© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

CHAPTER 2

Basic Elements of Research, Exposure, and Outcome Assessment

LEARNING OBJECTIVES

After completing this chapter, you should be able to:

1. Distinguish between a research problem, a research question, and a hypothesis.
2. Define "variable" and distinguish between dependent and independent variables.
3. Know the primary types of data.
4. Describe direct and indirect methods of exposure and outcome assessment.
5. Define hypothesis.

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

Introduction

In epidemiologic research, once the health problem has been empirically established, an investigation follows, employing scientific reasoning and methods. The scientific method involves beginning with information or data obtained through observation of the phenomenon of interest. Hypotheses are formed and then tested by further observation or experimentation. Significant results then provide an important piece of evidence used in causal inference. The scientific method is intended to define, classify, or categorize events and their relationships with potential causes; identify causal associations; provide a basis for predicting the effects of certain exposures; and use this information to prevent and control health problems.

Prior to the investigation, the researcher develops a statement of the research problem and identifies appropriate variables, data, and hypotheses. These basic elements of research help transform ideas into concrete research operations. In this chapter these terms are defined and discussed in the context of environmental epidemiology.

Research Problem

The general epidemiology question is how a susceptible host, agent of disease or injury, and permissive environment interact in time and space to produce disease or injury. Formulating this interrelationship is the essence of the **research problem**. The research problem should be clear and specific; otherwise the purpose and methods of the research become meaningless. A good research statement of a health problem is often simple, considering one exposure and one outcome variable such as organic solvents and brain cancer. This does not mean that the economic, social, cultural, and/or political contexts of the problem should not be considered, but these contexts are secondary to the primary problem. The research problem should also be specific with regards to the study subjects and variables to be employed in the analysis. Who is the specified population of interest and what variables will allow us to effectively assess the suspected exposure and health outcome? The research problem should be written prior to the outset of the study in order to focus the research efforts on a specific objective and provide a strong basis for interpreting the results.

Once the research problem is formulated, it is followed by the research question and hypothesis. **Research questions** ask about associations between exposures and health outcomes. A hypothesis is then formulated to predict the result. For

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

example, the research problem may be that there is an elevated rate of hospital admissions for children with respiratory problems in a community that is located near a coke-works site.[1] The research question could be: "Is proximity of the residents to the coke-works site associated with hospital admissions of children for respiratory problems?" The research hypothesis could be that "the risk of hospital admissions for respiratory problems for children increases with proximity to the coke-works site."

Variables

The research process begins by forming concepts to describe health problems in specified populations. A concept is a general idea or understanding derived from specific instances or occurrences.[2] Movement from the conceptual to the empirical involves converting concepts into variables. An empirical approach emphasizes direct observation and experimentation, and variables are used to evaluate hypotheses. A **variable** is a characteristic that differs from one observation to the next and can be measured or categorized. For example, "sex" is a variable because it can be male or female; "severity rating" is a variable because it can take on values like mild, moderate, high; "number of children" is a variable that is limited to integer values (0, 1, 2, etc.); "distance" in meters from one or more hazardous waste sites is a variable because it has different values per individual; micrograms per liter of arsenic in drinking water is a variable because it can take on any value. Consider the variable "temperature," which can take on any value (e.g., 75.3544 . . .°). On the other hand, "hot" is not a variable, but a description.

Evidence from observation provides knowledge about potential environmental exposures, health risks, and disease occurrence. In environmental health we are often concerned with identifying the change in one variable, called the response or **dependent variable** (e.g., heart disease), as it relates to a given change in another variable (e.g., ground-level ozone), called the explanatory or **independent variable**. The distinction between these two types of variables relates to the research purpose; that is, the researcher chooses how to view the variables, basing the decision on the research problem. Variables may also be included in an analysis to reduce the risk of attributing explanatory power to the suspected exposure variable that in fact may have no association with a disease outcome (i.e., confounding).

An environmental epidemiologic study considers those variables appropriate for addressing the research question. Data are then obtained on those variables under consideration. The compilation of data over all the persons in the study makes up

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

the data set used in the analysis portion of the study. An exception is the ecologic study design where the unit of analysis is on the population level; that is, the unit of analysis represents aggregated data rather than individual level data.

