An Introduction to Epidemiology

Wolfgang Ahrens, Klaus Krickeberg, Iris Pigeot

1 Epidemiology and Related Areas 3

Definition and Purpose of Epidemiology 3
Epidemiology in Relation to Other Disciplines 5
Overview 7

2 Development of Epidemiology 9

Historical Background 9
Milestones in Epidemiological Research 12
Methodological Limits 14

3 Concepts and Methodological Approaches in Epidemiology 16

Concepts 16
Study Designs 17
Data Collection 20

4 Statistical Methods in Epidemiology 21

Principles of Data Analysis 22
Statistical Thinking 23
Multivariate Analysis 24
Handling of Data Problems 27
Meta-Analysis 28

5 Applications of Epidemiological Methods and Research Areas in Epidemiology 29

Description of the Spectrum of Diseases 29
Identification of Causes of Disease 29

Application of Epidemiological Knowledge 32
Ethical Aspects 35

References 36

Epidemiology and Related Areas 1

Various disciplines contribute to the investigation of determinants of human health and disease, to the improvement of health care, and to the prevention of illness. These contributing disciplines stem from three major scientific areas, first from basic biomedical sciences such as biology, physiology, biochemistry, molecular genetics, and pathology, second from clinical sciences such as oncology, gynecology, orthopedics, obstetrics, cardiology, internal medicine, urology, radiology, and pharmacology, and third from public health sciences with epidemiology as their core.

Definition and Purpose of Epidemiology 1.1

One of the most frequently used definitions of epidemiology was given by MacMahon and Pugh (1970):

> Epidemiology is the study of the distribution and determinants of disease frequency in man.

The three components of this definition, i.e. frequency, distribution, and determinants embrace the basic principles and approaches in epidemiological research. The measurement of disease *frequency* relates to the quantification of disease occurrence in human populations. Such data are needed for further investigations of patterns of disease in subgroups of the population. This involves "… describing the *distribution* of health status in terms of age, sex, race, geography, etc., …" (MacMahon and Pugh 1970). The methods used to describe the distribution of diseases may be considered as a prerequisite to identify the *determinants* of human health and disease.

This definition is based on two fundamental assumptions: First, the occurrence of diseases in populations is not a purely random process, and second, it is determined by causal and preventive factors (Hennekens and Buring 1987). As mentioned above, these factors have to be searched for systematically in populations defined by place, time, or otherwise. Different ecological models have been used to describe the interrelationship of these factors, which relate to host, agent, and environment. Changing any of these three forces, which constitute the so-called epidemiological triangle (Fig. 1.1), will influence the balance among them and thereby increase or decrease the disease frequency (Mausner and Bahn 1974).

Thus, the search for etiological factors in the development of ill health is one of the main concerns of epidemiology. Complementary to the epidemiological triangle the triad of time, place, and person is often used by epidemiologists to describe the distribution of diseases and their determinants. Determinants that influence health may consist of behavioral, cultural, social, psychological, biological, or physical factors. The determinants by time may relate to increase/decrease

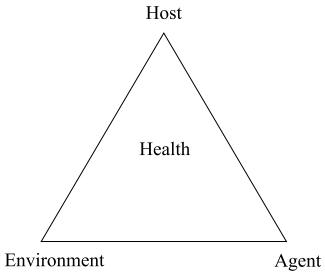


Figure 1.1. The epidemiological triangle

over the years, seasonal variations, or sudden changes of disease occurrence. Determinants by place can be characterized by country, climate zone, residence, and more general, by geographic region. Personal determinants include age, sex, ethnic group, genetic traits, and individual behavior. Studying the interplay between time, place, and person helps to identify the etiologic agent and the environmental factors as well as to describe the natural history of the disease, which then enables the epidemiologist to define targets for intervention with the purpose of disease prevention (Detels 2002). This widened perspective is reflected in a more comprehensive definition of epidemiology as given by Last (2001):

> The study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control of health problems.

In this broader sense, health-related states or events include "diseases, causes of death, behaviors such as use of tobacco, reactions to preventive regimens, and provision and use of health services" (Last 2001). According to this definition, the final aim of epidemiology is to promote, protect, and restore health. Hence, the major goals of epidemiology may be defined from two overlapping perspectives. The first is a biomedical perspective looking primarily at the etiology of diseases and the disease process itself. This includes

- the description of the disease spectrum, the syndromes of the disease and the disease entities to learn about the various outcomes that may be caused by particular pathogens,
- the description of the natural history, i.e. the course of the disease to improve the diagnostic accuracy which is a major issue in clinical epidemiology,
- the investigation of physiological or genetic variables in relation to influencing factors and disease outcomes to decide whether they are potential risk factors, disease markers or indicators of early stages of disease,
- the identification of factors that are responsible for the increase or decrease of disease risks in order to obtain the knowledge necessary for primary prevention,
- the prediction of disease trends to facilitate the adaptation of the health services to future needs and to identify research priorities,

- the clarification of disease transmission to control the spread of contagious diseases e.g. by targeted vaccination programs.

Achievement of these aims is the prerequisite for the second perspective, which defines the scope of epidemiology from a public health point of view. Especially in this respect, the statement as given in Box 1 was issued by the IEA (International Epidemiological Association) Conference already in 1975.

Box 1. Statement by IEA Conference in 1975 (White and Henderson 1976)

"The discipline of epidemiology, together with the applied fields of economics, management sciences, and the social sciences, provide the essential quantitative and analytical methods, principles of logical inquiry, and rules for evidence for:

- …;
- diagnosing, measuring, and projecting the health needs of community and populations;
- determining health goals, objectives and priorities;
- allocating and managing health care resources;
- assessing intervention strategies and evaluating the impact of health services."

This list may be complemented by the provision of tools for investigating consequences of disease as unemployment, social deprivation, disablement, and death.

Epidemiology in Relation to Other Disciplines 1.2

Biomedical, clinical and other related disciplines sometimes claim that epidemiology belongs to their particular research area. It is therefore not surprising that biometricians think of epidemiology as a part of biometry and physicians define epidemiology as a medical science. Biometricians have in mind that epidemiology uses statistical methods to investigate the distribution of health-related entities in populations as opposed to handling single cases. This perspective on distributions of events, conditions, etc. is statistics by its very nature. On the other hand, physicians view epidemiology primarily from a substantive angle on diseases and their treatment. In doing so, each of them may disregard central elements that constitute epidemiology.

Moreover, as described at the beginning, epidemiology overlaps with various other domains that provide their methods and knowledge to answer epidemiological questions. For example, measurement scales and instruments to assess subjective well-being developed by psychologists can be applied by epidemiologists to investigate the psychological effects of medical treatments in addition to classical clinical outcome parameters. Social sciences provide indicators and methods of field work that are useful in describing social inequality in health, in investigating social determinants of health, and in designing population-based

prevention strategies. Other examples are methods and approaches from demography that are used to provide health reports, from population genetics to identify hereditary factors, and from molecular biology to search for precursors of diseases and factors of susceptibility.

Of course, epidemiology does not only borrow methods from other sciences but has also its own methodological core. This pertains in particular to the development and adaptation of study designs. It is also true for statistical methods. In most cases they can directly be applied to epidemiological data, but sometimes peculiarities in the data structure may call for the derivation of special methods to cope with these requirements. This is in particular currently the case in genetic epidemiology when e.g. modeling gene-environment interactions is needed.

The borderline between epidemiology and related disciplines is often blurred. Let us take clinical medicine as an example. In clinical practice, a physician decides case-by-case to diagnose and treat individual patients. To achieve the optimal treatment for a given subject, he or she will classify this patient and then make use of knowledge on the group to which the person belongs. This knowledge may come from randomized clinical trials but also from (clinical) epidemiological studies. A randomized clinical trial is a special type of a randomized controlled trial (RCT). In a broad sense, a RCT is an epidemiological experiment in which subjects in a population are randomly allocated into groups, i.e. a study group where intervention takes place and a control group without intervention. This indicates an overlap between clinical and epidemiological studies, where the latter focus on populations while clinical trials address highly selected groups of patients. Thus, it may be controversial whether randomized clinical trials for drug approval (i.e. phase III trials) are to be considered part of epidemiology, but it is clear that a follow-up concerned with safety aspects of drug utilization (so-called phase IV studies) needs pharmacoepidemiological approaches.

When discussing the delimitation of epidemiology the complex area of public health plays an essential role. According to Last's definition (Last 2001) public health has to do with the health needs of the population as a whole, in particular the prevention and treatment of disease. More explicitly, "Public health is one of the efforts organized by society to protect, promote, and restore the people's health. It is the combination of sciences, skills, and beliefs that is directed to the maintenance and improvement of the health of all the people through collective or social actions. (…) Public health … goals remain the same: to reduce the amount of disease, premature death, and disease-produced discomfort and disability in the population. Public health is thus a social institution, a discipline, and a practice." (Last 2001). The practice of public health is based on scientific knowledge of factors influencing health and disease, where epidemiology is, according to Detels and Breslow (2002), "the core science of public health and preventive medicine" that is complemented by biostatistics and "knowledge and strategies derived from biological, physical, social, and demographic sciences".

In conclusion, epidemiology cannot be reduced to a sub-division of one of the contributing sciences but it should be considered as a multidisciplinary science giving input to the applied field of public health.

1.3 Overview

The present handbook intends to reflect all facets of epidemiology, ranging from basic principles (Part I) through statistical methods typically applied in epidemiological studies (Part II) to the majority of important applications (Part III) and to special fields of research (Part IV). Within these four parts, its structure is to a large extent determined by various natural subdivisions of the domain of epidemiology. These correspond mostly to the elements of the definition of epidemiology as given by Last and quoted above, namely study, distribution, determinants (factors, exposures, explanatory variables), health-related states or events (outcomes), populations, applications.

For instance, the concepts of a study and of determinants lead to the distinction of observational epidemiology on the one hand and experimental epidemiology on the other. In the first area, we study situations as they present themselves without intervening. In particular, we are interested in existing determinants within given populations. A typical example would be the investigation of the influence of a risk factor like air pollution on a health-related event like asthma. In experimental epidemiology, however, determinants are introduced and controlled by the investigator in populations which he or she defines by himself or herself, often by random allocation; in fact, experimental epidemiology is often simply identified with RCTs. Clinical trials to study the efficacy of the determinant "treatment" are a special type within this category. They are to be distinguished from trials of preventive interventions, another part of experimental epidemiology.

The idea of the purpose of a study gives rise to another, less clearly defined, subdivision, i.e. explanatory vs. descriptive epidemiology. The objective of an explanatory study is to contribute to the search of causes for health-related events, in particular by isolating the effects of specific factors. This causal element is lacking or at least not prominent in purely descriptive studies. In practice this distinction often amounts to different, and contrasting, sources of data: In descriptive epidemiology they are routinely registered for various reasons whereas in explanatory or analytic epidemiology they are collected for specific purposes. The expression "descriptive epidemiology" used to have a more restrictive, "classical" meaning that is also rendered by the term "health statistics" where as a rule the determinants are time, place of residence, age, gender, and socio-economic status.

"Exposure-oriented" and "outcome-oriented" epidemiology represent the two sides of the same coin. Insofar this distinction is more systematic rather than substantive. If the research question emphasizes disease determinants, e.g. environmental or genetic factors, the corresponding studies usually are classified as exposure-oriented. If, in contrast, a disease or another health-related event like lung cancer or osteoarthritis is the focus, we speak of "outcome-oriented" studies, in which risk factors for the specific disease are searched for. Finally, some subfields of epidemiology are defined by a particular type of application such as prevention, screening, and clinical epidemiology.

