Exposure Assessment

I.11

Sylvaine Cordier, Patricia A. Stewart

11.1 Introduction

Accurate exposure assessment is a prerequisite for an efficient study design, more than ever before, because of the increasing challenges that epidemiology has to face to demonstrate low increases in risk, to disentangle mixed potential risk factors in disease causation, and to provide exposure-response relationships for policy makers.

Exposure assessment is the process that leads to establishing a dichotomy between exposed and non-exposed subjects, and/or introducing a level of classification between subjects. A prerequisite for any epidemiologic study is that there is variability of exposure to the agent of interest within a population and that this variability between subjects (inter-individual variability) will overcome individual variation of exposure (intra-individual variability).

This chapter will describe what choices have to be made for a proper exposure assessment depending on the pathological process under study, give an overview of the different instruments available for this assessment and highlight some specific difficulties in this process (retrospective assessment, ecological measurement or multiple exposures). Finally, measurement errors and ways for controlling them will be described.

11.2 Definition of Exposure and Exposure Assessment

Exposure can be defined as a contact of an individual with an agent through any medium or environment. An agent can also be thought of as a susceptibility characteristic. The agent is not necessarily considered to be harmful (e.g., exercise or fiber in the diet). *Exposure assessment* aims to identify whether a person is exposed or not (a dichotomous classification) to a particular agent and if the individual is exposed, to develop a ranking of subjects by exposure level.

11.2.1 Types of Exposure

An exposure may be to a chemical, a biologic, a physical, or a societal agent in the external environment (e.g., cadmium, endotoxin, ionizing radiation, and the existence of a support system, respectively). It may be a characteristic of an individual (e.g., weight or physical activity) or a perception of an individual (e.g., lack of control in the workplace). Finally, it may be a biologic agent in the body (e.g., herpes virus), a metabolite of an external agent (e.g., 1-hydroxypyrene, a metabolite of polycyclic aromatic hydrocarbons), a substance representing a pathway of action (e.g., DNA-PAH adducts), or the presence of a polymorphism (e.g., NAT wildtype). In this chapter we use the term exposure to apply to all of these, rather than

separating external agents from internal agents. The concept of dose is discussed later in this chapter.

True Risk Factor or Surrogate 11.2.2

Ideally, an exposure assessment should focus on the *true risk factor*. When true risk factors have been confirmed, protective measures and monitoring of exposure can then be implemented. Medical surveillance in the work place, which usually includes some kind of biological monitoring of compounds known for their toxicity (e.g.: urinary cadmium), may be required. In many situations, however, a *surrogate* must be evaluated because the true risk factor has not yet been identified or only a surrogate can be measured. For example, the causal role of inhaled benzo[a]pyrene in the carcinogenicity of cigarette smoking for the lung may never be formally proven because the true risk factor (i.e., the total amount of inhaled benzo[a]pyrene over a period covering many decades) is impossible to measure (Rothman and Greenland 1998). The International Agency for Research on Cancer (IARC) has classified certain work environments as probably carcinogenic to humans, without identifying the specific compound(s) responsible for this health effect (e.g., the process of refining nickel). Thus, although the true risk factor(s) linked to a health effect may not yet be identified or quantified (e.g., nickel refining, tobacco smoking), measurement of a surrogate remains very useful for research and public health purpose. A surrogate is useful for identifying factors of variation for the exposure, establishing presumptive causal associations and exposure-response relationships, and narrowing the search for the true risk factor(s).

Dose versus Exposure 11.2.3

The term *exposure* usually refers to contact with an agent in the external environment. (As indicated above, common nomenclature also may include agents in the body). Measuring an external agent should, but may not, take into account all the exposure sources (e.g., at home, at work, and leisure time), the time spent in each (i.e., activity patterns), and the individual susceptibility to this agent (e.g., due to physical exercise, diet, and physiological and genetic characteristics). These variables will affect the internal *dose* measured in human tissue or fluid. A biological marker of internal dose therefore comes closer to the relevant measure of exposure in some circumstances than an external exposure. This will be discussed more in Sect. 11.3.1. In the rest of the text, the term *exposure* will be used to describe agents that are being estimated for use in exposure- (or dose-) response relationships in an epidemiologic study. Dose will be used to describe the level of the true risk factor at the target organ.

Selection of Metric 11.2.4

Once the agent or a scenario to be investigated in the epidemiologic study has been selected, the relevant dimensions to quantify this exposure need to be de-

termined. The appropriate quantification of exposure (metric) should reflect the toxic mechanism of action for the agent and disease of interest. The choice of this metric depends on the knowledge about the supposed biological mechanism inducing the health effect. Chronic diseases such as cancer, for example, are thought to be a result of lifetime exposure, so that the exposure metric often studied is *cumulative exposure*; whereas acute diseases such as asthma are thought to be due to recent high exposures, so that the metric often studied is *peak exposures*.

If there is a biological level above which detoxification processes of the organism are impaired (threshold), the *dose-rate* (*average*) of an exposure or a peak exposure may be more relevant than cumulative exposure, because exposures below such a threshold would not cause any deleterious effect.

Oftentimes, however, the biological mechanism of the disease process is not known. In such cases, it is useful to explore multiple metrics such as cumulative (life-time), highest, average (dose-rate), highest short-term (peak) exposure, and components of these (e.g., cumulative exposure level or time above a particular exposure level). For example, the induction of carcinogenesis by a mutagenic compound is, theoretically, initiated at any dose, but the mechanism necessitates a long (sometimes several decades) induction period (*latency*). In this case, recent exposure (immediately preceding diagnosis) is not pertinent, and often measurement of past exposure is "lagged", i.e. exposure occurring in years just before diagnosis of the disease is not taken into account. The exposure metric, then, may incorporate a lagged latency. When an adverse effect is expected to occur only above a certain dose (*threshold*), for instance in acute toxicity, a metric representing a quantitative level above the threshold would be more appropriate than a metric estimating the total exposure.

Often, the total exposure to a given compound received over a particular time period (cumulative exposure) is the relevant parameter in a pathological process. There are, however, several ways to receive the same cumulative exposure: a high intensity for a short period of time or a lower intensity over a longer period. For instance, the history of tobacco smoking is often summarized by a cumulative index (pack-years), i.e., the number of years of smoking times the average number of packs of cigarettes smoked every day during the smoking period. This index, or any equivalent based on the product of duration of exposure by an intensity level, does not distinguish between the roles of *duration* of exposure, irrespective of the rate of exposure, and *intensity* of exposure at every instant.

Selection of an exposure metric that does not appropriately describe the pattern of exposure to the agent being investigated as it relates to the disease of interest will result in misclassification and loss of statistical power (see Sect. 11.5).

11.3 Exposure Data

Because exposures can have different natures, the sources of data used in exposure assessments differ. Exposure data can be thought to be of two types: measurement

data (*direct*) and *indirect* information (e.g., questionnaire information, diaries, and records of surrogate information).

Measurement Data

11.3.1

Measurement data are generally considered the most accurate type of exposure data because they are objective measures of exposure. Measurement data include measurements of chemical hazards on the skin and chemical or radiation hazards in the food, air, or water in the general environment or in the workplace. They may be measures of quality of life, such as levels of stress. They also include measurements of human health, such as physical activity levels, physiologic measurements, such as blood pressure or weight, or measurements of agents in biologic tissues, such as drugs or nutrients. They also include measures of internal exposure or effect, such as blood lead levels and DNA adducts, respectively. For more examples of biological markers please refer to Chap. III.6 of this handbook.