Data

The growth of empirical data collection associated with the environment and health has been astonishing in recent decades. Data about environmental conditions and environmental health risk factors have been widely disseminated in the media and the professional literature. Data motivate and support research hypotheses. **Data** may be thought of as observations or measurements of a phenomenon of interest such as exposure to environmental contaminants or disease information collected about a patient, family, or community. Data are usually established by observation, measurement, or experiment for a select number of variables.

One of the simplest forms of data is **nominal data**, which fall into unordered categories (see Table 2.1). Nominal data are sometimes called qualitative data because they describe the quality of a thing or a person. Distinct levels differ in quality, not quantity. The categories are often represented by numbers. For example, researchers of a study may assign males a value of 1 and females a value of 2. Despite the fact that the sexes are labeled with numbers, both the order and the magnitudes of the numbers are not important. In addition, information that can be captured that is nonnumerical in nature is qualitative data. Examples include text (spoken, written, etc.) and action (video).

When the ordering among categories is important, the data are referred to as **ordinal data**. For example, researchers may assign levels of tension that study par-

TABLE 2.1 Types of data

	Description	Examples
Nominal	Categorical – unordered categories	
	Two levels – dichotomous	Sex, disease (yes, no)
	More than two levels – multichotomous	Race, marital status, education status
Ordinal	Categorical – ordering informative	Preference rating (e.g., agree, neutral, disagree)
Discrete	Quantitative – integers	Number of cases
Continuous	Quantitative – values on a continuum	Dose of ionizing radiation

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

ticipants experience in their shoulders as 1 never, 2 rarely, 3 occasionally, 4 often, and 5 constantly. A natural ordering exists in the data, with higher numbers representing a higher level of tension. However, as in the case of nominal data, the magnitude of the numbers is not important. We could have reversed the numbers, letting 5 represent never and 1 represent constantly. Consequently, it is not appropriate to apply many arithmetic operations to nominal or ordinal data.

For **discrete data**, both ordering and magnitude are important. Here the data represent quantities (integer values), not just labels. A natural order exists among the data. Because it is meaningful to measure the distance between levels of discrete observations, application of arithmetic computations is appropriate, such as the mean, median, and mode. Unlike discrete data, which is restricted to take on only certain integer values, **continuous data** represent measurable values on a continuum. The difference between two levels of a continuous variable can be arbitrarily small.

If a lesser degree of detail is desired than is available with continuous data, we may transform the data to a discrete, ordinal, or dichotomous form. For example, it is usually not necessary to know a person's exact age (e.g., 18.73 . . .), but to obtain the person's age as an integer value (e.g., 18); some continuous environmental exposures may have a danger threshold (e.g., safe versus dangerous); and it may be adequate to merely identify someone as ill or not ill rather than obtaining the level of the illness (e.g., his or her temperature).

Risk Assessment

Risk assessment is a tool to integrate exposure and health effects in order to identify the potential health hazards in humans. Risk assessment is also used to associate exposure and ecological effects in order to characterize the potential for other hazards in the environment. The definition of risk assessment used by the U.S. Environmental Protection Agency (EPA) is "[r]isk assessment is a process in which information is analyzed to determine if an environmental hazard might cause harm to exposed persons and ecosystems."[3]

Risk assessment involves an array of techniques to measure or estimate whether the exposure poses a threat to health or the ecosystem. Exposure is described more fully in the next section. Later in this chapter, direct and indirect measures of exposure are also discussed. As presented in the previous chapter, factors that influence the toxicity severity of a substance that enters the body include route of exposure, duration of exposure, and concentration of exposure, among other things

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

(see the subsection entitled Toxicokinetics in Chapter 1). Risk assessment is the formal process that is used to establish regulations and standards by the EPA and other organizations. For example, risk assessment has led the Occupational Safety and Health Administration to set a limit of 0.2 milligram of polycyclic aromatic hydrocarbons (PAHs) per cubic meter of air (0.2 mg/m^3). The permissible exposure limit for mineral oil mist containing PAHs is 5 mg/m^3, averaged over 8 hours of exposure. The National Institute for Occupational Safety and Health recommends that coal tar products should not exceed 0.1 mg/m^3, averaged over 10 hours of exposure. Other limits have also been set for workplace exposure to coal, mineral oil, and other things containing PAHs.[4]