Let us now have a short look at the chapters of the handbook. Part I contains general concepts and methodological approaches in epidemiology: After introducing

the philosophical background and the conceptual building blocks of epidemiology such as models for causation and statistical ideas (Chap. I.1), Chap. I.2 deepens the latter aspect by giving an overview of various risk measures usually asked for in epidemiological studies. These measures depend heavily on the study type chosen for obtaining the data required to answer the research question. Various designs can be thought of to collect the necessary information. These are described in Chaps. I.3 to I.8. Descriptive studies and disease registries provide the basic information for health reporting. Experimental studies like cohort and case-control studies, modern study designs, and intervention trials serve to examine associations and hypothesized causal relationships. Chapter I.9 discusses in detail the two concepts of interaction and confounding, which are, on the one hand, very technical, but on the other hand fundamental for the analysis of any epidemiological study that involves several determinants. They allow us to describe the synergy of several factors and to isolate the effect of any of them. Chapters I.10 to I.13 concern practical problems to be handled when conducting an epidemiological study: field data collection in Chap. I.10, difficulties specific to exposure assessment in Chap. I.11, some key aspects of the planning of studies in general in Chap. I.12, and quality control and related aspects in Chap. I.13.

Due to the large variety of epidemiological issues, methodological approaches, and types of data, the arsenal of statistical concepts and methods to be found in epidemiology is also very broad. Chapter II.1 treats the question of how many units (people, communities) to recruit into a study in order to obtain a desired statistical precision. Chapter II.2 focuses on the analysis of studies where exposures and/or outcomes are described by continuous variables. Since the relationships between exposures and outcomes, which are the essence of epidemiology, are mostly represented by regression models it is not surprising that Chap. II.3 that is devoted to them is one of the longest of the whole handbook. Chapter II.4 discusses in detail the models used when the outcome variables are in the form of a waiting time until a specific event, e.g. death, occurs. Given that in practice data are often erroneous or missing, methods to handle the ensuing problems are presented in Chaps. II.5 and II.6. Meta-analysis is the art of drawing joint conclusions from the results of several studies together in order to put these conclusions on firmer ground, in particular, technically speaking, to increase their statistical power. It is the subject of Chap. II.7. The last chapter on statistical methodology, Chap. II.8, concerns the analysis of spatial data where the values of the principal explanatory variable are geographic locations. The topic of this chapter is closely related to the fields of application in Part III.

Although each epidemiological study contains its own peculiarities and specific problems related to its design and conduct, depending on the field of application, common features may be identified. Many important, partly classical, partly recent applications of epidemiology of general interest to public health are defined by specific exposures, and hence Part III starts with the presentation of the main exposure-oriented fields: social (III.1), occupational (III.2), environmental (III.3), nutritional (III.4), and reproductive epidemiology (III.5), but also more recent applications such as molecular (III.6) and genetic (III.7). Clinical epidemiology

(III.8) and pharmacoepidemiology (III.9) are large areas where knowledge about the interplay between many types of exposures, e.g. therapies, and many types of outcomes, usually diseases, is being exploited. A similar remark applies to the classical domains of screening in view of early detection of chronic diseases (Chap. III.10) and community-based health promotion, which mostly aims at prevention (Chap. III.11). These fields extend to public health research and build the bridge to the final part of this handbook.

Intensive research is going on in all of the foregoing areas, hence the selection of the topics for Part IV might appear a bit arbitrary, but in our opinion these seem to be currently the subject of particular efforts and widespread interest. The first four are outcome-oriented and deal with diseases of high public health relevance: infectious diseases (Chap. IV.1), cardiovascular diseases (Chap. IV.2), cancer (Chap. IV.3), and muscoloskeletal disorders (Chap. IV.4). The public health perspective is not restricted to these outcome-oriented research areas. The results of epidemiological studies may have a strong impact on political decisions and the health system, an area that is described for developed countries in Chap. IV.5. The particular problems related to health systems in developing countries and the resulting special demands for epidemiological research are addressed in Chap. IV.6. The handbook closes with the very important issue of human rights and responsibilities that have to be carefully considered at the different stages of an epidemiological study. These are discussed in Chap. IV.7 on ethical aspects.

Development of Epidemiology

2

Historical Background

2.1

The word "epidemic", i.e. something that falls upon people (ἐπί upon; δῆμος people), which was in use in ancient Greece, already reflected one of the basic ideas of modern epidemiology, namely to look at diseases on the level of *populations*, or *herds* as they also have been called, especially in the epidemiology of infectious diseases. The link with the search for *causes* of illness was present in early writings of the Egyptians, Jews, Greeks, and Romans (Bulloch 1938). Both Hippocrates (ca. 460–ca. 375 BC) and Galen (129 or 230–200 or 201) advanced etiological theories. The first stressed atmospheric conditions and "miasmata" but considered nutrition and lifestyle as well (Hippocrates 400 BC). The second distinguished three causes of an "epidemic constitution" in a population: an atmospheric one, susceptibility, and lifestyle. The basic book by Coxe (1846) contains a classification of Galen's writings by subject including the subject "etiology". For a survey on the various editions of Galen's work and a biography see the essay by Siegel (1968).

Regarding more specific observations, the influence of dust in quarries on chronic lung diseases was mentioned in a Roman text of the first century. Paracelsus in 1534 published the first treatise on *occupational diseases*, entitled "Von der Bergsucht" (On miners' diseases); see his biography in English by Pagel (1982).

Ramazzini (1713) conjectured that the relatively high incidence of breast cancer among nuns was due to celibacy. Sixty-two years later, Percival Pott (1775) was among the first ones to phrase a comparative observation in quantitative terms. He reported that scrotal cancer was very frequent among London chimney sweeps, and that their death rate due to this disease was more than 200 times higher than that of other workers.

The most celebrated early observational epidemiological study is that of John Snow on cholera in London in 1853. He was able to record the mortality by this disease in various places of residence under different conditions of water supply. And by comparison he concluded that deficient quality of water was indeed the cause of cholera (Snow 1855).

Parallel to this emergence of observational epidemiology, three more currents of epidemiological thinking have been growing during the centuries and interacted among them and with the former, namely the debate on *contagion* and *living causal agents*, *descriptive* epidemiology in the classical sense of health statistics, and *clinical trials*.

A contagion can be suspected from recording cases and their location in time, space, families, and the like. The possibility of its involvement in epidemics has therefore no doubt been considered since time immemorial; it was alluded to in the early writings mentioned at the beginning. Nevertheless, Hippocrates and Galen did not admit it. It played an important role in the thinking about *variolation*, and later on *vaccination* as introduced by Jenner in 1796 (Jenner 1798). The essay by Daniel Bernoulli on the impact of variolation (Bernoulli 1766) was the beginning of the theory of *mathematical modeling* of the spread of diseases.

By contrast to a contagion itself, the existence of *living* pathogens cannot be deduced from purely epidemiological observations, but the discussion around it has often been intermingled with that about contagion, and has contributed much to epidemiological thinking. Fracastoro (1521) wrote about a *contagium animatum*. In the sequel the idea came up again and again in various forms, e.g. in the writings of Snow. It culminated in the identification of specific parasites, fungi, bacteria, and viruses as agents in the period from, roughly, 1840 when Henle, after Arabian predecessors dating back to the ninth century, definitely showed that mites cause scabies, until 1984 when the HIV was identified.

As far as we know, the term "epidemiology" first appeared in Madrid in 1802. From the late 19th century to about the middle of the 20th, it was restricted to *epidemical infectious* diseases until it took its present meaning (see Sect. 2.2 and Greenwood 1932).

Descriptive epidemiology had various precursors, mainly in the form of church and military records on one hand (Marshall and Tulloch 1838), life tables on the other (Graunt 1662; Halley 1693). In the late 18th century, local medical statistics started to appear in many European cities and regions. They took a more systematic turn with the work of William Farr (1975). This lasted from 1837 when he was appointed to the General Register Office in London until his retirement in 1879. In particular, he developed classifications of diseases that led to the first International List of Causes of Death, to be adopted in 1893 by the International

Statistical Institute. Farr took also part in the activities of the London Epidemiological Society, founded in 1850 with him and Snow as founding members, and apparently the oldest learned society featuring the word "epidemiological" in its name.

Geographic epidemiology, i.e. the presentation of health statistics in the form of maps, also started in the 19th century (Rupke 2000).

If we mean by a clinical trial a *planned*, *comparative*, and *quantitative* experiment on humans in order to learn something about the efficacy of a curative or preventive treatment in a clinical setting, James Lind is considered having done the first one. In 1747 he tried out six different supplements to the basic diet of 12 sailors suffering from scurvy, and found that citrus fruits, and only these, cured the patients (Lind 1753). Later he also compared quinine to treat malaria with less well-defined control therapies (Lind 1771).

The first more or less rigorous trial of a preventive measure was performed by Jenner with 23 vaccinated people, but he still used what is now being called "historical controls," i.e. he compared these vaccinated people with unvaccinated ones of the past who had not been specially selected beforehand for the purpose of the trial (Jenner 1798).

In the 19th century some physicians began to think about the general principles of clinical trials and already emphasized probabilistic and statistical methods (Louis 1835; Bernard 1865). Some trials were done, for example on the efficacy of bloodletting to treat pneumonia, but rigorous methods in the modern sense were established only after World War II (see Sect. 2.2), beginning in 1948 with the pioneer trial on the treatment of pulmonary tuberculosis by streptomycin as described in Hill (1962).

Let us conclude this all too short historical sketch with a few remarks on the history of applications of epidemiology.

Clinical trials have always been tied, by their very nature, to immediate applications as in the above mentioned examples; hence we will not dwell on this anymore.

Observational epidemiology, including classical descriptive epidemiology, has led to hygienic measures. In fact, coming back to a concept of Galen (1951), one might define *hygiene* in a modern and general sense as applied observational epidemiology, its task being to diminish or to eliminate causal factors of any kind. For example, the results of Snow's study on cholera found rapid applications in London but not in places like Hamburg where 8600 people died in the cholera epidemic of 1892.

Hygiene was a matter of much debate and activity during the entire 19th century, although, before the identification of living pathogens, most measures taken were necessarily not directed against a known specific agent, with the exception of *meat inspection* for trichinae. This was made compulsory in Prussia in 1875 as proposed by Rudolf Virchow, one of the pioneers of modern hygiene and also an active politician (Ackerknecht 1953).

Hygienic activities generally had their epidemiological roots in the descriptive health statistics mentioned above. These statistics usually involved only factors like

time, place of residence, sex, and age, but Virchow, for example, analyzed during the years 1854–1871 the mortality statistics for the city of Berlin and tried to link those factors with social factors like poverty, crowded dwellings, and dangerous professions, thus becoming a forerunner of *social epidemiology*.

As a result of such reflections as well as of political pressure, large *sewage systems* were built in Europe and North America, the *refuse disposal* was reorganized and the *water supply* improved. Other hygienic measures concerned the structure and functioning of *hospitals*, from reducing the number of patients per room and dispersing wards in the form of pavilions to antiseptic rules. The latter had mainly been inspired by more or less precise epidemiological observations on infections after the treatment of wounds and amputations (Tenon 1788; Simpson 1868–1869, 1869–1870; Ackerknecht 1967), and on puerperal fever (Gordon 1795; Holmes 1842–1843; Semmelweis 1861).

2.2 Milestones in Epidemiological Research

The initiation of numerous epidemiological studies after the Second World War accelerated the research in this field and led to a systematic development of study designs and methods. In the following some exemplary studies are introduced that served as role models for the design and analysis of many subsequent investigations. It is not our intention to provide an exhaustive list of all major studies since that time, if at all feasible, but to exhibit some cornerstones marking the most important steps in the evolution of this science. Each of them had its own peculiarities with a high impact both on methods and epidemiological reasoning as well as on health policies.