Measurements may be taken for purposes of an epidemiologic study or may be available from existing records. Although individual measurement data are often thought to be the gold standard, they can be subject to substantial biases. Measurements may not represent the intensity of exposure during the relevant time window, e.g., current levels of physical activity may not reflect earlier levels of physical activity. The number of measurements on any individual is generally small, and because the variability of some exposures is large (e.g., in air and in water), one or a few measurements may not reflect the metric of interest, such as long-term exposure levels.

In addition, historical measurement data in records may not represent the true exposure level, because the purpose of the data collection was taken for reasons other than to obtain an estimate of the exposure metric of interest to the study investigator. For example, measurements of agents in the workplace often have been taken to evaluate compliance with exposure regulations, and it has been speculated that such data may reflect higher exposures than the true long-term exposure level. Moreover, the analytical method may not have measured the true risk factor (e.g., historical measurements of cholesterol did not distinguish between high and low density cholesterol, and many historical measurements of dust in the air did not distinguish respirable dust from inhalable dust).

Biological measurements of exposure (e.g., carbon monoxide in the breath) *or of effect* (e.g., cholinesterase levels in the blood) are generally thought to be the gold standard, because they most closely reflect the dose received by the target organ. (Note that biologic measurements can be both exposure data and the outcome, depending on the study design. Here, only biologic measurements used as exposure data are discussed.) There are many limitations to this type of measurement, however. The variability of the concentration of an agent in the body is often greater than that seen in the external environment, so that if the number of measurements is limited, a mean of those measurements may not accurately reflect the average exposure. Some biologic measurements may not reflect the dose at the target organ. Instead, they may reflect the amount of agent that was not

received by the target organ (e.g., if the agent was measured in the urine) or the amount that was metabolized in the body (including by organs other than the target organ). In such cases it is assumed that the amount measured and the amount in the target organ are highly correlated, but this correlation is likely to vary by agent or by organ and may vary considerably by individual. There are, in addition, no long-term biomarkers for most agents, and current levels may not reflect long-term exposure levels. Moreover, biologic measurements reflect the body burden at one point in time. Even if the agent has a long half-life, the measurement may not be an accurate reflection of the total amount received due to metabolism and elimination over time (e.g., McGrail, Stewart and Schwartz 1995).

Biological measurements are often invasive and costly. For some known risk factors, only invasive techniques are available for biomonitoring, and exposure assessment, therefore, still relies on more traditional instruments. For example, asbestos is a recognized potent carcinogen. One way to evaluate asbestos exposure would be to measure the asbestos in broncho-alveolar lavage specimens. This invasive and expensive technique, however, is not routinely feasible, nor is it appropriate, because it does not reflect past exposure, which is the most relevant for cancer induction. In this example, exposure assessment must rely on indirect methods of measurement such as questionnaires or records.

If the measurement data were taken after the onset of disease (which is very difficult to determine because the onset may not be detectable), the measurements may be an effect of the disease, rather than a precursor. An example of such a measurement is serum levels of androgens and prostate cancer (Hsing 2001).

Because of their cost, biologic measurements are used more often in case-control or cross-sectional studies or in a sample of a cohort, rather than for an entire cohort. In spite of these limitations, biologic measurements can provide key insights into the toxicologic mechanisms of the agent and can be useful in estimating exposure levels if used judiciously. They can be useful in estimating recent or chronic exposure levels that have low variability over time. In addition, they represent concentrations received from all sources of exposure, so that the total amount of exposure received is better estimated. This advantage is especially important when individual work practices, such as hand washing before eating, can affect internal concentrations.

Measurements of the external environment are thought to be a lower gold standard than biologic measurements because they do not measure the internal dose received. They too represent only one point in time. This type of measurement often reflects only one source of exposure when several sources may be contributing to a study subject's overall exposure (e.g., pesticide exposures can occur from application at work, in the house and garden, from contamination of the soil from nearby farming operations and from consumption of pesticide-contaminated food and water). Thus, measurement of only one source may cause other important sources to be missed. Measuring exposures from a single source, therefore, without considering other sources, can result in lower estimates of

exposure and an overestimation of disease risk. In addition, external environmental measurements do not provide an estimate of internal dose. There are several advantages of this type of measurement over biological measurements, however. External environmental measurements are non-invasive and less expensive and the number of agents for which there are analytical methods is larger. The variability of the concentration of an agent in the external environment usually is lower than the intra-individual variability in the body, meaning that when a small number of measurements is available, a small number of environmental measurements on a group of similarly exposed workers is likely to result in a better estimate of the true exposure level than a small number of biologic measurements.

Finally, when measurements are taken for the purpose of an epidemiologic study, investigators should ensure that the data are collected in a way to reflect the metric being investigated. The sampling strategy should be developed to reflect the goals of the study (e.g., randomly or randomly within strata). Strict quality control methods should be followed. When records of measurements are being used, investigators should review the collection, analytic, and quality control methods to determine the accuracy of the data and how the measurements compare to the metric being assessed in the epidemiologic study.

Indirect Exposure Data

11.3.2

The second type of exposure information, indirect data, is derived from questionnaires, diaries, or records identifying measurements of exposure surrogates. Questionnaires may describe measurement data, e.g., cigarettes consumed per week or more subjective measures, such as the perception of control at the workplace. Examples of indirect data from diaries or records of surrogates are the amount of milk products consumed or distance of a residence from a hazardous waste site, respectively. As with measurement data, information from questionnaires, diaries or records may be problematic.

Questionnaires are developed by study investigators to ensure that information is collected in a structured, standardized approach to reduce differential questioning of cases and controls and to ensure that the data are as complete as possible.

The circumstances under which the questionnaire is administered (in person, telephone, mail, at home or in a hospital) may reflect the level of response. Development and administration of the questionnaire and data entry and clean up is costly and time-consuming. Computer-assisted personal and telephone interviews (CAPI and CATI, respectively) have substantially reduced data entry and cleanup costs, but their development is more expensive than using a paper copy. They can, however, include logic checks within the questionnaire to catch errors immediately, rather than long after the interview has taken place (cf. Chap. I.10 of this handbook). Questionnaires are usually administered by professional interviewers rather than by scientists knowledgeable of the areas being investigated, so that if a respondent asks for clarification or provides a response that is unclear or inappropriate, the interviewer may not be able to respond in a way to increase the quality

of the data. Interviewer training and inclusion of probing questions are means to reduce this problem. In spite of these limitations, oftentimes questionnaires are the only way to collect information on exposures.

Designing a questionnaire consists of establishing a list of questions in a predefined order, aimed at eliciting the presence of and often the amount of a given exposure. A questionnaire is defined by its content, the time span it covers in a subject's life, and its format and wording. Common sense principles should guide the construction of a questionnaire. Thus, each question and the flow of the questionnaire should be clear and subject to minimal misinterpretation. Administration of the question should not be a substantial burden to the subject, either in regards to the amount of time spent answering the questionnaire, the complexity of the information being collected, or the sensitivity of the questions. One hour is usually considered the maximum amount of time that respondents retain interest, but it may be much less. Aids can be used to help the respondent accurately recall information, such as lists of pesticides, logos, trademarks of products used, and pictures of medication bottles.

The list of questions in the questionnaire should include only those that the respondent can answer and that will ensure an accurate assessment of exposures. As the questions are developed, an analytical strategy also should be developed on how the responses will be used. A minimum set of questions should be asked that ensure maximum efficiency, but a small number of additional questions may be included for cross checking data. A few "red herring" questions (i.e., questions that are included to determine the accuracy of the responses, such as inserting in a list of real products, a product with a fake name) are often useful to evaluate the responses. More details on conducting interviews can be found in Chap. I.10 of this handbook.

