

Bias Analysis

Sander Greenland and Timothy L. Lash

Introduction	345	Confounder Misclassification	359
The Need for Bias Analyses	346	Misclassification of Multiple Variables	360
Caveats about Bias Analysis	347	A Matrix Adjustment Method	360
Analysis of Unmeasured Confounders	348	Use of Validation Data	361
External Adjustment	348	Selection Bias	362
Relation of Unadjusted to Adjusted Odds Ratios	351	Multiple-Bias Analyses	363
Combination with Adjustment for Measured Confounders	352	Probabilistic Bias Analysis	364
Analysis of Misclassification	352	Probabilistic Sensitivity Analysis	364
Exposure Misclassification	353	Analysis of Uncontrolled Confounding	365
Nondifferentiality	355	Some Easily Generated Distributions	370
Application to the Resins-Lung Cancer Example	356	Trapezoidal Distributions	371
Relation of Predictive Values to Sensitivity and Specificity	357	Generating Dependent Variables	371
Disease Misclassification	358	Analysis of Misclassification	372
		Analysis of Selection Bias	375
		Comparative and Combined Analyses	376
		Bayesian and Semi-Bayesian Analysis	378
		Conclusion	380

INTRODUCTION

This chapter provides an introduction to quantitative methods for evaluating potential biases (systematic errors) in individual studies. The first half of this chapter covers basic methods to assess sensitivity of results to confounding by an unmeasured variable, misclassification, and selection bias. These methods quantify systematic errors by means of *bias parameters*, which in a sensitivity analysis are first fixed at hypothetical values and then varied to see how the results vary with the parameters.

The second half of this chapter extends basic bias-sensitivity analysis by assigning a prior probability distribution to the bias parameters (probabilistic bias modeling) to produce *distributions* of results as output. The methods are typically implemented via simulation, and their outputs have a natural interpretation as semi-Bayesian posterior distributions (Chapter 18) for exposure effects. They are “semi-” Bayesian because they use prior distributions for the bias parameters but not for the effect under study. We focus on the special case in which the observed data can be represented in a 2×2 table of an exposure indicator X (coded 1 = present, 0 = absent) and a disease indicator D . Many of the basic principles and difficulties of bias analysis can be illustrated with this simple case, because the 2×2 table can be thought of as a stratum from a larger data set.

Most statistical methods also assume specific models for the form of effects (of exposure, modifiers, and confounders) and of random errors. Use of erroneous model forms is sometimes

called *specification error* and can lead to systematic errors known as *specification biases*. *Model-sensitivity analysis* addresses these biases by seeing how the results change as the model form is changed (Leamer, 1985; Draper, 1995; Saltelli et al., 2000). We do not cover model-sensitivity analysis, because it involves technical issues of model selection beyond the scope of this chapter; see Chapter 21 for further discussion.

We also do not address general *missing-data bias* (bias due to nonrandomly incomplete data); see Robins et al. (1994) and Little and Rubin (2002) for discussions, especially under the heading of *informatively or nonignorably missing data* (Lyles and Allen, 2002). All the problems we discuss can be viewed as extreme cases of missing-data bias: Uncontrolled confounding is due to missing data on a confounder; misclassification is due to missing data on the true variables; and selection bias is due to nonrandomly missing members of the source population.

All methods, whether conventional methods or those described here, consider the data as given; that is, they assume that we have the data and that the data have not been corrupted by miscodings, programming errors, forged responses, etc. Thus, the methods assume that there is no misclassification due to data processing errors or investigator fraud. Such problems arise from isolated events that may affect many records, and their correction depends entirely on detection (e.g., via logic checks, data examination, and comparison of data sources). Thus, we do not see such problems as falling within the sphere of sensitivity analysis.

THE NEED FOR BIAS ANALYSES

Our discussion of statistical methods has so far focused on accounting for measured confounders and random errors in the data-generating process. Randomization of exposure assignment is the conventional assumption of statistical methods for causal inference within a study cohort, for it makes confounding a chance phenomenon. Random sampling forms the analogous conventional assumption of statistical methods for inference from the sample to a larger population (e.g., from a case-control study to a source population), for it makes sampling error a chance phenomenon. Most methods assume that measurement error is absent, but those that account for errors assume that the errors are random (Carroll et al., 2006). Upon stratification, these assumptions (that confounding, sampling errors, and measurement errors are random) are made within levels of the stratifying variables. We will call methods based on these randomness assumptions *conventional methods*.

By assuming that all errors are random and that any modeling assumptions (such as homogeneity) are correct, all uncertainty about the effect of errors on estimates is subsumed within conventional standard deviations for the estimates (standard errors), such as those given in earlier chapters (which assume no measurement error), and any discrepancy between an observed association and the target effect may be attributed to chance alone. When the assumptions are incorrect, however, the logical foundation for conventional statistical methods is absent, and those methods may yield highly misleading inferences. Epidemiologists recognize the possibility of incorrect assumptions in conventional analyses when they talk of residual confounding (from nonrandom exposure assignment), selection bias (from nonrandom subject selection), and information bias (from imperfect measurement). These biases rarely receive quantitative analysis, a situation that is understandable given that the analysis requires specifying values (such as amount of selection bias) for which little or no data may be available. An unfortunate consequence of this lack of quantification is the switch in focus to those aspects of error that are more readily quantified, namely, the random components.

Systematic errors can be and often are larger than random errors, and failure to appreciate their impact is potentially disastrous. The problem is magnified in large studies and pooling projects, because in those studies the large size reduces the amount of random error, and as a result the random error may be only a small component of total error. In such studies, a focus on “statistical significance” or even on confidence limits may amount to nothing more than a decision to focus on artifacts of systematic error as if they reflect a real causal effect.

Addressing concerns about systematic errors in a constructive fashion is not easy, but is nonetheless essential if the results of a study are to be used to inform decisions in a rational fashion. The process of addressing bias quantitatively we shall call *bias analysis*. As described in a number of books (e.g., Eddy et al., 1992; National Research Council, 1994; Vose, 2000), the basic ideas have existed for decades under the topic of sensitivity analysis and the more general topics of *uncertainty*

analysis, risk assessment, and risk analysis. These topics address more sources of uncertainty than we shall address, such as model misspecification and informatively missing data. Here we focus only on the effects of basic validity problems.

A discomforting aspect of these analyses is that they reveal the highly tentative and subjective nature of inference from observational data, a problem that is concealed by conventional statistical analysis. Bias analysis requires educated guesses about the likely sizes of systematic errors, guesses that are likely to vary considerably across observers. The conventional approach is to make the guess qualitatively by describing the study's limitations. An assessment of the extent of bias, compared with the extent of exposure effects, therefore becomes an exercise in intuitive reasoning under uncertainty.

The ability to reason under uncertainty has been studied by cognitive psychologists and sociologists, who have found it susceptible to many predictable patterns of mistakes (Kahneman et al., 1982; Gilovich, 1993; Gilovich et al., 2002). This literature, where it deals with situations analogous to epidemiologic inference, indicates that the qualitative approach tends to favor exposure effects over systematic errors as an explanation for observed associations (Lash, 2007). Quantitative methods such as those described in this chapter offer a potential safeguard against these failures, by providing insight into the importance of various sources of error and by helping to assess the uncertainty of study results. For example, such assessments may argue persuasively that certain sources of bias cannot by themselves plausibly explain a study result, or that a bias explanation cannot be ruled out. As discussed in Chapters 2 and 18, and later in this section, the primary caution is that what appears "plausible" may vary considerably across persons and time.