The usefulness of descriptive study designs has been convincingly demonstrated by migrant studies comparing the incidence or mortality of a disease within a certain population between the country of origin and the new host country. Such observations offer an exceptional opportunity to distinguish between potential contributions of genetics and environment to the development of disease and thus make it possible to distinguish between the effects of nature and nurture. The most prominent examples are provided by investigations on Japanese migrants to Hawaii and California. For instance, the mortality from stomach cancer is much higher in Japan than among US inhabitants whereas for colon cancer the relationship is reversed. Japanese migrants living in California have a mortality pattern that lies between those two populations. It was thus concluded that dietary and other lifestyle factors have a stronger impact than hereditary factors, which is further supported by the fact that the sons of Japanese immigrants in California have an even lower risk for stomach cancer and a still higher risk for cancer of the colon than their fathers (Buell and Dunn 1965).

One of the milestones in the development of epidemiology was the case-control design, which facilitates the investigation of risk factors for chronic diseases with long induction periods. The most famous study of this type, although not the first, is the study on smoking and lung cancer by Doll and Hill (1950). As early

as 1943, the German pathologist Schairer published together with Schöniger from the Scientific Institute for Research into the Hazards of Tobacco, Jena, a case-control study comparing 109 men and women deceased from lung cancer with 270 healthy male controls as well as with 318 men and women who died from other cancers with regard to their smoking habits (Schairer and Schöniger 1943). Judged by modern epidemiological standards this study had several weaknesses, still, it showed a clear association of tobacco use and lung cancer. The case-control study by Doll and Hill was much more sophisticated in methodological terms. Over the whole period of investigation from 1948 to 1952 they recruited 1357 male and 108 female patients with lung cancer from several hospitals in London and matched them with respect to age and sex to the same number of patients hospitalized for non-malignant conditions. For each patient, detailed data on smoking history was collected. Without going into detail here, these data came up with a strong indication for a positive association between smoking and lung cancer. Despite the methodological concerns regarding case-control studies, Doll and Hill themselves believed that smoking was responsible for the development of lung cancer. The study became a landmark that inspired future generations of epidemiologists to use this methodology (cf. Chap. I.6 of this handbook). It remains to this day a model for the design and conduct of case-control studies, with excellent suggestions on how to reduce or eliminate selection, interview, and recall bias (cf. Chaps. I.9, I.10, I.12, I.13).

Because of the strong evidence they started a cohort study of 20,000 male British physicians in 1951, known as the British Doctors' Study. These were followed to further investigate the association between smoking and lung cancer. The authors compared mortality from lung cancer among those who never smoked with that among all smokers and with those who smoked various numbers of cigarettes per day (Doll and Hill 1954, 1964; Doll and Peto 1978).

Another, probably even more important cohort study was the Framingham Heart Study that was based on the population of Framingham, a small community in Massachusetts. The study was initiated in 1949 to yield insights into causes of cardiovascular diseases (CVD) (see Chap. IV.2 of this handbook). For this purpose, 5127 participants free from coronary heart disease (CHD), 30 to 59 years of age, were examined and then followed for nearly 50 years to determine the rate of occurrence of new cases among persons free of disease at first observation (Dawber et al. 1951; Dawber 1980). The intensive biennial examination schedule, long-term continuity of follow-up and investigator involvement, and incorporation of new design components over its decades-long history have made this a uniquely rich source of data on individual risks of CVD events. The study served as a reference and good example for many subsequent cohort studies in this field adopting its methodology. In particular, analysis of these data led to the development of the perhaps most important modeling technique in epidemiology, the multiple logistic regression (Truett et al. 1967; see Chap. II.3).

Two other leading examples of cohort studies conducted within a single population or for comparison of multiple populations to assess risk factors for cardio-

vascular events are the Whitehall Study of British civil servants (Rose and Shipley 1986; see also Chap. III.1) and the Seven Countries Study of factors accounting for differences in CHD rates between populations of Europe, Japan, and North America (Keys 1980; Kromhout et al. 1995; see Chap. IV.2).

In contrast to the above cohort studies that focused on cardiovascular diseases the U.S. Nurses' Health Study is an impressive example of a multipurpose cohort study. It recruited over 120,000 married female nurses, 30 to 55 years of age, in a mail survey in 1976. In this survey, information on demographic, reproductive, medical and lifestyle factors was obtained. Nurses were contacted every two years to assess outcomes that occurred during that interval and to update and to supplement the exposure information collected at baseline. Various exposure factors like use of oral contraceptives, post-menopausal hormone therapy, and fat consumption were related to different outcomes such as cancer and cardiovascular disease (Lipnick et al. 1986; Willett et al. 1987; Stampfer et al. 1985). The most recent results have had an essential impact on the risk-benefit assessment of post-menopausal hormone therapy speaking against its use over extended periods (Chen et al. 2002).

Final proof of a causal relationship is provided by experimental studies, namely intervention trials. The most famous and largest intervention trial was the so-called Salk vaccine field trial in 1954 where nearly one million school children were randomly assigned to one of two groups, a vaccination group that received the active vaccine and a comparison group receiving placebo. A 50 percent reduction of the incidence of paralytic poliomyelitis was observed in the vaccination group as compared to the placebo group. This gave the basis for the large-scale worldwide implementation of poliomyelitis vaccination programs for disease prevention.

In recent years, the accelerated developments in molecular biology were taken up by epidemiologists to measure markers of exposure, early biological effects, and host characteristics that influence response (susceptibility) in human cells, blood, tissue and other material. These techniques augment the standard tools of epidemiology in the investigation of low-level risks, risks imposed by complex exposures, and the modification of risks by genetic factors. The use of such biomarkers of exposure and effect has led to a boom of the so-called molecular epidemiology (Schulte and Perera 1998; Toniolo et al. 1997; Chap. III.6 of this handbook), a methodological approach with early origins. These developments were accompanied by the sequencing of the human genome and the advances in high-throughput genetic technologies that led to the rapid progress of genetic epidemiology (Khoury et al. 1993; Chap. III.7 of this handbook). The better understanding of genetic factors and their interaction with each other and with environmental factors in disease causation is a major challenge for future research.

2.3 Methodological Limits

The successes of epidemiology in identifying major risk factors of chronic diseases have been contrasted with many more subtle risks that epidemiologists have

seemingly discovered. Such risks are difficult to determine and false alarms may result from chance findings. Thus it is not surprising that in recent years many studies showed conflicting evidence, i.e. some studies seem to reveal a significant association while others do not. The uncritical publication of such contradictory results in the lay press leads to opposing advice and thus to an increasing anxiety in the public. This has given rise to a critical debate about the methodological weaknesses of epidemiology that culminated in the article "Epidemiology faces its limits" by Taubes (1995) and the discussions that it prompted.

In investigating low relative risks, say, below 2 or even below 1.5, the methodological shortcomings inherent in observational designs become more serious. Such studies are more prone to yield false positive or false negative findings due to the distorting effects of misclassification, bias, and confounding (see Chaps. I.9 and II.5 of this handbook). For instance, the potential effect of environmental tobacco smoke (ETS) on lung cancer was denied because misclassification of only a few active smokers as non-smokers would result in relative risks that might explain all or most of the observed association between ETS and the risk of lung cancer in non-smokers (Lee and Forey 1996). Validation studies showed that this explanation was unlikely (Riboli et al. 1990; Wells et al. 1998). Thus, the numerous positive findings and the obvious biological plausibility of the exposure-disease relationship support the conclusion of a harmful effect of ETS (Boffetta et al. 1998; Chan-Yeung and Dimich-Ward 2003; IARC Monograph on ETS 2004). This example also illustrates that the investigation of low relative risks is not an academic exercise but may be of high public health relevance if a large segment of the population is exposed.

It is often believed that large-scale studies are needed to identify small risks since such studies result in narrower confidence intervals. However, a narrow confidence interval does not necessarily mean that the overwhelming effects of misclassification, bias and confounding are adequately controlled by simply increasing the size of a study. Even sophisticated statistical analyses will never overcome serious deficiencies of the data base. The fundamental quality of the data collected or provided for epidemiological purposes is therefore the cornerstone of any study and needs to be prioritized throughout its planning and conduct (see Chap. I.13). In addition, refinement of methods and measures involving all steps from design over exposure and outcome assessment to the final data analysis, incorporating e.g. molecular markers, may help to push the edge of what can be achieved with epidemiology a little bit further.

Nevertheless a persistent problem is "The pressures to publish inconclusive results and the eagerness of the press to publicize them …" (Taubes 1995). This pressure to publish positive findings that are questionable imposes a particular demand on researchers not only to report and interpret study results carefully in peer reviewed journals but also to communicate potential risks also to the lay press in a comprehensible manner that accounts for potential limitations. Both authors and editors have to take care that the pressure to publish does not lead to a publication bias favoring positive findings and dismissing negative results.

3 Concepts and Methodological Approaches in Epidemiology

Extending the basic ideas of epidemiology presented above together with its definition, its scope and approaches will now be described further.

3.1 Concepts

Epidemiology may be considered as minor to physical sciences because it does not investigate the biological mechanism leading from exposure to disease as, e.g., toxicology does. However, the ability of identifying modifiable conditions that contribute to health outcome without also identifying the biological mechanism or the agent(s) that lead to this outcome is a strength of epidemiology: It is not always necessary to wait until this mechanism is completely understood in order to facilitate preventive action. This is illustrated by the historical examples of the improvements of environmental hygiene that led to a reduction of infectious diseases like cholera, that was possible before the identification of vibrio cholerae.

What distinguishes epidemiology is its perspective on groups or populations rather than individuals. It is this demographic focus where statistical methods enter the field and provide the tools needed to compare different characteristics relating to disease occurrence between populations. Epidemiology is a comparative discipline that contrasts diseases and characteristics relative to different time periods, different places or different groups of persons. The comparison of groups is a central feature of epidemiology, be it the comparison of morbidity or mortality in populations with and without a certain exposure or the comparison of exposure between diseased subjects and a control group. Inclusion of an appropriate reference group (non-exposed or non-diseased) for comparison with the group of interest is a condition for causal inference.

In experimental studies efficient means are available to minimize the potential for bias. Due to the observational nature of the vast majority of epidemiological studies bias and confounding are the major problems that may restrict the interpretation of the findings if not adequately taken into account (see Chaps. I.9 and I.12 of this handbook). Although possible associations are analyzed and reported on a group level it is important to note that only data that provide the necessary information on an individual level allow the adequate consideration of confounding factors (see Chap. I.3).

Most epidemiological studies deal with mixed populations. On the one hand, the corresponding heterogeneity of covariables may threaten the internal validity of a study, because the inability to randomize in observational studies may impair the comparability between study subjects and referents due to confounding. On the other hand the observation of "natural experiments" in a complex mixture of individuals enables epidemiologists to make statements about the real world and thus contributes to the external validity of the results. This population perspective

focuses epidemiology on the judgment of effectiveness rather than efficacy, e.g. in the evaluation of interventions.

Due to practical limitations, in a given study it may not be feasible to obtain a representative sample of the whole population of interest. It may even be desired to investigate only defined subgroups of a population. Whatever the reason, a restriction on subgroups may not necessarily impair the meaning of the obtained results; it may still increase the internal validity of a study. Thus, it is a misconception that the cases always need to be representative of all persons with the disease and that the reference group always should be representative of the general non-diseased population. What is important is a precise definition of the population base, i.e., in a case-control study, cases and controls need to originate from the same source population and the same inclusion/exclusion criteria need to be applied to both groups. This means that any interpretation that extends beyond the source population has to be aware of a restricted generalizability of the findings.