The time span of the questionnaire is important. Respondents can more easily report on current exposures than historical ones. Past exposures, however, may be more important than current exposures in the etiology of chronic diseases, but collecting varying information over many years is problematic. Recollection of important life events at the earlier age can improve recall.

The format of the questions will determine the response rate to the question and the accuracy of the response. Open-ended questions (e.g., "What type of exercise did you do when you were in your twenties?") often gather more information than closed-ended questions because respondents can identify important exposures that are not anticipated by the investigator. Open-ended questions, however, require extensive coding, and some information collected is likely to be useless. Furthermore, important exposures may not be recalled. Close-ended questions (e.g., "Did you do any of the following in your twenties: walk? jog? play tennis? etc.") take more time, but the respondent is less likely to forget one of the identified exposures, making the information collected generally more accurate. If, however, the respondent had an important exposure to an agent not on the list, it may not be reported. Open-ended questions may be used in pilot studies to develop more standardized closed formats. Wording should be geared to the educational level of the respondents. In the US, the reading level of a 14-year old is generally consid-

ered appropriate for general population studies. When developing questionnaires, the investigator should consult one of the many references on questionnaire design (Sudman and Bradburn 1982; Armstrong et al. 1994; cf. Chap. I.10 of this handbook).

Screening questions are useful to minimize the time spent on answering inapplicable questions. Screening questions may require a simple yes or no (e.g., "Did you ever take birth control pills?"), or they may be formatted to screen out the lower exposed individuals (e.g., "Did you ever take birth control pills for at least one year?").

Diaries are another source of exposure information and have been used most frequently for diet and to a lesser extent, physical activity. In a diary, the respondent reports the amount of exposure (e.g., red meat consumption) at some identified frequency (e.g., daily). Diaries are best used when exposure occurs frequently, because if the frequency is too low, the respondent is likely to forget to complete the diary. Time spent recording the information should be minimal (e.g., less than one minute) and the time covered by the diary should be short (e.g., one to two weeks) to maximize compliance. Diaries should be formatted in a way to ease data entry as much as possible (e.g., check boxes rather than open-ended questions).

Records are often needed for retrospective exposure assessment (see Sect. 11.4.3). Records of surrogate information (including geographic information systems (GIS)) are often used in ecologic studies of the environment. Thus, amount of corn grown in various counties may be used to rank individuals with presumed exposure to herbicides. The data in such records may or may not have been accurately collected, but even if the data were accurately collected, the design of the data collection may impact the usefulness of the data in an epidemiologic study. For example, the Toxic Release Inventory of the US Environmental Protection Agency collects emissions data from private businesses. These data can be used to identify geographic areas with significant releases of agents into the air, water, and ground. However, there is a minimum amount of contaminant that must be released into the environment before reporting is required. Companies releasing smaller amounts of agents into the environment are not identified. Thus, if there are many small companies of one type in an area, the emissions reported in the database may suggest very low levels that may not, in fact, be low at all. In such cases, there may be no better data available for use in a study, but the protocol and quality control measures for the data collection should be carefully evaluated prior to use of such data, so that the investigator is aware of the strengths and limitations of the data. It may be useful to compare such data to other records systems as well. For example, a study of farmers' responses on pesticide use found reasonably good agreement with suppliers' information on pesticides bought by the farmer (Blair and Zahm 1993).

In summary, the choice of a measurement instrument is determined by knowledge of the disease (what is the true risk factor?), the feasibility of the measurement (its invasiveness and the ease of use in the exposure assessment), the cost, and its validity and reproducibility characteristics (see Sect. 11.5).

11.4 The Process of Exposure Assessment

The process of exposure assessment aims at the construction of an individual exposure estimate, from exposure data available, in order to produce a valid and efficient classification of subjects. Exposure data are usually imperfect, however, and there is a need for exposure *assessment* (rather than measurement), in order to approach the relevant dose.

The main steps for building exposure estimates and classification of subjects are described below. The specific problems resulting from the retrospective character of exposure assessment, the use of ecological estimates and the handling of multiple correlated exposures will also be presented, where ecological estimate refers to estimating an exposure level for a group of individuals, rather than for each individual separately.

The process of exposure assessment can be straightforward to relatively complicated, depending on the level of detail and the accuracy of the exposure data (e.g., surrogates of exposure may warrant less-intensive exposure assessment efforts than accurate and detailed exposure information on the true risk factor), the goal of the study (e.g., hypothesis-generating or hypothesis-testing), and the resources of the investigator.

11.4.1 Creating an Exposure Estimate

Some exposure data need little processing such as information obtained directly from answers to a questionnaire, for example smoking habits or intake of some kind of nutrients. In other investigations, some type of processing is needed. In the case of diet, for example, food composition tables allow the computation of the amount of nutrients across food groups (e.g., total vitamin A from various fruits, vegetables, meats, etc.). These tables take into account the mode of preparation and of preservation of the food. They are usually country-specific and need regular updating for an accurate translation from food groups into nutrients. For more details on assessment of micronutrients we refer to Chap. III.4 of this handbook.

Similarly, exercise can be measured using an accelerometer that measures movement, so that the total amount of energy expended can be estimated for an individual getting several types of exercise (Ainsworth et al. 1999).

In environmental studies (cf. Chap. III.3 of this handbook), the estimation process often is more complicated. These types of studies often make use of recognized pollutant dispersion models using exposure data reported by the subjects as well as exposure data from other records systems. Investigators of a study of respiratory symptoms developed exposure estimates from a model using type of vehicle, mean traffic density, emission exhaust rates, local topography, and meteorologic conditions to estimate airborne nitrogen dioxide levels (Oosterlee et al. 1996). Estimates of tricholoroethylene were developed for a municipal water system in a study of neurobehavioral effects using information on piping, flow input, water demand, and other variables, and a geographic information system (GIS) on the water distribution systems (Reif et al. 2003).

Occupational epidemiology (cf. Chap. III.2 of this handbook) also tends to estimate exposure from multiple pieces of exposure information, but to date, there are no recognized methods. In the past, experts have based their estimates on job titles and industry with little documentation as to how these estimates were derived. Recently, more attention has been paid to identifying *determinants of exposure* (e.g., factors that affect exposure) (Vermeulen et al. 2002). Examples of determinants include the presence of ventilation, the use of protective equipment, and the quantity of the contaminant in the workplace. Models to estimate an exposure score can be developed by simply assigning weights to the values of the determinants. For example, for a study of man-made mineral fibre, type of emission (active, passive), handling of fibres, presence of controls, protective equipment, and other variables were identified as affecting exposures (Cherrie et al. 1996). Variations in these variables across jobs resulted in the assignment of different exposure scores. Use of these determinants in statistical models allows for a more rigorous and transparent estimation process, such as for a study of paving workers where measurement data and determinants such as the type of paving (oil, mastic) and the use of tar were used to develop a estimation model for benzo(a)pyrene exposures (Burstyn et al. 2000).

Establishing a Level of Classification 11.4.2

In deciding on a classification, a decision must be made as to whether it will be qualitative (yes/no or ever/never), semi-quantitative or ordinal (e.g., low, medium, or high, or scores of say, 1–3, with or without the quantitative levels associated with the categories identified) or quantitative (with units of measurements). This decision is usually based on the quality of the exposure data.

Continuous data (i.e., quantitative) have greater statistical power to find an association than categorical data. Continuous data, however, also provide an impression of higher quality of exposure data than categorical data do, so that if the exposure data are poor, it may be better to describe the exposures categorically.

