Confounding and Interaction

I.9

Neil Pearce, Sander Greenland

9.1 Introduction

All epidemiologic studies are (or should be) based on a particular *source population* followed over a particular *risk period*. The goal is usually to estimate the effect of one or more exposures on one or more health outcomes. When we are estimating the effect of a specific exposure on a specific health outcome, *confounding* can be thought of as a mixing of the effects of the exposure being studied with the effect(s) of other factor(s) on the risk of the health outcome of interest. *Interaction* can be thought of as a modification, by other factors, of the effects of the exposure being studied on the health outcome of interest, and can be subclassified into two major concepts: *biological dependence of effects*, also known as synergism; and *effect-measure modification*, also known as heterogeneity of a measure. Both confounding and interaction can be assessed by stratification on these other factors (i.e. the potential confounders or effect modifiers). The present chapter covers the basic concepts of confounding and interaction and provides a brief overview of analytic approaches to these phenomena. Because these concepts and methods involve far more topics than we can cover in detail, we provide many references to further discussion beyond that in the present handbook, especially to relevant chapters in Modern Epidemiology by Rothman and Greenland (1998).

9.2 Confounding

9.2.1 Basic Concepts

Confounding occurs when the exposed and non-exposed subpopulations of the source population have different background disease risks, which is to say: these subpopulations would have different disease risks even if exposure had been absent from both subpopulations (Greenland and Robins 1986; Rothman and Greenland 1998 Chap. 4; Greenland et al. 1999a,b). When we estimate the effect of exposure on the exposed by comparing the frequency of disease in the exposed and non-exposed groups, we assume that the disease frequency in the non-exposed group provides a valid estimate of what the disease frequency would have been in the exposed group if it had not been exposed. If this assumption is incorrect, i.e. if the exposed and non-exposed groups would have had different disease frequencies in the counterfactual situation in which the exposed group had not been exposed, then we say that the comparison of the exposed group to the non-exposed group is confounded.

More generally, confounding can arise when the exposed and non-exposed group are not completely comparable or "exchangeable" with respect to their exposure response; that is, for at least one level of exposure, the exposed and unexposed groups would exhibit different risks even if they both had experienced that exposure level (Greenland and Robins 1986). Note that the earlier definition

takes this level to be that of non-exposure because it is presumed there that the effect of exposure on the exposed is the effect of interest. If instead we were interested in the effect of non-exposure on the non-exposed (as might be the case in a study of a preventive factor), we would have confounding if the exposed group failed to exhibit the risk that the non-exposed would have had if they had been exposed.

This problem of non-comparability (non-exchangeability) can also occur in randomized trials because randomization may fail, leaving the treatment groups with different characteristics (and different baseline disease risk) at the time that they enter the study, or because of differential loss and non-compliance across treatment groups. However, there is more concern about non-comparability in observational epidemiology studies because of the absence of randomization. Randomization prevents certain sources of confounding (e.g., confounding due to physician selection of treatment based on patient characteristics, also known as "confounding by indication"); also, bias due to differential loss and non-compliance can be at least partially controlled by using the randomization indicator (the "intent-to-treat" variable) as an instrumental variable (Sommer and Zeger 1991; Greenland 2000a). These benefits of randomization are not available in observational studies, and in fact confounding should be expected to occur as a by-product of ordinary life events and choices.

As an example, if we compare the risk of lung cancer in people with a low dietary beta carotene intake compared with people with a high dietary beta carotene intake, it is very likely that these two groups will differ with respect to other risk factors for lung cancer such as tobacco smoking, because people who are less health conscious are more likely to smoke as well as to neglect dietary recommendations. If this is the case (e.g. if a greater percentage of people smoke in the low beta carotene intake group than in the high beta carotene intake group), then smoking will confound the association between beta carotene intake and lung cancer: The higher smoking prevalence among those with a low beta carotene diet will lead to a higher lung-cancer risk among them compared to those with a high beta carotene diet, even if beta carotene intake itself has no effect on lung cancer risk.

Any variable that affects disease in the absence of exposure has the potential to confound the exposure-disease relationship. It will confound that relationship if, in the absence of exposure, it would have a distribution that is sufficiently different across exposure groups to produce a difference in risk across those groups even if exposure were absent. This was the case for smoking in the beta carotene example. A confounder, if not adequately controlled, will bias the estimated effect of exposure on disease. The bias will be upward if the higher-risk levels of the confounder occur more frequently among the exposed; conversely, the bias will be downward if the higher-risk levels of the confounder occur more frequently among the unexposed. Confounding may even reverse the apparent direction of an effect in extreme situations. Confounding may also occur when the main exposure under study has no effect on the risk of disease – a spurious association may be observed which is entirely due to confounding. Factors associated with confounders can also act like confounders and serve as surrogates for confounders, provided that they are not

affected by exposure or disease. For example, socioeconomic status may serve as a surrogate measure of causal factors (living conditions, lifestyle, lack of preventive care, etc.) that are potential confounders.

Three conditions are traditionally given as necessary (but not sufficient) for a factor to be a confounder (Rothman and Greenland 1998, Chap. 8). First, to produce confounding, a factor has to be predictive of disease in the absence of the exposure under study. Note that a confounder need not be a genuine cause of the disease under study, but merely "predictive" within exposure levels apart from chance relations. Hence, surrogates for causal factors (e.g., ethnicity, gender, socioeconomic status) may be regarded as potential confounders, even if they are not direct causal factors. It is not always clear from the data whether an observed relation between the factor and disease represents a genuine (replicable) predictive quality, as opposed to (say) chance. In making such a determination, prior information as opposed to statistical testing should play a dominant role (Greenland and Neutra 1980; Miettinen and Cook 1981; Robins and Morgenstern 1987). This is why one almost always sees adjustment made for age and sex: these factors are known to be predictive of risk of most diseases. When prior information is not available one must of course turn to the data collected for the study as a guide as to whether the factor is predictive of disease in the source population; even in these cases, however, there are better strategies for confounder selection than those based on statistical testing. We will return to this topic below.

Second, a confounder has to be associated with the study exposure in the source population. It may occur that when participants in a case-control study are selected from the source population, then due to chance a factor may be associated with exposure in the study, even though it was not associated with exposure in the source population. In this situation, the factor is not a confounder (Miettinen and Cook 1981; Robins and Morgenstern 1987). Although in practice it is common to use the data actually collected to decide whether a factor is associated with exposure, more commonly the data are used to decide whether adjustment for the factor makes an important difference in the estimated exposure effect, a practice we will discuss below. In a case-control study, one should expect a confounder to be associated with exposure among the controls (at least if the controls are selected with no bias). If the factor is not associated with exposure among controls, an association may still occur among the cases simply because the study factor and a potential confounder are both risk factors for the disease, but this is a consequence of those effects and so does not cause confounding. A factor-exposure association will only indicate confounding by the factor if it reflects the association in the source population.

Third, a variable that is affected by the exposure or by the disease, e.g., an intermediate in the causal pathway between exposure and disease, or conditions that are caused by the health outcome of interest, should not be treated as a confounder because to do so could introduce serious bias into the results (Greenland and Neutra 1980; Robins and Morgenstern 1987; Robins and Greenland 1992; Weinberg 1993; Rothman and Greenland 1998, Chap. 8; Cole and Hernan 2002). For example, in a study of obesity and death from coronary heart disease, it would be inap-

propriate to control for hypertension if it was considered that hypertension was a consequence of obesity, and hence a part of the causal chain leading from obesity to death from coronary heart disease. On the other hand, if hypertension itself was of primary interest, then this would be studied directly, and obesity would be regarded as a potential confounder if it also involved exposure to other risk factors for death from coronary heart disease.