Rarely a single positive study will provide sufficient evidence to justify an intervention. Limitations inherent in most observational studies require the consideration of alternative explanations of the findings and confirmation by independent evidence from other studies in different populations before preventive action is recommended with sufficient certainty. The interpretation of negative studies deserves the same scrutiny as the interpretation of positive studies. Negative results should not hastily be interpreted to prove an absence of the association under investigation (Doll and Wald 1994). Besides chance, false negative results may easily be due to a weak design and conduct of a given study.

Study Designs 3.2

Epidemiological reasoning consists of three major steps. First, a statistical association between an explanatory characteristic (exposure) and the outcome of interest (disease) is established. Then, from the pattern of the association a hypothetical (biological) inference about the disease mechanism is formulated that can be refuted or confirmed by subsequent studies. Finally, when a plausible conjecture about the causal factor(s) leading to the outcome has been acknowledged, alteration or reduction of the putative cause and observation of the resulting disease frequency provide the verification or refutation of the presumed association.

In practice these three major steps are interwoven in an iterative process of hypothesis generation by descriptive and exploratory studies, statistical confirmation of the presumed association by analytical studies and, if feasible, implementation and evaluation of intervention activities, i.e. experimental studies. An overview of the different types of study and some common alternative names are given in Table 1.1.

A first observation of a presumed relationship between exposure and disease is often done at the group level by correlating one group characteristic with an outcome, i.e. in an attempt to relate differences in morbidity or mortality of population groups to differences in their local environment, living habits or other factors. Such correlational studies that are usually based on existing data (see

Table 1.1. Types of epidemiological studies

Type of study	Alternative name	Unit of study
Observational		
Ecological	Correlational	Populations
Cross-sectional	Prevalence; survey	Individuals
Case-control	Case-referent	Individuals
Cohort	Follow-up	Individuals
Experimental	*Intervention studies*	
Community trials	Community intervention studies	Communities
Field trials		Healthy individuals
Randomized controlled trials	RCT	Individuals
Clinical trials	Therapeutic studies[a]	Individual patients

[a] Clinical trials are included here since conceptually they are linked to epidemiology, although they are often not considered as epidemiological studies. Clinical trials have developed into a vast field of its own because of methodological reasons and their commercial importance.

Chap. I.4) are prone to the so-called "ecological fallacy" since the compared populations may also differ in many other uncontrolled factors that are related to the disease. Nevertheless, ecological studies can provide clues to etiological hypotheses and may serve as a gateway towards more detailed investigations. In such studies the investigator determines whether the relationship in question is also present among individuals, either by asking whether persons with the disease have the characteristic more frequently than those without the disease, or by asking whether persons with the characteristic develop the disease more frequently than those not having it. The investigation of an association at the individual level is considered to be less vulnerable to be mixed up with the effect of a third common factor. For a detailed discussion of this issue we refer to Sect. 4.2.5 of Chap. I.3 of this handbook.

Studies that are primarily designed to describe the distribution of existing variables that can be used for the generation of broad hypotheses are often classified as descriptive studies (cf. Chap. I.3 of this handbook). Analytical studies examine an association, i.e. the relationship between a risk factor and a disease in detail and conduct a statistical test of the corresponding hypothesis. Typically the two main types of epidemiological studies, i.e. case-control and cohort, belong to this category (see Chaps. I.5 and I.6 of this handbook). However, a clear-cut distinction between analytical and descriptive study designs is not possible. A case-control study may be designed to explore associations of multiple exposures with a disease. Such "fishing expeditions" may better be characterized as descriptive rather than analytical studies. A cross-sectional study is descriptive when it surveys a community to determine the health status of its members. It is analytic when the association of an acute health event with a recent exposure is analyzed.

Cross-sectional studies provide descriptive data on prevalence of diseases useful for health care planning. Prevalence data on risk factors from descriptive studies also help in planning an analytical study, e.g. for sample size calculations. The design is particularly useful for investigating acute effects but has significant drawbacks in comparison to longitudinal designs because the temporal sequence between exposure and disease usually cannot be assessed with certainty, except for invariant characteristics like blood type. In addition, it cannot assess incident cases of a chronic disease (see Chap. I.3). Both case-control and cohort studies are in some sense longitudinal because they incorporate the temporal dimension by relating exposure information to time periods that are prior to disease occurrence. These two study types – in particular when data are collected prospectively – are therefore usually more informative with respect to causal hypotheses than cross-sectional studies because they are less prone to the danger of "reverse causality" that may emerge when information on exposure and outcome relates to the same point in time. The best means to avoid this danger are prospective designs where the exposure data are collected prior to disease. Typically these are cohort studies, either concurrent or historical, as opposed to retrospective studies, i.e. case-control studies where information on previous exposure is collected from diseased or non-diseased subjects. For further details of the strengths and weaknesses of the main observational designs see Chap. I.12 of this handbook.

The different types of studies are arranged in Table 1.2 in ascending order according to their ability to corroborate the causality of a supposed association. The criteria summarized by Hill (1965) have gained wide acceptance among epidemiologists as a checklist to assess the strength of the evidence for a causal relationship. However, an uncritical accumulation of items from such a list cannot replace the critical appraisal of the quality, strengths and weaknesses of each study. The weight of evidence for a causal association depends in the first place – at least in part – on the type of study, with intervention studies on the top of the list (Table 1.2) (see Chap. I.8). The assessment of causality has then to be based on a critical judgment of evidence by conjecture and refutation (see Chap. I.1 for a discussion of this issue).

Table 1.2. Reasoning in different types of epidemiological study

Study type	Reasoning
Ecological	Descriptive; association on group level may be used for development of broad hypotheses
Cross-sectional	Descriptive; individual association may be used for development and specification of hypotheses
Case-control	Increased prevalence of risk factor among diseased may indicate a causal relationship
Cohort	Increased risk of disease among exposed indicates a causal relationship
Intervention	Modification (reduction) of the incidence rate of the disease confirms a causal relationship

3.3 Data Collection

Data are the foundation of any empirical study. To avoid any sort of systematic bias in the planning and conduct of an epidemiological study is a fundamental issue, be it information or selection bias. Errors that have been introduced during data collection can in most cases not be corrected later on. Exceptions from this rule are for example measurement instruments yielding distorted measurements where the systematic error becomes apparent so that the individual measurement values can be adjusted accordingly. In other instances statistical methods are offered to cope with measurement error (see Chap. II.5). However, such later efforts are second choice and an optimal quality of the original data must be the primary goal. Selection bias may be even worse as it cannot be controlled for and may affect both the internal and the external validity of a study. Standardized procedures to ensure the quality of the original data to be collected for a given study are therefore crucial (see Chap. I.13).

Original data will usually be collected by questionnaires, the main measurement instrument in epidemiology. Epidemiologists have neglected for a long time the potential in improving the methods for interviewing subjects in a highly standardized way and thus improving the validity and reliability of this central measurement tool. Only in the last decade it has been recognized that major improvements in this area are not only necessary but also possible, e.g. by adopting methodological developments from the social sciences (Olsen et al. 1998). Chapter I.10 of this handbook is devoted to the basic principles and approaches in this field. Prior to the increased awareness related to data collection methods, the area of exposure assessment has developed into a flourishing research field that provided advanced tools and guidance for researchers (Armstrong et al. 1992; Kromhout 1994; Ahrens 1999; Nieuwenhuijsen 2003; Chap. I.11 of this handbook). Recent advances in molecular epidemiology have introduced new possibilities for exposure measurement that are now being used in addition to the classical questionnaires. However, since the suitability of biological markers for the retrospective assessment of exposure is limited due to the short half-life of most agents that can be examined in biological specimens, the use of interviews will retain its importance but will change its face. Computer-aided data collection with built-in plausibility-checks – that is more and more being conducted in the form of telephone interviews or even using the internet – will partially replace the traditional paper and pencil approach.

Often it may not be feasible to collect primary data for the study purpose due to limited resources or because of the specific research question. In such cases, the epidemiologist can sometimes exploit existing data bases such as registries (see Chap. I.4). Here, he or she usually has to face the problem that such "secondary data" may have been collected for administrative or other purposes. Looking at the data from a research perspective often reveals inconsistencies that had not been noticed before. Since such data are collected on a routine basis without the claim for subsequent systematic analyses they may be of limited quality. The degree of standardization that can be achieved in collection, doc-

umentation, and storage is particularly low if personnel of varying skills and levels of training is involved. Moreover, changes in procedures over time may introduce additional systematic variation. Measures for assessing the usefulness and quality of the data and for careful data cleaning are then of special importance.

The scrutiny, time and effort that need to be devoted to any data, be it routine data or newly collected data, before it can be used for data analysis are rarely addressed in standard textbooks of epidemiology and often neglected in study plans. This is also true for the coding of variables like diseases, pharmaceuticals or job titles. They deserve special care with regard to training and quality assurance. Regardless of all efforts to ensure an optimal quality during data collection, a substantial input is needed to guarantee standardized and well documented coding, processing, and storing of data. Residual errors that remain after all preceding steps need to be scrutinized and, if possible, corrected (see Chap. I.13). Sufficient time has to be allocated for this workpackage that precedes the statistical analysis and publication of the study results. Finally, all data and study materials have to be stored and documented in a fashion that allows future use and/or sharing of the data or auditing of the study. Materials to be archived should not only include the electronic files of raw data and files for the analyses, but also the study protocol, computer programs, the documentation of data processing and data correction, measurement protocols, and the final report. Both, during the conduct of the study as well as after its completion, materials and data have to be stored in a physically safe place with limited access to ensure safety and confidentiality even if the data have been anonymized.

Statistical Methods in Epidemiology

The statistical analysis of an empirical study relates to all its phases. It starts at the planning phase where ideally all details of the subsequent analysis should be fixed (see Chap. I.12 of this handbook). This concerns defining the variables to be collected and their scale, the methods how they should be summarized e.g. via means, rates or odds, the appropriate statistical models to be used to capture the relationship between exposures and outcomes, the formulation of the research questions as statistical hypotheses, the calculation of the necessary sample size based on a given power or vice versa the power of the study based on a fixed sample size, and appropriate techniques to check for robustness and sensitivity. It is crucial to have in mind that the study should be planned and carried out in such a way that its statistical analysis is able to answer the research questions we are interested in. If the analysis is not already adequately accounted for in the planning phase or if only a secondary analysis of already existing data can be done, the results will probably be of limited validity and interpretability.

4.1 Principles of Data Analysis

Having collected the data, the first step of a statistical analysis is devoted to the cleaning of the data set. Questions to be answered are: "Are the data free of measurement or coding errors?" "Are there differences between centers?" "Are the data biased, already edited or modified in any way?" "Have data points been removed from the data set?" "Are there outliers or internal inconsistencies in the data set?" A sound and thorough descriptive analysis enables the investigator to inspect the data. Cross-checks based e.g. on the range of plausible values of the variables and cross-tabulations of two or more variables have to be carried out to find internal inconsistencies and implausible data. Graphical representations such as scatter plots, box plots, and stem-and-leaf diagrams help to detect outliers and irregularities. Calculating various summary statistics such as mean compared to median, standard deviation compared to median absolute deviation from the median is also reasonable to reveal deficiencies in the data. Special care has to be taken to deal with measurement errors and missing values. In both cases, statistical techniques are available to cope with such data (see Chaps. II.5 and II.6 of this handbook).

After having cleaned the data set, descriptive measures such as correlation coefficients or graphical representations will help the epidemiologist to understand the structure of the data. Such summary statistics need, however, to be interpreted carefully. They are descriptive by their very nature and are not to be used to formulate statistical hypotheses that are subsequently investigated by a statistical significance test based on the same data set.