Risk assessment is only one step in the **risk management** process. Risk management involves the integration of recognized risk, risk assessment, development of strategies to manage risk, and mitigation of risk through managerial resources.[5] The aim of risk management is to reduce risks to levels deemed acceptable by the community. Risk management includes those types of threats to human health caused by environmental factors. The risk management process is extensive and goes beyond the scope of this book. The interested reader should refer to other sources on this topic, such as the EPA's General Risk Management Program Guidance (http://yosemite.epa.gov/oswer/ceppoweb.nsf/content/EPAguidance.htm#General) or the Risk Management Standard (http://www.theirm.org/publications/PUstandard.html).

Exposure Data

An **exposure** may represent an actual exposure (e.g., toxic chemical or microorganism), a behavior (e.g., where one works or socializes), or an individual attribute (e.g., age, sex, race). A causal association may be implied in a descriptive study without a direct measure of exposure. However, greater confidence in causal associations between environmental factors and human health requires accurate identification of the primary mechanism (physical, chemical, biological, psychosocial) of the environmental contaminant or stress, and determination of how and in what form the environmental contaminant or stress comes in contact with people. It is necessary to establish whether there is evidence that the environmental contaminant or stress is capable of harming human health. This requires an accurate assessment of exposure.

Measuring the intensity and duration of exposure is often necessary for supporting causal association. Exposure may involve an intense dose over a relatively

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

short period of time, or a low-level prolonged dose over a period from weeks to years. The effects of acute, high-dose environmental exposures may appear within hours or days (e.g., sunburn). On the other hand, the effects of chronic, low-dose exposures may not appear until years later (e.g., cancer).

Identifying an association between dose and an adverse health outcome provides support for causality. The quality of the exposure measurements influences the validity of the study. Validity refers to the "truthfulness" of a measure. A valid measure of a concept measures what it claims to measure. This is often a challenge when the environmental exposure, such as a hazardous chemical or radiation fallout, occurred in the distant past. Reconstructing past exposures may be complicated by limited recall, incomplete measurements, inaccurate records, and variability of exposure from person to person. In this situation, a direct measure of the past exposure may not be possible and can require estimation through modeling.

An exposure may be a specific event and relatively easy to measure, such as exposure to a chemical leak. Other exposures can be subdivided into dose or duration (e.g., number of glasses of water, number of years worked in a coal mine). A disease may require a minimal level of exposure and increase in probability with longer exposure. Such a relationship between exposure and disease may be missed with a dichotomous measure characterizing the presence or absence of the exposure. In many cases it is more appropriate to use ordinal or continuous measures of the exposure, especially when trying to assess a dose–response relationship. It may be useful to restrict the study cohort to those who are most likely exposed or to those with the most years of exposure. This may increase the probability of finding a dose-related effect while increasing the efficiency of the study by requiring fewer numbers of participants.

Although it is often desirable to isolate the association between a given exposure and disease outcome, assessing the combined effects of multiple exposures on human health is a potentially important approach from a public health perspective because public health interventions often target an overall exposure scenario.[6] Combined effects may reflect contaminants where the total effect differs from the sum of the individual effects. For example, automotive exhaust can produce volatile organic compounds and nitrogen oxides that when combined with sunlight produce ground-level ozone.

Measurement of an exposure variable on a continuous scale is the most informative for evaluating associations. Continuous data allow us to measure a dose–response relation between variables. However, in some cases exposure information is available only on a nominal scale—exposed versus unexposed or likely exposed

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

versus unlikely exposed. This may be the only alternative when there are perceived versus documented exposures or the exposure occurred in the past and direct measurement is not feasible. There also exists the basic question as to what to measure; for example, should the peak exposure or the cumulative exposure be measured?[7] When determining what to measure, consider what relates most closely to the incidence and magnitude of a biological response. Is it maximum concentration, average concentration, minimum concentration, or total dose of the biologically active contaminant?[8]