Similarly, we should avoid controlling for health outcomes that may be part of the pathogenic disease process, such as reduced pulmonary function following exposure to a respiratory hazard in a study of chronic obstructive lung disease (Checkoway et al. 2004). We would, however, be justified in controlling for baseline (i.e. pre-exposure) lung function if there were reasons to believe that baseline lung function was associated with subsequent exposure level. Evaluating whether certain factors are exposure or health outcome intermediates in causal pathways requires information external to the study. Intermediate variables can sometimes be included in the analysis, although special techniques are then required to avoid adding bias (Robins 1989; Robins and Greenland 1994; Robins et al. 1992, 2000). In no case would control of a variable affected by the disease be valid, however (Greenland et al. 1999a; Pearl 2000).

Thus, an assessment of confounding by a factor that is not an intermediate involves consideration of whether the exposed and non-exposed groups are "comparable" in the source population with respect to their disease risk in the absence of exposure. In practice, we often focus on specific potential confounders – variables that are risk predictive of disease in the absence of exposure (such as age and sex) and assess whether they are associated with exposure in the source population on which the study was based. If such an association is present, it is evidence that the two groups are not comparable or exchangeable with respect to baseline risk. If such an association is absent, however, it does not mean that the groups are comparable, because there may be other uncontrolled risk factors that confound the observed association, or the association may have been obscured by measurement error.

Because it involves judgments about causal as well as temporal ordering, the property of being a confounder cannot be determined from data alone (Miettinen and Cook 1981; Greenland and Robins 1986; Greenland et al. 1999a; Pearl 2000; Robins 2001; Hernan et al. 2002). Once that ordering is established, however, it is common to assess confounding by seeing whether the main effect estimate changes when the potential confounder is controlled in the analysis. In this approach, near-equality of the crude and adjusted effect estimates is taken as evidence that there is no confounding by the factor, and conversely, an important difference is taken as evidence of confounding by the factor. Many epidemiologists prefer to make a decision based on the basis of this "collapsibility" or "change-in-estimate" criterion (rather than the criterion of "exchangeability"), although this approach can be misleading, particularly if (as usual) there is misclassification of the adjustment factors or the exposure (Greenland 1980; Greenland and Robins 1985; Savitz and Baron 1989; Marshall and Hastrup 1996, 1999) or if the outcome is common and the measure is an odds ratio or rate ratio (Miettinen and Cook 1981; Greenland 1996;

Greenland et al. 1999b); also, this criterion does not exhibit good statistical properties, although it is no worse than significance-testing procedures (Maldonado and Greenland 1993).

The decision to control for a presumed confounder can be made with more confidence if there is supporting prior knowledge that the factor is predictive of disease, independent of its association with exposure. Such prior knowledge is usually available for well-studied factors such as age, sex, and tobacco smoking. At the very least, it is usually known if the factor is affected by exposure (in which case it is not a potential confounder and should not be controlled, at least not by conventional methods). If even this much is uncertain, the decision to control or not control a variable may be controversial, in which case analyses both with and without its control may be presented (Greenland and Neutra 1980).

As a final caution, in studies involving aggregate-level effect (such ecologic and multilevel studies), a factor at one level may, if not controlled, confound effect estimates at another level, and a factor may modify and confound effects differently at different levels of aggregation. For example, both the income of an individual and the income of his or her neighbourhood may separately predict risk of an outcome, possibly in opposite directions. Robbery rates are often higher in low-income neighbourhoods, yet within neighbourhoods it could still be that an individual's risk of robbery went up as his or her income went up. In that case both neighbourhood income and individual income could be confounders, but would confound effect estimates in opposite directions if both were positively and separately associated with the exposure under study. Thus, regardless of level of interest (e.g., country, neighbourhood, individual), it is often essential to measure and adjust for variables on other levels (Greenland 2001a).

9.2.2 Example of Confounding

Table 9.1 presents a hypothetical example of confounding in a cross-sectional study of asthma. Overall, one-half of the study participants are smokers and one-half are not. However, two-thirds of the exposed group are smokers compared with one-

Table 9.1. Hypothetical example of confounding by tobacco smoking in a study of occupational asthma

	Smokers		Non-smokers		Total	
	Exposed	Non-exposed	Exposed	Non-exposed	Exposed	Non-exposed
Asthma cases	800	400	200	400	1000	800
Non-cases	1200	600	800	1600	2000	2200
Total	2000	1000	1000	2000	3000	3000
Prevalence (%)	40	40	20	20	33.3	26.7
Prevalence ratio	1.0		1.0		1.25	

third of the non-exposed workers. Thus, although exposure is not associated with asthma either among smokers (the prevalence of asthma is 40% in the exposed and 40% in the non-exposed, PR = 1.0) or in non-smokers (the prevalence of asthma is 20% in the exposed and 20% in the non-exposed, PR = 1.0), it is associated with asthma (PR = 1.25) when the two subgroups are combined. This occurs because smoking is associated with the exposure in the source population, and is an independent risk factor for asthma. In this hypothetical example, the two stratum-specific estimates are each 1.00, thus the adjusted estimate will also be 1.00 (or very close to 1.00) whatever weights are used. Thus, the crude prevalence ratio is 1.25, whereas the adjusted prevalence ratio is 1.00, indicating that confounding has occurred (provided that there has not been biased selection of the study participants from the source population).

Control in the Study Design 9.2.3

Confounding can be controlled in the study design, in the analysis, or both. There are three common methods for control at the design stage (Rothman and Greenland 1998). The first is randomization – random allocation of participants to exposure categories. However, this is usually only an option for potentially beneficial exposures, e.g. it would be impractical and unethical to conduct a randomized trial of the health effects of smoking, and as mentioned earlier randomization may fail to prevent all confounding.

A second method of control at the design stage is to restrict the study to narrow ranges of values of the potential confounders, e.g., by restricting the study to white males aged 35–54. This approach has a number of conceptual and computational advantages, but may severely restrict the number of potential study subjects and ultimately limit the generalizability of the study.

A third method of control involves matching study subjects on potential confounders. For example, in a cohort study one could match a white male non-exposed subject aged 35–39 with an exposed white male aged 35–39. This will prevent age-gender-ethnicity confounding in a cohort study, but is seldom done because it is expensive and time-consuming. In case-control studies, matching does not prevent confounding, but does facilitate its control in the analysis in that matching on a strong confounder will usually increase the precision of effect estimates. However, matching may reduce precision in a case-control study if it is done on a factor which is associated with exposure, but is not a risk factor for the disease of interest. The matching process effectively turns such a factor into a confounder, which must then be controlled in the analysis, thus reducing precision and increasing analytical complexity. For example, in a case-control study of power-frequency electromagnetic field (EMF) exposure and childhood cancer, choosing sibling controls (i.e. for each case choosing a sibling as a control) would mean that in almost every instance, the case and control would have lived in the same house and would have similar EMF exposure, resulting in almost no exposure-discordant pairs and almost no precision in the resulting matched-pair estimates.

As already mentioned, matching may be expensive and time-consuming. Finding suitable controls becomes increasingly difficult as the number of matching factors increases beyond two or three. Moreover, when it occurs the increase in precision from matching is often modest, typically involving a 5 to 15 percent reduction in the variance of the effect estimate (Schlesselman 1982; Thomas and Greenland 1983). Therefore, although many discussions of matching stress issues of statistical efficiency, practical considerations (such as ease of finding controls) are often more important (Rothman and Greenland 1998, Chap. 10).

9.2.4 Control in the Analysis

Confounding can also be controlled in the analysis by adjusting simultaneously for all confounding factors or a sufficient subset of them. This presumes, of course, that a sufficient subset has been accurately measured, which is often not the case. Methods for controlling confounding in the analysis are discussed in more depth in the chapters on specific study designs (Chaps. I.5, I.6 and I.8 of this handbook), in Part II of this handbook and in many textbooks (e.g. Rothman and Greenland 1998, Chaps. 15–21). In the simplest situation, control of confounding in the analysis involves stratifying the data according to the levels of the confounder(s) and calculating an effect estimate that summarizes the association across strata of the confounder(s). As an example, controlling for age (grouped into 5 categories) and gender (with 2 categories) might involve grouping the data into the $10 (= 5 \times 2)$ confounder strata and calculating a summary effect estimate, which is a weighted average of the stratum-specific effect estimates. It is usually not possible to control simultaneously for more than 2 or 3 confounders in a stratified analysis, since finer stratification will often lead to many strata containing no exposed or no non-exposed persons. Such strata are uninformative; thus, too fine stratification is wasteful of information. This problem can be mitigated to some extent by the use of regression modeling (cf. Chap. II.3 of this handbook), which allows for simultaneous control of more confounders by "smoothing" the data across confounder strata.