There are several reasons why quantitative methods that take account of uncontrolled biases have traditionally seen much less development than methods for addressing random error. First, until recently, randomized experiments supplied much of the impetus for statistical developments. These experiments were concentrated in agriculture, manufacturing, and clinical medicine and often could be designed so that systematic errors played little role in the final results. A second reason is that most uncontrolled biases cannot be analyzed by conventional methods (i.e., without explicit priors for bias parameters) unless additional "validation" data are available. Such data are usually absent or very limited. Furthermore, validation studies may themselves be subject to systematic errors beyond those present in the main study, such as potentially biased selection for validation (e.g., if validation requires further subject consent and participation). As a result, investigators must resort to less satisfactory partial analyses, or quantify only the uncertainty due to random error.

Editors and reviewers in the health sciences seldom call on authors to provide a quantitative assessment of systematic errors. Because of the labor and expertise required for a bias analysis, and the limited importance of single studies for policy issues, it makes little sense to require such an analysis of every study. For example, studies whose conventional 95% confidence limits exclude no reasonable possibility will be viewed as inconclusive regardless of any further analysis. It can be argued that the best use of effort and journal space for single-study reports is to focus on a thorough description of the study design, methods, and data, to facilitate later use of the study data in reviews, meta-analyses, and pooling projects (Greenland et al., 2004).

On the other hand, any report with policy implications may damage public health if the claimed implications are wrong. Thus it is justifiable to demand quantitative bias analysis in such studies. Going further, it is arguably an ethical duty of granting agencies and editors to require a thorough quantitative assessment of relevant literature and of systematic errors to support claimed implications for public policy or medical practice. Without the endorsement of these gatekeepers of funding and publication, there is little motivation to collect validation data or to undertake quantitative assessments of bias.

CAVEATS ABOUT BIAS ANALYSIS

As noted above, results of a bias analysis are derived from inputs specified by the analyst, a point that should be emphasized in any presentation of the methods or their results. These inputs are constructed from judgments, opinions, or inferences about the likely magnitude of bias sources or parameters. Consequently, bias analyses do not establish the existence or absence of causal effects any more than do conventional analyses. Rather, they show how the analysts developed their output judgments (inferences) from their input judgments.

An advantage of bias analysis over a qualitative discussion of study limitations is that it allows mathematics to replace unsound intuitions and heuristics at many points in judgment formation. Nonetheless, the mathematics should not entice researchers to phrase judgments in objective terms that mask their subjective origin. For example, a claim that “our analysis indicates the conventional results are biased away from the null” would be misleading. A better description would say that “our analysis indicates that, *under the values we chose for the bias parameters*, the conventional results *would be* biased away from the null.” The latter description acknowledges the fact that the results are sensitive to judgmental inputs, some of which may be speculative. The description, by its nature, should also encourage the analyst to present evidence that the values chosen for the bias parameters cover the range of reasonable combinations of those parameters.

The more advanced methods of this chapter require input distributions (rather than sets of values) for bias parameters, but the same caveat holds: The results of the bias analysis apply only under those chosen distributions. The analysis will be more convincing if the analyst provides evidence that the chosen distributions assign high probability to reasonable combinations of the parameters.

To some extent, similar criticisms apply to conventional frequentist and Bayesian analyses (Chapters 13–18 and 20–21), insofar as those analyses require many choices and judgments from the investigators. Examples include choice of methods used to handle missing data (Chapter 13), choice of category boundaries for quantitative variables (Chapter 13 and 17), choice of methods for variable selection (Chapter 21), and choice of priors assigned to effects under study (Chapter 18). As the term “conventional” connotes, many choices have default answers (e.g., a binomial model for the distribution of a dichotomous outcome). Although the scientific basis for these defaults is often doubtful or lacking (e.g., missing-data indicators; percentile boundaries for continuous variables; stepwise variable selection; noninformative priors for effects), deviations from the defaults may prompt requests for explanations from referees and readers.

Bias analysis requires more input specifications than do conventional analyses, and as yet there is no accepted convention regarding these specifications. As a result, input judgments are left entirely to the analyst, opening avenues for manipulation to produce desired output. Thus, when examining a bias analysis, a reader must bear in mind that other reasonable inputs might produce quite different results. This input sensitivity is why we emphasize that bias analysis is a collection of methods for explaining and refining subjective judgments in light of data (like the subjective Bayesian methods of Chapter 18), rather than a method for detecting nonrandom data patterns. In fact, a bias analysis can be made to produce virtually any estimate for the study effect without altering the data or imposing an objectionable prior on that effect. Such outcome-driven analyses, however, may require assignment of values or distributions to bias parameters that have doubtful credibility. Therefore, it is crucial that the inputs used for a bias analysis be described in detail so that those inputs can be examined critically by the reader.

ANALYSIS OF UNMEASURED CONFOUNDERS

Sensitivity analysis and external adjustment for confounding by dichotomous variables appeared in Cornfield et al. (1959) and were further elaborated by Bross (1966, 1967), Yanagawa (1984), Axelson and Steenland (1988), and Gail et al. (1988). Extensions of these approaches to multiple-level confounders are available (Schlesselman, 1978, correction by Simon, 1980b; Flanders and Khouri, 1990; Rosenbaum, 2002); see Chapter 33 for an example. Although most of these methods assume that the odds ratios or risk ratios are constant across strata, it is possible to base external adjustment on other assumptions (Yanagawa, 1984; Gail et al., 1988). Practical extensions to multiple regression analyses typically involve modeling the unmeasured confounders as latent (unobserved) variables (e.g., Lin et al., 1998; Robins et al., 1999a; Greenland, 2005b; McCandless et al., 2007)).

EXTERNAL ADJUSTMENT

Suppose that we have conducted an analysis of an exposure X and a disease D , adjusting for the recorded confounders, but we know of an unmeasured potential confounder and want to assess the possible effect of failing to adjust for this confounder. For example, in a case-control study of occupational exposure to resin systems (resins) and lung cancer mortality among male workers at a transformer-assembly plant, Greenland et al. (1994) could adjust for age and year of death, but

TABLE 19–1

Crude Data for Case-Control Study of Occupational Resins Exposure (X) and Lung Cancer Mortality (Greenland et al., 1994); Controls Are Selected Noncancer Causes of Death

	$X = 1$	$X = 0$	Total
Cases ($D = 1$)	$A_{1+} = 45$	$A_{0+} = 94$	$M_{1+} = 139$
Controls ($D = 0$)	$B_{1+} = 257$	$B_{0+} = 945$	$M_{0+} = 1202$

Odds ratio after adjustment for age and death year: 1.77.

Age-year adjusted conventional 95% confidence limits for OR_{DX} : 1.18, 2.64.

they had no data on smoking. Upon adjustment for age and year at death, a positive association was observed for resins exposure and lung cancer mortality ($OR = 1.77$, 95% confidence limits = 1.18, 2.64). To what extent did confounding by smoking affect this observation?

For simplicity, suppose that resins exposure and smoking are treated as dichotomous: $X = 1$ for resin-exposed, 0 otherwise; $Z = 1$ for smoker, 0 otherwise. We might wish to know how large the resins/smoking association has to be so that adjustment for smoking removes the resins-lung cancer association. The answer to this question depends on a number of parameters, among them (a) the resins-specific associations (i.e., the associations within levels of resins exposure) of smoking with lung cancer, (b) the resins-specific prevalences of smoking among the controls, and (c) the prevalence of resins exposure among the controls. Resins prevalence is observed, but we can only speculate about the first two quantities.