There are both direct and indirect types of data for approximating exposure (see Table 2.2). Data most appropriate for assessing risk factors are those obtained from personal monitoring and use of biological markers. Personal measurement allows for assessment of the contaminant. Biological markers are useful for representing total dose to the body from multiple routes of exposure. These data can provide exposure measures on a continuous scale, which is useful for identifying adverse health outcomes according to dose and whether a threshold exists. Data are also useful that reflect information obtained from quantifying the concentration of toxic contaminants in a specific environment (air, water, soil, and food). These provide direct measures of dose. The remaining types of data listed in the table provide indirect measurements of dose. These data are easier to

TABLE 2.2 Direct and indirect measurements of dose

Direct
Quantified personal measurement
Quantified area or ambient measurements in the vicinity of the residence or other sites of activity
Indirect
Estimates of drinking water use, food use, etc.
Distance from site **and** duration of residence
Distance from site **or** duration of residence
Residence or employment in geographic area in reasonable proximity to site where exposure can be assumed
Residence or employment in defined geographical area of the site (e.g., a county)

Source: Modified from National Research Council, 1997, p. 120.[9]

For a list of advantages and disadvantages of various exposure methods see Armstrong, White, and Saracci (1992).[10]

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

obtain but obviously less precise. Limited time and money may make it necessary to rely on indirect measures of exposure. Although causal inference is strongest when exposure information is directly measured, surrogate measures of exposure can also provide important insights into causality.

Exposure Assessment Methods

Direct Measures

Direct measures of exposure include personal monitoring and use of biologic markers. Personal monitoring involves quantitative measurements of personal exposure to environmental physical stresses, chemical or biological agents, and psychosocial milieu. Personal monitoring devices are worn by individuals while they pursue their normal activities, most often in the workplace. For example, an individual may wear a dosimeter to estimate total exposure to radiation in the workplace through the air, water, and food. A personal air monitor may be worn to measure exposure to air pollutants in the home.

Biologic markers are those specific anatomic, physiologic, biochemical, or molecular characteristics used to measure the presence and severity of a disease or condition. Biological monitoring can involve measurements of concentrations in human tissues (blood lead), metabolic products (dimethylarsinic acid in urine after arsenic exposure), or markers of physiologic effects (e.g., protein adducts induced by beta-naphthylamine in cigarette smoke).[6] Biologic markers can be measured through physical examinations, laboratory assays, and medical imaging. For example, heavy metals and some pesticides can accumulate in the body. Over time the risk of human harm increases. The pollutants, which reflect the amount of pollution in the environment, leave residues in the body that are usually measured in the blood or urine. The approach of measuring pollutant levels in tissue or fluid samples is called **biomonitoring.**[11]

Indirect Measures

Indirect methods for obtaining exposure information include questionnaires, surrogates, existing records, and diaries.

Questionnaires translate the research objectives into specific questions. Answers to these questions provide the data used in data analysis. Questionnaires may be administered through face-to-face interviews, over the telephone, and through the mail or Internet. In a population-based case-control study involving female

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

nonsmokers in Hong Kong, face-to-face interviews were conducted using a standardized questionnaire. There were 200 cases and 285 controls. Cumulative exposure to cooking by various forms of frying increased the risk of lung cancer. The increased risk of lung cancer was greatest when cooking involved deep-frying, followed by frying, and then stir-frying.[12] A telephone-conducted questionnaire of 1,009 American veterans (65% response rate) deployed and not deployed to the Gulf War found that 6% of non–Gulf War veterans reported being exposed to biological or chemical warfare compared with 64% of the Gulf War veterans. Veterans tended to associate exposure with having adverse physical symptoms and receiving an alert from the military.[13] In another questionnaire-based study, data were collected through the mail. This study involved 1,456 Australian Gulf War veterans (80.5% response rate) and a comparison group (56.8% response rate). The study found that the Australian Gulf War veterans had a higher than expected risk of respiratory symptoms and asthma and bronchitis.[14]