9.2.5 Assessment of Confounding

When one lacks data on a suspected confounder, and thus cannot control confounding directly, it is still desirable to assess the likely direction and magnitude of the confounding. In particular, it may be possible to obtain information on a surrogate for the confounder of interest. For example, social class is associated with many lifestyle factors such as smoking, and may therefore be a useful surrogate for some lifestyle-related confounders. Even though confounder control will be imperfect in this situation, it is still possible to examine whether the exposure effect estimate changes when the surrogate is controlled in the analysis, and to assess the strength and direction of the change. For example, suppose the relative risk relating low dietary beta carotene intake to lung cancer actually increases (e.g. from 2.0 to 2.3) or remains stable (e.g., at 2.0) when social

class is controlled. This might be taken as evidence that the observed excess risk is not entirely due to smoking, because social class is correlated with smoking (Kogevinas et al. 1997), and control for social class involves partial control for smoking. The strength of this evidence depends of course on what exposure is being studied and what sort of classification errors or other sources of bias are present.

Even if it is not possible to obtain confounder information for any study participants, it may still be possible to estimate how strong confounding is likely to be for particular risk factors. For example, this is often done in occupational studies, where tobacco smoking is a potential confounder, but smoking information is rarely available. In fact, although smoking is the strongest risk factor for lung cancer, with relative risks of 10-fold or more, it appears that smoking rarely exerts a confounding effect of greater than 1.5 times in studies of occupational disease (Axelson 1978, 1989; Siemiatycki et al. 1988; Kriebel et al. 2004) (although this degree of confounding may still be important in some contexts). There are several approaches to the assessment of potential confounding by factors such as cigarette smoking when data are lacking or incomplete. One approach is to conduct an analysis of smoking-related diseases other than the disease of primary interest (Steenland et al. 1984). If mortality from such diseases (e.g., nonmalignant respiratory disease) is not elevated, this may suggest that any excess for the disease of interest is unlikely to be due to smoking. Similarly, one might be less inclined to attribute an excess of an alcohol-related cancer to unusually high drinking prevalence among the exposed if liver cirrhosis mortality is not elevated.

When detailed individual risk factor information is not available on a potential confounder, it may be possible to assess the impact of this factor on risk estimates by conducting a type of *sensitivity analysis* that estimates the potential direction and extent of confounding (Cornfield et al. 1959; Bross 1967; Axelson 1978, 1989; Schlesselman 1978; Checkoway and Waldman 1985; Axelson and Steenland 1988; Flanders and Khoury 1990; Rothman and Greenland 1998, Chap. 19). In this sensitivity analysis, the magnitude of the effect of the potential confounder on the disease should be known with some confidence, and the prevalence of the potential confounder among the exposed and comparison groups should be estimable, within reasonable bounds. Then, a range of confounding effects, including a "worst case scenario," can be calculated (Checkoway et al. 2004).

Consider the incidence rate, I, of disease in a population as consisting of two components: one being the incidence among those without the confounder, and the other the incidence among those with the confounder (assume in this simple case that the confounder is dichotomous) (Axelson 1978). Then:

$$I = I_0 \left(1 - p_c\right) + \mathrm{RR}_c \left(I_0\right) \left(p_c\right)$$

where: I = incidence rate overall, RR_c = relative risk due to the confounder, I_0 = incidence rate among those who are confounder negative, and p_c = prevalence of the confounder in the source population from which the cases arose.

This expression can be expanded to include several levels of a confounder, for example: light, moderate, and heavy smoking. By applying the equation to two (or more) study groups (e.g., exposed and non-exposed subgroups) for which p_c is assumed to differ, one can calculate a confounding bias factor B_c comparing these two exposed groups and due to confounding alone. If there is no effect of exposure, then B_c is the magnitude of effect one would observe due to differences in p_c alone. Then, if RR_{Obs} is the observed relative risk comparing exposed to non-exposed,

$RR_{Adj} = RR_{Obs}/B_c$ is the adjusted relative risk, controlling for confounding.

To illustrate, suppose that one is concerned about smoking confounding a finding that $RR_{Obs} = 3.2$ for lung cancer and some dichotomous occupational exposure. Individual smoking data are not known. We might assume that smoking habits in the non-exposed approximate those of other typical blue-collar workers whose smoking habits have been studied. To estimate the most extreme confounding that might reasonably be expected, we might assume that the exposed were heavier smokers, with habits more like the 90th percentile of blue-collar workers. These assumptions would imply that the non-exposed might have been 50% non-smokers, 40% moderate smokers, and 10% heavy smokers, whereas the exposed were 20% non-smokers, 55% moderate smokers, and 25% heavy smokers. Assuming that moderate smoking confers a relative risk of smoking of 10 compared to non-smokers, and heavy smoking a relative risk of 20, the confounding due to smoking is $B_c = 1.65$. Thus, one would observe a relative risk of 1.65 comparing exposed to unexposed due to these smoking differences alone. One can then calculate that $RR_{Adj} = 3.2/1.65 = 1.9$ as a hypothetically "adjusted" exposure effect under a plausible, but unlikely scenario for smoking differences among exposed groups. If RR_{Adj} is elevated, one might conclude that confounding is unlikely to be the entire explanation for the elevated risk (Checkoway et al. 2004).

This method allows one to place limits on the degree of confounding that can result from failure to adjust for an unmeasured risk factor that is associated with the exposure under study. Its application is restricted however to control for factors whose risks are well established quantitatively, and for which the confounder prevalence in the population can be estimated fairly reliably. Cornfield et al. (1959) and others (Bross 1967; Flanders and Khoury 1990) showed that the relative risk that would result from differences in the prevalence of a covariate, such as smoking or alcohol consumption, may be quite limited, even in the absence of complete knowledge about the covariate. In particular, Flanders and Khoury (1990) showed that:

$$1 < B_c < \min\left\{OR, RR_c, 1/p_c, RR_c/\left(q_c + RR_c * p_c\right), OR/\left(q_c + RR_c * p_c\right)\right\}$$

where B_c, RR_c and p_c are as above, $q_c = 1-p_c$, and OR is the odds ratio measuring the association between exposure and the covariate (both measured dichotomously in this example). Using this equation, rather severe limits can be placed on the range of possible values of B_c with information about p_c and RR_c alone, making no assumption about OR at all. For example, if $p_c = 0.5$, and RR_c for esophageal cancer from 1 pack/day smoking = 5, then $B_c = 1.7$. That is, given $RR_c = 5$ an

observed RR_{obs} for an exposure effect is unlikely to be confounded by more than 1.7 times. With a reasonable assumption about the likely values of OR and RR_c, the association between exposure and the confounder, the maximum value for B_c could be lower or higher. This method can be extended for the situation in which there are multiple levels of exposure and multiple levels of covariates (Schlesselman 1978; Flanders and Khoury 1990), and to other measures of association such as odds ratios (Yanagawa 1984).

A more sophisticated version of sensitivity analysis places uncertainty distributions (priors) on the unknown quantities in the sensitivity formula, repeatedly draws values from these distributions and corrects the data or estimates based on the drawn values, adds in a random-error correction, and presents the resulting distribution of corrected estimates. Such Monte-Carlo sensitivity analysis has long been a staple of risk analysis and is now finding application in epidemiology (e.g., see Greenland 2003, 2004a, 2005; Phillips 2003; Lash and Fink 2003).

