time to evaluate the disease's natural history.[43]

The latter type, more formally described as self-controlled case-series studies, divide individual patient follow-up time into exposed and unexposed periods and use fixed-effects Poisson regression processes to compare the incidence rate of a given outcome between exposed and unexposed periods. This technique has been extensively used in the study of adverse reactions to vaccination and has been shown in some circumstances to provide statistical power comparable to that available in cohort studies.[citation needed]

Case-control studies [edit]

Case-control studies select subjects based on their disease status. It is a retrospective study. A group of individuals that are disease positive (the "case" group) is compared with a group of disease negative individuals (the "control" group). The control group should ideally come from the same population that gave rise to the cases. The case-control study looks back through time at potential exposures that both groups (cases and controls) may have encountered. A 2×2 table is constructed, displaying exposed cases (A), exposed controls (B), unexposed cases (C) and unexposed controls (D). The statistic generated to measure association is the odds ratio (OR), which is the ratio of the odds of exposure in the cases (A/C) to the odds of exposure in the controls (B/D), i.e. $OR = (AD/BC)$.

Cases Controls Exposed A B Unexposed C D

If the OR is significantly greater than 1, then the conclusion is "those with the disease are more likely to have been exposed," whereas if it is close to 1 then the exposure and disease are not likely associated. If the OR is far less than one, then this suggests that the exposure is a protective factor in the causation of the disease. Case-control studies are usually faster and more cost-effective than cohort studies but are sensitive to bias (such as recall bias and selection bias). The main challenge is to identify the appropriate control group; the distribution of exposure among the control group should be representative of the distribution in the population that gave rise to the cases. This can be achieved by drawing a random sample from the original population at risk. This has as a consequence that the control group can contain people with the disease under study when the disease has a high attack rate in a population.

A major drawback for case control studies is that, in order to be considered to be statistically significant, the minimum number of cases required at the 95% confidence interval is related to the odds ratio by the equation:

$$\text{total cases} = A + C = 1.96^2 (1 + N) \left(\frac{1}{\ln(OR)} \right)^2 (OR + 2\sqrt{OR} + 1) \approx 15.5 (1 + N) \left(\frac{1}{\ln(OR)} \right)^2 \left(\frac{OR + 2\sqrt{OR} + 1}{\sqrt{OR}} \right)$$

where N is the ratio of cases to controls. As the odds ratio approaches 1, the number of cases required for statistical significance grows towards infinity; rendering case-control studies all but useless for low odds ratios. For instance, for an odds ratio of 1.5 and cases = controls, the table shown above would look like this:

Cases	Controls	Exposed	103	84	Unexposed	84	103
-------	----------	---------	-----	----	-----------	----	-----

For an odds ratio of 1.1:

Cases	Controls	Exposed	1732	1652	Unexposed	1652	1732
-------	----------	---------	------	------	-----------	------	------

Cohort studies [edit]

Cohort studies select subjects based on their exposure status. The study subjects should be at risk of the outcome under investigation at the beginning of the cohort study; this usually means that they should be disease free when the

cohort study starts. The cohort is followed through time to assess their later outcome status. An example of a cohort study would be the investigation of a cohort of smokers and non-smokers over time to estimate the incidence of lung cancer. The same 2x2 table is constructed as with the case control study. However, the point estimate generated is the relative risk (RR), which is the probability of disease for a person in the exposed group, $P_e = A / (A + B)$ over the probability of disease for a person in the unexposed group, $P_u = C / (C + D)$, i.e. $RR = P_e / P_u$.

..... Case Non-case Total Exposed A B (A + B) Unexposed C D (C + D)

As with the OR, a RR greater than 1 shows association, where the conclusion can be read "those with the exposure were more likely to develop the disease."

Prospective studies have many benefits over case control studies. The RR is a more powerful effect measure than the OR, as the OR is just an estimation of the RR, since true incidence cannot be calculated in a case control study where subjects are selected based on disease status. Temporality can be established in a prospective study, and confounders are more easily controlled for. However, they are more costly, and there is a greater chance of losing subjects to follow-up based on the long time period over which the cohort is followed.

Cohort studies also are limited by the same equation for number of cases as for cohort studies, but, if the base incidence rate in the study population is very low, the number of cases required is reduced by 1/2.

Causal inference [edit]

Although epidemiology is sometimes viewed as a collection of statistical tools used to elucidate the associations of exposures to health outcomes, a deeper understanding of this science is that of discovering causal relationships.

"Correlation does not imply causation" is a common theme for much of the epidemiological literature. For epidemiologists, the key is in the term inference. Correlation, or at least association between two variables, is a necessary but not sufficient criterion for the inference that one variable causes the other. Epidemiologists use gathered data and a broad range of biomedical and psychosocial theories in an iterative way to generate or expand theory, to test hypotheses, and to make educated, informed assertions about which relationships are causal, and about exactly how they are causal.