Oftentimes, investigators believe that *categorical data* are more accurate than continuous data. In one sense, this may be true. It generally is easier to assign a study subject to one of three categories than to estimate a quantitative level. The use of categories, however, does not reduce the error of the exposure assessment because all individuals within the category are assigned the same value. To illustrate this point, when categories are used, either a score is assigned to the category or the median of the range the category represents is used. It would be rare, however, that all individuals within an exposure category actually have the same exposure level. There are likely to be some individuals exposed at the median level of the category who are therefore appropriately assigned. There are also likely to be some individuals on both the low and the high ends of the category who will be assigned the same value as those individuals at the median level. Moreover, the individuals on the edges of adjacent exposure categories (e.g., the individuals on the high end of the low exposure category and the individuals on the low end of the adjacent higher category) are assigned to different exposure categories and

therefore to different median values, although they may be very similar in exposure levels. Thus, within any category of exposure, there is variability in exposure levels, and this variability will reduce the ability of the investigator to identify exposure response-relationships.

Another consideration in selecting the level of classification is the underlying assumption of the exposure-response relationship (cf. Chap. II.2 of this handbook). Using a continuous measurement of exposure in regression modelling (cf. Chap. II.3 of this handbook) assumes a linear increase of disease risk (or a transformed scale such as logit) for one unit of exposure. Use of categories of exposure, at least as a first approach, will, instead, fit observed values more closely without requiring any hypothesis about the shape of the exposure-response relationship. Categories must be developed, however, keeping in mind the limitations described above.

Grouping Strategies

Exposure groups are subsets of the population being studied that are viewed as being similarly exposed and therefore assigned the same exposure level. Exposure groups may be defined during questionnaire development, the exposure assessment process, or the analytical stage. When developing questionnaires, exposure groups are defined when responses to the questions are provided in categories. For example, if the possible responses to "At what age did you get your first menstrual period?" are $<$ 10, 10–12, 13–14, $\geq$ 15 years of age, these categories result in four exposure groups. In some studies, exposure groups are developed during the exposure assessment process. Thus, in an environmental study a question may be asked, "How far did you live from the ABC waste site?" The exposure data that will be used in the exposure assessment may be described in three categories, e.g., concentrations of an agent within a mile, 2–5 miles, and $\geq$ 5 miles. The investigator, then, may develop three exposure groups: one of subjects who report living $\leq$ 1 mile, one of subjects living 2–5 miles, and one of subjects living $\geq$ 5 miles. Alternatively, the exposure data may be continuous (e.g., concentrations at various distances). In this case, the investigator may leave the question open-ended. Alternatively, he/she may prefer to use the same three response categories as indicated above because the investigator may believe that the subjects can more accurately identify the correct category than estimate a continuously measured distance. Finally, during the analytical stage, investigators may decide to group individuals into quartiles or other arbitrary or ad hoc categories. An advantage of this strategy is that categories can be developed using differing cutpoints to allow comparisons with other studies.

The definition of exposure groups is important in an epidemiologic study because the variability of exposure level within and across groups affects the power to observe an exposure-response relationship (see also Sect. 11.5). There are three types of variability in epidemiologic studies. The first is *intra-individual* or day-to-day variability. For example, a subject with a mean alcohol consumption of two glasses a day may have no drinks some days and four drinks other days. The

epidemiologist has no control over this variability, but it is important to appreciate that there is variability of most exposures of individuals, which could be important when investigating threshold effects.

Intragroup variability is the variability that occurs within the exposure group. Thus, within an exposure group consuming 2–4 drinks/day, there will be some individuals who average two, some who average three and some who average four drinks/day. *Intergroup variability* is the variability across the groups (for example, with categories of 0, $\leq$ 1–2, 3–4, 5–6 and $>$ 6, the range is 0 $\rightarrow$ 6 drinks/day). The more intragroup variability there is compared to the intergroup variability, the more likely that an exposure-response relationship will be missed. The goal, therefore, is to have narrow ranges of exposure levels within the groups (with little to no overlap across other groups due to misreporting) and as wide a range across groups as possible. For example, in a study investigating coal dust and change in lung function (forced expiratory ventilation in one second (FEV_1)), four different exposure groups were evaluated for intragroup and intergroup variability and the effect of variability on the FEV_1. The intragroup variance ranged from 0.18–0.35 and the intergroup variance ranged from 0.20–0.23 (Heederik and Attfield 2000). The FEV_1 coefficient (in ml per mg/m^3 of coal dust) ranged from −2.0 to −5.9. The exposure group with the lowest intragroup variance (0.18) and the highest intergroup variance (0.23) was associated with the highest loss of FEV_1 per unit of dust exposure (−5.9 ml/mg/m^3 of dust). Intragroup and intergroup variability can be evaluated using analysis of variance techniques (e.g., Burstyn et al. 2000).

Retrospective Exposure Assessment 11.4.3

The challenges of using instruments to measure current (i.e., recent) exposures are compounded when investigating chronic disease. Because historical measurements are often lacking, investigators may collect current measurements and assume that historic levels were similar or extrapolate historic levels from the current measurements. Similarly, exposure information is often asked in questionnaires in reference to a single point in time (e.g., 20 years ago or when the subject was at a certain age), which is equivalent to having only one historical measurement. For example, in the area of nutrition, questionnaires used to investigate chronic disease have traditionally collected only information on current diet. Because diets have changed over time, current diet is not necessarily highly correlated to diets of 20 to 30 years ago.

In contrast, in the occupational investigations, however, complete work histories are often collected, which is likely to result in more accurately historical exposure estimates than using only current job. There is a whole body of literature relative to retrospective exposure assessment using job exposure matrices (JEM) or expert assessment from a panel of experts (Benke et al. 2001). A JEM is a cross tabulation of jobs (or job/industry combinations) and agents by time that automatically assigns the same exposure level to all individuals having the same job. Used in association with a subject's complete work history, JEMs or expert evaluation provide an individual probability of exposure to a given agent.

11.4.4 Ecological versus Individual Exposure Assessment

Measurement data may not be available on the actual study subject, but rather on individuals thought to be similarly exposed as the individual under study. These types of measurements are called *ecologic assessments*. In contrast, assessment of individual exposures takes into account the personal characteristics of the individual. An example of an ecologic assessment is assigning the same level of trihalomethanes in a public water supply system to all individuals on that water supply, in spite of the recognition that the concentration of trihalomethanes can vary within a system. Assigning the same exposure level to individuals with different exposure levels will result in misclassification of study subjects, because in the same (macro) environment, subjects are likely, in fact, to have different exposure levels. For example, subjects living in an area with a polluted public water supply will be exposed differently to a pollutant in the water depending on whether their water resources come from a public supply or from a private well, the amount of tap water they drink, their use of tap water for cooking, etc.

An ecological evaluation is used when exposure data or resources are limited. Ecological estimates are the rule in areas such as air pollution epidemiology, where individual exposures are often defined by atmospheric measurements at the sampling location nearest to the individual's residence, or more broadly, at the city level. Ecological estimates are also popular in occupational epidemiology, where job exposure matrices have been developed. In these examples, investigators of air pollution or workplace exposures usually do not have measurement data on the individuals or individual-specific parameters such as individual work practices and protective equipment. The ecologic evaluation, therefore, assigns the same exposure value to a group of subjects sharing the same (macro) environment.

Ecologic evaluations can result in substantial misclassification of exposure levels. In the field of occupation, even among individuals thought by occupational health professionals to have similar exposure levels, the exposure level can be up to three to six times larger or smaller than estimated, as indicated by geometric standard deviations often found (van der Woord et al. 1999). It seems reasonable to assume that similar degrees of misclassification occur among other types of environmental exposures. Extrapolation of measurement data from one individual to another or from a system to an individual therefore must be done with caution.

