Epidemiology

Textbook of Epidemiology, Second Edition. Lex Bouter, Maurice Zeegers and Tianjing Li.

Companion website: www.wiley.com/go/Bouter/TextbookofEpidemiology

1.1 What is epidemiology?

1.1.1 Definition

When a new infectious disease (such as COVID-19) is discovered, epidemiologists work with other scientists to find out who has it, why they have it, and what can be done about it. Epidemiology is the study of the distribution (frequency, pattern), and explanatory factors (causes, risk factors) of health-related states and events (health outcomes) in specified populations at a particular time. Using a COVID-19 outbreak as an example, epidemiologists identified the source of the outbreak by conducting field investigations in the community and health facilities in Wuhan, China. They collected specimens from the nose and throat for lab analyses. These field investigations informed those who were infected, when they became sick, and where they had been just before they became sick, which ultimately led to the identification of the likely source. Epidemiologists monitor and track the disease by using surveillance systems that report data on new cases of infection, symptoms, hospitalizations, treatments, deaths, and demographic information. Epidemiologists further investigate the disease to find out how contagious COVID-19 is, how to test for the presence of infection, what risk factors for severe illness are, what the natural and treated course of the disease on average is, and which treatments are most effective. Finally, epidemiologists contribute to developing guidance for preventive actions (e.g. face-mask wearing and social distancing) to slow the spread of the disease and lessen its impact.

Following on from the above, we can characterize epidemiology as follows:

- Health outcomes are the main focus of epidemiological research.
- Epidemiology is the study of health outcomes in human populations.
- Epidemiologists usually look at health outcomes in relation to other phenomena. These are factors suspected to have influenced the development of the health outcome in question (etiological factors), to give an indication of the presence of a particular health outcome (diagnostic factors), to be associated with the course of the health outcome (prognostic factors), or to mitigate the impact of the health outcome (interventions).
- The application of epidemiological study findings is meant to improve the wellbeing of the population.

Case 1.1 Scurvy and citrus fruits

LANDMARK PAPER

James Lind, a Scottish ship's doctor, carried out an inventive experiment on scurvy patients in 1747. Scurvy (or scorbutus) is a disease characterized by bleeding gums, internal bleeding, stiff limbs and rough skin, among other things. It was very common in those days among crews of ocean-going ships. Long after Lind made his observations, it was found that scurvy is caused primarily by a deficiency of vitamin C (ascorbic acid), which is needed to synthesize collagen, a substance that strengthens blood vessels. James Lind selected 12 cases in the same stage of the disease on his ship. In addition to the common diet all participants received as their primary source of nutrition, he compared six “treatments”: two patients were prescribed cider, two elixir of vitriol, two sea water, two a mixture of nutmeg and a commonly used medicine based on garlic and mustard seed, and two were given two oranges and one lemon a day. Only the patients who were given the orange and lemon treatment experienced a rapid cure. The mechanism by which scurvy can be cured and prevented using citrus fruit was not known at that time. However, it did not present an obstacle to taking effective measures. It took many years for citrus fruit to be introduced as a prophylactic for this dreaded disease on ships. Nearly 50 years later (1 year after Lind's death), the Admiralty of England made issue of lemon juice compulsory in its navy. As a nod to Lind's pioneering work, the Institute of Naval Medicine includes a lemon tree in its crest.

1.1.2 Health outcomes

The basic variable in epidemiology, and in any epidemiological study, is a health outcome. Epidemiologists are interested in the extent to which diseases or more generally health outcomes occur (i.e. their frequency) among the population. This defines the object of epidemiological research, namely health outcomes in human populations. The word “disease” needs to be interpreted broadly in this context: it can mean a broad variety of medical conditions, both infectious and non-infectious, acute and chronic, somatic and mental. In this context “health outcome” can refer to all sorts of phenomena on a continuum from full health to death from a condition. It can also refer to a disability, injury due to trauma, quality of life, or a physiological measure. As there is no term that encompasses all these aspects, we generally use the phrase “health outcome” – or simply “outcome” – in this book. Chapter 2 provides more detail on measuring various aspects of health. For the purpose of epidemiological research, we need to define the health outcome of interest as precisely as possible so as to enable us to detect relationships with explanatory factors.