It is this speculation, or educated guessing, that forms the basis for sensitivity analysis. We will assume various plausible combinations of values for the smoking/lung cancer association and resins-specific smoking prevalences, then see what values we get for the smoking-adjusted resins-lung cancer association. If all the latter values are substantially elevated, we have a basis for doubting that the unadjusted resins-lung cancer association is due entirely to confounding by smoking. Otherwise, confounding by smoking is a plausible explanation for the observed resins-lung cancer association.

We will use the crude data in Table 19–1 for illustration. There is no evidence of important confounding by age or year in these data, probably because the controls were selected from other chronic-disease deaths. For example, the crude odds ratio is 1.76 versus an age-year adjusted odds ratio of 1.77, and the crude 95% confidence limits are 1.20 and 2.58 versus 1.18 and 2.64 after age-year adjustment. If it is necessary to stratify on age, year or both, we could repeat the computations given below for each stratum and then summarize the results across strata, or we could use regression-based adjustments (Greenland, 2005b).

Consider the general notation for the expected stratified data given in Table 19–2. We will use hypothesized values for the stratum-specific prevalences to fill in this table and solve for an assumed

TABLE 19–2

General Layout (Expected Data) for Sensitivity Analysis and External Adjustment for a Dichotomous Confounder Z

	$Z = 1$			$Z = 0$		
	$X = 1$	$X = 0$	Total	$X = 1$	$X = 0$	Total
Cases	A_{11}	A_{01}	M_{11}	$A_{1+} - A_{11}$	$A_{0+} - A_{01}$	$M_{1+} - M_{11}$
Controls	B_{11}	B_{01}	M_{01}	$B_{1+} - B_{11}$	$B_{0+} - B_{01}$	$M_{0+} - M_{01}$

common odds ratio relating exposure to disease within levels of Z ,

$$\begin{aligned} OR_{DX} &= \frac{A_{11}B_{01}}{A_{01}B_{11}} \\ &= \frac{(A_{1+} - A_{11})(B_{0+} - B_{01})}{(A_{0+} - A_{01})(B_{1+} - B_{11})} \end{aligned}$$

Suppose that the smoking prevalences among the exposed and unexposed populations are estimated or assumed to be P_{Z1} and P_{Z0} , and the odds ratio relating the confounder and disease within levels of exposure is OR_{DZ} (i.e., we assume odds-ratio homogeneity). Assuming that the control group is representative of the source population, we set $B_{11} = P_{Z1}B_{1+}$ and $B_{01} = P_{Z0}B_{0+}$. Next, to find A_{11} and A_{01} , we solve the pair of equations

$$OR_{DZ} = \frac{A_{11}(B_{1+} - B_{11})}{(A_{1+} - A_{11})B_{11}}$$

and

$$OR_{DZ} = \frac{A_{01}(B_{0+} - B_{01})}{(A_{0+} - A_{01})B_{01}}$$

These have solutions

$$A_{11} = OR_{DZ}A_{1+}B_{11}/(OR_{DZ}B_{11} + B_{1+} - B_{11}) \quad [19-1]$$

and

$$A_{01} = OR_{DZ}A_{0+}B_{01}/(OR_{DZ}B_{01} + B_{0+} - B_{01}) \quad [19-2]$$

Having obtained data counts corresponding to A_{11} , A_{01} , B_{11} , and B_{01} , we can put these numbers into Table 19-2 and compute directly a Z -adjusted estimate of the exposure-disease odds ratios OR_{DX} . The answers from each smoking stratum should agree.

The preceding estimate of OR_{DX} is sometimes said to be “indirectly adjusted” for Z , because it is the estimate of OR_{DX} that one would obtain if one had data on the confounder Z and disease D that displayed the assumed prevalences and confounder odds ratio OR_{DZ} . A more precise term for the resulting estimate of OR_{DX} is “externally adjusted,” because the estimate makes use of an estimate of OR_{DZ} obtained from sources external to the study data. The smoking prevalences must also be obtained externally; occasionally (and preferably), they may be obtained from a survey of the underlying source population from which the subjects were selected. (Because we assumed the odds ratios are constant across strata, the result does not depend on the exposure prevalence.)

To illustrate external adjustment with the data in Table 19-1, suppose that the smoking prevalences among the resins exposed and unexposed are 70% and 50%. Then

$$B_{11} = P_{Z1}B_{1+} = 0.70(257) = 179.9$$

and

$$B_{01} = P_{Z0}B_{0+} = 0.50(945) = 472.5$$

Taking $OR_{DZ} = 5$ for the resins-specific smoking/lung cancer odds ratio, equations 19-1 and 19-2 yield

$$A_{11} = 5(45)179.9/[5(179.9) + 257 - 179.9] = 41.45$$

and

$$A_{01} = 5(94)472.5/[5(472.5) + 945 - 472.5] = 78.33$$

Putting these results into Table 19-2, we obtain the stratum-specific resins-lung cancer odds ratios

$$OR_{DX} = \frac{41.45(472.5)}{179.9(78.33)} = 1.39$$

TABLE 19–3

Sensitivity of Externally Adjusted Resins–Cancer Odds Ratio OR_{DX} to Choice of P_{Z1} and P_{Z0} (Smoking Prevalences among Exposed and Unexposed), and OR_{DZ} (Resins-Specific Smoking–Cancer Odds Ratio)

P_{Z1}	P_{Z0}	OR_{XZ}	OR_{DZ}		
			5	10	15
0.40	0.30	1.56	$OR_{DX} = 1.49$	$OR_{DX} = 1.42$	$OR_{DX} = 1.39$
0.55	0.45	1.49	1.54	1.49	1.48
0.70	0.60	1.56	1.57	1.54	1.53
0.45	0.25	2.45	1.26	1.13	1.09
0.60	0.40	2.25	1.35	1.27	1.24
0.75	0.55	2.45	1.41	1.35	1.33

and

$$OR_{DX} = \frac{(45 - 41.45)(945 - 472.5)}{(257 - 179.9)(94 - 78.33)} = 1.39$$

which agree (as they should). We see that confounding by smoking could account for much of the crude resins odds ratio if there were a much higher smoking prevalence among the resin exposed relative to the unexposed.

In a sensitivity analysis, we repeat the above external adjustment process using other plausible values for the prevalences and the confounder effect (see Sundararajan et al., 2002; Marshall et al., 2003; and Maldonado et al., 2003 for examples). Table 19–3 presents a summary of results using other values for the resins-specific smoking prevalences and the smoking odds ratio. The table also gives the smoking-resins odds ratio

$$OR_{XZ} = O_{Z1}/O_{Z0} = P_{Z1}(1 - P_{Z0})/(1 - P_{Z1})P_{Z0}$$

where $O_{Zj} = P_{Zj}/(1 - P_{Zj})$ is the odds of $Z = 1$ versus $Z = 0$ when $X = j$. There must be a substantial exposure-smoking association to remove most of the exposure-cancer association. Because there was no reason to expect an exposure-smoking association at all, Table 19–3 supports the notion that the observed resins-cancer association is probably not due entirely to confounding by the dichotomous smoking variable used here. We would have to consider a polytomous smoking variable to further address confounding by smoking.