Ecological measurements are often derived from existing records (air quality monitoring records, occupational measurements surveys) and are much cheaper to obtain and estimate than individual measurements. Using ecological measurements instead of individual measurements makes sense if the contrast of exposure between the groups (e.g., cities or jobs) is greater than variability of exposures among individuals in the same group. Studies based on ecologic measurements may also be useful for hypothesis-generation.

Individual assessment generally requires a greater assessment effort but is likely to result in less misclassification. Considerations for selecting one approach over

the other include: time and financial resources, availability of exposure data and its quality and quantity, and the purpose of the study (e.g., hypothesis-generating or -testing, and investigation of an exposure-response relationship).

11.4.5 Dealing with Multiple Exposures

In many situations, exposures to various potential risk factors in human populations tend to aggregate for an individual, due to individual behaviour. An example is the correlated habits of smoking, alcohol and coffee drinking among some individuals. Similarly, in the outdoors environment, humans are exposed to mixtures of compounds originating from the same source (e.g., mercury, polychlorinated bi-phenyls (PCBs), and other organochlorines from eating fish) or from various sources (e.g., carbon monoxide from automobile and truck exhaust).

Epidemiological studies have proved to be informative about many complex mixtures such as cigarette smoke or air pollution. However, identification of the component(s) responsible for the health effects (and their joint effects) observed is still required for a better understanding of disease causation, cost-effective monitoring of the hazard, and an efficient strategy of prevention of disease.

The situation of the mixed exposures cannot be treated as a classical problem of confounding because the exposures are highly correlated. Stratified analysis or multivariate modelling is, in general, inefficient because such analytical approaches do not allow the presence of a high colinearity among different exposures. In addition, the presence of one or several agents "representative" of mixed exposures or the occurrence of interaction among exposures is not merely a statistical problem. It also requires a strategy that recognizes the different underlying biological hypotheses of the various components of the mixtures. Much of the insight about multiple exposures comes from epidemiology (for instance tobacco smoke or outdoor air pollution) because toxicological experiments often cannot replicate complex mixtures to which people are exposed across time, and such experiments are usually limited to single components or suitably chosen combinations.

To illustrate the problem of complex mixtures, we describe as an example environmental exposure to PCBs. Similar examples, however, are found in many other areas of study, including diet and occupational exposures. PCBs are a persistent type of industrial compound that includes 209 different chemical members referred to as congeners. The commercial product always is a mixture of correlated congeners, so that studying the toxicity of these compounds is not easy. For example, some PCBs act like dioxins by binding to the aryl hydrocarbon AhR receptor, and may result in cancer (Longnecker et al. 1997). Experimental work has shown the highest dioxin-like activity occurs for congeners with no chlorine in the *ortho* position. It has been speculated that neurologic effects of PCBs, on the other hand, may be caused by congeners with chlorine in the *ortho* position.

Samet (1995) has proposed five general strategies for studying such complex mixtures efficiently: (1) treating the mixture as a single agent; (2) selecting an indi-

cator component; (3) creating a summary index; (4) identifying the separate effects of the mixture's individual components; and (5) characterizing the independent and joint effects of the components. We review these strategies with application to the problem of the toxicity of PCBs.

(1) Treating the Mixture as a Single Agent. The early studies in Japan and Taiwan that recognized the neurotoxicity of PCBs, and the later studies in Michigan, relied on total PCBs. At that time congener-specific data were not available (Schantz et al. 2003). The exposure measurements taken in these studies were powerful enough to strongly suggest the neurotoxic potential of PCBs. There is still, however, a debate about discrepancies in health effects among studies in different countries. These discrepancies may be due to different analytical procedures, different patterns of congeners, or different co-exposures to other organochlorines, such as dioxins or furans, which have similar environmental pathways (Longnecker et al. 1997). In summary, treating the mixture of PCBs as a single agent has proved efficient for hazard identification in early work, but exposure misclassification limits the interpretation of the discrepant findings.

(2) Selecting an Indicator Component. Several recent large studies have focussed on a small number of congeners present in relatively high concentrations (e.g.: PCB 153). The congeners present in high concentrations, however, are not necessarily the most toxic. As a rule, "a single component of a mixture may be an appropriate index of toxicity if the component mirrors the dosimetry and toxicity of other components relevant to the health effects of concern" (Samet 1995).

(3) Creating a Summary Index. Creating a summary index implies the attribution of some type of weighting to the individual concentrations of the different components of a mixture. The weight assigned to each congener is defined according to an underlying hypothesis about the biological activity of each component. If one assumes that endocrine disruption is a relevant biological mechanism of toxicity for PCBs, a measurement of the total estrogenic xenobiotic burden in adipose tissue could provide an integrated biomarker of xeno-hormonal activity resulting from exposure to a given mixture of compounds (Soto et al. 1997). Another example of biological activity, the dioxin-like activity of a PCB congener, can be calculated using a toxic equivalency factor (TEF) (Ahlborg et al. 1994), which is assigned relative to the toxicity of the dioxin 2, 3, 7, 8 TCDD. The total toxic equivalency (TEQ) of a mixture of PCBs can then be estimated by summing across all compounds, the product of the concentration and TEF for each compound. It is likely, however, that the weighting is dependent on the state of knowledge about the relative potency of the different components at the time of calculation, and that over time it would be necessary to modify the summary index as more information becomes available.

(4) Separating Effects of the Mixture's Components. Creating one summary index does not reflect the heterogeneity of the mixture. There is a trade-off between measuring concentrations of the individual compounds in the mixture (which is usually time-consuming and expensive) and summarizing the mixture of highly correlated congeners. Analyzing concentrations of 38 PCBs congeners from 497 human milk samples from Canada in 1992, Gladen et al. (2003) distinguished three groups of congeners: one group of the congeners, including most of the major congeners, that were highly correlated, meaning that their individual biologic effects realistically could not be separated in an epidemiologic study; another group of congeners quantifiable in only a small fraction of the population by the assay methods used and therefore an epidemiologic analysis would be uninformative; and a third group quantifiable in a reasonable fraction of samples and not correlated with the bulk of major congeners. The authors concluded the components of this last group are worth studying separately and are good candidates for individual determination and inclusion in epidemiologic studies.

(5) Characterizing the Independent and Joint Effects of Components. Measurements of selected congeners allow the evaluation of health effects related to single or joint exposures. Correlations, however, exist not only between concentrations of PCBs congeners, but also with other common organochlorines, metals, and pesticides and there are strong suspicions of possible interactions among these compounds at the molecular level that affect neurobehavioral function in particular (Carpenter et al. 2002). The strategies presented earlier provide some guidelines for studying these joint effects in epidemiological studies.

Two other points regarding mixtures are appropriate. It should be recognized that while some agents within a mixture may cause a disease, it is possible that other agents in that same mixture reduce the likelihood of the disease by deactivating the active compound. For instance there is an active discussion around the beneficial impact on birthweight of seafood consumption during pregnancy, which brings high amounts of fatty acids and selenium, relative to the potential toxicity of seafood from contaminants such as mercury (Grandjean et al. 2001). This situation complicates the determination of causality in epidemiologic studies. Also, individual characteristics of the study subjects (e.g., polymorphisms) may intensify or reduce the effect of the agent. Currently, our ability to tease out these situations is limited, but investigators should at least recognize that they may be possible.