Questionnaire data rely on individual recall and knowledge and are thus subject to error. Bias may be introduced by the interviewer's inflections, expressions, gender, and appearance. Telephone interviews are becoming more difficult to conduct because of caller identification, cell phones, and a decreasing tolerance of telemarketing in the population. Historically, many national health surveys in the United States have relied on telephone surveys, yet their response rates have been steadily declining in recent years. For example, the average response rates for the Behavior Risk Factor Surveillance System survey, a national survey conducted every year to monitor health risk-factor behavior, fell consistently between 1997 and 2005, from 68.3% to 51.1%.[15]

Mailed questionnaires avoid interviewer influences but are subject to low response rates. In addition, they exclude individuals who cannot read and do not allow the responder to obtain item clarifications. Electronic mail questionnaires are becoming an increasingly popular way to obtain information because of their relative speed, low cost, and ability to attach pictures and sound files; they often stimulate higher response levels than "snail" mail surveys. However, some challenges to email surveys include obtaining (or purchasing) a list of email addresses, nonresponse to unsolicited email (which may be higher than unsolicited regular mail), and obtaining a representative sample of the general population.

General questions that may be used for characterizing exposure include:

- What is the mechanism (physical, chemical, biological, psychosocial)?
- How potent or intense was the exposure?
- What was the duration of exposure?

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

- Was there a relation between duration of exposure and disease?
- What is the exposure pathway (air, water, soil, food)?
- How did the pollutant enter the body?
- What clinical signs, if any, are associated with the exposure?

Some exposures may be represented by surrogate measures. Some examples include years of employment, census track, carbon monoxide in indoor air at home, trihalomethanes in water coming out of the tap, self-reported water consumption, and hot shower use. However, such measures are crude and subject to errors. A direct measure would be more accurate, but is often financially prohibitive. Halperin (2002) suggests that cost may be reduced by using an indirect measure of dose, like years of employment in a cohort, and then performing a nested case-control study using a direct measure.[16]

It may also be possible to obtain exposure information from existing records, such as hospital admission or discharge records, pathology records, and crisis assessment prevention intervention services. This approach avoids the problems of interviewer bias, recall bias, and response bias.

Several studies have used diaries to identify exposure. For example, in a prospective cohort of newlywed couples in two districts (Tiexi and Dadong) in Shenyang, China, investigators examined the effects of various environmental and occupational exposures on reproductive outcomes. The study consisted of 165 newlywed, nonsmoking Chinese women with no past history of dysmenorrhea (cramps or painful menstruation) at the time of enrollment. Enrollment began after couples obtained permission to become pregnant. Daily diaries were used to record exposure to environmental tobacco smoke until the occurrence of clinical pregnancy or for up to 1 year. Environmental tobacco smoke was defined as "the mean number of cigarettes smoked per day at home by household members over an entire menstrual cycle before the menstrual period." The adjusted odds ratios of dysmenorrhea associated with tertile groupings of environmental tobacco smoke exposure compared with no exposure were 1.1 (95% CI, 0.5–2.6), 2.5 (CI, 0.9–6.7), and 3.1 (CI, 1.2–8.3), respectively. These data indicate an increased risk of dysmenorrhea among women exposed to environmental tobacco smoke, more so with higher levels of exposure.[17]

Modeling

Pollutants are released from multiple sources (e.g., treatment storage and disposal facilities, industry, government facilities, households, and others). The **fate and transport** of contaminants is an important issue in managing hazardous

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

pollutants. Fate and transport involves groundwater, soil, gas, and atmospheric transport of chemicals. Of primary concern in fate and transport is determination of the transport speed and synergistic effects of chemicals in their environments. Considerable effort has gone into understanding the state and science of the fate and transport of selected substances like mercury in aquatic and terrestrial systems, their transformation processes, and approaches to effectively manage ecological and human exposures to selected substances.[18]

A number of models have been developed to estimate the magnitude of pollutants in the air, water, and soil:

