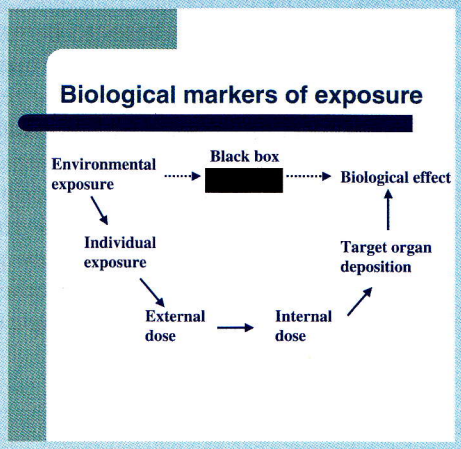

CHAPTER 21

ENVIRONMENTAL EPIDEMIOLOGY

JESPER B. NIELSEN AND TINA KOLD JENSEN
University of Southern Denmark

CONTENTS

I. INTRODUCTION

Epidemiology is the study of the occurrence of disease in populations. It originates from an observational discipline that describes changes in the prevalence or incidence of a specific disease—changes that may be observed over time, between geographical regions, or between populations. Thus, basic epidemiology delivers numbers with no explanation. Prevalences or incidences are, however, only really useful if associated with explanatory variables. These variables may relate to genetics, lifestyle, age, gender, occupation, environment, etc. Environmental epidemiology is therefore the study of associations between environmental exposures and the occurrence of disease within a population.

Few environmental diseases are pathognomonic in the sense that only one specified exposure may cause a certain disease. In most cases, several chemical exposures may cause the same disease, aggravate an existing disease, or in some situations even offer a certain degree of protection. Likewise, several sociodemographic factors and occupational exposures may affect exposure as well as disease. Proof of causation in epidemiological studies is therefore seldom, and associations between exposure and disease may often be biased.

This chapter does not replace epidemiological textbooks, but it is intended to introduce and discuss some more basic features related to study design and measures of exposure and outcome, as well as bias. The purpose is that non-epidemiologists should be able to critically read and understand most epidemiological studies, know strengths and weaknesses of different common study designs, and be able to recognize the more general types of bias occurring in health and exposure assessment.

The chapter will present some common problems related to environmental epidemiology and primarily use three exposure scenarios (case 1–3) based on recently published scientific articles to illustrate some of these problems. Most problems are general in nature and an inherent consequence of the chosen study design. The three examples have been chosen to illustrate different study designs, different outcomes or health effects, and different types of exposures. The

studies selected are a case-control study of residential radon exposure and lung cancer (Barros-Dios et al., 2002), a cohort study on malignant mesothelioma and environmental exposure to asbestos (Metintas et al., 2002), and an ecological study on adverse pregnancy outcomes and exposure to arsenic in drinking water (Yang et al., 2003). All three articles are available as full text articles free of charge on the Internet.

II. STUDY DESIGN

Study designs can be broadly categorized according to whether they are describing distributions of a health outcome (descriptive studies) or elucidating its determinants (analytical studies). Descriptive studies describe general characteristics of the distribution of an outcome in relation to person, place, and time. Analytical studies are used to test specific hypotheses and infer that exposure precedes outcome. They can be categorized into case-control or cohort studies according to whether the study subjects are selected on the basis of outcome or exposure (Figure 1). This section will briefly introduce the different study designs and discuss their strengths and weaknesses. For further reading and more specified details, the reader is referred to epidemiological textbooks (e.g., Rothman & Greenland, 1998).

A. Descriptive Studies

Descriptive studies describe general characteristics of the distribution of an outcome in relation to person, place, and time. The identification of descriptive characteristics is an important first step in the search for determinants or risk factors for specific outcomes, and thereby for the formulation of hypotheses to be tested in analytical studies. Descriptive studies use information from diverse sources such as census data, disease registers, and vital and clinical records, as well as national figures on consumption of food, drinking water, etc. Because this information is already available, descriptive studies are generally far less expensive and time-consuming than analytical studies. Usually they preclude the ability to test epidemiological hypotheses. Descriptive studies can be categorized into ecological or cross-sectional studies.

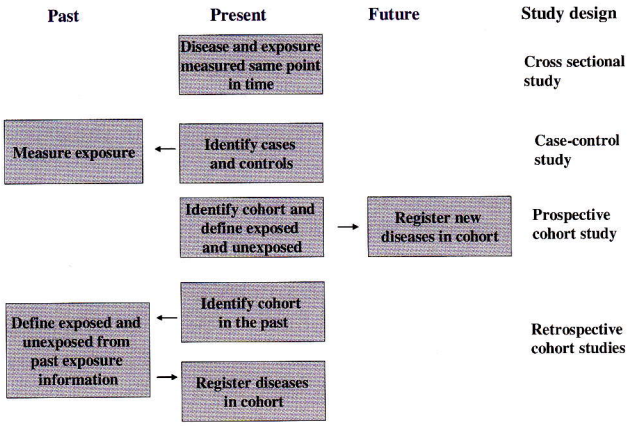

FIGURE 1 A schematic view of different study designs.