Multiple exposures can be evaluated using interaction analysis, but can also be grouped using hierarchical cluster analysis (e.g., see Hines et al. 1995 for an example). In this study fabrication workers in a semi-conductor company were exposed to multiple chemicals. Hierarchical analysis allowed the investigators to identify groups of workers exposed to the same pattern of exposures (e.g., various glycol ethers).

11.5 Measurement Errors

All types of exposure assessment in every area of investigation will have some error. Chapter II.5 of this handbook describes statistical methods to cope with measurement errors. Appreciation of the types and degree of error allows for a more appropriate interpretation of the study results. Knowing the sources of error can also provide areas for methodologic investigation within the study to allow quantification of the error. This, in turn, can allow the investigator to estimate the effect of the error on the epidemiologic findings.

11.5.1 Types of Measurement Errors

There are two types of errors that arise from measurements: random and systematic. Random error will result in the measurements being randomly distributed around the mean. Systematic error, or bias, will result in an overall mean that is erroneously high or low compared to the true mean. Both types of error are of concern in exposure assessment and they are described in terms of precision and validity. *Precision* measures random error and refers to the reproducibility or reliability of the measure. *Validity* measures systematic error and refers to the distance between the exposure measured and the target variable (ideally, the true risk factor, but practically, the surrogate).

A measurement instrument must be *reproducible*. Under ideal conditions this means that if the instrument is administered under the varying conditions, it should provide the same response within a reasonable level of variation. Generally, however, reproducibility more practically is defined as providing the same response within a reasonable level of variation under the same circumstances. Reproducibility is a necessary condition to accurately evaluate intraindividual and intragroup variability, but somewhat less necessary to accurately evaluate intergroup variability. In addition, to be useful, the measurement instrument must also be *valid* (i.e. it should measure the exposure it is supposed to measure and identify the true quantity present).

Historically, measurement error more often has been associated with categorical assessments than quantitative, probably because quantitative assessments have been limited in the past. Measurement error in either type of assessment will result in *misclassification error* when estimating the exposure levels of study subjects. For example, if a subject was assigned to a high fruit intake category, rather than a medium fruit intake category, the subject is misclassified. Misclassification of confounders can be also a serious problem since it will usually reduce the degree to which confounding can be controlled. For instance in many studies it is essential to obtain a complete smoking history including detailed periods of smoking or quitting, and quantity smoked during each period, because tobacco smoking is a risk factor, and therefore a potential confounder, for many diseases. When studying lifestyle factors associated with smoking, such as alcohol consumption, misclassification of smoking habits will result in in-

complete adjustment and residual confounding. In the context of an epidemiologic study misclassification is characterized as nondifferential or differential, depending on whether it affects the comparison groups (i.e., the diseased and non-diseased subjects) similarly. *Differential misclassification*, which results from there being a different amount of error for the diseased compared to the non-diseased, can lead to underestimation or overestimation of the association between the exposure and disease. In the latter situation, misclassification can induce spurious statistically significant results. *Nondifferential misclassification* of exposure usually will bias estimates of relative risks towards the null. There are examples, however, occurring in extreme conditions, where nondifferential misclassification of exposure can produce bias away from the null (Rothman and Greenland 1998). Thus, both types of misclassification can result in incorrect conclusions.

Sources of Measurement Errors 11.5.2

Armstrong et al. (1994) classified sources of measurement error in five categories: faulty design of the instrument, errors or omissions in the protocol regarding the use of the instrument, poor execution of the protocol during data collection, limitations due to subject characteristics (e.g. poor memory of past exposures or day-to-day variability in biological characteristics), and errors during data entry and analysis. They have provided an extensive list of circumstances in which these errors may occur and these sources should be carefully evaluated before attempting to use any type of instrument.

Measurement instruments and analytical methods (such as for an air or biological measurement, blood pressure, etc.) generally are designed to be as accurate and reproducible as possible when used under similar conditions, i.e., with the same protocol. Two possible sources of systematic differences that can occur are from the measurement/analytical method itself and from the interference of other substances present in the measured environment. Reduction of these errors in the investigation of disease risks can be made by following the manufacturer's/laboratory recommendations, calibrating the instrument under the conditions being measured, using spiked and blank samples, and following other quality control procedures (cf. Chap. I.13). Random error can arise from a lack of technical precision of the instrumentation, variation introduced by the laboratory technicians, and the analytical procedures themselves. This inherent limitation of the instrument and analytical methods, however, explains only part of the variability. Other sources of variation include weather conditions, presence of other exposures, the actual concentration being out of the range of the instrument's measurement range, and the timing of the instrument's response in the relation to a change in concentration. The sources of error need to be identified in order to decrease, or at least, recognize and quantify the variability.

Questionnaires, because they also can suffer from the two types of misclassification, systematic and random, can be viewed similarly. Systematic differences can result from incorrect phrasing of questions (such that all respondents misinterpret

the question similarly) or from inappropriate or misleading response categories. Random sources of misclassification can result from poor phrasing (such that different respondents interpret a question differently) and lack of interest on the part of the respondent. To collect quality data, questionnaires should be standardized, so as to ensure that all study subjects are asked the same questions. Questions must be clearly phrased, without ambiguity and use terms that are understandable to respondents. Respondents must be able to remember the events being asked about and be able to correctly respond to the questions. Thus, reporting of events that took place many years ago or that require mathematical calculations (e.g., estimating "average" amount of foods eaten on a seasonal basis) is likely to be subject to more random error than reporting of more recent events or events that do not require calculations (e.g., Bradburn et al. 1987 and Subar et al. 1995). Pilot testing of questions should be conducted on a group of individuals with the characteristics of the group who will be receiving the questions because respondents often interpret questions very differently from investigators, even if the questions were carefully developed. Questions should also be tested under the conditions that the questionnaire will be administered (e.g., in the home). Following these procedures should decrease bias and increase precision.

Diaries are prone to both systematic and random errors from the same sources as questionnaires. Records, in contrast, may have systematic and random error similar to measurement data or questionnaire data, depending on the type of record.

Both systematic and random errors may result from limited data. For example, systematic error could result in missing information from asking about sensitive issues, such as the number of sexual partners (Lindzey and Aronson 1985). Subjects may be more inclined to respond with a "don't know" if the number of partners exceeds what they consider to be acceptable. Cases with workplace-induced cancer may be so sick that proxies are used as the respondents. Proxies generally know little about workplaces of the subject. In contrast, many of the control subjects would be able to provide detailed information about the workplace.

Having limited exposure information can result in misclassification of subjects by exposure level. In the environmental area, Brunekreef et al. (1987) illustrated the effect of limited data on misclassification in a study of the relationship between environmental exposure to lead and blood lead levels in children. He found that averaging four measurements of lead on home floors increased the regression coefficient explaining blood lead levels by 69%, compared to the model using a single home floor measurement. Having only one measurement, therefore, would have increased the misclassification of subjects. Generally non-differential misclassification due to limited data will result in random error.

The problem of limited data also is evident in the use of questionnaires. For example, often investigators restrict the workplace exposure information collected to jobs, industries and dates. From these limited data, they apply job exposure matrices to assign occupational exposure estimates. When applying the matrix, individuals holding a job are considered non-exposed if the exposure occurs only in a small proportion of workers in the job. This procedure will, however, inevitably

result in classifying among the "unexposed" individuals, a small proportion of workers who are, in reality, exposed. Similarly, individuals having jobs entailing a high probability of exposure will be considered exposed, even if they belong to the small proportion of nonexposed workers on this job. Detailed descriptions of tasks and work conditions of the jobs held by individual study subjects and evaluation of these data on the individual subject level are necessary for a better assessment. Thus, limited exposure data can contribute to misclassification, in that the available data (from which exposure is characterized) may not be representative of the individual's actual exposure level. This problem is more related to selection bias, is a general problem in epidemiology, and is not unique to exposure variables. The concept of bias is treated in Chap. I.12 of this handbook.