Sensitivity analysis can also be useful in certain situations in which confounder information has been collected, but the validity or precision of those data are weak. Sometimes smoking data are available on only a subset of the members of a cohort. One option is to conduct an analysis that controls for smoking directly on the subset with smoking data. However, the precision of this analysis and the generalizability of findings to the entire cohort may be questionable. Instead, one might apply the data relating the confounder to exposure among this subset in a sensitivity analysis for the entire cohort (Fingerhut et al. 1991). A third option is to adjust the entire cohort based on the data from the subset, using two-stage methods or missing-data methods (Rothman and Greenland 1998, Chap. 15).

Relationship Between Confounding and Other Biases 9.2.6

In this chapter confounding has been defined as non-comparability (non-exchangeability) of the exposed and non-exposed subgroups in the source population with respect to their risk of the disease outcome in the absence of exposure. Confounding is thus a property of the source population rather than of the specific group of study participants.

Selection bias involves biases arising from the procedures by which the study participants are selected (or select themselves) from the source population. Thus, selection bias is not an issue in a cohort study (or cross-sectional study) involving complete recruitment and follow-up because in this instance the study group comprises the entire source population. However, selection bias can occur if participation in the study or follow-up is incomplete. Selection bias is usually more of an issue in case-control studies in that selection bias may occur if the case group does not include (or is not representative of) all cases in the source population, or if the control group is not representative of the population-at-risk that the cases came from.

If control selection involves only sampling at random from the entire source population with complete cooperation, *selection bias* is a minor concern. More realistically, however, selection bias is likely if response rates differ according

to exposure level and disease status. When controls are selected from among persons with other diseases, considerable care must be taken in specifying the diseases that form the control group. In particular, a specific disease may not correctly reflect the exposure pattern in the source population, especially if it is caused by the study exposure. One strategy is to include only diseases that are thought to be unrelated to the exposure(s) of interest (Miettinen 1985; Rothman and Greenland 1998, Chap. 7), but this requirement may be difficult to satisfy in practice because adequate evidence for the absence of exposure effects on many diseases is frequently not available (Axelson et al. 1982). An alternative method is to select as controls a sample of all other diseases. This approach is reliable if one can be sure the factor under study does not markedly increase risk of numerous or the most common diseases. It has become common practice to exclude diseases known to be related to exposure from the pool of potential controls; however, even this restriction will not always eliminate bias (Pearce and Checkoway 1988).

Selection bias and confounding are not always clearly demarcated. In particular, selection bias in the form of non-response at baseline of a cohort can be viewed as a source of confounding, since it may produce associations of exposure with other risk factors in the study cohort and thus turn those factors into confounders (Hernan et al. 2004). A similar phenomenon occurs in case-control studies when selection is affected by a factor that itself affects exposure. An example occurs when matching on a factor that is associated with exposure in the source population; even if the factor is not a risk factor for disease in the absence of exposure, matching may turn it into a confounder which must be controlled in the data analysis, because matching will create the necessary factor-disease association if exposure affects disease (Rothman and Greenland 1998, Chap. 10). Unfortunately, as discussed earlier, if selection is affected by exposure and associated with case-control status (e.g. selection bias due to inappropriate selection of controls from persons with other diseases, or selection on factors affected by exposure), stratification on the selection-related factors will rarely produce valid estimates, and hence this type of selection bias should not be viewed as confounding.

Information bias is the result of misclassification of study participants with respect to disease or exposure status. Thus, the concept of information bias refers to those people actually included in the study, whereas selection bias refers to the selection of the study participants from the source population, and confounding generally refers to non-comparability of subgroups within the source population. There are many methods to adjust for misclassification (e.g. Copeland et al. 1977; Greenland and Kleinbaum 1983; Espeland and Hui 1987; Greenland 1988; Armstrong et al. 1992; Thomas et al. 1993; Armstrong 1998). These require estimates of sensitivity and specificity, or the reliability of the measurement, based on prior information or validation data. Estimates based on prior information are often only best guesses that may not apply to the population under study. However, it is an informative exercise to conduct sensitivity analyses that explore the range of results that might have occurred under various scenarios (Rothman and Greenland 1998, Chap. 19), and again, Monte-Carlo sensitivity analyses may be applied (Lash and Fink 2003; Phillips 2003; Greenland 2004a, 2005).

Some consider any bias that can be controlled in the analysis as confounding, but this definition is too general because any bias control requires background information for proper execution, and any bias can be controlled in the analysis given enough background information. Confounding is distinguished in that it represents a mixing or confusion of the effects of other factors with the effects of the study exposure, a concept that goes back at least as far as the writings of John Stuart Mill (Greenland et al. 1999b). Other biases are then be categorized according to whether they arise from the selection of study subjects (selection bias), or their classification (information bias). Most observational studies suffer from more than one form of bias, and the effects of multiple biases may compound error. Perhaps the best-appreciated situation is when there is misclassification of a confounder, in which case attempts to control for the confounder will not fully control that confounder and may actually increase bias (Greenland 1980; Greenland and Robins 1985; Savitz and Baron 1989; Marshall and Hastrup 1996, 1999). When (as is often the case) multiple biases are present, many complex and counterintuitive phenomena can occur, and a clear picture of the net effects of bias will require analyses that account for these bias interactions (Lash and Fink 2003; Phillips 2003; Greenland 2005).

Interaction 9.3

Basic Concepts 9.3.1

The concept of *interaction* (*effect modification*), also known as effect heterogeneity and effect variation, refers to a condition where the effect of exposure on the outcome under study varies by some other factor. In other words, in order to estimate the effect of exposure on an outcome (such as a disease time, a disease risk, or a disease rate), we must first know whether or not another factor is present (or what the level of this other factor is). This concept can be subclassified into two major concepts: *biological dependence of effects*, also known as synergism; and *effect-measure modification*, also known as heterogeneity of a measure. With regard to the latter, all secondary risk factors modify either the rate ratio or the rate difference, and uniformity over one measure implies non-uniformity over the other (Steenland and Thun 1986; Rothman and Greenland 1998, Chap. 4), e.g., an apparent additive joint effect implies a departure from a multiplicative model. A further source of ambiguity is that the term "effect modification" implies that one factor in some way biologically "modifies" the effect of the other factor but this is not necessarily the case. For this reason, the term "effect-measure modification" and "effect-measure variation" are more accurate terms, and are logically equivalent to the definition of "interaction" used in most statistics books and computer programs (Rothman and Greenland 1998, Chaps. 4 and 18).

The concepts of interaction and confounding are quite distinct. An effect-measure modifier may or may not be a confounder and a confounder may or

may not be an effect-measure modifier (Miettinen 1974; Rothman and Greenland 1998, Chap. 4). For example, if we are comparing exposed and non-exposed subgroups of a population, and the percentage of people who smoke (and the intensity of smoking, and the length of time that each person has been smoking) is the same in both groups, then smoking is not a confounder. However, the rate ratio for the exposure effect may still vary by smoking status, e.g. the exposure may double the risk of disease in smokers but have no effect in non-smokers. In this situation, smoking would not be a confounder, but would be an effect-measure modifier.

9.3.2 Example of Effect-measure Modification

Table 9.2 presents a hypothetical example of effect-measure modification in a cross-sectional study of asthma. The overall findings are the same as for the study presented in Table 9.1, but the stratum-specific findings are different. Now there is no confounding by smoking because the percentage of smokers is the same in the exposed and non-exposed groups. However, there is effect modification since the prevalence ratio (for the association of exposure with disease) is 1.5 in smokers and 1.0 in non-smokers. Thus, whereas the assessment of confounding involved the comparison of the crude and adjusted effect estimates, the assessment of effect modification involves the comparison of the stratum-specific effect estimates with each other.