1. Ecological Studies

In an ecological study data from entire populations are used to compare outcome frequencies between different groups during the same time period or in the same population at different points in time. It is not possible to link exposure information to the occurrence of outcome in a particular individual. Furthermore, the studies are unable to control for confounding. Therefore, they cannot be used to test hypotheses or infer causality. They are, however, quick and inexpensive and use already available information (Table I).

Case 1 is an ecological study where birthweight distributions in two different regions of Taiwan with different levels of arsenic in drinking water are compared. No individual exposure information was collected and the place of birth determined the exposure status. The women may, however, have moved to that area just before the delivery and therefore not have been exposed to the drinking water of that region in pregnancy at all. Furthermore, no information about the actual intake of water in pregnancy was available and it was not known whether the women with high water intake were those who delivered prematurely. In addition, no information on how much water the women drank at home was available, and the study was unable to control for confounding as, for example, the women's smoking habits. The study is, however, good for generating hypotheses to be tested in analytical studies (see also Chapters 11 and 22, this volume).

Case 1

Prevalence of adverse pregnancy outcome among 18,259 first-parity singleton live births in Taiwan was linked to place of birth. Two geographic regions with different median levels of arsenic in the drinking water

TABLE I. Strengths and Limitations in Different Types of Epidemiological Studies

Study type	*Ecological study*	*Cross-sectional study*	*Case-control study*	*Cohort study*
Strengths	Quick and inexpensive often using already available information	Quick and inexpensive Provide information about health status of great public health relevance	Optimal for evaluation of rare diseases Can examine multiple etiologic factors for a single disease Relatively quick and inexpensive compared to cohort studies Well suited for evaluation of diseases with long latency periods	Valuable for rare exposures Can examine multiple effects of a single exposure Can elucidate temporal relationship between exposure and disease Minimizes bias in exposure assessment Allows direct incidence rates to be calculated
Limitations	Unable to link exposure with disease in particular individuals Unable to control for confounding No individual exposure information	Cannot determine whether exposure preceded or resulted from the disease Considered prevalent and will reflect determinants of etiology as well as survival	Inefficient for evaluation of rare exposures Cannot compute incidence rates in exposed and unexposed individuals The temporal relationship between exposure and disease may be difficult to establish Prone to bias, particularly recall and selection bias	Inefficient for evaluation of rare diseases Prospective: extremely expensive and time-consuming Retrospective: requires the availability of adequate records Losses of followup can affect results

were included. Children from the arsenic-endemic area had on average a 30-g lower birthweight (statistical significant) and the rate of preterm deliveries was increased by 9% (insignificant). No data on individual exposures were available and the exposure to arsenic in drinking water in the arsenic-endemic area varied between <0.15 ppb and 3.6 ppm (20,000-fold), whereas the exposure in the area of comparison was below 0.9 ppb. In the arsenic-endemic area, 83% of drinking water resources had arsenic concentrations above 0.9 ppb.

Design: Ecological study
Outcome: Preterm delivery and birthweight
Exposure: Arsenic in drinking water

Reference: Yang et al., Environmental Research, 2003.

2. *Cross-Sectional Studies*

In cross-sectional studies the status of an individual with respect to the presence or absence of both exposure and outcome is assessed simultaneously (Figure 1). Thus, a cross-sectional study provides information about the frequency and characteristics of an outcome by a "snapshot" of the population at a specific time. Such data are of great value to public health administrators when assessing the health status or health-care needs of a population. However, as exposure and outcome are assessed at the same point in time, cross-sectional surveys cannot always distinguish whether the exposure preceded the outcome development or whether the presence of disease affected the individual's level of exposure. It is, in other words, not possible to determine whether the exposure preceded or was caused by the disease (Table I). Thus, cross-sectional studies have found that infertile couples report more psychological distress symptoms, which implies that stress therefore causes infertility. It is, however, not known whether the couples became infertile because of the stress or whether the infertility and its consequences and treatment caused the stress. Cross-sectional studies are like ecological studies, which are valuable for raising a question of the presence of an association rather than for testing a hypothesis.