One can often recognize the circumstances in which differential misclassification may occur. Diseased subjects may have reflected more on their past exposures than the nondiseased (recall bias) or may take more care in providing correct responses. Differential bias will potentially occur when the exposure measurement instrument uses a human intermediary (e.g., the subject himself and/or an interviewer) aware of (or thinks he/she is aware of) the disease status. Thus, face-to-face interviews involve a substantial risk of producing interviewer effects. If a bias results from a different attitude of the interviewer toward the diseased compared with the non-diseased subjects, it is called interviewer bias. Self-administered questionnaires are generally believed to be less vulnerable to influences of response bias; however, the appearance of the questionnaire, the introductory letter, and the research group may all have an impact on response. The likelihood of bias from telephone interviews falls between these two data collection methods. Computer-assisted telephone interviewing has become the method of choice in many studies, and often has a high response rate and few missing data (Nybo Andersen and Olsen 2002).

Quantification of Measurement Errors – Reproducibility Studies 11.5.3

Evaluation of the reproducibility of measurement instruments can be done by comparing the same instrument under the same conditions over time or by comparing various instruments under the same conditions at the same time. An example of the first type of study evaluated the reproducibility of a self-administered lifetime physical activity questionnaire (Chasan-Taber et al. 2002). Subjects reconstructed physical activity at four ages, starting at menarche, twice in the same mail questionnaire administered one year apart. All intraclass correlation coefficients used to measure reproducibility ranged from 0.78 to 0.87, with a value of 0.83 for total lifetime estimate of exposure.

The area of nutritional epidemiology (cf. Chap. III.4 of this handbook) is one in which the design of proper questionnaire instruments has been extensively investigated. Subar et al. (2001) compared a new food frequency questionnaire and two widely used dietary questionnaires using telephone 24-hour recalls. Despite

substantial differences in the length and the design of the questionnaires, correlations obtained for dietary composition (i.e., total energy intake and 26 nutrients) were very similar. This comparison provides evidence that carefully designed self-administered food frequency questionnaires can provide reasonably reproducible measures of current nutrient intakes in epidemiologic applications. There are still, however, questions about the validity of these instruments, and probably only the comparison of questionnaires with a truly uncorrelated error, such as a biochemical indicator of diet, will resolve these validity issues.

11.5.4 Quantification of Measurement Errors – Validation Studies

Ideally, an instrument should be evaluated by comparing it to a standard under the conditions the instrument is used. In evaluating the validity of any measurement instrument, the choice of the gold standard is a critical issue. Biochemical indicators of internal exposure provide an independent assessment for which measurement errors are not likely to be correlated with errors in air or water measurements or questionnaires. Biologic measurements may represent historical exposures only if the chemical of interest has a sufficiently long biological half-life and may represent recent exposures only if the chemical has a relatively short half-life. In both cases, for the biologic measure to be useful, the body burden cannot be affected by the disease or its treatment. In other situations, the biomarker may not measure the target agent of interest. Other challenges of biologic monitoring can be found in Sect. 11.3.1 of this chapter. Biochemical indicators of dietary intake have a great appeal as the gold standard to assess the validity of dietary questionnaires (Willett 1990). There are limitations, however, in that the indicators may not reflect only dietary intake, and there are many dietary factors of interest for which there is no biomarker.

Practically, however, a gold standard often does not exist, especially when exposure has to be assessed retrospectively (e.g., historical tobacco consumption of individuals). For some exposures, however, a partial validation may be possible, by comparing questionnaire results to pre-existing records. For example, reported jobs can be compared to employers' records, and smoking consumption can be compared to past medical records. Identification of gold standards that are "alloyed", and how to account for this error has been discussed (Wacholder et al. 1995). The validation of the instrument is also often measured by its ability to predict disease risk in prospective studies (Willett 1998). This approach is somewhat problematic, however, in that the epidemiologic outcome is used to test the instrument. Nonetheless, a good instrument should produce better risk estimates than a poor one (Tielemans et al. 1998).

11.5.5 Methods for Correcting Measurement Errors

The effect of a systematic difference between the actual concentration and the concentration measured can be reduced or minimized simply by applying a correction factor reflecting the difference to the exposure estimate if the difference

is known. Internal validation studies have been proposed to reduce the impact of measurement error. In one approach, exposure is measured, although imperfectly, from everyone in the study, and, simultaneously, a more accurate but more expensive measurement is collected on only a small subset of cases and controls selected randomly. Sophisticated statistical methods can then be applied in order to infer the corrected odds ratio from measurement error models fitted to the parallel exposure measurements from the validation sample (Stürmer et al. 2002). These so-called two-phase designs are among others investigated by Schill et al. (1993, 1997) and have been applied by Pohlabeln et al. (2002). This method, however, has not yet been routinely implemented, and further research is needed to establish the robustness of the procedures in realistic settings and to determine optimal designs for selecting a validation sample. As quoted by Chatterjee and Wacholder (2002) in a recent commentary, "the best way to reduce bias from measurement error is to improve tools for measuring exposures including biological markers, environmental samples and questionnaires".

A second approach that is gaining popularity is to conduct an uncertainty analysis (or sensitivity analysis; Rothman and Greenland 1998). In this approach, investigators identify the uncertainty around a point estimate (e.g., 2 drinks of wine a day). For example, if a question asked "How many glasses of wine do you drink?" and the responses were < 1/day, 1–3/day, 4–5/day, > 5/day, the uncertainty ranges of these responses could be 0–0.9, 1–3, 4–5 and 6–10, respectively. Monte Carlo or other statistical simulations allow a better understanding of the uncertainty around the disease risk estimates.

11.6 Conclusions

The demand for accurate exposure assessment implies the need for development of validated and reliable tools in parallel with reduced costs and increased applicability in field studies. Sophisticated techniques are now available for direct measurement of chemicals in most mediums with excellent sensitivity and reproducibility. Similarly, questionnaires are being developed in various fields with considerable effort being put into their validation.

In some areas, such as occupational or environmental epidemiology, improvement is dependent upon additional knowledge on exposure determinants both at the personal and population levels, and on objective comparisons of the quality of various available methods for exposure assessment (Liljelind et al. 2003). Quantitative estimates of exposure using statistical modelling are currently being developed, mainly for risk assessment purposes, but their applicability to epidemiological studies has not been fully explored.

To solve the problem of mixed exposures, the trend is towards building exposure indices summarizing several exposures according to biological hypotheses about their joint mechanisms of action. In the near future, new biotechnologies (e.g.,

genomics, proteomics) will contribute to the development of biomarkers of gene expression, intermediate between markers of exposure and markers of early effects that will summarize the joint action of mixed exposures at the molecular level (Henry et al. 2002, cf. Chap. III.6 of this handbook). The applicability of these techniques in epidemiological studies opens a whole new area of research.