Table 9.2. Hypothetical example of effect modification by tobacco smoking in a study of asthma prevalence

	Smokers		Non-smokers		Total	
	Exposed	Non-exposed	Exposed	Non-exposed	Exposed	Non-exposed
Asthma cases	600	400	400	400	1000	800
Non-cases	900	1100	1100	1100	2000	2200
Total	1500	1500	1500	1500	3000	3000
Prevalence (%)	40	26.7	26.7	26.7	33.3	26.7
Prevalence ratio	1.5		1.0		1.25	

9.3.3 Concepts of Interaction

Although at first glance, the assessment of interaction is relatively straightforward, there are considerable hidden complexities. Some of the analytic issues in studying effect-measure modification will be illustrated with data (Table 9.3) from a study by Selikoff et al. (1980) of lung cancer death rates per 100,000 person-years at risk in relation to exposure to cigarette smoke and asbestos (Steenland and Thun

Table 9.3. Example of joint effects: lung cancer mortality rates per 100,000 person-years at risk in a cohort of asbestos workers compared to those in other blue collar occupations. Source: (Steenland and Thun 1986)

	Rate in smokers (RR)	Rate in non-smokers	Rate ratio
Asbestos	935.8 (RR_{11} = 32.7)	500.5 (RR_{01} = 17.5)	1.9
Non-asbestos	199.5 (RR_{10} = 7.0)	28.6 (RR_{00} = 1.0)	7.0
Rate ratio	4.7	17.5	
Rate difference	736.3	471.9	

1986). The rate difference due to asbestos exposure is 472 per 100,000 person-years in non-smokers and 736 per 100,000 person-years in smokers. Thus, the rate difference for the effect of asbestos exposure on lung cancer mortality is lower in non-smokers than in smokers. On the other hand, the rate ratio for the same effect is higher in non-smokers (asbestos rate ratio = 17.5) than in smokers (asbestos rate ratio = 4.7). Thus, both the rate difference and the rate ratio are subject to effect-measure modification in that the effect estimate depends on the presence or absence (or more generally, the level) of another factor (i.e. smoking), but the dependencies are in opposite directions: the rate difference is larger in smokers and the rate ratio is larger for non-smokers.

Several authors (Kupper and Hogan 1978; Walter and Holford 1978) have taken this dependence of interaction on the underlying effect measure to imply that the assessment of interaction is "model-dependent". Thus the authors equate all uses of the term "interaction" with effect-measure modification. In contrast to these statistically-based definitions, other authors (e.g. (Rothman and Greenland 1998, Chaps. 2 and 18)) adopt a definition of interaction in which two factors are said to exhibit interdependent effects or "biologic interaction"or "synergism" if they are component causes in the same sufficient cause (Rothman 1976; cf. Chap. I.1 of this handbook), or if individual patterns of response (the potential or counterfactual outcomes) to exposure change when the other "interacting" factor is changed (Greenland and Poole 1988; Rothman and Greenland 1998, Chap. 18). This concept of dependence of effects implies that additivity of risks will arise when no biologic interaction is present. With this concept in hand, one can show that the presence and degree of effect-measure modification depend to a large extent on the prevalence of causal cofactors of exposure, as well as the actual biologic mechanisms at work (Rothman 1976; see Chap. I.1 of this handbook for further explanation).

There was originally some confusion about the relation of biologic interaction to nonadditivity (Koopman 1977). If two factors (A and B) belong to different sufficient causes, but a third factor (C) belongs to both sufficient causes, then A and B are competing for a single pool of susceptible individuals (those who have C). Consequently the joint effect of A and B will be less than additive. Miettinen (1982) reaches a similar conclusion based on a model of individual outcomes. However, this phenomenon can be incorporated directly into the sufficient-cause model by

clarifying a previous ambiguity in the description of antagonism in the model's terms. Specifically, the absence of B before A can be included in the sufficient cause involving A, and vice versa. Then, two factors would exhibit interaction, specifically antagonism, if the presence of one factor and absence of the other factor were component causes in the same sufficient cause (Greenland and Poole 1988; Rothman and Greenland 1998, Chap. 18).

Under a potential-outcomes (counterfactual) model of causation, two factors are said to exhibit interaction if the response schedule (response type) of any individual to one of the factors depended on the level of the other factor (Greenland and Poole 1988; Rothman and Greenland 1998, Chap. 18). This definition leads to the same operational (statistical) criterion for identifying the presence of interaction as that derived from the sufficient cause model, namely, departure from risk additivity.

It should be stressed that this concept of independent effects is distinct from from certain other biological concepts of no interaction. For example, Siemiatycki and Thomas (1981) give a definition in which two factors have biologically independent effects "if the qualitative nature of the mechanism of action of each is not affected by the presence of absence of the other"; this concept does not lead to an unambiguous definition of dependent effects, however (Siemiatycki and Thomas 1981), and thus does not produce clear analytic implications. In contrast, under the sufficient-cause and potential-outcome (counterfactual) models, a particular biologic model, rather than being accepted as the "baseline", is itself evaluated in terms of the co-participation of factors in a sufficient cause, or in modification of individual response. For example, two factors which act at different stages of a multistage process have dependent effects because they are joint components of at least one sufficient cause. This occurs irrespective of whether they affect each other's qualitative mechanism of action (the ambiguity in Siemiatycki and Thomas' formulation stems from the ambiguity of this concept).

9.3.4 Additive and Multiplicative Models

The sufficient-component and potential-outcome definitions of interaction (co-participation in a sufficient cause, or change in response schedule) are attractive because they are based on an explicit causal model that leads to an unambiguous definition of independence of effects, and because they lead to the additive model as the baseline for assessing interactions, just as obtained through public health (cost-benefit) considerations (Rothman et al. 1980; Rothman and Greenland 1998, Chap. 18). However, the analytic implications of these concepts are not straightforward, since assessing independence of effects is usually only one of the analytic goals of an epidemiological study. There are several other considerations which often favor the use of multiplicative models.

One is that multiplicative models have convenient statistical properties. Estimation in non-multiplicative models may have problems of convergence, and inference based on the asymptotic standard errors may be flawed unless the study size is very large (Moolgavkar and Venzon 1987). Another is that, if it is desired to keep interaction (effect-measure modification, corresponding to product terms in

a regression model) to a minimum, then a multiplicative model is often most effective. It is not uncommon for joint effects to appear closer to multiplicative than to additive (Saracci 1987). In this situation there may be less masking of heterogeneity in calculating an overall rate ratio than in calculating an overall rate difference. Although there are also many instances of non-multiplicative departures from additivity (Selikoff et al. 1980; Saracci 1987), even in these cases multiplicative-model summaries are more often closer to a population-average (standardized) measure than are additive-model summaries (Greenland and Maldonado 1994). Finally, additive-risk models are not identical to additive relative-risk models when the model includes terms for confounder adjustment; unfortunately, in typical case-control studies only the latter models can be fit, thus rendering it difficult or impossible to make unconfounded assessments of risk additivity (Greenland 1993a,b). In contrast, departures from multiplicativity can be assessed in the same fashion from cohort and case-control data.

9.4 Detecting Interactions

Determining whether or not a factor is an effect-measure modifier is often done by estimating an effect measure (e.g., relative risk) for the exposure of interest separately for each level of the presumed effect modifier and testing for equality of these measures across the modifier strata (Rothman and Greenland 1998, Chap. 15). This approach lacks power, however, and so it can be quite misleading to conclude modification is absent just because the test yields a large *P*-value. Because of such power problems and other problems due to sample size limitations, when there are multiple possible effect-measure modifiers, such as age, ethnicity, gender, or previous employment in a hazardous industry, effect modification is usually examined for each potential modifying variable separately, or else through use of modeling methods that allow continuous modification by quantitative variables such as age. Prior selection of potential effect modifiers of greatest interest can simplify the task. Then, assessing effect-measure modification for a subset of modifying variables might be carried out, with adjustment made for other variables.

A major obstacle to interaction as well as confounding assessment is misclassification. Misclassification of any of the variables in the analysis (whether the exposure, disease, confounder, or modifier) can make a measure appear to vary across strata when in reality it does not, or make it appear nearly constant when in reality it does vary (Greenland 1980). Similarly, measurement errors can spuriously create or mask the need for product terms (interactions) in a statistical model (Greenland 1993b; cf. Chap. II.3 of this handbook). In an analogous fashion, variation in a measure across strata may be spuriously created or masked by differences in other biases (such as residual confounding or selection bias) across strata. Again, such problems can be explored using sensitivity analysis.