B. Analytical Studies

1. Case-Control Studies

In a case-control study, subjects are selected on the basis of whether they have (cases) or do not have (controls) a specific outcome. In its most basic form, cases with the outcome of interest are selected from hospitals or the general population and compared with a group (controls) without the outcome. More refined study designs exist, and interested readers are referred to Rothman (1998, 2002). The proportions with the exposure of interest in each group are compared (Figure 1). The case-control design is a good way to study diseases with long latency periods, because investigators can identify affected and unaffected individuals and assess antecedent exposures rather than waiting a number of years for the disease to develop. Therefore, case-control studies are time and cost efficient. In addition, by selecting the cases on the basis of outcome, the study can identify an adequate number of affected and unaffected individuals. Consequently, this strategy is particularly well suited for rare diseases, which in cohort studies would need inclusion of very large numbers of individuals in order to accumulate a sufficient number of cases with the outcome of interest. Finally, case-control studies allow evaluation of a range of potential etiologic exposures and their effect on the outcome. The case-control design can therefore be used to test specific a priori hypotheses or explore the effect of a range of different exposures.

The major drawback of case-control studies is that both the exposure and the outcome have already occurred at the time when the participants enter the study. This may affect the motivation to participate and the way that participants remember and report their exposures. This study design is therefore particularly vulnerable to selection and information bias, especially recall bias (see below). Furthermore, case-control studies are not efficient for rare exposures, as too few cases would then be exposed. In addition, only one outcome can be studied because the cases and controls are selected on the basis of that outcome. In case-control studies, no absolute measures of risk or incidence can be calculated. Instead the odds ratio estimates the relative risk or incidence rate ratio. This is, however, not a reason for not conducting case-control studies, as they offer advantages mentioned before and provide answers to hypotheses relatively fast (Table I).

One of the first issues to be considered in the evaluation of a case-control study is the definition of disease or outcome of interest. It is important that the definition of disease (outcome) is as homogeneous as possible, because very similar manifestations of disease may have very different etiologies. For example, congenital malformations which encompass many different diseases such as congenital heart malformations, cleft-palate, or neural tube defects are often compiled into one outcome because of the rare nature of each of these disease categories. They do, however, have very different etiologies and combining them does not give clues to the risk factors of each particular outcome. It is therefore important to establish strict diagnostic criteria for the disease under study.

The selection of appropriate controls is perhaps the most difficult and critical issue in a case-control study. Controls are necessary to evaluate whether the exposure observed in the case group differs from what would have been expected in a comparable group of individuals without the disease. Controls must be selected, not to represent the entire non-diseased population, but the population of individuals who would have been identified and included as cases had they also developed the disease. They can be chosen from hospitals or the general population. Hospital controls are selected from people admitted to the same hospital as the cases but with a different disease. The advantage of this approach is that people admitted to hospitals are easy to identify, motivated, and more likely to be aware of antecedent exposures. The disadvantages are that there might be different selection factors leading to admission to that hospital for different diseases. Furthermore, they differ from healthy individuals and may therefore not represent the exposure distribution in the population from which the cases derived. Controls can also be chosen from the general population. This can be done in a number of ways including canvassing households in the targeted neighborhood, random digit telephone dialing, or identification from population registers or voting lists. This is, however, usually more costly and time-consuming. Furthermore, the quality of the information obtained and the participation rate from cases and controls may differ as healthy individuals from the general population do not recall exposures with the same level of accuracy and they are less motivated to participate.

It is often argued that cases should be representative of all persons with the disease. This is, however, not true and case-control studies can be restricted to a particular type of case from whom complete and reliable information on exposure and disease can be obtained (for example, in a limited age range). Then control subjects should be selected to be comparable to the cases. Such case-control studies will provide a valid estimate of the association between exposure and disease and a judg-

ment of the generalizability of the findings can then be safely made.

Case 2

A total of 163 cases of primary lung cancer (response rate 70%) and 241 cancer-free controls (response rate 62%) were included. Cases were on average 8 years older than controls and had a 40% higher rate of cancer within the family. Close to 92% of cases were smokers as compared to 55% of the controls. Residential radon was measured in 98% of the homes for an average of 150 days. Residential radon exposure was close to 20% higher among cases than controls. This study concludes that residential radon exposure at levels below official guidelines of 148–200 Bq/m^3 may lead to a 2.5-fold increase in lung cancer risk. Further, synergism (an effect greater than that expected by their separate actions) between residential radon exposure and smoking was demonstrated.

Design: Population-based case-control study
Outcome: Confirmed primary lung cancer
Exposure: Indoor radon concentration
Reference: Barros-Dios et al., 2002

Case 2 is an example of a population-based, case-control study where lung cancer patients were compared with healthy controls from the same area in Spain. Controls were proportionally stratified randomly but excluded if they had respiratory tract disease, lived in the area less than 5 years, or were younger than 35 years of age. Exposure information was obtained from next of kin if the case or control had died. Information about radon exposure was measured, so no recall bias was present. The length that the participants lived at their current address was, however, not taken into account. In addition the participation rates were 10% higher among cases than controls and cases were approximately 8 years older, which may have introduced selection bias (see below). Moreover, more than 90% of the cases were smokers as compared to 55% of controls.