RELATION OF UNADJUSTED TO ADJUSTED ODDS RATIOS

An equivalent approach to that just given uses the following formulas for the ratio of the unadjusted to Z -adjusted odds ratios (Yanagawa, 1984):

$$\begin{aligned} \frac{OR_{DX\text{-unadjusted}}}{OR_{DX\text{-adjusted}}} &= \frac{(OR_{DZ}O_{Z1} + 1)(O_{Z0} + 1)}{(OR_{DZ}O_{Z0} + 1)(O_{Z1} + 1)} \\ &= \frac{(OR_{DZ}OR_{XZ}O_{Z0} + 1)(O_{Z0} + 1)}{(OR_{DZ}O_{Z0} + 1)(OR_{XZ}O_{Z0} + 1)} \\ &= \frac{OR_{DZ}OR_{XZ}P_{Z0} + 1 - P_{Z0}}{(OR_{DZ}P_{Z0} + 1 - P_{Z0})(OR_{XZ}P_{Z0} + 1 - P_{Z0})} \\ &= \frac{OR_{DZ}P_{Z1} + 1 - P_{Z1}}{OR_{DZ}P_{Z0} + 1 - P_{Z0}} \end{aligned} \quad [19-3]$$

Assuming that Z is the sole uncontrolled confounder, this ratio can be interpreted as the degree of bias due to failure to adjust for Z . This series of equations shows that when Z is not associated with the disease ($OR_{DZ} = 1$) or is not associated with exposure ($OR_{XZ} = 1$), the ratio of the unadjusted and adjusted odds ratios is 1, and there is no confounding by Z . In other words, a confounder must be associated with the exposure and the disease in the source population (Chapter 9). Recall, however, that these associations are not sufficient for Z to be a confounder, because a confounder must also satisfy certain causal relations (e.g., it must not be affected by exposure or disease; see Chapters 4, 9, and 12). The equations in 19–3 also show that the ratio of unadjusted to adjusted odds ratios depends on the prevalence of $Z = 1$; i.e., the degree of confounding depends not only on the magnitude of the associations but also on the confounder distribution.

In many circumstances, we may have information about only one or two of the three parameters that determine the unadjusted/adjusted ratio. Nonetheless, it can be seen from 19–3 that the ratio cannot be further from 1 than are $OR_{DZ}/(OR_{DZ}P_{Z0} + 1 - P_{Z0})$, $OR_{XZ}/(OR_{XZ}P_{Z0} + 1 - P_{Z0})$, $1/(OR_{DZ}P_{Z0} + 1 - P_{Z0})$, or $1/(OR_{XZ}P_{Z0} + 1 - P_{Z0})$; the ratio is thus bounded by these quantities. Furthermore, because the bound $OR_{DZ}/(OR_{DZ}P_{Z0} + 1 - P_{Z0})$ cannot be further from 1 than OR_{DZ} , the ratio cannot be further from 1 than OR_{DZ} , and similarly cannot be further from 1 than OR_{XZ} (Cornfield et al., 1959; Bross, 1967). Thus, the odds-ratio bias from failure to adjust Z cannot exceed the odds ratio relating Z to D or to X .

These methods readily extend to cohort studies. For data with person-time denominators T_{ji} , we use the T_{ji} in place of the control counts B_{ji} in the previous formulas to obtain an externally adjusted rate ratio. For data with count denominators N_{ji} , we use the N_{ji} in place of the B_{ji} to obtain an externally adjusted risk ratio (Flanders and Khoury, 1990). Bounds analogous to those above can be derived for risk differences (Kitagawa, 1955). Improved bounds can also be derived under deterministic causal models relating X to D in the presence of uncontrolled confounding (e.g., Maclehose et al., 2005). There is also a large literature on bounding causal risk differences from randomized trials when uncontrolled confounding due to noncompliance may be present; see Chapter 8 of Pearl, 2000 for references.

COMBINATION WITH ADJUSTMENT FOR MEASURED CONFOUNDERS

The preceding equations relate the unadjusted odds ratio to the odds ratio adjusted only for the unmeasured confounder (Z) and thus ignore the control of any other confounders. If adjustment for measured confounders has an important effect, the equations must be applied using bias parameters conditioned on those measured confounders. To illustrate, suppose that age adjustment was essential in the previous example. We should then have adjusted for confounding by smoking in the age-adjusted or age-specific odds ratios. Application of the previous equations to these odds ratios will require age-specific parameters, e.g., P_{Z0} will be the age-specific smoking prevalence among unexposed noncases, OR_{DZ} will be the age-specific association of smoking with lung cancer among the unexposed, and OR_{XZ} will be the age-specific association of smoking with resins exposure among noncases.

Although most estimates of confounder-disease associations are adjusted for major risk factors such as age, information adjusted for other parameters is often unavailable. Use of unadjusted parameters in the preceding equations may be misleading if they are not close to the adjusted parameters (e.g., if the unadjusted and age-adjusted odds ratios associating smoking with exposure are far apart). For example, if age is associated with smoking and exposure, adjustment for age could partially adjust for confounding by smoking, and the association of smoking with exposure will change upon age adjustment. Use of the age-unadjusted smoking-exposure odds ratio (OR_{XZ}) in the preceding equations will then give a biased estimate of the residual confounding by smoking after age adjustment. More generally, proper external adjustment in combination with adjustments for measured confounders requires information about the unmeasured variables that is conditional on the measured confounders.

ANALYSIS OF MISCLASSIFICATION

Nearly all epidemiologic studies suffer from some degree of measurement error, which is usually referred to as classification error or *misclassification* when the variables are discrete. The effect of

even modest amounts of error can be profound, yet rarely is the error quantified (Jurek et al., 2007). Simple situations can be analyzed, however, using basic algebra (Copeland et al., 1977; Greenland, 1982a; Kleinbaum et al., 1984), and more extensive analyses can be done using software that performs matrix algebra (e.g., SAS, GAUSS, MATLAB, R, S-Plus) (Barron, 1977; Greenland and Kleinbaum, 1983; Greenland, 1988b). We will focus on basic methods for dichotomous variables. We will then briefly discuss methods that allow use of validation study data, in which classification rates are themselves estimated from a sample of study subjects.

EXPOSURE MISCLASSIFICATION

Consider first the estimation of exposure prevalence from a single observed category of subjects, such as the control group in a case-control study. Define the following quantities in this category:

- $X = 1$ if exposed, 0 if not
- $X^* = 1$ if *classified* as exposed, 0 if not
- PVP = probability that someone classified as exposed is truly exposed
= predictive value of an exposure “positive” = $\Pr(X = 1|X^* = 1)$
- PVN = probability that someone classified as unexposed is truly unexposed
= predictive value of an exposure “negative” = $\Pr(X = 0|X^* = 0)$
- B_1^* = number classified as exposed (with $X^* = 1$)
- B_0^* = number classified as unexposed (with $X^* = 0$)
- B_1 = expected number truly exposed (with $X = 1$)
- B_0 = expected number truly unexposed (with $X = 0$)

If they are known, the predictive values can be used directly to estimate the numbers truly exposed (B_1) and truly unexposed (B_0) from the misclassified counts B_1^* and B_0^* via the expected relations

$$\begin{aligned} B_1 &= PVP \cdot B_1^* + (1 - PVN)B_0^* \\ B_0 &= PVN \cdot B_0^* + (1 - PVP)B_1^* \end{aligned} \quad [19-4]$$

Note that the total M_0 is not changed by exposure misclassification:

$$\begin{aligned} M_0 &= B_1 + B_0 \\ &= PVN \cdot B_0^* + (1 - PVP)B_1^* + PVP \cdot B_1^* + (1 - PVN)B_0^* \\ &= (PVP + 1 - PVP)B_1^* + (PVN + 1 - PVN)B_0^* \\ &= B_1^* + B_0^* \end{aligned}$$

Thus, once we have estimated B_1 , we can estimate B_0 from $B_0 = M_0 - B_1$. From the preceding equations we can estimate the true exposure prevalence as $P_{e0} = B_1/M_0$. Parallel formulas for cases or person-time follow by substituting A_1 , A_0 , A_1^* , and A_0^* or T_1 , T_0 , T_1^* , and T_0^* for B_1 , B_0 , B_1^* , and B_0^* in equations 19-4. The adjusted counts obtained by applying the formula to actual data are only estimates derived under the assumption that the true predictive values are PVP and PVN and there is no other error in the observed counts (e.g., no random error). To make this clear, one should denote the solutions in equations 19-4 by \hat{B}_1 and \hat{B}_0 instead of B_1 and B_0 ; for notational simplicity, we have not done so.