1.1.3 Frequency measures

Epidemiology is the study of health outcomes and explanatory factors in humans. Animal experiments and observations in cell or organ cultures do not fall into the domain of epidemiology. Although the measurements are usually carried out on individuals, the results of epidemiological research always relate to groups of people. As the disease frequency for a group can be interpreted as the mean disease risk for each member of the group, the knowledge gained from epidemiological research is valuable not only at the group level (for public health) but also at the individual level (in healthcare). Epidemiologists ascertain whether each individual has the health outcome and then count the number of individuals with the health outcome in the group as a whole, thus yielding the epidemiological fraction.

$$\frac{\text{Number of cases}}{\text{Total number of persons in the group from which these cases are taken}}$$

This fraction provides the basis for all epidemiological measures of frequency (see Chapter 2). For example, we refer to incidence when counting new cases of a disease in a group at risk of contracting that disease (the at-risk population). Prevalence, on the other hand, relates to the number of existing cases of the disease. Incidence is also a measure of risk: it indicates the average risk of developing the health outcome during the specified period.

Health and disease are not equally distributed among the population; and it is this fact that lends epidemiology its raison d'être. The purpose of epidemiology is first and foremost to identify differences in health outcomes between populations. The distribution of a health outcome among the population becomes clear when we investigate the differences in frequency between groups of people at different times, in different places, and with different characteristics. Differences in time can manifest themselves between seasons or over years or decades. Differences between geographical areas can relate to continents, countries, regions of a country, urban versus rural areas, or districts of a town or city. Examples of individual characteristics that can be associated with differences in disease frequency are age, sex, genetic predisposition, occupation, and lifestyle characteristics such as smoking, drinking, and physical activities. Categorizing the population into subpopulations based on time, place, or individual characteristics gives us an understanding of the distribution of the disease risk, and enables us to identify risk periods, risk areas, or at-risk groups. Identifying the distribution pattern of a health outcome falls under the domain of descriptive epidemiology. The branch of epidemiology that aims to identify causal factors underlying the distribution of a health outcome is known as analytical epidemiology.

1.1.4 Explanatory factors

In addition to the health outcomes, epidemiology studies the factors related to the occurrence of health outcomes. These factors fall into three categories: etiological, diagnostic, and prognostic. The term “explanatory factors” is used in this book as a generic term for these factors. Individuals can be exposed to several factors either at the same time or successively. Epidemiologists are interested first and foremost in factors that are causally responsible or co-responsible for the development of a disease (etiological factors or risk factors) or that influence the course of the disease (prognostic factors). They may also look for factors that distinguish people who have a particular disease from those who do not have it (diagnostic factors).

The factors that influence the development or course of a disease or a health outcome can be broadly divided into three categories: genetics, lifestyle, and environment. Genetic properties and biological

1

characteristics of genetic origin are important factors, but they are as of yet, very difficult to manipulate. Interventions to influence the risk of disease are more likely to be available for environmental and lifestyle factors to which people are exposed voluntarily or involuntarily (e.g. diet, smoking, alcohol, drugs, sexual habits, microorganisms, environmental, and occupational exposures). Preventive, diagnostic, and therapeutic interventions can also be regarded as explanatory factors of disease and prognosis, for example, screening programs and various medical and paramedical interventions such as smoking cessation programs, physical activities for weight loss, or surgical and pharmaceutical therapies for a condition.

1.1.5 Epidemiological function

Epidemiology contributes to medicine and public health by discovering what factors are associated with a health outcome, how strong the link is, and what contribution these factors make to the occurrence of the health outcome. Usually, multiple exposures are involved in every health outcome at all stages. This can be expressed as follows:

$$P(O) = f(F_i)$$

This epidemiological function states that the occurrence of the health outcome (O) is a mathematical function of a series of k factors (F_i, where $i = 1, \ldots, k$). In this equation, the health outcome is the dependent variable or outcome variable. The factors are the independent variables in the formula. Various measures of association can be used to express the strength of the relationship between the occurrence of the factor and the health outcome. The most important measures of association will be discussed in Chapter 3.