2. *Cohort Studies*

In a cohort study, a group of individuals are defined on the basis of presence or absence of exposure. At the time of exposure classification, subjects must be free from the outcome under investigation. Participants are then followed over a period of time to assess the occurrence of the specified outcome among those who are exposed and unexposed (Figure 1). Most often, the followup period must be at least several years to allow an adequate number to develop the outcome so that meaningful comparisons of disease frequency between exposed and unexposed individuals can be made. As the participants by inclusion criteria are free of disease at the time when their exposure status is defined and the study initiated, the temporal sequence between exposure and disease can be more clearly defined.

For many exposures the proportion of exposed individuals with the outcome is too small to make meaningful comparisons between exposed and unexposed. Therefore, cohort studies are particularly well suited for assessing the effect of rare exposures. Thus, cohort studies can enroll participants on the basis of their exposures and thereby include a large number of exposed, for example, among a cohort of heavily exposed workers. Furthermore, cohort studies offer less potential for selection bias and direct measurement of association (incidence rates) can be calculated among the exposed and unexposed. Finally, cohort studies allow the examination of multiple outcomes of a single exposure.

As cohorts studies often involve a large number of individuals followed for many years, they are time-consuming and expensive. Furthermore, only a proportion of those eligible actually participate in the study, and they often differ from the non-participants in motivation and attitudes toward health. As outcome is compared among exposed and unexposed, this does not usually affect the relationship except when non-response is related to both exposure and outcome. A way to address this problem is by comparing participants with non-participants with respect to basic available information such as age and socioeconomic status. In addition, losses to followup may seriously affect the results, especially if it differs between exposed and unexposed individuals or is related to exposure or outcome or both. Losses to followup should therefore be minimized and for those lost to followup, attempts to gain information about outcome from independent sources should be made (for example, through death or disease registers).

Cohort studies are often categorized into prospective or retrospective studies according to whether the outcome of interest has occurred at the time the study is initiated (Figure 1). In a prospective cohort study, the cohort is identified and categorized according to exposure. After a followup period the frequency of the outcome among exposed and unexposed is compared. In a retrospective cohort study both exposure and outcome have occurred at the time of the start of the investigation. A historical cohort is identified at the start

of the study and past exposures in the cohort are identified from already existing information. Then the frequency of outcomes (which has occurred) is determined. Prospective or retrospective solely refers to whether the outcome has occurred at the start of the study. Case-control studies can also be both prospective and retrospective, but they are most often retrospective, i.e., the outcome defining the case has occurred when the study is initiated. Retrospective cohort studies can usually be conducted more quickly and cheaply than their prospective counterparts, because all relevant events have already occurred at the time the study is initiated. They do, however, depend on routine availability of relevant exposure data in adequate details from pre-existing records. Because these data were collected for other purposes, the quality is often not optimal. Moreover, information on potential confounding factors is often unavailable (see also Chapter 18, this volume).

Case 3

In a cohort of 1886 villagers in a rural area in Turkey, the incidence of malignant pleural mesothelioma (MPM) was studied. The villagers were environmentally exposed to asbestos dust due to the use of asbestos-contaminated white soil. The soil was used as a whitewash or plaster material for walls, as insulation, and also in pottery. Exposure was assessed on a subgroup level through measurement of airborne fiber concentrations both indoors and outdoors. During a 10-year observation period, 24 cases of MPM were diagnosed within the cohort corresponding to an annual incidence rate close to 130/100,000. This incidence rate exceeds the expected in the general Turkish population by more than 100-fold, and is comparable to risks of MPM observed in occupational settings with much higher exposures.

Design: Cohort-study
Outcome: Malignant pleural mesothelioma
Exposure: Inhalation of dust from asbestos in soil
(Metintas et al. 2002)

Case 3 is an example of a retrospective cohort study examining the incidence of malignant mesothelioma among people in villages exposed to white soil containing asbestos compared to the incidence among the background population in Turkey. It is an example on how rare exposures can be studied if highly exposed cohorts are chosen. The authors sampled 11 out of 403 villages, and it is difficult to rule out if a selection bias is present. Exposure levels were measured in white soil and information on potential confounding factors was obtained by interview with relatives. As it is often the problem with retrospective studies, the information about confounders is limited and no information on smoking habits was obtained.