Unfortunately, predictive values are seldom available, and when they are, their applicability is highly suspect, in part because they depend directly on exposure prevalence, which varies across populations (see formulas 19-9 and 19-10). For example, those study participants who agree to participate in a much more extensive validation substudy of food intake or medication usage (highly cooperative subjects) may have different patterns of intake and usage than other study participants. Owing to variations in exposure prevalence across populations and time, predictive values from a different study are even less likely to apply to a second study. Even when one can reliably estimate predictive values for a study, these estimates must be allowed to vary with disease and confounder levels, because exposure prevalence will vary across these levels.

These problems in applying predictive values lead to alternative adjustment methods, which use classification parameters that do not depend on true exposure prevalence. The following four probabilities are common examples of such parameters:

$$\begin{aligned}
 \text{Se} &= \text{probability that someone exposed is classified as exposed} \\
 &= \text{sensitivity} = \Pr(X^* = 1|X = 1) \\
 \text{Fn} &= \text{probability that someone exposed is classified as unexposed} \\
 &= \text{false-negative probability} = \Pr(X^* = 0|X = 1) = 1 - \text{Se} \\
 \text{Sp} &= \text{probability that someone unexposed is classified as unexposed} \\
 &= \text{specificity} = \Pr(X^* = 0|X = 0) \\
 \text{Fp} &= \text{probability that someone unexposed is classified as exposed} \\
 &= \text{false-positive probability} = \Pr(X^* = 1|X = 0) = 1 - \text{Sp}
 \end{aligned}$$

The following equations then relate the expected misclassified counts to the true counts:

$$\begin{aligned}
 B_1^* &= \text{expected number of subjects classified as exposed} \\
 &= \text{Se } B_1 + \text{Fp } B_0
 \end{aligned} \tag{19-5}$$

and

$$\begin{aligned}
 B_0^* &= \text{expected number of subjects classified as unexposed} \\
 &= \text{Fn } B_1 + \text{Sp } B_0
 \end{aligned} \tag{19-6}$$

Note that $\text{Se} + \text{Fn} = \text{Sp} + \text{Fp} = 1$, showing again that the total is unchanged by the exposure misclassification:

$$\begin{aligned}
 M_0 &= B_1 + B_0 = (\text{Se} + \text{Fn})B_1 + (\text{Sp} + \text{Fp})B_0 \\
 &= \text{Se } B_1 + \text{Fp } B_0 + \text{Fn } B_1 + \text{Sp } B_0 = B_1^* + B_0^*
 \end{aligned}$$

In most studies, one observes only the misclassified counts B_1^* and B_0^* . If we assume that the sensitivity and specificity are equal to Se and Sp (with Fn = 1 - Se and Fp = 1 - Sp), we can estimate B_1 and B_0 by solving equations 19-5 and 19-6. From equation 19-6, we get

$$B_0 = (B_0^* - \text{Fn } B_1)/\text{Sp}$$

We can substitute the right side of this equation for B_0 in equation 19-5, which yields

$$B_1^* = \text{Se } B_1 + \text{Fp}(B_0^* - \text{Fn } B_1)/\text{Sp}$$

We then solve for B_1 to get

$$\begin{aligned}
 B_1 &= (\text{Sp } B_1^* - \text{Fp } B_0^*)/(\text{Se Sp} - \text{Fn Fp}) \\
 &= (B_1^* - \text{Fp } M_0)/(\text{Se} + \text{Sp} - 1) \\
 B_0 &= M_0 - B_1 = (B_0^* - \text{Fn } M_0)/(\text{Se} + \text{Sp} - 1).
 \end{aligned} \tag{19-7}$$

From these equations we can also estimate the true exposure prevalence as $P_{e0} = B_1/M_0$. Again, the B_1 and B_0 obtained by applying equation 19-7 to actual data are only estimates derived under the assumption that the true sensitivity and specificity are Se and Sp.

Sensitivity analysis for exposure classification proceeds by applying equation 19-7 for various pairs of classification probabilities (Se, Sp) to the observed noncase counts B_1^* and B_0^* . To construct a corrected measure of association, we must also apply analogous equations to estimate A_1 and A_0 from the observed (misclassified) case counts A_1^* and A_0^* :

$$A_1 = (A_1^* - \text{Fp } M_1)/(\text{Se} + \text{Sp} - 1) \tag{19-8}$$

from which we get $A_0 = M_1 - A_1$, where M_1 is the observed case total. These formulas may be applied to case-control, closed-cohort, or prevalence-survey data. For person-time follow-up data, equation 19-7 can be modified by substituting T_1 , T_0 , T_1^* , and T_0^* for B_1 , B_0 , B_1^* , and B_0^* .

The formulas may be applied within strata of confounders as well. After application of the formulas, we may compute “adjusted” stratum-specific and summary effect estimates from the estimated true counts. Finally, we tabulate the adjusted estimates obtained by using different pairs (Se , Sp), and thus obtain a picture of how sensitive the results are to various degrees of misclassification.

Formulas 19–7 and 19–8 can yield negative adjusted counts, which are impossible values for the true counts. One way this can arise is if $Se + Sp < 1$, which implies that the classification is assigning values worse than randomly, in the following sense: Imagine that we conduct a coin toss with a probability p of heads to decide whether someone was exposed or not, setting $X^* = 1$ when the coin toss yielded heads (p may be any number between 0 and 1). The sensitivity and specificity of this completely random classification are then p and $1 - p$, respectively, and $Se + Sp = 1$. We will henceforth assume that our actual classification method is better than a coin toss, in the sense that $Se + Sp > 1$.

Even with this assumption, the solution B_1 to equation 19–7 will be negative if $Fp > B_1^*/M_0$, i.e., if the assumed false-positive probability exceeds the observed prevalence of exposure in the noncases, or, equivalently, if $Sp < B_0^*/M_0$. In parallel, B_0 will be negative if $Fn > B_0^*/M_0$ (equivalently, $Se < B_1^*/M_0$), A_1 will be negative if $Fp > A_1^*/M_1$, and A_0 will be negative if $Fn > A_0^*/M_1$. A negative solution indicates that either other errors (e.g., random errors) have distorted the observed counts, the value chosen for Se or for Sp is wrong, or some combination of these problems.

Although sensitivity and specificity do not depend on the true exposure prevalence, they are influenced by other characteristics. Because predictive values are functions of sensitivity and specificity (see formulas 19–9 and 19–10, later), they too will be affected by these characteristics, as well as by any characteristic that affects prevalence. For example, covariates that affect exposure recall (such as age and comorbidities) will alter the classification probabilities for self-reported exposure history and may vary considerably across populations. In such situations, sensitivity and specificity may not generalize well from one population to another (Begg, 1987). This lack of generalizability is one reason why varying classification probabilities in the formulas (sensitivity analysis) is crucial even when estimates are available from the literature.