Epidemiologists emphasize that the "one cause â€" one effect" understanding is a simplistic mis-belief.[44] Most outcomes, whether disease or death, are caused by a chain or web consisting of many component causes.[45] Causes can be distinguished as necessary, sufficient or probabilistic conditions. If a necessary condition can be identified and controlled (e.g., antibodies to a disease agent, energy in an injury), the harmful outcome can be avoided (Robertson, 2015). One tool regularly used to conceptualize the multicausality associated with disease is the causal pie model.[46]

Bradford Hill criteria [edit]

In 1965, Austin Bradford Hill proposed a series of considerations to help assess evidence of causation,[47] which have come to be commonly known as the "Bradford Hill criteria". In contrast to the explicit intentions of their author, Hill's considerations are now sometimes taught as a checklist to be implemented for assessing causality.[48] Hill himself said "None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required sine qua non." [47]

Strength of Association: A small association does not mean that there is not a causal effect, though the larger the association, the more likely that it is causal.[47] Consistency of Data: Consistent findings observed by different persons in different places with different samples strengthens the likelihood of an effect.[47] Specificity: Causation is likely if a very specific population at a specific site and disease with no other likely explanation. The more specific an association between a factor and an effect is, the bigger the probability of a causal relationship.[47] Temporality: The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, then the effect must occur after that delay).[47] Biological gradient: Greater exposure should generally lead to greater incidence of the effect. However, in some cases, the mere presence of the factor can trigger the effect. In other cases, an inverse proportion is observed: greater exposure leads to lower incidence.[47] Plausibility: A plausible mechanism between cause and effect is helpful (but Hill noted that knowledge of the mechanism is limited by current knowledge).[47] Coherence: Coherence between epidemiological and laboratory findings increases the likelihood of an effect. However, Hill noted that "... lack of such [laboratory] evidence cannot nullify the epidemiological effect on associations".[47] Experiment: "Occasionally it is possible to appeal to experimental evidence".[47] Analogy: The effect of similar factors may be considered.[47]

Legal interpretation [edit]

Epidemiological studies can only go to prove that an agent could have caused, but not that it did cause, an effect in any particular case:

Epidemiology is concerned with the incidence of disease in populations and does not address the question of the cause of an individual's disease. This question, sometimes referred to as specific causation, is beyond the domain of the science of epidemiology. Epidemiology has its limits at the point where an inference is made that the relationship between an agent and a disease is causal (general causation) and where the magnitude of excess risk attributed to the agent has been determined; that is, epidemiology addresses whether an agent can cause disease, not whether an agent did cause a specific plaintiff's disease.[49]

In United States law, epidemiology alone cannot prove that a causal association does not exist in general. Conversely, it can be (and is in some circumstances) taken by US courts, in an individual case, to justify an inference that a causal association does exist, based upon a balance of probability.

The subdiscipline of forensic epidemiology is directed at the investigation of specific causation of disease or injury in individuals or groups of individuals in instances in which causation is disputed or is unclear, for presentation in legal settings.

Population-based health management [edit]

Epidemiological practice and the results of epidemiological analysis make a significant contribution to emerging population-based health management frameworks.

Population-based health management encompasses the ability to:

Assess the health states and health needs of a target population;

Implement and evaluate interventions that are designed to improve the health of that population; and

Efficiently and effectively provide care for members of that population in a way that is consistent with the community's cultural, policy and health resource values.

Modern population-based health management is complex, requiring a multiple set of skills (medical, political, technological, mathematical, etc.) of which epidemiological practice and analysis is a core component, that is unified with management science to provide efficient and effective health care and health guidance to a population. This task requires the forward-looking ability of modern risk management approaches that transform health risk factors, incidence, prevalence and mortality statistics (derived from epidemiological analysis) into management metrics that not only guide how a health system responds to current population health issues but also how a health system can be managed to better respond to future potential population health issues.[50]

Examples of organizations that use population-based health management that leverage the work and results of epidemiological practice include Canadian Strategy for Cancer Control, Health Canada Tobacco Control Programs, Rick Hansen Foundation, Canadian Tobacco Control Research Initiative.[51][52][53]

Each of these organizations uses a population-based health management framework called Life at Risk that combines epidemiological quantitative analysis with demographics, health agency operational research and economics to perform:

Population Life Impacts Simulations : Measurement of the future potential impact of disease upon the population with respect to new disease cases, prevalence, premature death as well as potential years of life lost from disability and death;

: Measurement of the future potential impact of disease upon the population with respect to new disease cases, prevalence, premature death as well as potential years of life lost from disability and death; Labour Force Life Impacts Simulations : Measurement of the future potential impact of disease upon the labour force with respect to new disease cases, prevalence, premature death and potential years of life lost from disability and death;

: Measurement of the future potential impact of disease upon the labour force with respect to new disease cases, prevalence, premature death and potential years of life lost from disability and death; Economic Impacts of Disease Simulations: Measurement of the future potential impact of disease upon private sector disposable income impacts (wages, corporate profits, private health care costs) and public sector disposable income impacts (personal income tax, corporate income tax, consumption taxes, publicly funded health care costs).