III. EXPOSURE ASSESSMENT

Outcome and exposure assessment are equal partners in a well-balanced epidemiological study. Accordingly, the very same questions on validity, bias, or confounding should be considered. Further, the representability of the exposure assessment with respect to individual, time, and place should be scrutinized. Thus, exposure assessments include qualitative as well as quantitative questions. This section will illustrate some common problems related to exposure assessment, but for a more in-depth discussion of this theme, readers are referred to more specific literature (a starting point is included in the Further Reading list).

A. What?

The qualitative questions relate to the validity of the analytical methods: (1) what is measured; (2) what is the specificity of the method, and (if the exposure is a mixture), (3) does the mixture change qualitatively over time or between areas included in the study. In relation to exposure to metals, the analysis of the total concentration of metal is often insufficient as different metal species will have specific toxicological profiles. Thus, for the assessment of intake of mercury or lead from soil by children, it would be relevant to know the species and salts occurring, as the intestinal absorption as well as toxicity of these metals depends on these features. Likewise, the authors in Case 3 use a well-validated method for fiber collection and only analyze and report the fraction of fibers (>5 um) relevant for the outcome (mesothelioma). This is a relevant approach, but it is not clear from the description in the article what proportion of the larger fibers analyzed were asbestos fibers. Neither is the exact type of asbestos described, which could be expected to influence the risk for the exposed individuals. These queries are especially important if the exposure is expected to vary qualitatively over time or between geographic regions included in a study.

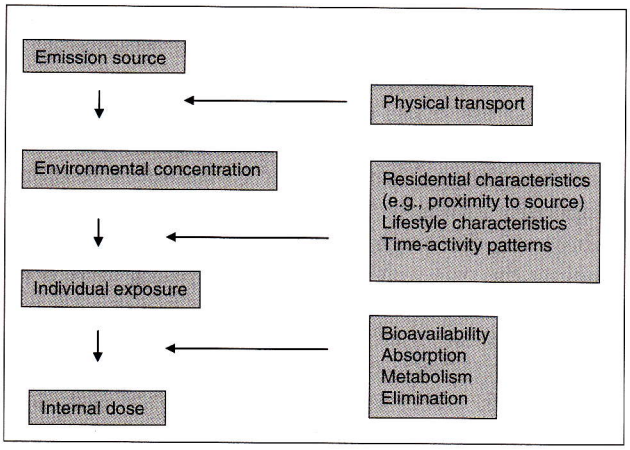

FIGURE 2 Sources of variability in exposure assessment.

B. Who?

Exposure may be assessed through measurements at different levels, beginning with very unspecific measurements at the emission source and ending with specific measurements of internal dose (Figure 2). Ecological studies (like Case 1) will use exposure assessments at a very crude level, which causes severe uncertainty with respect to the individual exposure. Thus, in Case 1, 83% of the drinking water resources in the arsenic-endemic region had arsenic concentrations above 0.9 ppb. It is, however, not known whether the women with adverse pregnancy outcomes actually consumed water from these wells or from the 17% with uncontaminated drinking water nor how much water they drank. Advancing the exposure assessment will clarify different factors of variability and increase the knowledge of individual exposure (Figure 2). An important point is, however, that the interindividual differences in exposure will not necessarily decrease with more specific exposure data. Instead, the uncertainty will be replaced by variability that may potentially be used in modeling or otherwise taken into account. In cohort or case-control studies, the possibility of getting more specific information on exposure is better as illustrated in Cases 2 and 3, where the exposure was assessed in the homes of the participants. However, the inherent problem that the exposure assessment will also have to be done retrospectively needs to be addressed. Prospective studies will have the advantage that exposure assessment can be planned in advance. Individual exposure and information on internal dose requires personal monitoring equipment or the use of biomarkers of exposure. These approaches are, however, often time-consuming and expensive, and they may not be an option in larger cohort studies.

C. How Much?

Using exposure assessments without individual exposure data raises some general questions relating to the representativity of the concentrations measured in the environment. At their best, these concentrations will represent an average exposure, but may over- as well as underestimate individual exposures. In Case 3, the concentration of airborne fibers was measured in a few homes in each village, but it is not known whether these exposure levels represent the exposure in the remaining houses. In Case 2, radon was measured in 98% of the homes, but not related to the time spent indoors. Besides averaging exposures between persons and locations, a measurement of a contaminant in the environment will be a snapshot in time. A single measurement or even a few will not be very informative regarding variation in exposure levels over time. The ideal exposure assessment in an epidemiological study is seldom achieved, but the key information that allows for a useful exposure assessment is quantitative information on changes in exposure over time and between locations. This will help exchange uncertainty with variability.

D. When?

Knowledge of variations in exposure over time is critical for several outcome measures. This information is needed if the health outcome depends on specific windows of susceptibility within the exposed population. Thus, for adverse pregnancy outcomes, neither information on exposures dating 20 years back nor information on exposure after birth is relevant. Likewise, when the outcome appears after a delay in time, e.g., cancer, current exposure is not relevant. Thus, in Case 3, the most relevant information on exposure to asbestos fibers probably dates 10–20 years back.