Valid variances for adjusted estimates cannot be calculated from the adjusted counts using conventional formulas (such as those in Chapters 14–18), even if we assume that sensitivity and specificity are known or are unbiased estimates from a validation study. This problem arises because conventional formulas do not take account of the data transformations and random errors in the adjustments. Formulas that do so are available (Selén, 1986; Espeland and Hui, 1987; Greenland, 1988b, 2007c; Gustafson, 2003; Greenland and Gustafson, 2006). Probabilistic sensitivity analysis (discussed later) can also account for these technical issues, and for other sources of bias as well.

NONDIFFERENTIALITY

In the preceding description, we assumed nondifferential exposure misclassification, that is, the same values of Se and Sp applied to both the cases (equation 19–8) and the noncases (equation 19–7). To say that a classification method is nondifferential with respect to disease means that it has identical operating characteristics among cases and noncases, so that sensitivity and specificity do not vary with disease status. We expect this property to hold when the mechanisms that determine the classification are identical among cases and noncases. In particular, we expect nondifferentiality when the disease is unrelated to exposure measurement. This expectation is reasonable when the mechanisms that determine exposure classification precede the disease occurrence and are not affected by uncontrolled risk factors, as in many cohort studies, although even then it is not guaranteed to hold (Chapter 9). Thus, to say that there is nondifferential misclassification (such as when exposure data are collected from records that predate the outcome) means that neither disease nor uncontrolled risk factors result in different accuracy of response for cases compared to noncases.

Put more abstractly, nondifferentiality means that the classification X^* is independent of the outcome D (i.e., the outcome conveys no information about X) conditional on the true exposure X and adjustment variables. Although this condition may seldom be met exactly, it can be examined on the basis of qualitative mechanistic considerations. Intuition and judgment about the role of the outcome in exposure classification errors are the basis for priors about measurement behavior. Such judgments provide another reason to express such priors in terms of sensitivity and specificity, as we will do later, rather than predictive values.

As discussed in Chapter 9, differentiability should be expected when exposure assessment can be affected by the outcome. For example, in interview-based case-control studies, cases may be more likely to recall exposure (correctly or falsely) than controls, leading to higher sensitivity or lower specificity among cases relative to controls (*recall bias*). When differential misclassification is a reasonable possibility, we can extend the sensitivity analysis by using different sensitivities and specificities for cases and noncases. Letting Fp_1 , Fp_0 be the case and noncase false-positive probabilities, and Fn_1 , Fn_0 the case and noncase false-negative probabilities, the corrected odds ratio for a single 2-by-2 table simplifies to

$$\frac{(A_1^* - Fp_1 M_1)(B_0^* - Fn_0 M_0)}{(A_0^* - Fn_1 M_1)(B_1^* - Fp_0 M_0)}$$

This formula is sensible, however, only if all four parenthetical terms in the ratio are positive.

APPLICATION TO THE RESINS-LUNG CANCER EXAMPLE

As a numerical example, we adjust the resins-lung cancer data in Table 19–1 under the assumption that the case sensitivity and specificity are 0.9 and 0.8, and the control sensitivity and specificity are 0.8 and 0.8. This assumption means that exposure detection is somewhat better for cases. (Because this study is record-based with deaths from other diseases as controls, it seems unlikely that the actual study would have had such differential misclassification.) From equations 19–7 and 19–8, we obtain

$$B_1 = [257 - 0.2(1,202)]/[0.8 + 0.8 - 1] = 27.67$$

$$B_0 = 1,202 - 27.67 = 1,174.33$$

$$A_1 = [45 - 0.2(139)]/[0.8 + 0.9 - 1] = 24.57$$

$$A_0 = 139 - 24.57 = 114.43$$

These yield an adjusted odds ratio of $24.57(1,174.33)/114.43(27.67) = 9.1$. This value is much higher than the unadjusted odds ratio of 1.8, despite the fact that exposure detection is better for cases.

By repeating the preceding calculation, we obtain a resins-misclassification sensitivity analysis for the data in Table 19–1. Table 19–4 provides a summary of the results of this analysis. As can be seen, under the nondifferential misclassification scenarios along the descending diagonal, the adjusted odds-ratio estimates (2.34, 2.42, 10.7, 11.0) are always further from the null than

TABLE 19–4

Adjusted Resins-Lung Cancer Mortality Odds Ratios under Various Assumptions about the Resins Exposure Sensitivity (Se) and Specificity (Sp) among Cases and Controls

		Controls			
		Cases	Se: 0.90 Sp: 0.90	0.80 0.90	0.90 0.80
Se	Sp				
0.90	0.90		2.34 ^a	2.00	19.3
0.80	0.90		2.83	2.42 ^a	23.3
0.90	0.80		1.29	1.11	10.7 ^a
0.80	0.80		1.57	1.34	12.9
					11.0 ^a

^a Nondifferential misclassification.

the unadjusted estimate computed directly from the data (1.76, which corresponds to the adjusted estimate assuming $Se = Sp = 1$, no misclassification). This result reflects the fact that, if the exposure is dichotomous and the misclassification is better than random, nondifferential, and independent of all other errors (whether systematic or random), the bias produced by the exposure misclassification is toward the null. We caution, however, that this rule does not extend automatically to other situations, such as those involving a polytomous exposure (see Chapter 9).

In one form of recall bias, cases remember true exposure more than do controls, i.e., there is higher sensitivity among cases (Chapter 9). Table 19–4 shows that, even if we assume that this form of recall bias is present, adjustment may move the estimate away from the null; in fact, three adjusted estimates (2.00, 16.5, 9.1) are further from the null than the unadjusted estimate (1.76). These results show that the association can be considerably diminished by misclassification, even in the presence of recall bias. To understand this apparently counterintuitive phenomenon, one may think of the classification procedure as having two components: a nondifferential component shared by both cases and controls, and a differential component reflecting the recall bias. In many plausible scenarios, the bias toward the null produced by the nondifferential component overwhelms the bias away from the null produced by the differential component (Drews and Greenland, 1990).

Table 19–4 also shows that the specificity is a much more powerful determinant of the observed odds ratio than is the sensitivity in this example (e.g., with $Se = 0.8$ and $Sp = 0.9$, the adjusted estimate is 2.42, whereas with $Se = 0.9$ and $Sp = 0.8$, the adjusted estimate is 10.7), because the exposure prevalence is low. In general, when exposure prevalence is low, the odds-ratio estimate is more sensitive to false-positive error than to false-negative error, because false positives arise from a larger group and thus can easily overwhelm true positives.

Finally, the example shows that the uncertainty in results due to the uncertainty about the classification probabilities can be much greater than the uncertainty conveyed by conventional confidence intervals. The unadjusted 95% confidence interval in the example extends from 1.2 to 2.6, whereas the misclassification-adjusted odds ratios range above 10 if we allow specificities of 0.8, even if we assume that the misclassification is nondifferential, and to as low as 1.1 if we allow differential misclassification. Note that this range of uncertainty does not incorporate random error, which is the only source of error reflected in the conventional confidence interval.