In the next step parameters of interest such as relative risks or incidences should be estimated. The calculated point estimates should always be supplemented by their empirical measures of dispersion like standard deviations and by confidence intervals to get an idea about their stability or variation, respectively. In any case, confidence intervals are more informative than the corresponding significance tests. Whereas the latter just lead to a binary decision, a confidence interval also allows the assessment of the uncertainty of an observed measure and of its relevance for epidemiological practice. Nevertheless, if p-values are used for exploratory purposes, they can be considered as an objective measure to "decide" on the meaning of an observed association without declaring it as "statistically significant" or "non-significant". In conclusion, Rothman and Greenland (1998, p. 6) put it as follows: "The notion of statistical significance has come to pervade epidemiological thinking, as well as that of other disciplines. Unfortunately, statistical hypothesis testing is a mode of analysis that offers less insight into epidemiological data than alternative methods that emphasize estimation of interpretable measures."

Despite the justified condemnation of the uncritical use of statistical hypothesis tests, they are widely used in the close to final step of an analysis to confirm or reject postulated research hypotheses (cf. the next section). More sophisticated techniques such as multivariate regression models are applied in order to describe the functional relationship between exposures and outcome (see Chaps. II.2

and II.3). Such techniques are an important tool to analyze complex data but as it is the case with statistical tests their application might lead to erroneous results if carried out without accounting for the epidemiological context appropriately. This, of course, holds for any statistical method. Its blind use may be misleading with possibly serious consequences in practice. Therefore, each statistical analysis should be accompanied by sensitivity analyses and checks for model robustness. Graphical tools such as residual plots, for instance, to test for the appropriateness of a certain statistical model should also routinely be used.

The final step concerns the adequate reporting of the results and their careful interpretation. The latter has to be done with the necessary background information and substantive knowledge about the investigated epidemiological research field.

Statistical Thinking 4.2

According to the definitions quoted in Sect. 1.1, epidemiology deals with the distribution and determinants of health-related phenomena in *populations* as opposed to looking at *individual* persons or cases. Studying distributions and their determinants in populations in a quantitative way is the very essence of statistics. In this sense, epidemiology means statistical thinking in the context of health including the emphasis on causal analysis as described in Chap. I.1 and the manifold applications to be found all-over in this handbook. However, this conception of epidemiology has started to permeate the field relatively late, and, at the beginning, often unconsciously.

The traditional separation of statistics into its descriptive and its inferential component has existed in epidemiology until the two merged conceptually though not organizationally. The *descriptive* activities, initiated by people like Farr (see Sect. 2.1) continue in the form of *health statistics*, *health yearbooks* and similar publications by major hospitals, some research organizations, and various health administrations like national Ministries of Health and the World Health Organization, often illustrated by graphics. The visual representation of the geographic distribution of diseases has recently taken an upsurge with the advent of *geographical information systems* (Chap. II.8).

Forerunners of the use of *inferential* statistics in various parts of epidemiology are also mentioned in Sect. 2.1. Thus, in the area of *clinical trials*, the efficacy of citrus fruit to cure scurvy was established by purely statistical reasoning. In the realm of *causal factors* for diseases the discovery of water contamination as a factor for cholera still relied on quite rudimentary statistical arguments whereas the influence of the presence of a doctor at child birth on maternal mortality was confirmed by a quantitative argument coming close to a modern test of significance. The basic idea of statistics that one needs to *compare frequencies* in populations with different levels of the factors (or "determinants") to be studied was already present in all of these early investigations. The same is true for statistics in the domain of *diagnosis* where statistical thinking expresses itself by concepts like *sensitivity* or *specificity* of a medical test although it seems that this was only recently

conceived of as a branch of epidemiology on par with the others, indispensable in particular for developing areas like computer-aided diagnosis or tele-diagnosis.

The big "breakthrough" of statistical thinking in epidemiology came after the elaboration of the theory of *hypothesis testing* by Neyman and Pearson. No self-respecting physician wrote any more a paper on health in a population without testing some hypotheses on the significance level 5% or without giving a p-value. Most of these hypotheses were about the efficacy of a curative treatment or, to a lesser degree, the etiology of an ailment, but the efficacy of preventive treatments and diagnostic problems were also concerned.

However, the underlying statistical thinking was often deficient. Non-acceptance of the alternative hypothesis was frequently regarded as acceptance of the null hypothesis. The meaning of an arbitrarily chosen significance level or of a p-value was not understood, and in particular several simultaneous trials or trials on several hypotheses at a time were not handled correctly by confusing the significance level of each part of the study with the overall significance level. Other statistical procedures that usually provide more useful insights like *confidence bounds* were neglected. Above all, *causal interpretations* were often not clear or outright wrong and hence erroneous practical conclusions were drawn. A statistical result in the form of a hypothesis accepted either by a test or by a correlation coefficient far from 0 was regarded as final evidence and not as one element that should lead to further investigations, usually of a biological nature.

Current statistical thinking expresses itself mainly in the study of the effect of several factors on a health phenomenon, be it a causal effect in etiologic research (Chap. I.1), a curative or preventive effect in clinical or intervention trials (Chaps. I.8, III.8, III.9, and IV.1), or the effect of a judgment, e.g. a medical test or a selection of people in diagnosis and screening (Chaps. III.8 and III.10). Such effects are represented in quantitative, statistical terms, and relations between the action of several factors as described by the concepts of interaction and confounding play a prominent role (Chap. I.9). The use of modern statistical ideas and tools has thus allowed a conceptual and practical *unification* of the many parts of epidemiology. The *same* statistical models and methods of analysis (Chaps. II.1 to II.8) are being used in all of them. Let us conclude with a final example of this global view. The concept of the etiological fraction (Chap. I.2) may represent very different things in different contexts: In causal analysis it is the fraction of all cases of a disease *due* to a particular factor whereas in the theory of prevention it means the fraction of all cases prevented by a particular measure, the most prominent application being the *efficacy of a vaccination* in a given population.

4.3 Multivariate Analysis

An epidemiological study typically involves a huge number of variables to be collected from the study participants, which implies a high-dimensional data set that has to be appropriately analyzed to extract the essential information. This curse

of dimensionality becomes especially serious in genetic or molecular epidemiological studies due to genetic and familial information obtained from the study subjects. In such situations, statistical methods are called for to reduce the dimensionality of the data and to reveal the "true" underlying association structure. Various multivariate techniques are at hand depending on the structure of the data and the research aim. They can roughly be divided into two main groups. The first group contains methods to structure the data set without distinguishing response and explanatory variables, whereas the second group provides techniques to model and test for postulated dependencies. Although these multivariate techniques seem to be quite appealing at first glance they are not a statistical panacea. Their major drawback is that they cannot be easily followed by the investigator which typically leads to a less deep understanding of the data. This "black box" phenomenon also implies that the communication of the results is not as straightforward as it is when just showing some well-known risk measures supplemented by frequency tables. In addition, the various techniques will usually not lead to a unique solution where each of those obtained from the statistical analysis might be compatible with the observed data. Thus, a final decision on the underlying data structure should not be made without critically reflecting the results based on the epidemiological context, on additional substantive knowledge, and on simpler statistical analyses such as stratified analyses perhaps restricted to some key variables that hopefully support the results obtained from the multivariate analysis.

Multivariate analyses with the aim to structure the data set comprise factor analysis and cluster and discriminant analysis. Factor analysis tries to collapse a large number of observed variables into a smaller number of possibly unobservable, i.e. latent variables, so-called factors, e.g. in the development of scoring systems. These factors represent correlated subgroups of the original set. They serve in addition to estimate the fundamental dimensions underlying the observed data set. Cluster analysis simply aims at detecting highly interrelated subgroups of the data set, e.g. in the routine surveillance of a disease. Having detected certain subgroups of, say, patients, their common characteristics might be helpful e.g. to identify risk factors, prevention strategies or therapeutic concepts. This is distinct to discriminant analysis, which pertains to a known number of groups and aims to assign a subject to one of these groups (populations) based on certain characteristics of this subject while minimizing the probability of misclassification. As an example, a patient with a diagnosis of myocardial infarction has to be assigned to one of two groups, one consisting of survivors of such an event and the other consisting of non-survivors. The physician may then measure his/her systolic and diastolic blood pressure, heart rate, stroke index, and mean arterial pressure. With these data the physician will be able to predict whether or not the patient will survive. A more detailed discussion of these techniques would be beyond the scope of this handbook. We refer instead to classical text books in this field such as Dillon and Goldstein (1984), Everitt and Dunn (2001), and Giri (2004).

However, in line with the idea of epidemiology, epidemiologists are mostly not so much interested in detecting a structure in the data set but in explaining

the occurrence of some health outcome depending on potentially explanatory variables. Here, it is rarely sufficient to investigate the influence of a single variable on the disease as most diseases are the result of the complex interplay of many different exposure variables including socio-demographic ones. Although it is very helpful to look first at simple stratified 2×2 tables to account for confounders such techniques become impractical for an increasing number of variables to be accounted for and a restricted number of subjects. In such situations, techniques are needed that allow the examination of the effect of several variables simultaneously for adjustment, but also for prediction purposes.

This is realized by regression models that offer a wide variety of methods to capture the functional relationship between response and explanatory variables (see Chap. II.3 of this handbook). Models with more than one explanatory variable are usually referred to as multiple regression models, multivariable or multivariate models where the latter might also involve more than one outcome. Using such techniques one needs to keep in mind that a statistical model rests on assumptions like normality, variance homogeneity, independence, and linearity that have all to be checked carefully in a given data situation. The validity of the model depends on these assumptions which might not be fulfilled by the data. Various models are therefore available from which an adequate one has to be selected. This choice is partly based on the research question and the a priori epidemiological knowledge on the relevant variables and their measurement. Depending on the scale, continuous or discrete, linear or logistic regressions might then be used for modeling purposes. Even more complex techniques such as generalized linear models can be applied where the functional relationship is no longer assumed to be linear (see Chaps. II.2 and II.3). Once the type of regression model is determined one has to decide which and how many variables should be included in the model where in case that variables are strongly correlated with each other only one of them should be included. Many software packages offer automatic selection strategies such as forward or backward selection, which usually lead to different models that are all consistent with the data at hand. An additional problem may occur due to the fact that the type of regression model will have an impact on the variables to be selected and vice versa. The resulting model may also have failed to recognize effect modification or may have been heavily affected by peculiarities of this particular data set that are of no general relevance. Thus, each model obtained as part of the statistical analysis should be independently validated.

Further extensions of simple regression models are e.g. time-series models that allow for time-dependent variation and correlation, Cox-type models to be applied in survival analysis (see Chap. II.4) and so-called graphical chain models which try to capture even more complex association structures. One of their features is that they allow in addition for indirect influences by incorporating so-called intermediate variables that simultaneously serve as explanatory and response variables. The interested reader is referred to Lauritzen and Wermuth (1989), Wermuth and Lauritzen (1990), Whittaker (1990), Lauritzen (1996), and Cox and Wermuth (1996).

4.4 Handling of Data Problems

Data are the basic elements of epidemiological investigation and information. In the form of values of predictor variables they represent levels of factors (risk factors and covariates), which are the *determinants* of health-related states or events in the sense of the definition of epidemiology quoted in Sect. 1.1. As values of response (outcome) variables they describe the health-related phenomena themselves. Measuring these values precisely is obviously fundamental in any epidemiological study and for the conclusions to be drawn from it. The practical problems that arise when trying to do this are outlined in Chaps. I.10 to I.13. However, even when taking great care and applying a rigorous quality control, some data as registered may still be erroneous and others may be missing. The question of how to handle these problems is the subject of Chaps. II.5 and II.6.