References

Ahlborg UG, Becking GC, Birnbaum LS, Brouwer A, Derks HJGM, Feeley M, Golor G, Hanberg A, Larsen JC, Liem AKD, Safe SH, Schlatter C, Waern F, Younes M, Yrjänheikki E (1994) Toxic equivalency factors for dioxin-like PCBs. Chemosphere 28:1049–1067

Ainsworth BE, Richardson MT, Jacobs DR Jr, Leon AS, Sternfeld B (1999) Accuracy of recall of occupational physical activity by questionnaire. J Clin Epidemiol 52:219–227

Armstrong BK, White E, Saracci R (1994) Principles of exposure measurement in epidemiology. Monographs in Epidemiology and Biostatistics, vol 21. Oxford University Press, Oxford

Benke G, Sim M, Fritschi L, Alfred G, Forbes A, Kauppinen T (2001) Comparison of occupational exposure using three different methods: hygiene panel, job exposure matrix (JEM) and self reports. Appl Occup Environ Hyg 16: 84–91

Blair A, Zahm SH (1993) Patterns of pesticide use among farmers: implications for epidemiologic research. Epidemiology 4:55–62

Bradburn NM, Rips LJ, Shevell SK (1987) Answering autobiographical questions: The impact of memory and inference on surveys. Science 236:157–161

Brunekreef B, Noy D, Clausing P (1987) Variability of exposure measurements in environmental epidemiology. Am J Epidemiol 125:892–898

Burstyn I, Kromhout H, Kauppinen T, Heikkila P, Boffetta P (2000) Statistical modelling of the determinants of historical exposure to bitumen and polycyclic aromatic hydrocarbons among paving workers. Ann Occup Hyg 44: 54–56

Carpenter DO, Arcaro K, Spink DC (2002) Understanding the human health effects of chemical mixtures. Environ Health Perspect 110(suppl 1):25–42

Chasan-Taber L, Erickson JB, McBride JW, Nasca PC, Chasan-Taber S, Freedson PS (2002) Reproducibility of a self-administered lifetime physical activity questionnaire among female college alumnae. Am J Epidemiol 155:282–289

Chatterjee N, Wacholder S (2002) Validation studies: bias, efficiency and exposure assessment. Epidemiology 13:503–506

Cherrie JW, Schneider T, Spankie S, Quinn M (1996) A new method for structured, subjective assessments of past concentrations. Occ Hyg 3:75–83

Gladen BC, Doucet J, Hansen LG (2003) Assessing human polychlorinated biphenyl contamination for epidemiologic studies: lessons from patterns of congener concentrations in Canadians in 1992. Environ Health Perspect 111:437–443

Grandjean P, Bjerve KS, Weihe P, Steuerwald U (2001) Birthweight in a fishing community: significance of essential fatty acids and marine food contaminants. Int J Epidemiol 30:1272–1278

Heederik D, Attfield M (2000) Characterization of dust exposure for the study of chronic occupational lung disease: A comparison of different exposure assessment strategies. Am J Epidemiol 151:982–990

Henry CJ, Phillips R, Carpanini F, Corton JC, Craig K, Igarashi K, Leboeuf R, Marchant G, Osborn K, Pennie WD, Smith LL, Teta MJ, Vu V (2002) Use of genomics in toxicology and epidemiology: findings and recommendations of a workshop. Environ Health Perspect 110:1047–1050

Hines CJ, Selvin S, Samuels SJ, Hammond SK, Woskie SR, Hallock MF, Schenker MB (1995) Hierarchical cluster analysis for exposure assessment of workers in the semiconductor health study. Am J Ind Med 28:713–722

Hsing AW (2001) Hormones and prostate cancer: what's next. Epidemiol Rev 23: 42–58

Liljelind I, Rappaport S, Eriksson K, Andersson J, Bergdahl IA, Sunesson AL, Jarvholm B (2003) Exposure assessment of monoterpenes and styrene: a comparison of air sampling and biomonitoring. Occup Environ Med 60:599–603

Lindzey G, Aronson E (1985) Handbook of social psychology, 3rd edn, vol I. Random House, New York

Longnecker MP, Rogan WJ, Lucier G (1997) The human health effects of DDT (dichlorodiphenyl-trichloroethane) and PCBs (polychlorinated biphenyls) and an overview of organochlorines in public health. Annu Rev Public Health 18:211–244

McGrail MP, Stewart W, Schwartz BS (1995) Predictors of blood lead levels in organolead manufacturing workers. J Occup Environ Med 37:1224–1229

Nybo Andersen A-M, Olsen J (2002) Do interviewer's health beliefs and habits modify responses to sensitive questions? A study using data collected from pregnant women by means of computer-assisted telephone interviews. Am J Epidemiol 155:95–100

Oosterlee A, Drijver M, Lebret E, Brunekreef B (1996) Chronic respiratory symptoms in children and adults living along streets with high traffic density. Occup Environ Med 53:241–247

Pohlabeln H, Wild P, Schill W, Ahrens W, Jahn I, Bolm-Audorff U, Jöckel K-H (2002) Asbestos fibreyears and lung cancer: a two-phase case-control study with expert exposure assessment. Occupational and Environmental Medicine 59:410–414

Reif JS, Burch JB, Nuckols JR, Metzger L, Ellington D, Anger WK (2003) Neurobehavioral effects of exposure to trichloroethylene through a municipal water supply. Environ Res 93:248–258

Rothman KJ, Greenland S (eds) (1998) Modern epidemiology, 2nd edn. Lippincott-Raven, Philadelphia, pp 115–134

Samet JM (1995) What can we expect from epidemiologic studies of chemical mixtures? Toxicology 105:307–314

Schantz SL, Widholm JJ, Rice DB (2003) Effects of PCB exposure on neuropsychological function in children. Environ Health Perspect 111:357–376

Schill W, Jöckel K-H, Drescher K, Timm J (1993) Logistic analysis in case-control studies under validation sampling. Biometrika 80:339–352

Schill W, Drescher K (1997) The analysis of case-control studies under validation subsampling: a comparison of four approaches. Statistics in Medicine 16:117–132

Soto AM, Fernandez MF, Luizzi MF, Oles-Karasko AS, Sonnenschein C (1997) Developing a marker of exposure to xenoestrogen mixtures in human serum. Environ Health Perspect 105(suppl 3):647–654

Stürmer T, Thürigen D, Spiegelman D, Blettner M, Brenner H (2002) The performance of methods for correcting measurement error in case-control studies. Epidemiology 13:507–516

Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, McIntosh A, Rosenfeld S (2001) Comparative validation of the Block, Willett, and National Cancer Institute Food Frequency Questionnaires. The Eating at America's Table Study. Am J Epidemiol 154:1089–1099

Subar AF, Thompson FE, Smith AF, Jobe JB, Ziegler RG, Potischman N, Schatzkin A, Hartman A, Swanson C, Kruse L, Hayes RB, Lewis DR, Harlan LC (1995) Improving food frequency questionnaires: A qualitative approach using cognitive interviewing. J Am Diet Assoc 95:781–788

Sudman S, Bradburn NM (1982) Asking questions. Jossey-Bass Publishers, San Francisco

Tielemans E, Heederik D, Burdorf A, Vermeulen R, Veulemans H, Komhout H, Hartog K (1998) Assessment of occupational exposures in a general population: comparison of different methods. Occup Environ Med 56:145–151

Van der Woord MP, Kromhout H, Barregard L, Jonsson P (1999) Within-day variability of magnetic fields among electic utility workers: Consequences for measurement strategies. Am Ind Hyg Assoc J 60:713–719

Vermeulen R, Stewart P, Kromhout H (2002) Dermal exposure assessment in occupational epidemiologic research. Scand J Work Environ Health 28:371–385

Wacholder S, Hartge P, Dosemeci M, Armstrong B (1995) Validation studies using an alloyed gold standard (letter). Am J Epidemiol 141:277

Willett WC (1998) Invited commentary: comparison of food frequency questionnaires. Am J Epidemiol 148:1157–1159

Willett WC (1990) Reproducibility and validity of food-frequency questionnaires. In: Nutritional epidemiology. Oxford University Press, New York