RELATION OF PREDICTIVE VALUES TO SENSITIVITY AND SPECIFICITY

Arguments are often made that the sensitivity and specificity of an instrument will be roughly stable across similar populations, at least within levels of disease and covariates such as age, sex, and socioeconomic status. Nonetheless, as mentioned earlier, variations in sensitivity and specificity can occur under many conditions—for example, when the measure is an interview response and responses are interviewer-dependent (Begg, 1987). These variations in sensitivity and specificity will also produce variations in predictive values, which can be seen from formulas that relate the predictive values to sensitivity and specificity. To illustrate the relations, again consider exposure classification among noncases, where $M_0 = B_1 + B_0 = B_1^* + B_0^*$ is the noncase total, and let $P_{e0} = B_1/M_0$ be the true exposure prevalence among noncases. Then, in expectation, the predictive value positive among noncases is

$$\begin{aligned} PVP_0 &= (\text{number of correctly classified subjects in } B_1^*)/B_1^* \\ &= Se B_1/(Se B_1 + Fp B_0) \\ &= Se(B_1/M_0)/[Se(B_1/M_0) + Fp(B_0/M_0)] \\ &= SeP_{e0}/[SeP_{e0} + Fp(1 - P_{e0})] \end{aligned} \quad [19-9]$$

Similarly, in expectation, the predictive value negative among noncases is

$$PVN_0 = Sp(1 - P_{e0})/[FnP_{e0} + Sp(1 - P_{e0})] \quad [19-10]$$

Equations 19–9 and 19–10 show that predictive values are a function of the sensitivity, specificity and the unknown true exposure prevalence in the population to which they apply. When adjustments are based on internal validation data and those data are a random sample of the entire study, there

is no issue of generalization across populations. In such situations the predictive-value approach is simple and efficient (Marshall, 1990; Brenner and Gefeller, 1993). We again emphasize, however, that validation studies may be afflicted by selection bias, thus violating the randomness assumption used by this approach.

DISEASE MISCLASSIFICATION

Most formulas and concerns for exposure misclassification also apply to disease misclassification. For example, Equation 19–4 and 19–7 can be modified to adjust for disease misclassification. For disease misclassification in a closed-cohort study or a prevalence survey, PVP and PVN will refer to the predictive values for disease, and A , B , and N will replace B_1 , B_0 , and M_0 . For the adjustments using sensitivity and specificity, consider first the estimation of the incidence proportion from a closed cohort or of prevalence from a cross-sectional sample. The preceding formulas can then be adapted directly by redefining Se , Fn , Sp , and Fp to refer to disease. Let

$$D = 1 \text{ if diseased, } 0 \text{ if not}$$

$$D^* = 1 \text{ if classified as diseased, } 0 \text{ if not}$$

$$Se = \text{Probability someone diseased is classified as diseased}$$

$$= \text{Disease sensitivity} = \Pr(D^* = 1|D = 1)$$

$$Fn = \text{False-negative probability} = 1 - Se$$

$$Sp = \text{Probability someone nondiseased is classified as nondiseased}$$

$$= \text{disease specificity} = \Pr(D^* = 0|D = 0)$$

$$Fp = \text{False-positive probability} = 1 - Sp$$

Suppose that A and B are the true number of diseased and nondiseased subjects, and A^* and B^* are the numbers classified as diseased and nondiseased. Then equations 19–5 through 19–7 give the expected relations between A , B and A^* , B^* , with A , B replacing B_1 , B_0 ; A^* , B^* replacing B_1^* , B_0^* ; and $N = A + B = A^* + B^*$ replacing M_0 . With these changes, equation 19–7 becomes

$$A = (A^* - Fp N)/(Se + Sp - 1) \quad [19-11]$$

and $B = N - A$. These equations can be applied separately to different exposure groups and within strata, and “adjusted” summary estimates can then be computed from the adjusted counts. Results of repeated application of this process for different pairs of Se , Sp can be tabulated to provide a sensitivity analysis. Also, the pair Se , Sp can either be kept the same across exposure groups (nondifferential disease misclassification) or allowed to vary across groups (differential misclassification). As noted earlier, however, special variance formulas are required (Selén, 1986; Espeland and Hui, 1987; Greenland, 1988b; Greenland, 2007c; Gustafson, 2003).

The situation differs slightly for person-time follow-up data. Here, one must replace the specificity Sp and false-positive probability Fp with a different concept, that of the *false-positive rate*, Fr :

Fr = number of false-positive diagnoses (noncases diagnosed as cases) per unit person-time.

We then have

$$A^* = Se A + Fr T \quad [19-12]$$

where T is the true person-time at risk. Also, false-negatives (of which there are $Fn A$) will inflate the observed person-time T^* ; how much depends on how long the false-negatives are followed. Unless the disease is very common, however, the false negatives will add relatively little person-time and we can take T to be approximately T^* . Upon doing so, we need only solve equation 19–12 for A :

$$A = (A^* - Fr T^*)/Se \quad [19-13]$$

and get an adjusted rate A/T^* . Sensitivity analysis then proceeds (similarly to before) by applying equation 19–13 to the different exposure groups, computing adjusted summary measures, and repeating this process for various combinations of Se and Fr (which may vary across subcohorts).

The preceding analysis of follow-up data is simplistic, in that it does not account for possible effects if exposure lengthens or shortens the time from incidence to diagnosis. These effects have

generally not been correctly analyzed in the medical literature (Greenland, 1991a, Greenland, 1999a; see the discussion of standardization in Chapter 4). In these cases, one should treat time of disease onset as the outcome variable and adjust for errors in measuring this outcome using methods for continuous variables (Carroll et al., 2006).

Often studies make special efforts to verify case diagnoses, so that the number of false positives within the study will be negligible. If such verification is successful, we can assume that $Fp = 0$, $Sp = 1$, and equations 19–11 and 19–13 then simplify to $A = A^*/Se$. If we examine a risk ratio RR under these conditions, then, assuming nondifferential misclassification, the observed RR^* will be

$$RR^* = \frac{A_1^*/N_1}{A_0^*/N_0} = \frac{Se\ A_1/N_1}{Se\ A_0/N_0} = \frac{A_1/N_1}{A_0/N_0} = RR$$

In other words, with perfect specificity, nondifferential sensitivity of disease misclassification will not bias the risk ratio. Assuming that the misclassification negligibly alters person-time, the same will be true for the rate ratio (Poole, 1985) and will also be true for the odds ratio when the disease is uncommon. The preceding fact allows extension of the result to case-control studies in which cases are carefully screened to remove false positives (Brenner and Savitz, 1990).

Suppose now that the cases cannot be screened, so that in a case-control study there may be many false cases (false positives). It would be a severe mistake to apply the disease-misclassification adjustment equation 19–11 to case-control data if (as is almost always true) Se and Sp are determined from other than the study data themselves (Greenland and Kleinbaum, 1983), because the use of different sampling probabilities for cases and controls alters the sensitivity and specificity within the study relative to the source population. To see the problem, suppose that all apparent cases A_1^* , A_0^* but only a fraction f of apparent noncases B_1^* , B_0^* are randomly sampled from a closed cohort in which disease has been classified with sensitivity Se and specificity Sp . The expected numbers of apparent cases and controls selected at exposure level j is then

$$A_j^* = Se\ A_j + Fp\ B_j$$

and

$$f \cdot B_j^* = f(Fn\ A_j + Sp\ B_j)$$

The numbers of true cases and noncases at exposure level j in the case-control study are

$$Se\ A_j + f \cdot Fn\ A_j = (Se + f \cdot Fn)A_j$$

and

$$Fp\ B_j + f \cdot Sp\ B_j = (Fp + f \cdot Sp)B_j$$

whereas the numbers of correctly classified cases and noncases in the study are $Se\ A_j$ and $f \cdot Sp\ B_j$. The sensitivity and specificity in the study are thus

$$Se\ A_j / (Se + f \cdot Fn)A_j = Se / (Se + f \cdot Fn)$$

and

$$f \cdot Sp\ B_j / (Fp + f \cdot Sp)B_j = f \cdot Sp / (Fp + f \cdot Sp)$$

The study specificity can be far from the population specificity. For example, if $Se = Sp = 0.90$, all apparent cases are selected, and controls are 1% of the population at risk, the study specificity will be $0.01(0.90)/[0.1 + 0.01(0.90)] = 0.08$. Use of the population specificity 0.90 instead of the study specificity 0.08 in a sensitivity analysis could produce extremely distorted results.