Intuitively, it is clear that in both cases the approach to be taken depends on the particular situation, more precisely, on the type of *additional information* that may be available. We use this information either to correct or to supplement certain data individually or to correct the final results of the study.

Sometimes a naïve approach looks sensible. Here are two examples of the two types of correction. First, if we know that the data at hand represent the size of a tumor in consecutive months, we may be tempted to replace a missing or obviously out-of-range value by an interpolated one. Second, when monitoring maternal mortality in a developing country by studies done routinely on the basis of death registers, we may multiply the figures obtained by a factor that reflects the fact that many deaths in childbed are not recorded in these registers. This factor was estimated beforehand by special studies where all such deaths were searched for, e.g. by visits to the homes of diseased women and retrospective interviews. For example, in Guatemala the correcting factor 1.58 is being used.

Even with such elementary procedures, though, the problem of estimating the influence of their use on the *statistical quality* of the study, be it the power of a test or the width of a confidence interval, is not only at the core of the matter but also difficult. It should therefore not be surprising that the Chaps. II.5 and II.6 are more mathematical.

The basic idea underlying the rigorous handling of measurement errors looks like this. We represent the *true* predictor variables whose values we cannot observe exactly because of errors, via so-called *surrogate* predictor variables that can be measured error free and that are being used for "correcting the errors" or as surrogates for the true predictors. The way a surrogate and a predictor are assumed to be related and the corresponding distributional assumptions form the so-called *measurement error model*. Several types of such models have been suggested and explored, the goal always being to get an idea about the magnitude of the effect on the statistical quality of the study if we correct the final results as directed by the model. Based on these theoretical results, when planning a study, a decision about the model to be used must be taken before-

hand, subject to the demand that it be realistic and can be handled mathematically.

The general ideas underlying methods for dealing with missing values are similar although the technical details are of course quite different. The first step consists in jointly modeling the predictor and response variables and the missing value mechanism. This mechanism may or may not consist in filling in missing data individually (data imputation). Next, the influence of correcting under various models is investigated, and finally concrete studies are evaluated using one or several appropriate models.

4.5 Meta-Analysis

The use of meta-analyses to synthesize the evidence from epidemiological studies has become more and more popular. They can be considered as the quantitative parts of systematic reviews. The main objective of a meta-analysis is usually the statistical combination of results from several studies that individually are not powerful enough to demonstrate a small but important effect. However, whereas it is always reasonable to review the literature and the published results on a certain topic systematically, the statistical combination of results from separate epidemiological studies may yield misleading results. Purely observational studies are in contrast to randomized clinical trials where differences in treatment effects between studies can mainly be attributed to random variation. Observational studies, however, may lead to different estimates of the same effect that can no longer be explained by chance alone, but that may be due to confounding and bias potentially inherent in each of them. Thus, the calculation of a combined measure of association based on heterogeneous estimates arising from different studies may lead to a biased estimate with spurious precision. Although it is possible to allow for heterogeneity in the statistical analysis by so-called random-effects models their interpretation is often difficult. Inspecting the sources of heterogeneity and trying to explain it would therefore be a more sensible approach in most instances.

Nevertheless, a meta-analysis may be a reasonable way to integrate findings from different studies and to reveal an overall trend of the results, if existing at all. A meta-analysis from several studies to obtain an overall estimate of association, for instance, can be performed by pooling the original data or by calculating a combined measure of association based on the single estimates. In both cases, it is important to retain the study as unit of analysis. Ignoring this fact would lead to biased results since the variation between the different studies and their different within-variabilities and sample sizes would otherwise not be adequately accounted for in the statistical analysis.

Since the probably first application of formal methods to pool several studies by Pearson (1904) numerous sophisticated statistical methods have been developed that are reviewed in Chap. II.7 of this handbook.

Applications of Epidemiological Methods and Research Areas in Epidemiology 5

Epidemiology pursues three major targets: (1) to describe the spectrum of diseases and their determinants, (2) to identify the causal factors of diseases, and (3) to apply this knowledge for prevention and public health practice.

Description of the Spectrum of Diseases 5.1

Describing the distribution of disease is an integral part of the planning and evaluation of health care services. Often, information on possible exposures and disease outcomes has not been gathered with any specific hypothesis in mind but stems from routinely collected data. These descriptions serve two main purposes. First, they help in generating etiological hypotheses that may be investigated in detail by analytical studies. Second, descriptive data form the basis of health reports that provide important information for the planning of health systems, e.g. by estimating the prevalence of diseases and by projecting temporal trends. The approaches in descriptive epidemiology are presented in Chap. I.3 of this handbook.

Complementary descriptive information relates to the revelation of the natural history of diseases – one of the subjects of clinical epidemiology – that helps to improve diagnostic accuracy and therapeutic processes in the clinical setting. The understanding of a disease process and its intermediate stages also gives important input for the definition of outcome variables, be it disease outcomes that are used in classical epidemiology or precursors of disease and pre-clinical stages that are relevant for screening or in molecular epidemiology studies.

Identification of Causes of Disease 5.2

Current research in epidemiology is still tied to a considerable extent to the general methodological issues summarized in Sects. 3 and 4. These concern *any* kind of exposures (risk factors) and *any* kind of outcomes (health defects). However, the basic ideas having been shaped and the main procedures elaborated, the emphasis is now on more specific questions determined by a particular type of exposure (e.g. Chaps. III.1–III.4, III.7, III.9) or a special kind of outcome (e.g. Chaps. IV.1) or both (e.g. Chap. III.6).

Exposure-oriented Research

The search for extraneous factors that cause a disease is a central feature of epidemiology. This is nicely illustrated by the famous investigation into the causes of cholera by John Snow, who identified the association of ill social and hygienic conditions, especially of the supply with contaminated water, with the disease and thus provided the basis for preventive action. Since that time, the investigation of hygienic conditions has been diversified by examining infective agents

(Chap. IV.1), nutrition (Chap. III.4), pharmaceuticals (Chap. III.9), social conditions (Chap. III.1) as well as physical and chemical agents in the environment (Chap. III.3) or at the workplace (Chap. III.2). A peculiarity is the investigation of genetic determinants by themselves and their interaction with the extraneous exposures mentioned above (Chap. III.7).

Nutrition belongs to the most frequently studied exposures and may serve as a model for the methodological problems of exposure-oriented research and its potential for public health. There are few health outcomes for which nutrition does not play either a direct or an indirect role in causation, and therefore a solid evidence-base is required to guide action aiming at disease prevention and improvement of public health. Poor nutrition has direct effects on growth and normal development, as well as on the process of healthy ageing. For example, 40 to 70% of cancer deaths were estimated to be attributable to poor nutrition. The effect of poor diet on chronic diseases is complex, such as, for example, the role of micronutrients in maintaining optimal cell function and reducing the risk of cancer and cardiovascular disease. Foods contain more than nutrients, and the way foods are prepared may enhance or reduce their harmful or beneficial effects on health. Because diet and behavior are complex and interrelated, it is important, both in the design and the interpretation of studies, to understand how this complexity may affect the results. The major specific concern is how to define and assess with required accuracy the relevant measure of exposure, free from bias.

The latter is a general problem that exposure-oriented epidemiology is faced with, especially in retrospective studies (see Chap. I.11). The use of biological markers of exposure and early effect has been proposed to reduce exposure misclassification. In a few cases, biomarker-based studies have led to important advances, as for example illustrated by the assessment of exposure to aflatoxins, enhanced sensitivity and specificity of assessment of past viral infection, and detection of protein and DNA adducts in workers exposed to reactive chemicals such as ethylene oxide. In other cases, however, initial, promising results have not been confirmed by more sophisticated investigations. They include in particular the search for susceptibility to environmental carcinogens by looking at polymorphism for metabolic enzymes (Chap. III.6). The new opportunities offered by biomarkers to overcome some of the limitations of traditional approaches in epidemiology need to be assessed systematically. The measurement of biomarkers should be quality-controlled and their results should be validated. Sources of bias and confounding in molecular epidemiology studies have to be assessed with the same stringency as in other types of epidemiological studies.

Modern molecular techniques have made it possible to investigate exposure to genetic factors in the development or the course of diseases on a large scale. A familial aggregation has been shown for many diseases. Although some of the aggregation can be explained by shared risk factors among family members, it is plausible that a true genetic component exists for most human cancers and for the susceptibility to many infectious diseases. The knowledge of low-penetrance genes responsible for such susceptibility is still very limited, although research has currently focused on genes encoding for metabolic enzymes, DNA repair,

cell-cycle control, and hormone receptors. In many studies only indirect evidence can be used since the suspected disease-related gene (candidate gene) is not directly observable. To locate or to identify susceptibility genes, genetic markers are used either in a so-called whole genome scan or in the investigation of candidate genes (Chap. III.7). The latter can be performed through linkage studies, where the common segregation of a marker and a disease is investigated in pedigrees; and through association studies, where it is investigated whether certain marker alleles of affected individuals will be more or less frequent than in a randomly selected individual from the population. Both, population-based and family designs are complementary and play a central role in genetic epidemiological studies. In the case of low-penetrance genes, association studies have been successful in identifying genetic susceptibility factors. Given the lack of dependence of genetic markers from time of disease development, the case-control approach is particularly suitable for this type of investigation because their assessment is not prone to recall bias. More pronounced than in classic epidemiology, the three main complications in genetic epidemiology are dependencies, use of indirect evidence and complex data sets.

Outcome-oriented Research

Epidemiology in industrialized countries is nowadays dominated by research on chronic diseases, among them cardiovascular diseases (Chap. IV.2), cancer (Chap. IV.3) and musculoskeletal disorders (Chap. IV.4). Their epidemiology – especially the one of cancer – is characterized more than any other outcome-defined epidemiology by the abundance of observational studies to find risk factors of all kind.

Cardiovascular diseases have a multi-factorial etiology and confounding effects are especially intriguing. For example, clustering of coronary heart diseases in families could be due both to genetic factors and to common dietary habits. High blood pressure plays both the role of an outcome variable and of a risk factor. Typical features of the epidemiology of cardiovascular diseases are the existence of many long term prospective studies, of intervention programmes like those described in Chap. III.11, and of a decline of morbidity and mortality in some areas and population groups whose causes are manifold, too, including for example control of blood pressure and blood cholesterol.

In many respects cancer epidemiology exemplifies the strengths and the weaknesses of the discipline at large. Although it is a relatively young discipline, it has been the key tool to demonstrate the causal role of important cancer risk factors, like smoking, human papilloma virus infections in cervical cancer, solar radiation in skin cancer, and obesity in many neoplasms. Cancer epidemiology is an area in which innovative methodological approaches are developed as illustrated by the increasing use of biological and genetic markers pertaining to causal factors and early outcomes.

By comparison, the epidemiology of musculoskeletal disorders is less developed. Already the definition of the various disorders and the distinction between them are still subject to debate. Case ascertainment is often tricky. In spite of the high

prevalence of for example back pain or osteoarthritis and their enormous negative impact on quality of life, mortality caused by them is significantly lower than that by cancer or cardiovascular diseases. Even simple estimates of prevalence leave wide margins. Regarding established risk factors, we find for instance for osteoarthritis and depending on its location, genetic factors, gender, obesity, heavy physical workload, and estrogen use, but not much more seems to be known although certain nutritional factors have been mentioned like red meat and alcohol.