Applied field epidemiology [edit]

Applied epidemiology is the practice of using epidemiological methods to protect or improve the health of a population. Applied field epidemiology can include investigating communicable and non-communicable disease outbreaks, mortality and morbidity rates, and nutritional status, among other indicators of health, with the purpose of communicating the results to those who can implement appropriate policies or disease control measures.

Humanitarian context [edit]

As the surveillance and reporting of diseases and other health factors become increasingly difficult in humanitarian crisis situations, the methodologies used to report the data are compromised. One study found that less than half (42.4%) of nutrition surveys sampled from humanitarian contexts correctly calculated the prevalence of malnutrition and only one-third (35.3%) of the surveys met the criteria for quality. Among the mortality surveys, only 3.2% met the criteria for quality. As nutritional status and mortality rates help indicate the severity of a crisis, the tracking and reporting of these health factors is crucial.

Vital registries are usually the most effective ways to collect data, but in humanitarian contexts these registries can be non-existent, unreliable, or inaccessible. As such, mortality is often inaccurately measured using either prospective demographic surveillance or retrospective mortality surveys. Prospective demographic surveillance requires much manpower and is difficult to implement in a spread-out population. Retrospective mortality surveys are prone to selection and reporting biases. Other methods are being developed, but are not common practice yet.[54][55][56][57]

Validity: precision and bias [edit]

Different fields in epidemiology have different levels of validity. One way to assess the validity of findings is the ratio of false-positives (claimed effects that are not correct) to false-negatives (studies which fail to support a true effect). To take the field of genetic epidemiology, candidate-gene studies produced over 100 false-positive findings for each false-negative. By contrast genome-wide association appear close to the reverse, with only one false positive for every 100 or more false-negatives.[58] This ratio has improved over time in genetic epidemiology as the field has adopted stringent criteria. By contrast, other epidemiological fields have not required such rigorous reporting and are much less reliable as a result.[58]

Random error [edit]

Random error is the result of fluctuations around a true value because of sampling variability. Random error is just that: random. It can occur during data collection, coding, transfer, or analysis. Examples of random errors include poorly worded questions, a misunderstanding in interpreting an individual answer from a particular respondent, or a typographical error during coding. Random error affects measurement in a transient, inconsistent manner and it is impossible to correct for random error.

There is a random error in all sampling procedures. This is called sampling error.

Precision in epidemiological variables is a measure of random error. Precision is also inversely related to random error, so that to reduce random error is to increase precision. Confidence intervals are computed to demonstrate the precision of relative risk estimates. The narrower the confidence interval, the more precise the relative risk estimate.

There are two basic ways to reduce random error in an epidemiological study. The first is to increase the sample size of the study. In other words, add more subjects to your study. The second is to reduce the variability in measurement in the study. This might be accomplished by using a more precise measuring device or by increasing the number of measurements.

Note, that if sample size or number of measurements are increased, or a more precise measuring tool is purchased, the costs of the study are usually increased. There is usually an uneasy balance between the need for adequate precision and the practical issue of study cost.

Systematic error [edit]

A systematic error or bias occurs when there is a difference between the true value (in the population) and the observed value (in the study) from any cause other than sampling variability. An example of systematic error is if, unknown to you, the pulse oximeter you are using is set incorrectly and adds two points to the true value each time a measurement is taken. The measuring device could be precise but not accurate. Because the error happens in every instance, it is systematic. Conclusions you draw based on that data will still be incorrect. But the error can be reproduced in the future (e.g., by using the same mis-set instrument).

A mistake in coding that affects all responses for that particular question is another example of a systematic error.

The validity of a study is dependent on the degree of systematic error. Validity is usually separated into two components:

Internal validity is dependent on the amount of error in measurements, including exposure, disease, and the associations between these variables. Good internal validity implies a lack of error in measurement and suggests that inferences may be drawn at least as they pertain to the subjects under study.

External validity pertains to the process of generalizing the findings of the study to the population from which the sample was drawn (or even beyond that population to a more universal statement). This requires an understanding of which conditions are relevant (or irrelevant) to the generalization. Internal validity is clearly a prerequisite for external validity.