Although an epidemiological study usually focuses on one single explanatory factor of a health outcome, if we want to gain a good understanding of the relationship between that explanatory factor and the outcome, we will almost always need to include other factors in the study design, measure them and take them into account when analyzing the data.

The presence of these other factors can affect the association. We know, for example, that if we want to examine the effect of alcohol consumption on road accidents, we need to take the driving behaviors (e.g. speed and seatbelt use) into consideration, as the effect of alcohol is stronger for higher speed. In other words, the effect of alcohol on road accidents is modified by speed (effect modification). If we want to study the effect of exercise behavior on cardiovascular disease outcomes, we must not forget to adjust for the effect of healthy diet: people who are more active may also eat a healthier diet. Unless we include these extraneous factors, we cannot be certain whether the observed effect of exercise is actually due to healthy diet (confounding). More information on effect modification and confounding will be given in Chapter 5.

Case 1.2 Cannabis use and depression

In 2015, the United Nations Office on Drugs and Crime estimated the annual prevalence of cannabis use worldwide at around 2.8%–4.5%. This frequency varies between countries, with the highest prevalence (>10%) in Canada, the US, and Australia, and a prevalence between 2.5% and 10% in most European countries. With around 182.5 million users globally, cannabis ranks first among illicit drug use. Research has linked cannabis use to the development of depression. Most of these studies showed an increased risk of depression among cannabis users after adjusting for other possible risk factors for depression (confounding variables). There is also evidence of a dose–response relationship, since the risk of depression increases with higher levels of cannabis use.

Recently a US-cohort study was performed to study the effect of reducing cannabis use on depressive symptoms in young female adults. This cohort study included 332 young women of different ethnicities between the ages of 18 and 24 who had used cannabis at least three times in the previous three months. At baseline, and after three and six months of follow-up, the study participants filled out the Beck Depression Inventory-II questionnaire, which measures depression on a continuous scale with higher scores indicating more severe depressive symptoms. After controlling for alcohol use – which has been shown to be associated with depression – the results showed a significant reduction in depressive symptoms in participants who had stopped using cannabis. This reduction was more pronounced initially among moderately or severely depressed women compared to minimally depressed women. This cohort study supports the view that cannabis use may be a cause of depression. Since the research focused only on young women, the results may not be generalized to men directly. It may be worthwhile to study the existence of effect modification by sex in a mixed population. Before using these insights in the treatment of depressive patients who are also cannabis users, the effect of advice to stop using cannabis needs to be studied in a randomized trial (see Chapter 10).

1.1.6 The empirical cycle

The epidemiological function is a formal expression of the research question. The question whether cannabis use in humans increases the risk of depression (Case 1.2) can be represented by the following function:

$$P(\text{depression}) = f(\text{cannabis use})$$

How do epidemiologists come up with a research question? Research questions do not just arrive unbidden; often casual observations lead investigators to tackle a particular topic. Researchers' curiosity can also be aroused by results of previous research or reports of other people's results, and they will feel the need to confirm such results (verification), dispute them (contradiction), refute them (falsification), or make them more specific (elaboration). Often, they will feel the need to improve on weaknesses in the design and conduct of previous research. Subsequent studies on similar research questions will gradually increase our understanding of that particular aspect of reality. This is referred to as the empirical cycle. In its simplest form, the empirical cycle is shown in Figure 1.1.

Researchers start with a theory, a statement or coherent series of statements intended to have universal validity. In some cases, the theory is being handed down, but it is usually based partly on observations of phenomena, which do not by definition have to be systematic. Sometimes they come from incidental encounters of the researchers. Usually