Conventional statistical analysis strategies often assume it is not appropriate to calculate an overall effect estimate if interaction is present. However, this prin-

ciple is commonly ignored if the difference in stratum-specific effect estimates is not too great. In fact standardized rate ratios have been developed for precisely this situation, and will consistently estimate meaningful epidemiological parameters even under heterogeneity (Rothman and Greenland 1998, Chap. 15; Greenland 2004b). Furthermore, as mentioned earlier, rate ratios estimated from multiplicative models often approximate these standardized ratios (Greenland and Maldonado 1994).

As mentioned above, concluding that there is no interaction because the *P*-value is high can be misleading. Most studies are not designed to examine interaction, and as such, may have inadequate study sizes within strata of an effect-measure modifier to permit a useful statistical interpretation. Presentation of stratum-specific effect estimates and their confidence intervals can help to give a picture of whether the data allow any inference about effect-measure modification. Formal statistical tests may be most useful in situations where prior information suggests likely forms of effect modification (e.g., a harmful effect would only be anticipated among smokers) and the study is intentionally designed to accommodate an analysis of effect modification (e.g., sufficient numbers of smokers and non-smokers are selected).

Some authors (e.g. Kleinbaum et al. 1982) have developed modeling strategies in which the first step of an analysis involves testing for statistical interactions, where the latter are represented by product terms in the model. In the most extreme application this involves including all possible two-factor (and even three-factor) product terms in a preliminary model and retaining in subsequent analyses all products (and related lower-order terms) that meet the inclusion criterion (which might be having a *p*-value below a certain cut-off, such as 0.10, or having a point estimate larger than a particular magnitude). This approach often results in complex models with numerous product terms, which may lead to problems of convergence, bias in the parameter estimates, and difficulties in interpretation.

In fact, there is no logical necessity for the assessment of interaction to occur as the first step in an analysis, and there are several reasons why it can be preferable to evaluate confounding before considering interaction. One reason is that the initial aim of most analyses is to determine if there is any overall effect of exposure. It is necessary to control confounding to do this, but it is not essential to evaluate interaction when doing so (Rothman 1978). Although harmful effects in one stratum and protective effects in another stratum may yield an overall null effect, this phenomenon is presumably rare. A routine search may yield a high percentage of false positives; on the other hand, if there were a relevant a priori hypothesis then it would be appropriate to calculate stratum specific effect estimates irrespective of the value of the summary effect estimate.

Another reason to begin an analysis with confounding evaluation is that inclusion of extra stratification variables or extra product terms involving the main exposure complicates confounder assessment. With extra strata or terms, changes in either stratum-specific or in summary fitted measures must be examined; the stratum-specific measures may be numerous and unstable, and the summary of these measures can be difficult to construct from a fitted model that has product

terms involving the exposure, see Rothman and Greenland (1998, pp 413–416) and Greenland (2004b) for example formulas. Measures constructed from models that omit exposure product terms are often a reasonable approximation to the formally correct and more complicated measures that allow for interactions, and so can be adequate for confounding evaluation (Greenland and Maldonado 1994).

Even if subsequent analyses concentrate on specific subgroups, it may be preferable to evaluate confounding in the whole data set, since this provides the greatest precision. If a factor is a confounder overall, then it is a risk factor, and is also associated with exposure. Thus it is necessarily a confounder in some specific subgroups, and there may be little loss of precision from control in any subgroups in which it is not a confounder (although this cannot be guaranteed). Hence, it may be preferable to evaluate confounding first, and then adjust for the same confounders in each subgroup analysis.

Some qualifications should be noted. First, confounding may be evaluated purely on a priori considerations, and as mentioned earlier has an inescapable a priori (causal) component in observational studies. Because of this causal component, purely statistical selection procedures such as stepwise regression can be even more misleading for confounder selection than they are in pure prediction problems (Greenland 1980). Second, the entire selection process and the attendant problems can be avoided by switching to hierarchical regression methods (Rothman and Greenland 1998, Chap. 21), which we discuss further below. For general principles of data analysis we refer to Chapter II.2 of this handbook and Chaps. 12 and 13 of Rothman and Greenland (1998).

Assessment of Joint Effects 9.4.1

The above considerations imply an apparent dilemma. How can an analysis be conducted that combines the advantages of ratio measures of effect with the assessment of interaction in terms of a departure from additivity? If an excess risk is found (and assumed to be causal) then attention shifts to elaborating the nature of the effect. This naturally comes toward the end of the formal presentation of the findings. Typically, the last few tables of a manuscript might examine the joint effects of the main exposure with other factors of interest, and the discussion might relate these findings to current etiologic knowledge. As noted above, it is often sufficient to evaluate only those joint effects for which there is an a priori reason for interest.

As an example, when studying asbestos and lung cancer, interaction with smoking might be expected given the powerful effects of smoking. To examine the latter interaction, relative risks might be presented for smoking (in non-asbestos workers), asbestos exposure (in non-smokers) and exposure to both factors, relative to persons exposed to neither factor. These relative risks would be adjusted for all other factors (e.g. age) that are potential confounders, but not of immediate interest as effect modifiers. The relevant table (e.g. Table 9.3) can be derived from any form of model, including the statistically convenient multiplicative models, by including product (interaction) terms as appropriate.

The estimation of separate and joint effects may be difficult when the factors of interest are closely correlated. However, when it is feasible, this approach combines the best features of multiplicative models and additive interaction assessment; it also permits readers with other concepts of independence to draw their own conclusions. It can be illustrated with the data presented above (Table 9.3) on asbestos exposure, cigarette smoking, and lung cancer. In this example, the relative risk estimates (adjusted for age and calendar period) are 7.0 for asbestos exposure alone, 17.5 for smoking alone, and 32.7 for the joint effect of both exposures. Thus, the joint effect of asbestos and smoking is more than additive (the joint effect is 32.7 times, whereas it would be $1 + (7.0 - 1) + (17.5 - 1) = 23.5$ if it were additive). This is consistent with the hypothesis that asbestos and smoking are joint components in at least one sufficient cause (it might be argued that non-additivity refutes the hypothesis that asbestos and smoking never biologically interact, assuming as usual that there is no residual confounding or bias). If their joint effect were the sum of their separate effects, the result would have favored the hypothesis that they are not joint components of a sufficient cause and do not compete for a common pool of susceptibles. However, the latter interpretation is more restricted, since additivity could arise if two factors were components of the same sufficient cause, but also had antagonistic or competitive effects that balanced their synergistic effects. Thus, even in ideal circumstances, additivity does not refute the hypothesis that asbestos and smoking interact biologically in some people (Greenland and Poole 1988; Rothman and Greenland 1998, Chap. 18).

If it is provisionally accepted that smoking and asbestos do act together in a sufficient cause of lung cancer, then attention may shift to elaborating the effect with mathematical models deduced from biologic models of the interaction. For example, Doll and Peto (1978) have suggested that smoking acts at both an early stage (probably the 2nd) and the penultimate (5th) stage of a 6-stage carcinogenic process. Asbestos appears to act at one of the later stages, probably the 4th or 5th (Pearce 1988). If asbestos acted at the same late stage as smoking, then it could be expected that its effect would add onto the late stage effect of smoking, and multiply the early stage effect of smoking. The resulting joint effect would be intermediate between additive and multiplicative. This pattern has been observed in several studies (Selikoff et al. 1980) although there are, of course, other models which predict the same result (Saracci 1987).

When interaction evaluation occurs as the last stage of an analysis, the routine evaluation (screening) of a large number of joint effects increases the number of tables, but does not necessarily complicate other aspects of the presentation (Pearce 1989). It does however raise a number of statistical problems which have been the subject of much controversy and research. The first, lesser known problem is that exposure effect estimates may be biased away from the null when too many terms (such as product terms) are entered into a risk or rate regression (Greenland et al. 2000). The second is the multiple-comparisons problem. Although many epidemiologists have denied that such problems exist (e.g. Rothman 1990), their focus concerned situations in which despite many comparisons, the investigator was interested in just one or a few exposure-disease relations. Nonetheless, screening

a large number of effects (whether main effects or interactions) implies interest in many relations, and raises the issues of how one deals with the instability of the estimates and the high probability that some of the estimates are large simply because of large random errors (Greenland and Robins 1991). Classical multiple-comparisons procedures can be quite misleading, however, because they make no attempt to account for false negative error (in fact they inflate it tremendously), and are arguably inferior to making no adjustment at all if one is more concerned about false-negatives than false-positives.