Selection bias [edit]

Selection bias occurs when study subjects are selected or become part of the study as a result of a third, unmeasured variable which is associated with both the exposure and outcome of interest.[59] For instance, it has repeatedly been noted that cigarette smokers and non smokers tend to differ in their study participation rates. (Sackett D cites the example of Seltzer et al., in which 85% of non smokers and 67% of smokers returned mailed questionnaires.)[60] It is important to note that such a difference in response will not lead to bias if it is not also associated with a systematic difference in outcome between the two response groups.

Information bias [edit]

Information bias is bias arising from systematic error in the assessment of a variable.[61] An example of this is recall bias. A typical example is again provided by Sackett in his discussion of a study examining the effect of specific exposures on fetal health: "in questioning mothers whose recent pregnancies had ended in fetal death or malformation (cases) and a matched group of mothers whose pregnancies ended normally (controls) it was found that 28% of the former, but only 20% of the latter, reported exposure to drugs which could not be substantiated either in earlier prospective interviews or in other health records".[60] In this example, recall bias probably occurred as a

result of women who had had miscarriages having an apparent tendency to better recall and therefore report previous exposures.

Confounding [edit]

Confounding has traditionally been defined as bias arising from the co-occurrence or mixing of effects of extraneous factors, referred to as confounders, with the main effect(s) of interest.[61][62] A more recent definition of confounding invokes the notion of counterfactual effects.[62] According to this view, when one observes an outcome of interest, say $Y=1$ (as opposed to $Y=0$), in a given population A which is entirely exposed (i.e. exposure $X = 1$ for every unit of the population) the risk of this event will be R_{A1} . The counterfactual or unobserved risk R_{A0} corresponds to the risk which would have been observed if these same individuals had been unexposed (i.e. $X = 0$ for every unit of the population). The true effect of exposure therefore is: $R_{A1} - R_{A0}$ (if one is interested in risk differences) or R_{A1} / R_{A0} (if one is interested in relative risk). Since the counterfactual risk R_{A0} is unobservable we approximate it using a second population B and we actually measure the following relations: $R_{A1} - R_{B0}$ or R_{A1} / R_{B0} . In this situation, confounding occurs when $R_{A0} \neq R_{B0}$. [62] (NB: Example assumes binary outcome and exposure variables.)

Some epidemiologists prefer to think of confounding separately from common categorizations of bias since, unlike selection and information bias, confounding stems from real causal effects.[59]

The profession [edit]

Few universities have offered epidemiology as a course of study at the undergraduate level. One notable undergraduate program exists at Johns Hopkins University, where students who major in public health can take graduate-level courses, including epidemiology, during their senior year at the Bloomberg School of Public Health.[63]

Although epidemiologic research is conducted by individuals from diverse disciplines, including clinically trained professionals such as physicians, formal training is available through Masters or Doctoral programs including Master of Public Health (MPH), Master of Science of Epidemiology (MSc.), Doctor of Public Health (DrPH), Doctor of Pharmacy (PharmD), Doctor of Philosophy (PhD), Doctor of Science (ScD). Many other graduate programs, e.g., Doctor of Social Work (DSW), Doctor of Clinical Practice (DClinP), Doctor of Podiatric Medicine (DPM), Doctor of Veterinary Medicine

(DVM), Doctor of Nursing Practice (DNP), Doctor of Physical Therapy (DPT), or for clinically trained physicians, Doctor of Medicine (MD) or Bachelor of Medicine and Surgery (MBBS or MBChB) and Doctor of Osteopathic Medicine (DO), include some training in epidemiologic research or related topics, but this training is generally substantially less than offered in training programs focused on epidemiology or public health. Reflecting the strong historical tie between epidemiology and medicine, formal training programs may be set in either schools of public health and medical schools.

As public health/health protection practitioners, epidemiologists work in a number of different settings. Some epidemiologists work 'in the field'; i.e., in the community, commonly in a public health/health protection service, and are often at the forefront of investigating and combating disease outbreaks. Others work for non-profit organizations, universities, hospitals and larger government entities such as state and local health departments, various Ministries of Health, Doctors without Borders, the Centers for Disease Control and Prevention (CDC), the Health Protection Agency, the World Health Organization (WHO), or the Public Health Agency of Canada. Epidemiologists can also work in for-profit organizations such as pharmaceutical and medical device companies in groups such as market research or clinical development.

An April 2020 University of Southern California article noted that "The coronavirus epidemic... thrust epidemiology â€" the study of the incidence, distribution and control of disease in a population â€" to the forefront of scientific disciplines across the globe and even made temporary celebrities out of some of its practitioners." [64]

See also [edit]

References [edit]

Citations [edit]

Reference

[A Scientific Revolution: Ten Men and Women Who Reinvented American Medicine](#)

[Qualitative Data Collection Tools: Design, Development, and Applications
\(Qualitative Research Methods\)](#)