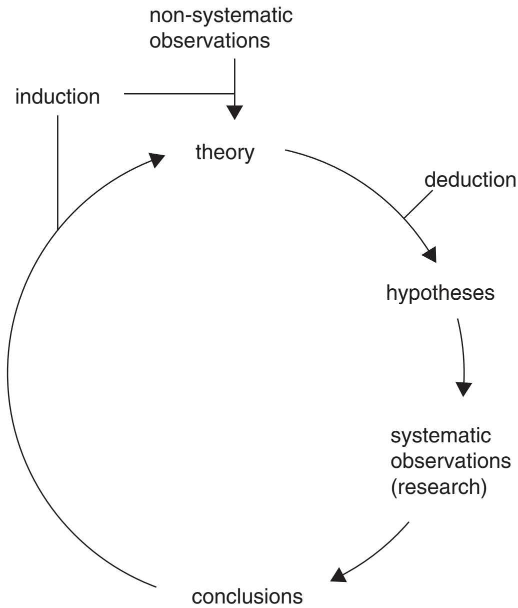

Figure 1.1 The empirical cycle.

systematic observations form the basis for theoretical explorations: results of previous research into the same topic conducted by the researchers themselves or other researchers. This previous research is by no means always epidemiological in nature. Incorporating these systematic and nonsystematic observations into a theory is no simple matter; it calls for a good deal of creativity and originality. This process of developing a more abstract, universal summary of reality from observations is known as induction. According to inductive reasoning, a new theory applies not only to the cases observed but also to all similar cases. To find out whether this is true, researchers carry out new, systematic observations. If the new observations tally with the expectations, confirmation is obtained that we are on the right track with our theory.

Another way of developing a theory is to try to prove that it is incorrect. This idea was introduced in the twentieth century by the philosopher of science Karl Popper (1902–1994) (Figure 1.2) and now forms a cornerstone of empirical science. It is impossible to confirm anything with certainty, but it is possible to falsify it. Popper therefore stated that every scientific theory must be falsifiable. Falsifiability means that we can prove that some proposition, statement, or theory is wrong. To do

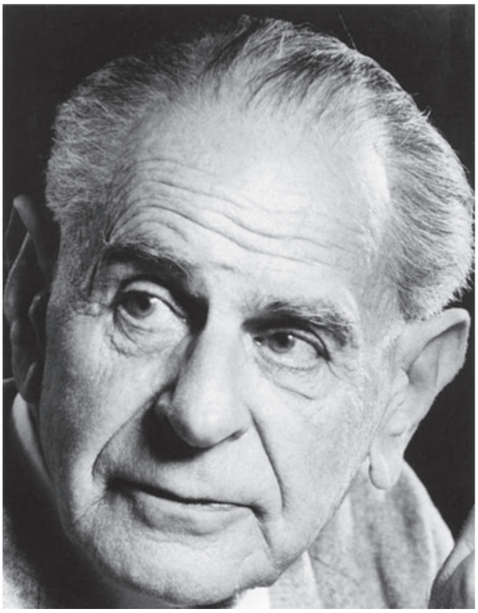

Figure 1.2 Kari Popper (1902–1994), philosopher of science and founding father of the falsifiability theory. Source: LSE Library / Flickr / Public domain.

this, we first need – taking the theory as our starting point – to formulate hypotheses, statements that can or cannot be refuted based on real-world observations.

This process of going from an abstract theory to one or more hypotheses for testing is known as deduction. In effect, the deduction process corresponds to translating a general research idea into one or more research questions. In contrast to what Figure 1.1 suggests, not every systematic study is designed to test a hypothesis, there is also research of a more exploratory nature. Exploratory research is conducted mainly on new and relatively unexplored problems, with the aim of generating promising hypotheses. The findings from this type of research also contribute to scientific theory.

As we have said, empirical research is needed to test a hypothesis or explore a new problem area. The phases of empirical research are as follows:

1. Formulate the research question
2. Draw up a study design:
   - Select the study population (specifying the sampling framework, the sampling procedure, the subpopulations to be compared, and the inclusion and exclusion criteria)
   - Decide what to measure, when to measure, and how to measure
   - Design a statistical analysis plan that addresses the research question.
3. Conduct the research (carry out the observations and collect data).
4. Analyze the results:
   - Tabulate the frequency and distribution of the variables measured (univariate analysis, descriptive in nature)
   - Examine relationships between the variables (bivariate and multivariate analysis, either descriptive, or explanatory).
5. Interpret the results and draw conclusions.