- The Atmospheric Sciences Modeling Division contains information about atmospheric models.
- The Center for Exposure Assessment Modeling provides predictive exposure models for aquatic, terrestrial, and multimedia pathways for organic chemicals and metals.
- The Division of Computational Toxicology applies mathematical and computer models to predict adverse effects and describe the mechanisms through which chemicals may induce harm.
- The Support Center for Regulatory Environmental Model provides model guidance, development, and application.
- The Support Center for Regulatory Air Models is a source of information on atmospheric dispersion (air quality) models.
- ADL Migration Exposure Model was developed to estimate the migration of chemicals from polymeric materials used in the home.
- The Landfill Air Emissions Estimation Model was developed to estimate emissions of methane, carbon dioxide, nonmethane organic compounds, and hazardous air pollutants from municipal solid-waste landfills.
- The Multi-Chamber Concentration and Exposure Model was designed to estimate average peak indoor air concentrations of chemicals from products or materials in homes.
- AQUATOX is a freshwater ecosystem simulation model that predicts the fate of selected pollutants such as organic toxicants and their effects on the ecosystem.
- CHEMFLO is a model for simulating water and chemical movement in unsaturated soils.
- PRESTO-EPA-POP is a computer model for evaluating radiation exposure from contaminated soil layers.[19]

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

A Physiologically Based Pharmacokinetic (PBPK) model is a physiologically based compartmental (e.g., lung, liver, rapidly perfused tissues, slowly perfused tissues, fat, and kidney) model used to characterize pharmacokinetic behavior (absorption, distribution, metabolism, and excretion) of a chemical. This model has become the tool of choice for predicting the fate of environmental contaminants in humans. Data on blood-flow rates, metabolic, and other processes that the chemical undergoes within each compartment are used to construct a mass-balance framework for the PBPK model.[20] PBPK models have been developed for methylmercury,[21] cadmium,[22] lead,[23] and methyl tert-butyl ether.[24]

Outcome Assessment

The **outcome** refers to the disease state, event, behavior, or condition associated with health that is under investigation. Outcome status refers to the presence or absence of the health-related state or event. Although outcome status is typically measured as a dichotomous variable, it may also be measured as an ordinal variable (e.g., severe, moderate, mild, no disease) or a continuous variable (e.g., concentrations of lead in the blood). The type of data considered is often determined by accessibility.

Accurate assessment of outcome status requires a standard case definition and adequate levels of reporting. A standard set of clinical criteria, or case definition, will ensure that cases are consistently diagnosed, regardless of where or when they were identified and who diagnosed the case. Whatever the criteria, they should be applied consistently and without bias to all those under investigation. A specific case definition will minimize misclassification resulting in bias.

The clinical criteria to be a case, particularly in the setting of an outbreak investigation, may be restricted by person (e.g., to children less than 5 years of age), place (e.g., to employees at a certain work site), and time (e.g., persons with onset of illness within the past 48 hours) variables. The clinical criteria may include laboratory confirmation. Yet acquiring the biological media and having the expertise and resources for assessment may prove difficult. Clinical criteria may also involve a combination of signs, symptoms, and other findings. Clinicians characterize disease status by examining and analyzing the specific symptoms and performing tests on the patient.

The availability of outcome data may also be influenced by the acceptability of the diagnostic test. Diagnostic tests predict the presence of an outcome. Ideally, a

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

diagnostic test will always give the right answer; for example, a positive test in the presence of disease and a negative test in the absence of disease. The test should also be quick, safe, painless, inexpensive, valid, and reliable. The validity of a test is determined by the sensitivity and specificity of the test. **Sensitivity** is the proportion of patients with a given outcome who have a positive test, and **specificity** is the proportion of individuals without a given outcome who have a negative test. Reliability refers to a test's performance over time.

Rarely is a diagnostic test ideal. Validity and reliability of a test typically result in some misclassification. Some occupational and environmental disorders are often difficult to diagnose. In such cases, the outcome of interest is likely to be underestimated. While the best diagnostic tests should be employed, further development of diagnostic procedures is needed, particularly for occupational and environmental disorders that are difficult to diagnose.

The Study (or Research) Hypothesis

A **hypothesis** is a tentative suggestion that certain associations exist in certain activities or a chain of events. The initial hypothesis is generally based on observation, which refers primarily to empirical findings from data systematically collected. Once the hypothesis is in place, it is necessary to determine how to design the study and formulate it in statistical terms. The hypothesis applies to a specified population of interest.