The investigation of infectious diseases is the most important historical root of epidemiology and is still of primary importance in developing countries. If a person suffers from a particular outcome like tuberculosis, the exposure "infection by the relevant micro-organism", i.e. by mycobacterium tuberculosis, must have been present by the very definition of the disease. However, it is not a sufficient condition for overt disease, and many analytical studies examine the influence of co-factors like social conditions, nutrition, and co-morbidities regarded as risk factors for opportunistic infections. Purely descriptive health statistics, too, play a very important role in controlling infectious diseases. Related activities are general *epidemic surveillance*, *outbreak studies* by tracing possible carriers, and the search for infectious sources like salmonella as sources of food poisoning or the various origins of *nosocomial* illness. The most specific features of the epidemiology of infectious diseases are *mathematical modeling* and *prevention by immunization*. Modeling is to be understood in the sense of population dynamics. What is being modeled is typically the time evolution of the incidence or prevalence of the disease in question. The model, be it deterministic or stochastic, describes the mechanism of the infection and depends on specific parameters like contact frequencies between infected and susceptible subjects and healing rates. It is interesting to note that the discoverer of the infectious cycle of malaria, Ronald Ross, also designed and analyzed a mathematical model for it that led him to conceive of the *threshold principle* (Bailey 1975; Diekmann and Heesterbeek 2000). Prevention of infectious diseases can in principle be done in three ways: by acting on co-factors of the type mentioned above; by interfering with the infectious process via hygiene, separation of susceptible persons from carriers or vectors, and elimination of vectors; or by raising the immunity of susceptible people by various measures like preventive drug treatment, the main method of immunization being a vaccination, though. The effect of a vaccination in a population can be modeled in its turn, which leads in particular to the basic epidemiological concept of *herd immunity*.

5.3 Application of Epidemiological Knowledge

Epidemiological knowledge concerns *populations*. There are two ways to use this knowledge. The first is *group-oriented*: It consists in applying knowledge about a specific population directly to this population itself. This is part of *Public Health*. The conceptually simplest applications of this kind concern the planning of the health system (Chap. IV.5) and of health strategies. For instance, epidemiological studies have shown that people exposed to inhaling asbestos fibers are prone to

develop asbestosis and its sequels like cor pulmonale. We apply this knowledge to the entire population by prohibiting the use of asbestos.

The second path is taken when we are confronted with an individual person, typically in a clinical setting: We can then regard this person as a member of an appropriate specific population for which relevant epidemiological knowledge is available, and deal with her or him accordingly. As an example, a physician confronted with a child suffering from medium dehydration due to acute diarrhea, knows from clinical trials that oral rehydration (see Chap. IV.6) will normally be a very efficient treatment. Hence she/he will apply it in this particular case.

Clinical epidemiology plays a major role for the second path, where epidemiological knowledge is applied in all phases of clinical decision making, i.e. in daily clinical practice, starting with diagnosis, passing to therapy, and culminating in prognosis and advice to the patient – including individual preventive measures.

Prevention

The first of the two preceding examples belongs to *population-based prevention* (see Chaps. I.8, III.11 and IV.6). The underlying idea is to diminish the influence of risk factors identified by previous observational epidemiological studies. These factors may be geographic, environmental, social, occupational, behavioral, nutritional, or genetic. Risks of transmission of infectious diseases have long played a particular role in Public Health: Their influence was reduced by public hygiene in the classical sense. Applying observational epidemiology in order to diminish or eliminate risk factors has therefore been termed *hygiene* in a modern, general sense. Preventive measures in this context are sometimes themselves subject to an a posteriori evaluation which may bear on one hand on the way they have been implemented and on the other hand on their effectiveness.

Population-based preventive measures can also be derived from results of experimental epidemiology. The most important applications of this kind are *vaccinations* performed systematically within a given population. They have to be subjected to rigorous efficacy trials before implementation. *Preventive drug treatments*, e.g. against malaria or cardiovascular events, fall into the same category.

In many cases the target population itself is determined by a previous epidemiological study. For instance, dietary recommendations to reduce cardiovascular problems, and vaccinations against hepatitis B, yellow fever or influenza, are usually given only to people that were identified as being of *high risk* to contract the disease in question.

Screening

Population-based *treatments* as a measure of Public Health are conceivable but hardly ever implemented. There exists, however, a population-based application of epidemiology in the realm of *diagnosis*, viz. screening (see Chap. III.10). Its purpose is to find yet unrecognized diseases or health defects by appropriate tests that can be rapidly executed within large population groups. The ultimate aim is mostly to allow a treatment at an early stage. Occasionally, screening was also performed in

order to isolate infected people, e.g. lepers. Classical examples of screening include mass X-ray examination to detect cases of pulmonary tuberculosis or breast cancer, and cytological tests to identify cancer of the cervix uteri. Screening programs may concentrate on *high risk groups* if it would be unfeasible, too expensive, or too dangerous to examine the entire original population. A striking example is the screening for pulmonary tuberculosis in Norway where most of the prevalent cases were found at an early stage by systematic X-ray examinations of only a small fraction of the population.

Case Management

The concept of the *individual* risk of a person (see Chaps. I.2 and I.5) that underlies the definition of risk groups represents a particular case of the second way of applications of epidemiology, viz. dealing with an individual person on the basis of epidemiological knowledge about populations to which she or he is deemed to belong. The most important application of this idea, however, is *clinical epidemiology* which was also called *statistics in clinical medicine*. It is the art of case-management in the most general sense: diagnosis using the epidemiological characteristics of medical tests like sensitivity and specificity, treatment using the results of clinical trials, prognosis for a specific case based again on relevant epidemiological studies. Chapter III.3 describes in detail fairly sophisticated procedures involving all aspects of case-management including the opinion of the patient or his/her relatives and considerations of cost, secondary effects, and quality of life as elements entering the therapeutic decision.

Health Services

Health services research (HSR) is a vast and multiform field. It has no concise and generally accepted definition but still there is a more or less general agreement about its essential ideas. Its purpose is to lay the *general*, *scientific* foundations for health policy in order to improve the health of people as much as possible under the constraints of society and nature. The subjects of HSR are, in the first place, the underlying *structures*, i.e. the basic elements concerned by questions of health and the relations between them, in the second place the *processes* of health care delivery, and in the third place the *effects* of health services on the health of the public.

On the methodological side, HSR means *analysis* and *evaluation* of all of these aspects. The tools are mainly coming from mathematics and statistics, economics, and sociology together with knowledge from clinical medicine and basic sciences like biology. Epidemiology plays a particularly important role.

Evaluation implies *comparison*: comparison of different existing health care systems, and comparison of an existing one with theoretical, ideal systems in order to design a better one. Comparison of health care systems of different countries has been a favorite subject. One of the main "factors" that distinguishes them is the way medical services are being paid for and the form of health insurance.

The basic elements are physicians, nurses and other personnel, hospitals, equipment, and money, but also the population getting into contact with the health

system and its health status. Relations between these elements comprise *health needs* and *access to services*, but also the *organization* of the health system.

Processes of health care delivery may of course mean the usual clinical curative treatment of patients but also person- or community-based preventive actions including environmental measures, health education, or health strategies like the one that led to the eradication of smallpox.

Finally, the effects of health services, i.e. the *output*, can be measured in many ways, e.g. by morbidity and mortality, life expectancy, quality of life preserved or restored, and economic losses due to illness. Questions of *effectiveness*, i.e. the value of outputs relative to (usually monetary) inputs, are in the limelight.

Epidemiology as a method serves two purposes. On the one hand, the results of epidemiological investigations enter the field as basic parameters. Some experimental epidemiological studies like intervention trials are even considered as *belonging* to health services research. On the other hand, many methods used in health services research that stem from mathematical statistics and whose goal is to study the influence of various factors on outcomes, are formally the same as those employed in epidemiology.

Given the enormity and complexity of the subject many different "approaches" and "models" have been proposed and tried out. Earlier ones were still fairly descriptive and static, focusing on the functioning of the health services or on health policy with a strong emphasis on the economic aspects. The input-output model where the effects of changes of essential inputs on the various outcomes of interest are studied, if possible in a quantitative way, is more recent. More than others it allows to a large extent a "modular" approach, separating from each other the investigation of different parts or levels of the health services.

Ethical Aspects 5.4

The protection of human rights is one of the most crucial aspects of all studies on humans. Although there are substantial differences between experimental and observational studies they both have to face the challenging task to protect the privacy of all individuals taking part in a study. This also implies as a basic principle that study subjects are asked for their informed consent.

Another ethical angle of epidemiological research concerns the study quality. Poorly conducted research may lead to unsubstantiated and wrong decisions in clinical practice or policy making in public health and may thus cause harm to individuals, but also to society as a whole. Therefore, guidelines have been prepared to maintain high study quality and to preserve human rights such as the "Good Epidemiological Practice" provided by the International Epidemiological Association in 1998.

Of course, the four general principles of the Declaration of Helsinki (World Medical Association 2000) have to be followed, i.e. autonomy (respect for individuals), beneficence (do good), non-maleficence (do no harm), and justice. These principles are of particular relevance in randomized controlled trials, where the intervention (or non-intervention) may involve negative consequences for participants.

Various recent developments in epidemiological research constitute a new challenge regarding ethical aspects. First, automated record linkage databases are now at least partly available that capture both exposure and outcome data on an individual level. Such databases have raised questions about confidentiality of patient's medical records, authorizing access to person-specific information, and their potential misuse. Second, the inclusion of molecular markers in epidemiological studies has led to a controversial debate on the potential benefit or harm of results gained by genetic and molecular epidemiological studies. This raises the following questions: Can knowledge on genetic markers be used in primary prevention programs? How should this knowledge be communicated to the study subjects who may be forced into the conflict between their individual "right to know" and their "right not to know"? A third driver of ethical questioning has been the discussion about integrity and conflict of interests, in particular in cases of sponsored epidemiological studies or when the results are contradictory.

As a consequence, an increasing awareness that ethical conduct is essential to epidemiological research can be observed among epidemiologists. Thus, it is not surprising that now basic principles of integrity, honesty, truthfulness, fairness and equity, respect for people's autonomy, distributive justice, doing good and not harm have become an integral part in the planning and conduct of epidemiological studies. Chapter IV.7 of this handbook is devoted to all of these aspects.

References

Ackerknecht EH (1953) Rudolf Virchow. Doctor – statesman – anthropologist. University of Wisconsin Press, Madison

Ackerknecht EH (1967) Medicine at the Paris hospital, 1794–1848. Johns Hopkins, Baltimore

Ahrens W (1999) Retrospective assessment of occupational exposure in case-control studies. ecomed Verlagsgesellschaft, Landsberg

Armstrong BK, White E, Saracci R (1992) Principles of exposure measurement in epidemiology. Oxford University Press, Oxford, New York, Tokio

Bailey NTJ (1975) The mathematical theory of infectious diseases and its applications. Griffin, London

Bernard C (1865) Introduction à l'étude de la médecine expérimentale. J.B. Baillière et Fils, Paris

Bernoulli D (1766) Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, & desavantages de l'inoculation pour la prévenir. Mém Math Phys Acad Roy Sci, Paris

Boffetta P, Agudo A, Ahrens W, Benhamou E, Benhamou S, Darby SC, Ferro G, Fortes C, Gonzalez CA, Jöckel KH, Krauss M, Kreienbrock L, Kreuzer M, Mendes A, Merletti F, Nyberg F, Pershagen G, Pohlabeln H, Riboli E, Schmid G, Simonato L, Tredaniel J, Whitley E, Wichmann HE, Saracci R (1998) Multicenter case-control study of exposure to environmental tobacco smoke and lung cancer in Europe. J Natl Cancer Inst 90:1440–1450

Buell P, Dunn JE (1965) Cancer mortality among Japanese Issei and Nisei of California. Cancer 18:656–664

Bulloch W (1938) The history of bacteriology. Oxford University Press, London

Chan-Yeung M, Dimich-Ward H (2003) Respiratory health effects of exposure to environmental tobacco smoke. Respirology 8:131–139

Chen WY, Colditz GA, Rosner B, Hankinson SE, Hunter DJ, Manson JE, Stampfer MJ, Willett WC, Speizer FE (2002) Use of postmenopausal hormones, alcohol, and risk for invasive breast cancer. Ann Intern Med 137:798–804

Cox D, Wermuth N (1996) Multivariate dependencies – models, analysis and interpretation. Chapman and Hall, London

Coxe JR (1846) The writings of Hippocrates and Galen. Lindsay and Blackiston, Philadelpia

Dawber TR (1980) The Framingham study: The epidemiology of atherosclerotic disease. Harvard University Press, Cambridge Mass.