Information on present exposure levels is mainly useful if the outcome appears without delay, in prospective studies, or when the exposure can be assumed to be unchanged over time. However, most exposures change with time. If risk is a function of time of exposure, exposure profiles including information on variation will be valuable (Figure 3). In these profiles, the exposure concentration or dose is plotted as a function of time. Concentration versus time is used to describe the exposure, while amount versus time characterizes dose. If the elimination rate of the chemical is known, dose characterization may be used to estimate accumulation of contaminants. Further, exposure profiles may be used to identify more limited time periods with higher than

Exposure/dose profiles are important
- When risk is a function of time of exposure
  - Reprotoxicity
  - Neurotoxicity
- When risk may occur following short peak exposures
- When a chemical is accumulated
  - Lead, mercury, etc.
- When outcomes appear after delay
  - Cancer

FIGURE 3 Importance of exposure/dose profiles.

average exposures, which may be relevant for some outcomes. Thus, a single short-term exposure to very high concentrations may induce adverse effects, even if the average exposure is much lower than an apparent no-effect-level. Such short-term peak exposures clearly remain unidentified if only average values are available. In other exposure scenarios, however, average values are sufficient to perform a valid exposure assessment. Therefore, an epidemiological study protocol should include careful considerations concerning the most relevant collection of data on exposure.

E. Modeling Exposure

Several models, e.g., Monte Carlo simulations, have been developed through recent decades to estimate risk and exposure, to assess changes in exposure over time, or to identify worst case scenarios. These models are often very useful as they are able to accommodate and use vast amounts of information on parameters of importance for modeling individual exposures. If physiological and behavioral parameters are included, these models may even estimate target organ deposition. One of the major achievements of these models is that they have enabled risk and exposure assessors to replace the often very conservative estimates of worst case scenarios with more realistic scenarios based on probability functions. It is, however, important to remember, that a model never gets more valid than the validity of the exposure information obtained and entered into the model.

IV. BIAS

Two types of errors may occur in epidemiological studies: random and systematic errors. Random errors are, as the word implies, random and are minimized when the study size or the precision of information is increased whereas systematic errors are unaffected by the size of the study. If a participant is weighed on an imprecise weighing scale, his or her weight may be over- or underestimated, but the error is random. Therefore, an increase in number of participants will reduce bias. If, however, the weighing scale is systematically overestimating the true weight of the participants, there is no effect of increasing the number of participants. It will still overestimate their weight, and a systematic error is introduced.

Systematic errors are often referred to as bias. Bias may be defined as any systematic error in an epidemiological study that causes an incorrect estimate of the association between exposure and outcome. Because epidemiological studies involve humans, even the most perfectly designed study will have the potential for one or more types of errors. Consequently, evaluating the role of bias as an alternative explanation for an observed association is a necessary step in the interpretation of any study. Therefore it is essential to discuss types of biases that might be present as well as the most likely direction and magnitude of their impact. A study can be biased because of the way in which the study subjects are selected (selection bias), the way the study variables are measured (information bias), or by the lack of measurement of other exposures related to the outcome (confounding).

A. Selection Bias

Selection bias is a systematic error in the study that occurs when the association between exposure and outcome differs for those who participate and those who do not participate in the study. The participation rate in a study is never 100%, therefore it is important to gain information about age and sociodemographic status from non-participants and compare these with the participants.

Selection bias may change the estimates both toward and away from the null hypothesis. Selection bias is of particular importance if the participation rate is low, varies between cases and controls or between exposed and unexposed. In Case 2, 70% of the lung cancer patients and 61% of the controls participated. Patients may be more interested in participating as they believe that the exposures studied may have caused their disease.

A special form of selection bias occurs when the prevalence of an outcome for a group of workers is com-

pared with the prevalence for the general population. This comparison is biased because the general population includes many people who cannot work because they are too ill. Consequently, the outcome is more frequent in the general population. This bias is often referred to as healthy worker effect.

B. Information Bias

Information bias is caused by systematic differences in the way data on exposure or outcome are obtained from the various study groups. The participants are thereby misclassified with respect to either exposure or disease. This misclassification can be either differential or non-differential. Consider the smoking information in Case 2. If both cases and controls underestimate the number of cigarettes that they smoke on average, this would lead to a categorization of heavy smokers as light smokers. The classification of exposure (smoking) is unrelated to the outcome (lung cancer) as both cases and controls underreport to the same extent. The information bias is therefore a non-differential misclassification. A non-differential misclassification will produce estimates of the effect that are diluted and will tend to support the null hypothesis. Now imagine that lung cancer patients underreport their smoking to a greater extent than controls. The classification of exposure (smoking) is then related to the outcome (lung cancer), and the bias is differential and may under- as well as overestimate the effect.