An analytic solution to both problems is to employ hierarchical modeling methods (also known as multilevel methods, penalized estimation, random-coefficient regression, shrinkage estimation, Stein estimation, empirical-Bayes regression, and semi-Bayes regression) (Greenland and Robins 1991; Rothman and Greenland 1998, pp 427–432; Greenland 2000b,c; Steenland et al. 2000). Such methods are demonstrably superior to either extreme (of no adjustment versus classical adjustment) in these situations, as shown by theory, simulations, and performance in real epidemiologic examples (Efron and Morris 1977; Greenland 1993c, 2000b,c, 2001b; Steenland et al. 2000; Witte et al. 2000). Furthermore, these methods can also be applied to control of multiple confounders in place of confounder selection methods (Greenland 2000c), and can be carried out with standard software (Witte et al. 2000; Greenland 2001b).

Conclusions

9.5

Confounding occurs when the exposed and non-exposed subpopulations of the source population have different background disease risks. When we make a comparison of the frequency of disease in the exposed and non-exposed groups, we would ideally wish to be able to assume that the disease frequency in the non-exposed group provides a valid estimate of what the disease frequency would have been in the exposed group if it had not been exposed. If this assumption is incorrect, i.e. if the exposed and non-exposed groups would have had different disease frequencies in the counterfactual situation in which the exposed group had not been exposed, then we say that the comparison of the exposed and unexposed groups is confounded. A related concept is that the exposed and non-exposed group are not "exchangeable", in that the estimated effects would have been different if the exposed group had not been exposed and the non-exposed group had been exposed (i.e. if the exposure status of the subjects had been exchanged).

Interaction usually means that the exposure effect on disease risk varies by some other factor. In other words, in order to estimate the effect of exposure, we must first know whether or not another factor is present (or what the level of this other factor is). This idea turns out to subsume two separate concepts: *effect-measure modification* (statistical interaction) and *biologic interaction*. When considering an exposure that has an effect, all other causal factors will modify either the rate ratio or the rate difference, and uniformity over one measure implies non-

uniformity over the other, e.g., an apparent additive joint effect implies a departure from a multiplicative model. Effect-measure modification is logically equivalent to the definition of "interaction" used in most statistics books and programs, and refers to a population measure of effect. In contrast, biologic interaction refers to effects in individuals although its absence implies absence of risk-difference modification.

In the simple case of a dichotomous main exposure (e.g. asbestos exposure), a dichotomous health outcome (e.g. lung cancer) and another categorized exposure (e.g. smoking vs. non-smoking), assessment of confounding involves stratifying on the potential confounder and assessing whether the stratum-specific effect estimates are similar to the (crude) overall effect estimate, e.g. how close are the relative risks in smokers and non-smokers (or a summary of these stratum-specific effect estimates) to the relative risk estimated when smoking is ignored? Assessment of effect-measure modification involves assessment of how the stratum specific effect estimates compare with each other, e.g. how does the relative risk in smokers compare with the relative risk in non-smokers? The two concepts are therefore often confused, because in this simple situation they are both assessed by stratification. However, confounding and interaction are completely different concepts. A factor may be a source of confounding, or effect-measure modification, or both, or neither.

Acknowledgements. The authors are grateful to Katherine J. Hoggatt and the editors for numerous helpful comments that improved this chapter. Funding for Neil Pearce's salary is from a Programme Grant from the Health Research Council of New Zealand.

References

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

Armstrong BG (1998) Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occupational & Environmental Medicine 55:651–656

Axelson O (1978) Aspects on confounding in occupational health epidemiology. Scandinavian Journal of Work, Environment & Health 4:85–89

Axelson O (1989) Confounding from smoking in occupational epidemiology. British Journal of Industrial Medicine 46:505–507

Axelson O, Steenland K (1988) Indirect methods of assessing the effects of tobacco use in occupational studies. American Journal of Industrial Medicine 13:105–118

Axelson O, Flodin U, Hardell L (1982) A comment on the reference series with regard to multiple exposure evaluations in a case-referent study. Scandinavian Journal of Work, Environment & Health 8(suppl 1):15–19

Bross IDJ (1967) Pertinency of an extraneous variable. Journal of Chronic Diseases 20:487–495

Checkoway H, Waldman GT (1985) Assessing the possible extent of confounding in occupational case-referent studies. Scandinavian Journal of Work, Environment & Health 11:131–133

Checkoway H, Pearce N, Kriebel D (2004) Research methods in occupational epidemiology. Oxford University Press, New York

Cole SR, Hernan MA (2002) Fallibility in estimating direct effects. International Journal of Epidemiology 31:163–165

Copeland KT, Checkoway H, McMichael AJ, Holbrook RH (1977) Bias due to misclassification in the estimation of relative risk. American Journal of Epidemiology 105:488–495

Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL (1959) Smoking and lung cancer: recent evidence and a discussion of some questions. Journal of the National Cancer Institute 22:173–203

Doll R, Peto R (1978) Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. Journal of Epidemiology & Community Health 32:303–313

Efron B, Morris C (1977) Stein's paradox in statistics. Scientific American 236:119–127

Espeland M, Hui SL (1987) A general approach to analyzing epidemiologic data that contain misclassification errors. Biometrics 43:1001–1012

Fingerhut MA, Halperin WE, Marlow DA, Piacitelli LA, Honchar PA, Sweeney MH, Greife AL, Dill PA, Steenland K, Suruda AJ (1991) Cancer mortality in workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. New England Journal of Medicine 324:212–218

Flanders WD, Khoury MJ (1990) Indirect assessment of confounding: graphic description and limits on effect for adjusting for covariates. Epidemiology 1:239–246

Greenland S (1980) The effect of misclassification in the presence of covariates. American Journal of Epidemiology 112:564–569

Greenland S (1988) Variance estimation for epidemiologic effect estimates under misclassification. Statistics in Medicine 7:745–757

Greenland S (1993a) Additive-risk versus additive relative-risk models. Epidemiology 4:32–36

Greenland S (1993b) Basic problems in interaction assessment. Environmental Health Perspectives 101(suppl 4):59–66

Greenland S (1993c) Methods for epidemiologic analyses of multiple exposures: A review and a comparative study of maximum-likelihood, preliminary testing, and empirical-Bayes regression. Statistics in Medicine 12:717–736

Greenland S (1996) Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. Epidemiology 7:498–501

Greenland S (2000a) An introduction to instrumental variables for epidemiologists. International Journal of Epidemiology 29:722–729

Greenland S (2000b) Principles of multilevel modelling. International Journal of Epidemiology 29:158–167