Dawber TR, Meadors GF, Moore FE Jr (1951) Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health 41:279–286

Detels R (2002) Epidemiology: The foundation of public health. In: Detels R, McEwen J, Beaglehole R, Tanaka H (eds) Oxford textbook of public health. Vol 2: The methods of public health, 4th edn. Oxford University Press, Oxford, New York, pp 485–491

Detels R, Breslow L (2002) Current scope and concerns in public health. In: Detels R, McEwen J, Beaglehole R, Tanaka H (eds) Oxford textbook of public health. Vol 1: The scope of public health, 4th edn. Oxford University Press, Oxford, New York, pp 3–20

Diekmann O, Heesterbeek JAP (2000) Mathematical epidemiology of infectious diseases. Wiley, Chichester

Dillon W, Goldstein M (1984) Multivariate analysis, methods and applications. Wiley, New York

Doll R, Hill AB (1950) Smoking and carcinoma of the lung: Preliminary report. Brit Med J 2:739–748

Doll R, Hill AB (1954) The mortality of doctors in relation to their smoking habits: A preliminary report. Brit Med J 1:1451–1455

Doll R, Hill AB (1964) Mortality in relation to smoking: Ten years' observation of British doctors. Brit Med J 1:1399–1410; 1460–1467

Doll R, Peto R (1978) Cigarette smoking and bronchial carcinoma: Dose and time relationships among regular smokers and lifelong non-smokers. J Epidemiol Comm Health 32:303–313

Doll R, Wald NJ (1994) Interpretation of Negative Epidemiological Evidence for Carcinogenicity. IARC Scientific Publications vol 65. International Agency for Research on Cancer (IARC), Lyon

Everitt B, Dunn G (2001) Applied multivariate data analysis, 2nd edn. Edward Arnold, London

Farr W (1975) Vital statistics: a memorial volume of selections from the reports and writings of William Farr. New York Academy of Sciences. Scarecrow Press, Metuchen NJ

Fracastoro G (1521) De contagione et contagiosis morbis et eorum curatione. English translation by Wright WC (1930). G P Putnam's Sons, New York

Galen (1951) De sanitate tuenda, English translation by Green R. C C Thomas, Springfield Ill

Giri NC (2004) Multivariate statistical analysis, 2nd edn. Marcel Dekker, New York

Gordon A (1795) A treatise on the epidemic puerperal fever of Aberdeen. London

Graunt J (1662) Natural and political observations mentioned in a following index, and made upon the Bills of mortality … with reference to the government, religion, trade, growth, ayre, diseases, and the several changes of the said city … Martin, Allestry, and Dicas, London

Greenwood M (1932) Epidemiology, historical and experimental. Johns Hopkins, Baltimore

Halley E (1693) An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslau; with an attempt to ascertain the price of annuities upon lives. Philos Trans Roy Soc London 17:596–610, 654–656

Hennekens CH, Buring JE (1987) Epidemiology in medicine. Little, Brown and Company, Boston, Toronto

Hill AB (1962) Statistical methods in clinical and preventive medicine. Livingstone, Edinburgh

Hill AB (1965) The environment and disease: association or causation? Proc R Soc Med 58:295–300

Hippocrates (400 BC, approximately) On airs, waters, and place. There exist several editions in English, among others in: (1846) Coxe, see above; (1939) The genuine works of Hippocrates. Williams and Wilkins, Baltimore; (1983) Lloyd (ed) Hippocratic writings. Penguin Classics, Harmondsworth (http://classics.mit.edu/Hippocrates/airwatpl.html) Accessed May 24, 2004

Holmes OW (1842–43) The contagiousness of puerperal fever. N Engl Quart J Med Surg 1:503–540

IARC (2004) Tobacco smoke and involuntary smoking. IARC Monographs vol 83. International Agency for Research on Cancer (IARC), Lyon

Jenner E (1798) An inquiry into the causes and effects of the variolae vaccinae, a disease known by the name of Cow Pox. Published by the author, London

Keys A (1980) Seven Countries: a multivariate analysis of death and coronary heart disease. Harvard University Press, Cambridge

Khoury MJ, Beaty TH, Cohen BH (1993) Fundamentals of genetic epidemiology. Oxford University Press, New York

Kromhout D, Menotti A, Bloemberg B, Aravanis C, Blackburn H, Buzina R, Dontas AS, Fidanza F, Giampaoli S, Jansen A, Karvonen M, Katan M, Nissinen A, Nedeljkovic S, Pekkanen J, Pekkarinen M, Punsar S, Räsänen L, Simic B, Toshima H (1995) Dietary saturated and trans fatty acids and cholesterol and 25-year mortality from coronary heart disease: the Seven Countries Study. Prev Med 24:308–315

Kromhout H (1994) From eyeballing to statistical modelling. Methods for assessment of occupational exposure. 1–210. 1994. Landbouwuniversiteit de Wageningen

Last JM (ed) (2001) A dictionary of epidemiology, 4th edn. Oxford University Press, New York, Oxford, Toronto

Lauritzen SL, Wermuth N (1989) Graphical models for associations between variables, some of which are qualitative and some quantitative. Ann Statist 17:31–57

Lauritzen SL (1996) Graphical models. Oxford University Press, Oxford

Lee PN, Forey BA (1996) Misclassification of smoking habits as a source of bias in the study of environmental tobacco smoke and lung cancer. Stat Med 15:581–605

Lind J (1753) A treatise on the scurvy. Sands, Murray, and Cochran, Edinburgh

Lind J (1771) An essay on diseases incidental to Europeans in hot climates with the method of preventing their fatal consequences, 2nd edn. T Becket and PA De Hondt, London

Lipnick RJ, Buring JE, Hennekens CH, Rosner B, Willett W, Bain C, Stampfer MJ, Colditz GA, Peto R, Speizer FE (1986) Oral contraceptives and breast cancer. A prospective cohort study. JAMA 255:58–61

Louis PCA (1835) Recherches sur les effets de la saignée dans quelques maladies inflammatoires et sur l'action de l'émétique et les vésicatoires dans la pneumonie. Baillière, Paris

MacMahon B, Pugh TF (1970) Epidemiology: principles and methods. Little, Brown and Company, Boston, Massachusetts

Marshall H, Tulloch AM (1838) Statistical report on the sickness, mortality and invaliding in The West Indies. Prepared from the records of the Army Medical Department and War Office returns. Clowes & Sons, London

Mausner JS, Bahn AK (1974) Epidemiology: An introductory text. W.B. Saunders Company, Philadelphia, London, Toronto

Nieuwenhuijsen JM (ed) (2003) Exposure assessment in occuppational and environmental epidemiology. Oxford University Press, Oxford

Olsen J, Ahrens W, Björner J, Grönvold M, Jöckel KH, Keiding L, Lauritsen JM, Levy-Desroches S, Manderson L, Olesen F (1998) Epidemiology deserves better questionnaires. Int J Epidemiol 27:935 (http://ije.oupjounals.org/cgi/reprint/27/6/935.pdf) Accessed May 24, 2004

Pagel W (1982) Paracelsus, 2nd edn. Karger, Basel

Pearson K (1904) Report on certain enteric fever inoculation statistics. Brit Med J 3:1243–1246

Pott P (1775) Chirurgical observations relative to the cataract, the polypus of the nose, the cancer of the scrotum, the different kinds of ruptures, and the mortification of the toes and feet. Hawes, Clarke and Collins, London

Ramazzini B (1713) De morbis artificum diatriba. Baptistam Conzattum, Padua

Riboli E, Preston-Martin S, Saracci R, Haley NJ, Trichopoulos D, Becher H, Burch JD, Fontham ETH, Gao YT, Jindal SK, Koo LC, Marchand LL, Segnan N, Shimizu H, Stanta G, Wu-Williams AH, Zatonski W (1990) Exposure of nonsmoking women to environmental tobacco smoke: a 10-country collaborative study. Cancer Causes Control 1:243–252

Rose G, Shipley M (1986) Plasma cholesterol concentration and death from coronary heart disease: 10 year results of the Whitehall Study. BMJ 293:306–307

Rothman KJ, Greenland S (1998) Modern epidemiology, 2nd edn. Lippincott-Raven, Philadelphia

Rupke N (ed) (2000) Medical geography in historical perspective. Medical History, Supplement No 20. The Wellcome Trust Centre for the History of Medicine at UCL, London

Schairer E, Schöniger E (1943) Lungenkrebs und Tabakverbrauch. Z Krebsforsch 54:261–269 (English translation: Schairer E, Schöniger E (2001) Lung cancer and tobacco consumption. Int J Epidemiol 30:24–27)

Schulte P, Perera FP (eds) (1998) Molecular epidemiology: Principles and practices, 2nd edn. Academic Press, New York

Semmelweis IP (1861) Die Aetiologie, der Begriff und die Prophylaxe des Kindbettfiebers. Hartleben, Pest, Wien, Leipzig

Siegel RE (1968) Galen's system of physiology and medicine. Karger, Basel

Simpson JY (1868–1869, 1869–1870) Our existing system of hospitalism and its effects. Edinburgh Mon J Med Sci 14:816–830, 1084–1115; 15:523–532

Snow J (1855) On the mode of communication of cholera, 2nd edn. Churchill, London

Stampfer MJ, Willett WC, Colditz GA, Rosner B, Speizer FE, Hennekens CH (1985) A prospective study of postmenopausal estrogene therapy and coronary heart disease. N Engl J Med 331:1044–1049

Taubes G (1995) Epidemiology faces its limits. Science 269:164–169

Tenon JR (1788) Mémoires sur les hôpitaux de Paris. Pierres, Paris

Toniolo P, Boffetta P, Shuker D, Rothman N, Hulka B, Pearce N (1997) Application of biomarkers to cancer epidemiology. IARC Scientific Publications No 142. International Agency for Research on Cancer, Lyon

Truett J, Cornfield J, Kannel E (1967) Multivariate analysis of the risk of coronary heart disease in Framingham. J Chronic Disease 20:511–524

Wells AJ, English PB, Posner SF, Wagenknecht LE, Perez-Stable EJ (1998) Misclassification rates for current smokers misclassified as smokers. Am J Public Health 88:1503–1509

Wermuth N, Lauritzen SL (1990) On substantive research hypotheses, conditional independence, graphs and graphical chain models. J Roy Statist Soc Series B 52:21–50

White KL, Henderson MM (1976) Epidemiology as a fundamental science. Its uses in health services planning, administration, and evaluation. Oxford University Press, New York

Whittaker J (1990) Graphical models in applied multivariate statistics. Wiley, Chichester

Willett WC, Stampfer MJ, Colditz GA, Rosner BA, Hennekens CH, Speizer FE (1987) Dietary fat and the risk of breast cancer. N Engl J Med 316:22–28

World Medical Association (2000) The revised Declaration of Helsinki. Interpreting and implementing ethical principles in biomedical research. Edinburgh: 52nd WMA General Assembly. JAMA 284:3043–3045