A common type of information bias is recall bias, which may occur in case-control studies where a subject is interviewed to obtain exposure information after the outcome has occurred. Cases then tend to have a different recall than controls as a result of their disease. In Case 2 lung cancer patients and controls were interviewed. If lung cancer cases remember and report their smoking differently than controls, this causes a recall bias. Case 2 was further complicated by the fact that some cases had died by the time of the investigation and next of kin were interviewed instead. They may not remember exposures as precisely as the cases themselves.

V. CONFOUNDING

Confounding is a mixing of effects. A confounder is an exposure other than the one investigated, which is associated with outcome, but unequally distributed between the groups compared. Furthermore, it must not be an intermediate step in the causal pathway from exposure to disease. Confounding can cause bias in either direction.

In Case 2, smoking is a confounder. It is associated with the outcome (lung cancer) and unequally distributed among cases and controls (92% of cases and 55% of controls smoked). Furthermore, it is not an intermediate step in the causal pathway between radon exposure and lung cancer. Therefore, if the study does not take smoking into account, the effect of radon exposure would be overestimated as the estimate would really measure the aggregated effect of smoking and radon on lung cancer.

Confounding may be controlled by restricting the study population, thus all participants are equal with respect to a potential confounder (e.g., restricting the study to a specific age category). In Case 2, the study could have included only non-smokers. Another way to deal with confounding is by matching the study subjects with respect to the confounder. In Case 2, a smoking control could be included every time a smoking case was included. Matching poses special challenges and is not discussed further in this chapter, but readers are referred to epidemiological textbooks (see Further Reading section). Confounding control may also be addressed during analysis of the data by multiple regression analysis or by stratifying data, i.e., study lung cancer and radon exposure among smokers and non-smokers separately.

VI. STATISTICS

The majority of statistical analyses involves comparisons between groups of subjects. Initially, a hypothesis, called the null hypothesis, states that there is no difference in the outcome of interest between the groups of subjects. Statistical analysis is then an evaluation whether to accept or reject this hypothesis. The selected study subjects are only subsamples of the entire population, and probabilities are used to describe the certainty by which the null hypothesis is rejected. This probability is given as a p value in most statistics. Thus, a p value of 0.05 means that there is 95% certainty that the null hypothesis is not true and should be rejected. There is, however, a remaining probability of 5% that the rejected null hypothesis was actually true. This is known as a false positive result and termed a type I

error. Thus, the risk of a type I error is determined by the size of the p value that is used as the level of rejection of the null hypothesis. False negative results may also occur, i.e., the acceptance of a null hypothesis that is not true, and this is called a type II error. In Case 1, the authors use a probability of $p < 0.001$ to conclude that the mean birthweight is different in the two regions. There is only a 0.1% risk of a type I error, e.g., that the difference observed is actually only a chance finding and not true. Type I and type II errors are interdependent. Thus, whenever the risk of a type I error is reduced, i.e., by decreasing the p value used as level of rejection of the null hypothesis, the risk of a type II error is increased and vice versa. For a thorough statistical explanation, readers are referred to statistical textbooks.

In many epidemiological studies large numbers of comparisons are made between different subgroups within the observed group of subjects. If a probability of $p = 0.05$ is used as the level for rejection of the null hypothesis, this means that for every 20 comparisons one will, just by chance, be a false positive finding. Therefore, most statistical packages include methods to reduce this risk of type I errors when doing large numbers of comparisons. Failing to apply these methods in multicomparison scenarios may invalidate conclusions.

VII. CHECK LIST FOR EVALUATING AN EPIDEMIOLOGICAL PAPER

Reading and understanding an epidemiological study report may be time-consuming and it is often difficult to evaluate its validity. We have tried to develop a checklist to guide the reader through the most pertinent questions relating to the validity, strengths, and weaknesses of an epidemiological paper. Answering the questions on the checklist will make the reader recognize possible problems in a paper.

1. What type of study design was used?

See strength and limitations for the different design in Table 1.

- Cross-sectional study—The problem with cross-sectional studies is that information about exposure and outcome is collected at the same point in time. It is therefore impossible to draw conclusions about causation.
- Ecological study—No individual exposure information is collected in an ecological study. It is therefore impossible to infer causation in these studies.
- Case-control study—Exposure among a group of cases selected on the basis of an outcome, i.e., disease, is compared to exposure among a control group without the outcome.

  Study Design:

  - How were the cases defined?
  - How were the controls selected? Community or hospital controls?
  - Were cases and controls from the same source population?
  - Were cases and controls matched?
  - What was the number of controls per case?