The empirical cycle can therefore be expanded as shown in Figure 1.3. The results of a single study can rarely make or break a theory or hypothesis. A theory represents current knowledge on a particular subject at a particular time. Plenty of research is usually needed before a theory can gain a satisfactory empirical basis. In effect each study means a fresh circuit of the empirical cycle, or rather the empirical spiral, working toward an ever-better description of reality. See also Figure 1.4.

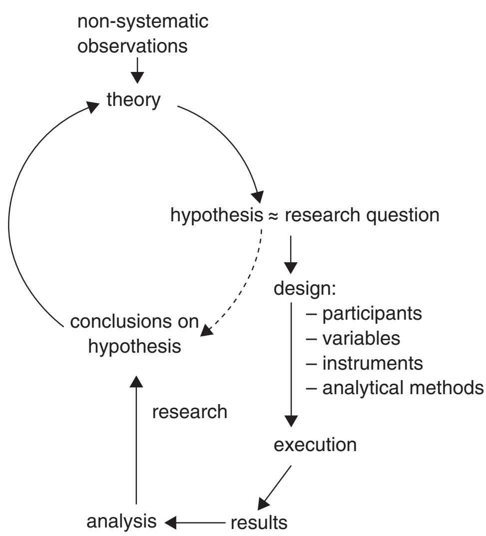

Figure 1.3 Research as an element in the empirical cycle.

Figure 1.4 Fokke and Sukke on the empirical cycle.

Figure 1.5 Hippocrates (circa 470–430 BC), epidemiologist *avant la lettre. Source: Welcome Images / Wikimedia Commons / CC BY 4.0.*

1.2 Developments in epidemiology

1.2.1 Historical examples

As an epidemiologist *avant la lettre*, Hippocrates (circa 470–430 BC) stressed the importance of describing cases of disease meticulously (Figure 1.5). Hippocrates demonstrated in his *book On Airs, Waters and Plac*es that health and disease are determined by all sorts of observable environmental factors. Subsequently, for more than two millennia, ideas on disease remained related to humoral pathology (water, earth, fire, and bile), with scarcely any attention being paid to empirical observation. Not until the end of the seventeenth century was the Hippocratic approach re-adopted, for example by the Italian clinician Bernardino Ramazzini (1633–1714), who stressed the importance of asking patients about their medical history, nutritional habits, and working conditions. He even demonstrated the added value of comparing similar cases and looking for patterns. The idea of recording, counting, and analyzing causes of death dates from the same period. The London physician John Graunt (1620–1674) is regarded as the founding father of this type of descriptive epidemiology (Figure 1.6).

The advent of empirical research, introduced by Galileo Galilei (1564–1642), William Harvey (1578–1657) and others, laid the foundation for the further

Figure 1.6 Captain John Graunt (1620–1674), founding father of descriptive epidemiology. Source: John Graunt / Wikimedia Commons / Public domain.

development of epidemiology. The systematic recording and analysis of causes of death yielded a wealth of information, on the basis of which researchers such as William Farr (1807–1883) presented impressive and highly relevant reports to the British Minister of Health. John Snow (1813–1858) studied the causes of cholera (see Case 1.3), presenting one of the first clear examples of the epidemiological approach (Figure 1.7).

In the eighteenth and early nineteenth century great progress was made in statistics, whose influence also grew in medical science. An important person in this regard is the French physician Pierre Louis (1787–1872), who introduced the "numerical method" in medicine and demonstrated with statistics that bloodletting was ineffective and even harmful. Empirical medical research became popular in the first half of the nineteenth century, an era with major developments in biology, pathology, and public health, and the first classifications of diseases as well. The German scientist Rudolf Virchow (1821–1902) played a vital role here, being not only an eminent pathologist but also a great advocate of a vigorous approach to public health problems. The work of the Frenchman Louis Pasteur (1822–1895) and the German Robert Koch (1843–1910) in the field of microorganisms uncovered the causes of many infectious diseases, including tuberculosis, which was widespread at that time. Thanks to Koch we have a first set of criteria for causality in epidemiological research. Around 1900, epidemiology was virtually synonymous with the epidemiology of infectious diseases. The epidemiological approach rose in popularity in the mid-twentieth century for the study of chronic conditions.