CONFOUNDER MISCLASSIFICATION

The effects of dichotomous confounder misclassification lead to residual and possibly differential residual confounding (Greenland, 1980; Chapter 9). These effects can be explored using the methods discussed previously for dichotomous exposure misclassification (Savitz and Baron, 1989; Marshall and Hastrup, 1996; Marshall et al., 1999). One may apply equations 19–7 and 19–8 to the confounder within strata of the exposure (rather than to the exposure within strata of the confounder) and then

compute a summary exposure–disease association from the adjusted data. The utility of this approach is limited, however, because most confounder adjustments involve more than two strata. We discuss a more general (matrix) approach below.

MISCLASSIFICATION OF MULTIPLE VARIABLES

So far, our analyses have assumed that only one variable requires adjustment. In many situations, age and sex (which tend to have negligible error) are the only important confounders, the cases are carefully screened, and only exposure remains seriously misclassified. There are, however, many other situations in which not only exposure but also major confounders (such as smoking level) are misclassified. Disease misclassification may also coexist with these other problems, especially when studying disease subtypes.

In examining misclassification of multiple variables, it is commonly assumed that the classification errors for each variable are independent of *errors* in other variables. This assumption is different from that of nondifferentiality, which asserts that errors for each variable are independent of the *true values* of the other variables. Neither, either one, or both assumptions may hold, and both have different implications for bias. As mentioned in Chapter 9, the generalization that “non-differential misclassification of exposure always produces bias toward the null” is false if the errors are dependent or if exposure has multiple levels.

If all the classification errors are independent across variables, we can apply the adjustment formulas in sequence for each misclassified variable, one at a time. For example, in a prevalence survey we may first obtain counts adjusted for exposure misclassification from equations 19–7 and 19–8, then further adjust these counts for disease misclassification using equation 19–11. If, however, the classification errors are dependent across variables, we must turn to more complex adjustment methods such as those based on matrix adjustment of counts (Greenland and Kleinbaum, 1983; Selén, 1986; Espeland and Hui, 1987; Greenland, 1988b). Dependent errors most easily arise when a common method, such as an interview or medical record review, is used to ascertain more than one variable involved in the analysis (Lash and Fink, 2003b). Matrix methods are also useful for adjustments of polytomous (multilevel) variables.

A MATRIX ADJUSTMENT METHOD

We now briefly outline one simple matrix approach that is the natural generalization of the formulas given earlier (Barron, 1977; Greenland and Kleinbaum, 1983; Kleinbaum et al., 1984; Greenland, 1988b), and can be applied under any misclassification setting, including dependent and differential misclassification of polytomous variables. Imagine that we have a multiway table of data classified by disease, one or more exposures, and one or more stratification variables (all of which may have multiple levels). Suppose that this table has K cells. We can list these cells in any order and index them by a subscript $k = 1, \dots, K$. Suppose that C_k^* subjects are classified into cell k , whereas C_k subjects truly belong in that cell. Next, define

p_{mk} = probability of being classified into cell m when the true cell is k .

Then, in expectation,

$$C_m^* = \sum_{k=1}^K p_{mk} C_k \quad [19-14]$$

This equation is a generalization of equations 19–5 and 19–6.

If we write C^* and C for the vectors of C_m^* and C_k , and P for the matrix of p_{mk} , then equation 19–14 reduces to $C^* = PC$. The adjusted counts, assuming that P contains the true classification probabilities, can then be found from the observed (misclassified) counts C^* by inverting P , to get $C = P^{-1}C^*$. The practical utility of this formula in a given application will of course depend on the information available to specify plausible values for the p_{mk} .

USE OF VALIDATION DATA

In the previous sensitivity formulas we assumed that Se and Sp were educated guesses, perhaps suggested by or estimated from external literature. Suppose instead that classification probabilities can be estimated directly from an internal validation subsample of the study subjects, and the latter subsample is itself free of bias. In particular, suppose that the sampling is random within levels of exposure, disease, and any adjustment covariates. A number of statistically efficient ways of using these data are then available, beginning with the predictive-value approaches described earlier, but also including two-stage and missing-data analysis methods. From such methods, correctly classified counts may be estimated using maximum-likelihood or related techniques, and full statistics (including confidence intervals) can be obtained for the resulting effect estimates (e.g., Tennenbein, 1970; Selén, 1986; Espeland and Hui, 1987; Greenland, 1988b; Marshall, 1990; Lyles, 2002; Greenland, 2007c). Regression methods may also be used to adjust for errors in continuous measurements (e.g., Rosner et al., 1989; Speigelman et al., 2000, 2001, 2005; Carroll et al., 2006; Freedman et al., 2004). Robins et al. (1994) and Carroll et al. (2006) review and compare a number of methods and their relation to missing-data techniques, and discuss general methods for continuous as well as discrete variables.

A general adjustment formula for the predictive-value approach follows if we have estimates of

$$q_{km} = \text{probability of truly belonging to cell } k \text{ when classified into cell } m$$

We can then obtain estimates of the true counts from the formula

$$C_k = \sum_{m=1}^M q_{km} C_m^*, \quad [19-15]$$

which is a generalization of equation 19-4. In the matrix algebra formulation, we can write this equation as $C = QC^*$, where Q is the matrix of q_{km} . Adjustment methods more efficient and more general than the preceding approach can be obtained using likelihood-based techniques; see Carroll et al. (2006) for a model-based treatment.

An important caution in interpreting the results from formal adjustment methods is that most methods assume the validation standard is measured without error. If, however, the validation measurement taken as the truth is itself subject to error, then the adjusted estimate will also be biased, possibly severely so (Wacholder et al., 1993). Many studies labeled as validation studies actually substitute one imperfect measure of a variable for another. They are thus measures of agreement, not validity. For binary exposures, the bias in the adjusted estimate can be kept below that of the unadjusted estimate if the validation measurement has higher sensitivity and specificity than the regular measurement, and the adjustment method does not assume nondifferentiality (Brenner, 1996). More generally, by introducing assumptions (models) for the joint relation of the available measurements to each other and to the true values, as well as assumptions about randomness of validation sampling, further adjustments can be made to account for errors in the validation measurements (Speigelman et al., 1997ab, 2000, 2001, 2005). These assumptions may then be subjected to sensitivity analyses of the impact of assumption violations.

We have underscored the random-sampling assumptions in validation methods because of the high potential for selection bias in many validation studies. Validation studies often require an additional respondent burden (e.g., to complete a diet diary in addition to a food frequency questionnaire), leading to questions about the representativeness of those who volunteer to participate. Even when no burden is imposed, selection bias may occur. For example, studies that validate self-reported history by record review sometimes treat those who refuse permission to review their records as if they were no different from those who permit review. Nonetheless, one reason to refuse permission may be to avoid detection of an inaccurate self-report. If this type of refusal is common enough, extrapolation of the results from those who permit review to the larger study group will exaggerate the validity of self-reported history. Sensitivity analysis of selection bias (discussed next) can be used to assess such problems.

In summary, when the validation study is itself subject to systematic error and bias, sensitivity analysis remains an important adjunct to more formal adjustment methods.