When developing a hypothesis, potential exposures should be described according to their toxicity (including carcinogenicity) and other effects when information is available from experimental studies involving animals and, if available, humans. Health risks associated with potential exposures can be obtained through epidemiologic studies showing the health effects of analogous contaminants in other circumstances. In some instances exposure is not accurately described or a causal factor is not even specified. However, if an excess in disease is observed, based on knowledge about exposure and disease from other sources, it may be possible to hypothesize about the nature of causation.[25]

The methods used in environmental epidemiology are generally observational and not experimental, making it more difficult to make causal inference. In observational studies there is less control over exposure and outcome measures because it is more difficult to control for confounding and there is greater susceptibility to bias. This means that the researcher must rely on inductive methods for making

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

inferences about his or her data. Fundamental to the development of hypothesis testing is inductive reasoning. This is the process leading from a set of specific facts to general statements that explain those facts. Inductive reasoning relies on

1. exact and correct observation;
2. accurate and correct interpretation of the facts in order to understand findings and their relationship to each other and to causality;
3. clear, accurate, and rational explanations of findings, information, and facts in reference to causality;
4. development based on scientific approaches, using facts in the analysis and in a manner that makes sense, based on rational scientific knowledge.[26]

A hypothesis is not necessary in descriptive studies where the distribution of a single variable is being described. On the other hand, a hypothesis is required when an association is being evaluated between exposure and outcome variable.

Key Issues

1. The general epidemiology question is "How do a susceptible host, agent of disease or injury, and permissive environment interact in time and space to produce disease or injury?" Formulating this interrelationship is the essence of a research problem.
2. Movement from the conceptual to the empirical involves converting concepts into variables. Variables are used to evaluate hypotheses.
3. Data may be thought of as observations or measurements of a phenomenon of interest such as exposure to environmental contaminants or disease information collected about a patient, family, or community. Data are usually established by observation, measurement, or experiment for a select number of variables. The four primary types of data are nominal, ordinal, discrete, and continuous.
4. A hypothesis is a tentative suggestion that certain associations exist in certain activities or a chain of events. The initial hypothesis is generally based on observation, which refers primarily to empirical findings from data systematically collected. Once the hypothesis is in place, it is necessary to determine how to design the study and formulate it in statistical terms.

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

Exercises

Key Terms

Define the following terms.

Biomonitoring
Continuous data
Data
Dependent variable
Discrete data
Exposure
Fate and transport
Hypothesis
Independent variable
Nominal data
Ordinal data
Outcome
Research problem
Research question
Risk assessment
Risk management
Sensitivity
Specificity
Variable

Study Questions

2.1. List some aspects for a good research problem.
2.2. How does a research question differ from a hypothesis?
2.3. What is a variable? How do dependent variables relate to independent variables in a study? Give examples of dependent and independent variables.
2.4. Describe the four types of data presented in the chapter.
2.5. Compare and contrast exposure and outcome data.
2.6. Distinguish between direct and indirect measures of exposure.
2.7. What is the meaning of the term "fate and transport"?
2.8. Explain the stages of hypothesis development.