Present-day epidemiology has been highly influenced by the studies into the relationship between smoking and lung cancer carried out around World War II. Following initial reports of a possible link in the late 1930s, three studies published in 1950 showed that smoking is likely to be causally linked to the risk of lung cancer. One of these studies was carried out by Richard Doll and Austin Bradford Hill, two researchers who continued their studies into the harmful effects of smoking on health during subse-

Figure 1.7 John Snow (1813–1858), one of the founding fathers of analytical epidemiology. Source: Thomas Jones Barker / Wikimedia Commons / Public domain.

Case 1.3 Cholera and drinking water

The London physician John Snow is often named as one of the founding fathers of epidemiology. A cholera epidemic broke out in the London district of Soho in 1848, and 500 residents of a single neighborhood died of the disease within 10 days. Snow started surveying the cholera cases systematically. He suspected that there was a connection with drinking water, as none of the monks at a nearby monastery – who drank no water, only beer – had cholera. Snow then compared the water supply provided by the Southwark and Vauxhall Company with that provided by the Lambeth Company. He concluded that the victims must have mainly used the water pump in Broad Street. At Snow's insistence the pump was sealed off and the cholera epidemic ended. Only years later did it become clear that cholera is actually caused by a bacterium, but for Snow this was not an obstacle to taking adequate preventive action. Close study of his writings demonstrates that he was guided by an explicit theory about the etiology of cholera despite the fact that the biological mechanism was far from understood.

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine ; Honorary Director of the Statistical Research Unit of the Medical Research Council

Figure 1.8 Doll and Hills' famous article on smoking and lung cancer.

quent decades. In 1964, they published the results of a major study among British physicians that clearly showed a causal relationship (Figure 1.8) (see also Case 4.5 in Chapter 4). It is no coincidence that soon afterwards (in 1965) Hill published a revised set of criteria for causality, which were applicable both to infectious diseases (as were Koch's criteria mentioned above) and non-infectious diseases (see Chapter 6). From the particularly heated debate about the relationship between smoking and lung cancer in the 1950s and 1960s we can infer that lack of understanding of the exact mechanisms of disease is both a strength and a weakness of the epidemiological approach. This convincing epidemiological example has certainly raised interest in the contribution of epidemiology to the solution of many other public health problems.

In addition, clinicians became interested in the potential of epidemiology for solving diagnostic, prognostic, and treatment problems. This development gained a major boost from the publication in 1972 of the book "Effectiveness and Efficiency: Random Reflections on Health Services," in which Archie Cochrane (1909–1988) ardently advocated the systematic use of randomized controlled trials (RCTs) to assess the efficacy of interventions (Figure 1.9). Present-day epidemiology can be regarded as the combination of classic observational epidemiological methods with experimental study designs rooted in statistical theory developed since the 1930s.

Figure 1.9 Archie Cochrane (1909–1988), who introduced the use of randomized controlled trials in epidemiology. Source: Cardiff University Library, Cochrane Archive, University Hospital Llandough.

1.2.2 Epidemiology today

Modern epidemiology has two key purposes: the development of epidemiological research methods and the production of biomedical knowledge by the application of these methods. Both have gained in strength enormously since World War II. By comparing recent and early editions of medical handbooks, you will discover what an enormous contribution epidemiology has made to our knowledge and understanding of the distribution of, explanatory factors of, and intervention options for major diseases, including ischemic heart disease, asthma, various types of cancer, and HIV/AIDS. The first book on epidemiological methodology was published in 1960; nowadays there are enough such books to fill a large bookcase. Novel study designs and data analysis methods have been developed over the past few decades to meet the needs of epidemiological practice. The rapid developments in science and information technology are also contributing to the development of the field, in particular:

- The incorporation of modern molecular biology techniques in epidemiological research, enabling not only phenotypes but also genotypes to be studied on a large scale. New developments in immunology, molecular biology, and genetics are being incorporated in epidemiological research: for example, using biomarkers (e.g. protein adducts containing exogenous toxins) instead of questionnaires to measure exposure. The study of individual genetic susceptibility will contribute to identifying the most vulnerable groups.
- Progress in clinical medicine, with a host of new diagnostic and treatment options and the adoption of evidence-based medicine.
- The rise of e-epidemiology – the use of digital media such as the internet, wearables, and mobile phones in epidemiology: for example, data collection through portable devices, recruitment through social media, predicting flu epidemics by analyzing Google search behavior and Twitter activity, and measuring air quality, personal physical activity, and food intake using personal wearables. These new methods have many advantages and are increasingly being seen as fully fledged alternatives to traditional methods. Nevertheless, we need to evaluate the internal and external validity of the findings they produce, safeguard privacy, and manage ethical issues related to new ways of recruiting participants and collecting data.
- Data have grown in variety, volume, and velocity. Big data, when properly tackled using epidemiological methods have the potential to revolutionize how we identify and intervene on the explanatory factors of population health.
- Increasing awareness that human health depends to a large extent on the quality of the physical and social environment. That quality has deteriorated precisely due to human action; epidemiology is relevant for identifying and tackling the public health problems.
- Recently, the issues of health equity, social, and racial justice have been brought to the forefront of public health. Health equity is when all members of society have a fair and just opportunity to be as healthy as possible. This requires removing obstacles to health such as poverty, discrimination, and their consequences. Epidemiology can contribute to identifying important health disparities and implementing interventions, practices, and policies to reduce inequities.

1.3 Developments in the discipline itself

Specialization is growing in epidemiology, just as in every maturing scientific discipline. Content and methodological development are growing fast with separate attention being paid to improving methods for exposure monitoring, measurement quality, dose–response models, longitudinal designs, and so on. Epidemiologists are also specializing in specific topic areas, either focusing on a particular category of diseases (e.g. cancer epidemiology, infectious disease epidemiology, or the epidemiology of aging), a particular category of explanatory factors (e.g. nutritional epidemiology, genetic epidemiology, or pharmacoepidemiology), a particular application (e.g. clinical epidemiology or forensic epidemiology), or a particular method (e.g. prediction modeling, systematic reviews, or machine learning). These types of specialization are an inevitable consequence of the successful development of the discipline, but they also entail the risk of losing sight of the wood for the trees, with few scholars still being able to tackle a complex public health problem using a general approach, and not enough results of epidemiological research finding their way into healthcare practice.

Case 1.4 The third-generation oral contraceptive pill and the risk of venous thrombosis

Soon after the first oral contraceptives (OCP) were introduced in 1960, there were reports of increased risk of thrombotic events, resulting in the development of new types of contraceptive pills with a lower estrogen content (but containing more progestogens). The first-generation OCP was followed by a second and then a third generation, the aim being to reduce the risk of arterial thrombosis. An increased risk of venous thrombosis remained, however. The World Health Organization concluded in 1998 that users of modern oral contraceptives (containing 30–40 μg of estradiol) are at a three to six times higher risk of venous thrombosis than similar women who are not on the pill. It should be noted, however, that the absolute risk is still low, from less than one case of venous thrombosis per 10,000 person-years in women not using the pill to 3 to 6 cases per 10,000 person-years in women using it. A host of epidemiological studies – both cohort studies and case–control studies – have meanwhile been published comparing the risk of venous thrombosis from "third-generation" OCP with that from second-generation OCP. These show that venous thrombosis is 1.5–4.0 times more frequent in users of third-generation pills than in users of second-generation pills. The risk of venous thrombosis is highest in the first few years of using this type of pill, approximately one case per 1,000 in the first years of use. This is thought to be due to the progestogens in the third-generation OCP having an adverse effect on certain blood coagulation factors. The advice, based on these findings, is not to prescribe a third-generation OCP to new users of oral contraceptives.