  Study population:

  - What was the target population?
  - What recruitment procedures were used?
  - Did the participation rate among cases and controls differ?

  Validity:

  - Over what time period was the study population recruited?
  - Were the cases and controls comparable with respect to characteristics, response rates and time of recruitment?
  - Was any information about non-responders obtained?
- Cohort study—A cohort is defined and categorized into exposed and unexposed, who are then followed in order to determine new occurrences of the outcome.

  Study Design:

  - Was the cohort defined retrospectively or prospectively?
  - How large was the cohort?
  - How many were exposed?
  - How many observed events were there?

  Study population:

  - What was the target population?
  - What recruitment procedures were used?
  - How many were lost to followup (percentages)?

  Validity:

  - Over what time period was the study population recruited?

- Was there any description of the losses to followup?
- Do the losses to followup introduce bias?

2. What were the hypotheses of the study?

- Was the study originally designed to test these specific hypotheses?

3. How was the data quality?

- Were the data collected for the purpose of the study or were they obtained from other sources, for example, registers or hospital files?

4. How was exposure assessed?

- Who provided the information about exposure (subject, family or others)?
- Was the quality of exposure information assessed?
- Were the subjects and/or interviewer blinded to the hypothesis?
- How was the exposure information linked to the cases?
- Is risk a function of time, and were exposure profiles included?

5. How was information about outcome measured?

- Self-reported, by health personnel or from registers

6. Was adequate statistical analysis used?

- Did the analyses control for potential confounders?
- How wide were the confidence intervals?
- Is a type 1 or type 2 error possible?

7. Bias

Selection and information bias:

- It should be detected from the questions asked under case-control and cohort studies.

Confounders:

- Was adequate information about the confounders obtained?
- How accurate and adequate was the information about confounders?
- Did the study control for confounding by restriction, matching, stratification, or multiple regressions?

8. Have other studies reported similar findings (consistency)?

- Have a number of studies conducted by different investigators, in different geographical areas, and among different cultures at various points in time using different methodology found similar results?
- Lack of consistency should lead to a high degree of caution at any causal interpretation of the findings.

9. What is the strength of the association?

- The magnitude of the observed association is useful to judge the likelihood that the exposure itself affects the risk of developing the disease, and therefore, the likelihood of a cause-effect relationship. Specifically, the stronger the association—that is the greater the magnitude of the increased (or decreased) risk observed—the less likely that it is merely due to the effect of unexpected and uncontrolled confounding. This does not imply that a weak association cannot be causal, merely that it is more likely to exclude alternative explanations.

10. Is there a plausible biological mechanism of action and do experimental studies show similar results?

- Because what is considered biologically plausible or tested in animal studies at a given time depends on the current knowledge, the lack of these criteria do not necessary mean that a relationship is not causal.

11. Did the exposure precede outcome?

- Many lifestyle factors are likely to be altered as the first symptoms of a disease appear.

SEE ALSO THE FOLLOWING CHAPTERS

Chapter 2 (Natural Distribution and Abundance of Elements) · Chapter 11 (Arsenic in Groundwater and the Environment) · Chapter 18 (Natural Aerosolic Mineral Dusts and Human Health) · Chapter 22 (Environmental Medicine) · Chapter 23 (Environmental Pathology)

FURTHER READING

Basic epidemiology:

Rothman, K. J. (2002). *Epidemiology. An Introduction*, Oxford University Press, New York.

Rothman, K. J., and Greenland, S. (1998). *Modern Epidemiology*, second edition, Lippincott-Raven, Philadelphia, PA.

Exposure assessment:

Paustenbach, D. J. (2000). The Practice of Exposure Assessment: A State-Of-The-Art Review, *J. Toxicol. Environ. Health*, Part B, 3, 179–291.

REFERENCES FOR CASE-STUDIES

Barros-Dios, J. M., Barreiro, M. A., Ruano-Ravina, A., and Figueiras, A. (2002). Exposure to Residential Radon and Lung Cancer in Spain: A Population-Based Case-Control Study, *Am. J. Epidemiol.*, 156, 548–555.

Metintas, S., Metintas, M., Ucgun, I., and Oner, U. (2002). Malignant Mesothelioma Due to Environmental Exposure to Asbestos—Follow-Up of a Turkish Cohort Living in a Rural Area, *Chest*, 122, 2224–2229.

Yang, C.-Y., Chang, C.-C., Tsai, S.-S., Chuang, H.-Y., Ho, C.-K., and Wu, T.-N. (2003). Arsenic in Drinking Water and Adverse Pregnancy Outcome in an Arseniasis-Endemic Area in Northeastern Taiwan, *Environ. Res.*, 91, 29–34.