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

References

1. Aylin P, Bottle A, Wakefield J, Jarup L, Elliott P. Proximity to coke works and hospital admissions for respiratory and cardiovascular disease in England and Wales. *Thorax*. 2001;56:228–233.
2. *Webster's Ninth New Collegiate Dictionary*. Springfield, MA: Merriam-Webster, Inc; 1991.
3. Environmental Protection Agency. Risk assessment principles & practices. Available at: http://www.epa.gov/OSA/pdfs/ratf-final.pdf. Accessed June 10, 2007.
4. Agency for Toxic Substances and Disease Registry. ToxFAQs: polycyclic aromatic hydrocarbons (PAHs). U.S. Department of Health and Human Services, Public Health Service Agency for Toxic Substances and Disease Registry. Available at: http://www.atsdr.cdc.gov/tfacts69.pdf. Accessed June 10, 2007.
5. Wikipedia: The free encyclopedia. *Risk Management*. Available at: http://en.wikipedia.org/wiki/Risk_management. Accessed June 10, 2007.
6. Hertz-Picciotto I. Environmental epidemiology. In: Rothman K, Greenland S, eds. *Modern Epidemiology*. Philadelphia, PA: Lippincott-Raven; 1998.
7. Bailar JC. Inhalation hazards: The interpretation of epidemiologic evidence. In: Bates DV et al., eds. *Assessment of Inhalation Hazards*. New York, NY: Springer-Verlag; 1989:39–48.
8. Gillette JR. Dose, species, and route extrapolation: general aspects. National Research Council, *Pharmacokinetics in Risk Assessment, Drinking Water and Health.* Vol 8. Washington, DC: National Academy Press; 1989:96–158.
9. National Research Council. *Environmental Epidemiology: Use of the Gray Literature and Other Data in Environmental Epidemiology.* Vol 2. Washington, DC: National Academy Press; 1997.
10. Armstong BK, White E, Saracci R. *Principles of Exposure Measurement in Epidemiology.* New York: Oxford University Press; 1992. Monographs in epidemiology and biostatistics, Vol 21.
11. Environmental Protection Agency. Measuring exposures to environmental pollution. Available at: http://www.epa.gov/indicate/roe/html/roeHealthMe.htm. Accessed April 6, 2007.
12. Yu IT, Chiu YL, Au JS, Wong TW, Tang JL. Dose-response relationship between cooking fumes exposures and lung cancer among Chinese nonsmoking women. *Cancer Res*. 2006;66(9):4961–4967.
13. Brewer NT, Lillie SE, Hallman WK. Why people believe they were exposed to biological or chemical warfare: A survey of Gulf War veterans. *Risk Anal.* 2006;26(2):337–345.
14. Kellsall HL, Sim MR, Forbes AB. Respiratory health status of Australian veterans of the 1991 Gulf War and the effects of exposure to oil fire smoke and dust storms. *Thorax*. 2004;59(1):897–903.
15. Centers for Disease Control and Prevention. Technical information and data: Summary data quality reports. Available at: http://www.cdc.gov/brfss/technical_infodata/quality.htm. Accessed January 22, 2006.

© Jones and Bartlett Publishers. NOT FOR SALE OR DISTRIBUTION

16. Halperin WE. Field investigations of occupational disease and injury. In: Gregg MB, ed. *Field Epidemiology.* 2nd ed. New York, NY: Oxford University Press; 2002:321.
17. Chen C, Cho S, Damokosh AI. Prospective study of exposure to environmental tobacco smoke and dysmenorrhea. *Environ Health Perspect*. 2000;108:1019–1022.
18. Environmental Protection Agency. Proceedings and summary report: workshop on the fate, transport, and transformation of mercury in aquatic and terrestrial environments. Available at: http://www.epa.gov/nrmrl/pubs/625r02005/625R02005.pdf. Accessed April 7, 2007.
19. Environmental Protection Agency. *Models*. Available at: http://www.epa.gov/epahome/models.htm. Accessed April 7, 2007.
20. Environmental Protection Agency. *Glossary archive*. Available at: http://www.epa.gov/iris/gloss8_arch.htm. Accessed April 7, 2007.
21. Shipp AM, Gentry PR, Lawrence G. Determination of a site-specific reference dose for methylmercury for fish-eating populations. *Toxicol Ind Health*. 2000;16(9–10):335–438.
22. Nordbert GF, Kjellstrom T. Metabolic model for cadmium in man. *Environ Health Perspect*. 1979;28:211–217.
23. O'Flaherty EJ. Physiologically based models for bone-seeking elements. IV. Kinetics of lead disposition in humans. *Toxicol Appl Pharmacol*. 1993;118(1):16–29.
24. Licata AC, Dekant W, Smith CE, Borghoff SJ. A physiologically based pharmacokinetic model for methyl tert-butyl ether in humans: implementing sensitivity and variability analyses. *Toxicol Sci*. 2001;62(2):191–204.
25. National Research Council. *Environmental Epidemiology: Public Health and Hazardous Wastes.* Vol 1. Washington, DC: National Academy Press; 1991.
26. Shindell S, Salloway JC, Oberembt CM. *A Coursebook in Health Care Delivery*. New York: Appleton-Century-Crofts; 1976.