Case 1.5 Prediction of live birth after IVF treatment

In-vitro fertilization (IVF) is a complex and costly assisted reproductive technology that involves a substantial emotional and physical burden for women and their families. Unfortunately, IVF treatment does not guarantee success. Couples would therefore like to know their prior probability of success before starting IVF treatment. The prediction models predict pregnancy followed by live birth after IVF are based on old data, and the model predictions often underestimate the clinical outcomes. These models include among other things the age of the women, previous pregnancies, and the use of donor cells. However, other factors associated with pregnancy outcome in women undergoing IVF, such as body-mass index (BMI), ovarian reserve, and ethnicity are not included in these models. Recently, data from a cohort study among 9915 women in the UK and Ireland derived and validated a novel prediction model of live birth for women undergoing IVF. The model includes age, duration of infertility, BMI, antral follicle count, previous miscarriage, previous live birth, cause of infertility and ethnicity. While the average probability of success in the cohort was 31.5%, individual predictions ranged between 18% and 41%, with overall good validity. This prediction model can be used to create a user-friendly decision aid to inform couples and their physicians about the likelihood of a successful pregnancy. The application of this new prediction model can increase the quality of the decisions and may reduce the number of unsuccessful IVF treatments.

Recommended reading

Bonita, R., Beaglehole, R., and Kjellstrom, T. (2006). *Basic Epidemiology*, 2e. Geneva: World Health Organization.

Carneiro, I. and Howard, N. (2005). *Introduction to Epidemiology*, 2e. Maidenhead: Open University Press.

Celentano, D.D., Szklo, M., and Gordis, L. (2019). *Gordis Epidemiology*, 6e. Philadelphia, PA: Elsevier. (Chapter 1).

Holland, W.W., Olsen, O., and du V Florey, C. (ed.) (2007). *The Development of Modern Epidemiology: Personal Reports from Those Who Were There*. Oxford: Oxford University Press.

Morabia, A. (2004). *A History of Epidemiologic Methods and Concepts*. Basel: Birkhäuser Verlag.

Szklo, M. and Nieto, F.J. (2019). *Epidemiology: Beyond the Basics*, 4e. Burlington, MA: Jones & Bartlett Learning. (Chapter 10).

Source references (cases)

Lind, J. (1957). *A Treatise of the Scurvy in Three Parts. Containing an Inquiry into the Nature, Causes and Cure of that Disease, Together with a Critical and Chronological View of What Has Been Published on the Subject*. London: A. Millar (Case 1.1).

LevRan, S., Roerecke, M., Le Foll, B. et al. (2014). The association between cannabis use and depression: a systematic review and meta-analysis of longitudinal studies. *Psychol. Med.* 44: 797–810. (Case 1.2).

Moitra, E., Bradley, J.A., and Stein, M.D. (2016). Reductions in cannabis use are associated with mood improvement in female emerging adults. *Depress. Anxiety* 33: 332–338. (Case 1.2).

Frost, W.H. (1936). *Snow on Cholera*. New York: The Commonwealth Fund (Case 1.3).

Vandenbroucke, J.P., Rosing, J., Bloemenkamp, K.W.M. et al. (2001). Oral contraceptives and the risk of venous thrombosis. *N. Engl. J. Med.* 344: 1527–1535. (Case 1.4).

Dhillon, R.K. (2016). Predicting the chance of live birth for women undergoing IVF: a novel pretreatment counselling tool. *Human Repr.* 31: 84–92. (Case 1.5).

Knowledge assessment questions

1. Are the following statements true or false?

1.1 Health outcomes are always binary (i.e. diseased or non-diseased).

1.2 Prevalence is a proportion.

1.3 Incidence is calculated using the number of new cases in the numerator.

1.4 All explanatory factors can be modified.

1.5 Deduction is the process by which one goes from theory to one or more hypotheses for testing.

2. Epidemiology, as defined in this chapter, would include which of the following activities? (select all that apply)
   A. Describe the demographic characteristics of persons with colorectal cancer in Denver, Colorado.
   B. Prescribe anti-hypertensive drugs to treat a patient with high blood pressure.
   C. Compare the demographic characteristics, family history, and environmental exposures of those with and without newly diagnosed glaucoma.
   D. Conduct a randomized study to compare two different surgical procedures for hip fractures.
3. John Snow's investigation of cholera is considered a model for epidemiological field investigations because it included a: (select all that apply)
   A. Biologically plausible hypothesis.
   B. Comparison of a health outcome among exposed and unexposed groups.
   C. Multivariate statistical modeling.
   D. Known biological mechanism.
   E. Recommendation for public health action.