Greenland S (2000c) When should epidemiologic regressions use random coefficients? Biometrics 56:915–921
Greenland S (2001a) Ecologic versus individual-level sources of confounding in ecologic estimates of contextual health effects. International Journal of Epidemiology 30:1343–1350
Greenland S (2001b) Putting background information about relative risks into conjugate prior distributions. Biometrics 57:663–670
Greenland S (2003) The impact of prior distributions for uncontrolled confounding and response bias: A case study of the relation of wire codes and magnetic fields to childhood leukemia. Journal of the American Statistical Association 98: 47–54
Greenland S (2004a) Interval estimation by simulation as an alternative to and extension of confidence intervals. International Journal of Epidemiology (in press)
Greenland S (2004b) Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. American Journal of Epidemiology (in press)
Greenland S (2005) Multiple-bias modeling for observational studies. Journal of the Royal Statistical Society, Series A (in press)
Greenland S, Kleinbaum D (1983) Correcting for misclassification in two-way tables and matched-pair studies. International Journal of Epidemiology 12:93–97
Greenland S, Maldonado G (1994) The interpretation of multiplicative model parameters as standardised parameters. Statistics in Medicine 13:989–999
Greenland S, Neutra RR (1980) Control of confounding in the assessment of medical technology. International Journal of Epidemiology 9:361–367
Greenland S, Poole C (1988) Invariants and noninvariants in the concept of interdependent effects. Scandinavian Journal of Work, Environment & Health. 14:125–129
Greenland S, Robins JM (1985) Confounding and misclassification. American Journal of Epidemiology 122:495–506
Greenland S, Robins JM (1986) Identifiability, exchangeability, and epidemiological confounding. International Journal of Epidemiology 15:413–419
Greenland S, Robins JM (1991) Empirical-Bayes adjustments for multiple comparisons are sometimes useful. Epidemiology 2:244–251
Greenland S, Pearl J, Robins JM (1999a) Causal diagrams for epidemiologic research. Epidemiology 10:37–48
Greenland S, Robins JM, Pearl J (1999b) Confounding and collapsibility in causal inference. Statistical Science 14:29–46
Greenland S, Schwartzbaum JA, Finkle WD (2000) Problems due to small samples and sparse data in conditional logistic regression analysis. American Journal of Epidemiology 151:531–539
Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA (2002) Causal knowledge as a prerequisite for confounding evaluation. American Journal of Epidemiology 155:176–184

Hernan MA, Hernandez-Diaz S, Robins JM (2004) A structural approach to selection bias. Epidemiology 15:615–625

Kleinbaum DG, Kupper LL, Morgenstern H (1982) Epidemiologic research. Principles and quantitative methods. Lifetime Learning Publication, Belmont, CA

Kogevinas M, Pearce N, Susser M, Boffetta P (1997) Social inequalities and cancer. IARC Scientific Publications No. 138, Lyon

Koopman JS (1977) Causal models and sources of interaction. American Journal of Epidemiology 106:439–444

Kriebel D, Zeka A, Eisen EA, Wegman DH (2004) Quantitative evaluation of the effects of uncontrolled confounding by alcohol and tobacco in occupational cancer studies. International Journal of Epidemiology 33:1–6

Kupper LL, Hogan MD (1978) Interaction in epidemiologic studies. American Journal of Epidemiology 108:447–453

Lash TL, Fink AK (2003) Semi-automated sensitivity analysis to assess systematic errors in observational epidemiologic data. Epidemiology 14:451–458

Maldonado G, Greenland S (1993) A simulation study of confounder-selection strategies. American Journal of Epidemiology 138:923–936

Marshall JR, Hastrup JL (1996) Mismeasurement and the resonance of strong confounders: uncorrelated errors. American Journal of Epidemiology 143:1069–1078

Marshall JR, Hastrup JL (1999) Mismeasurement and the resonance of strong confounders: correlated errors. American Journal of Epidemiology 150:88–96

Miettinen OS (1974) Confounding and effect-modification. American Journal of Epidemiology 100:350–353

Miettinen OS (1982) Causal and preventive interdependence. Elementary principles. Scandinavian Journal of Work, Environment & Health 8:159–168

Miettinen OS (1985) The "case-control" study: valid selection of subjects. Journal of Chronic Diseases 38:543–548

Miettinen OS, Cook EF (1981) Confounding: essence and detection. American Journal of Epidemiology 114:593–603

Moolgavkar SH, Venzon DJ (1987) General relative risk regression models for epidemiologic studies.[comment]. American Journal of Epidemiology 126:949–961

Pearce N (1988) Multistage modelling of lung cancer mortality in asbestos textile workers. International Journal of Epidemiology 17:747–752

Pearce N (1989) Analytical implications of epidemiological concepts of interaction. International Journal of Epidemiology 18:976–980

Pearce N, Checkoway H (1988) Case-control studies using other diseases as controls: problems of excluding exposure-related diseases.[comment]. American Journal of Epidemiology 127:851–856

Pearl J (2000) Causality: models, reasoning and inference. Cambridge University Press, Cambridge

Phillips CV (2003) Quantifying and reporting uncertainty from systematic errors. Epidemiology 14:459–466

Robins JM (1989) The control of confounding by intermediate variables. Statistics in Medicine 8:679–701
Robins JM (2001) Data, design, and background knowledge in etiologic inference. Epidemiology 12:550–560
Robins JM, Greenland S (1992) Identifiability and exchangeability for direct and indirect effects. Epidemiology 3:143–155
Robins JM, Greenland S (1994) Adjusting for differential rates of prophylaxis therapy for PCP in high-dose versus low-dose AZT treatment arms in an AIDS randomized trial. Journal of the American Statistical Association 89:737–749
Robins JM, Morgenstern H (1987) The foundations of confounding in epidemiology. Computers and Mathematics with Applications 14:869–916
Robins JM, Blevins D, Ritter G, Wulfsohn M (1992) G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology 3:319–336
Robins JM, Hernan MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. Epidemiology 11:550–560
Rothman KJ (1976) Causes. American Journal of Epidemiology 104:587–592
Rothman KJ (1978) Occam's razor pares the choice among statistical models. American Journal of Epidemiology 108:347–349
Rothman KJ (1990) No adjustments are needed for multiple comparisons. Epidemiology 1:43–46
Rothman KJ, Greenland S (1998) Modern epidemiology, 2nd edn. Lippincott-Raven, Philadelphia
Rothman KJ, Greenland S, Walker AM (1980) Concepts of interaction. American Journal of Epidemiology 112:467–470
Saracci R (1987) The interactions of tobacco smoking and other agents in cancer etiology. Epidemiologic Reviews 9:175–193
Savitz DA, Baron EA (1989) Estimating and correcting for confounder misclassification. American Journal of Epidemiology 129:1062–1071
Schlesselman JJ (1978) Assessing effects of confounding variables. American Journal of Epidemiology 99:3–8
Schlesselman JJ (1982) Case-control studies: design, conduct, analysis. Oxford University Press, New York
Selikoff IJ, Seidman H, Hammond EC (1980) Mortality effects of cigarette smoking among amosite asbestos factory workers. Journal of the National Cancer Institute 65:507–513
Siemiatycki J, Thomas DC (1981) Biological models and statistical interactions: an example from multistage carcinogenesis. International Journal of Epidemiology 10:383–387
Siemiatycki J, Wacholder S, Dewar R, Wald L, Begin D, Richardson L, Rosenman K, Gerin M (1988) Smoking and degree of occupational exposure: are internal analyses in cohort studies likely to be confounded by smoking status? American Journal of Industrial Medicine 13:59–69
Sommer A, Zeger SL (1991) On estimating efficacy from clinical trials. Statistics in Medicine 10:45–52

Steenland K, Thun M (1986) Interaction between tobacco smoking and occupational exposures in the causation of lung cancer. Journal of Occupational Medicine 28:110–118

Steenland K, Beaumont J, Halperin W (1984) Methods of control for smoking in occupational cohort mortality studies. Scandinavian Journal of Work, Environment & Health 10:143–149

Steenland K, Bray I, Greenland S, Boffetta P (2000) Empirical Bayes adjustments for multiple results in hypothesis-generating or surveillance studies. Cancer Epidemiology Biomarkers & Prevention 9:895–903

Thomas DC, Greenland S (1983) The relative efficiencies of matched and independent sample designs for case-control studies. Journal of Chronic Diseases 36:685–697

Thomas DC, Stram D, Dwyer J (1993) Exposure-measurement error: influence on exposure-disease relationships and methods of correction. Annual Review of Public Health 14:69–93

Walter SD, Holford TR (1978) Additive, multiplicative, and other models for disease risks. American Journal of Epidemiology 108:341–346

Weinberg CR (1993) Toward a clearer definition of confounding. American Journal of Epidemiology 137:1–8

Witte JS, Greenland S, Kim L-L, Arab L (2000) Multilevel modeling in epidemiology with GLIMMIX. Epidemiology 11:684–688

Yanagawa T (1984) Case-control studies: assessing the effects of a confounding factor. Biometrika 71:191–194