Sensitivity Analysis and Bias Analysis

19

Sander Greenland

Contents

S. Greenland
Department of Epidemiology, School of Public Health, University of California, Los Angeles, CA, USA

Department of Statistics, UCLA College of Letters and Science, Los Angeles, CA, USA

W. Ahrens, I. Pigeot (eds.) *Handbook of Epidemiology, 2nd edition*,
DOI 10.1007/978-0-387-09834-0_60,

19.1 Introduction

Over recent decades recognition has grown that the conventional statistical models used to analyze epidemiological data cannot be reasonably claimed to be correct in the way most textbooks treat them to be. In particular, conventional models for epidemiological data-generating processes cannot be credibly taken to represent targets of primary scientific interest. For example, a logistic model for the regression of an observed disease indicator on covariate measurements would only rarely correspond closely to the causal effects on disease of the risk factors represented by the measurements. The discrepancies between the statistical model parameters and the underlying target effects are often called systematic errors, biases, or bias sources. Large biases undermine the interpretation of both frequentist statistics (such as confidence intervals) and Bayesian statistics (such as posterior intervals).

The present chapter provides an overview of one useful family of approaches to analyzing and adjusting for bias that cannot be controlled by familiar stratification or regression adjustments on measured variables. It provides basic descriptions of familiar bias problems in the case of binary variables in 2×2 tables, with some simple illustrations. It then provides discussions of the bias problem in terms of statistical theory, compares ordinary sensitivity analysis to probabilistic bias analysis, and concludes by considering when these analyses might be helpful or essential.

An introduction to basic Bayesian concepts (e.g., Greenland 2008 and chapter ▸Bayesian Methods in Epidemiology of this handbook) is helpful in understanding concepts of bias analysis before reading the present chapter. More extensive coverage and examples of basics including computational details can be found in Fox et al. (2005), Greenland and Lash (2008), Greenland (2009a), and Lash et al. (2009), with software implementations in Steenland and Greenland (2004, appendices), Fox et al. (2005), Orsini et al. (2008), and Lash et al. (2009), while general theories can be found in Vansteelandt et al. (2006) and Greenland (2005, 2009b), among others. Throughout, variables will be denoted by capital letters, while unspecified values for a variable will be denoted by lowercase letters; for example, X may represent an exposure variable in which case x will represent a possible but unspecified value for X.

19.2 Fundamentals of Methodological Biases in 2 × 2 Table Analysis

Table 19.1 displays data on mother's report of antibiotic use during pregnancy (coded $X = 1$ yes, 0 no) and subsequent sudden infant death syndrome (SIDS, coded $Y = 1$ yes, 0 no) from a multicenter case-control study (Kraus et al. 1989); a plus sign ("+") in a subscript indicates summation over that subscript. Given the rarity of SIDS, the population risk ratio (relative risk) comparing newborns exposed to antibiotics in utero to those unexposed is approximated by the corresponding

Table 19.1 Data from case-control study of SIDS (Kraus et al. 1989). X indicates maternal recall of antibiotic use during pregnancy, and Y indicates SIDS ($Y = 1$ for cases, $Y = 0$ for controls). A plus sign ("+") in a subscript indicates summation over that subscript

	$X = 1$	$X = 0$	Totals
$Y = 1$	$A_{+11} = 173$	$A_{+01} = 602$	$A_{++1} = 775$
$Y = 0$	$A_{+10} = 134$	$A_{+00} = 663$	$A_{++0} = 797$

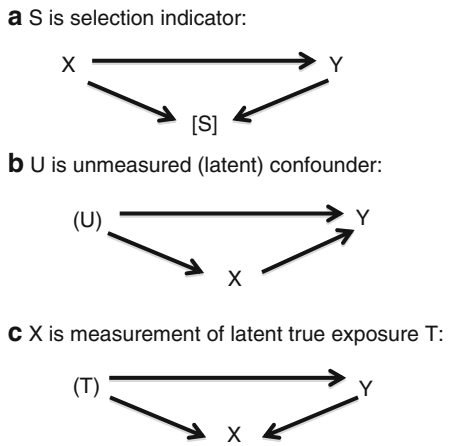

Fig. 19.1 Causal diagrams for three basic bias structures; *parentheses* indicate variable is unmeasured (unobserved), and *brackets* indicate observations are conditional on the variable

population odds ratio. If there were no bias, we could take this odds ratio OR_{XY} or its log, $\beta_{XY} = \ln(OR_{XY})$, as the target parameter.

The usual maximum-likelihood estimate (MLE) of $OR_{XY} = \exp(\beta_{XY})$ is the sample odds ratio $\widehat{OR}_{XY} = 173(663)/134(602) = 1.422$, with standard error for $\widehat{\beta}_{XY} = \ln(\widehat{OR}_{XY})$ of $(1/173+1/602+1/134+1/663)^{1/2} = 0.128$ and 95% confidence limits (CL) for $OR_{XY} = \exp(\beta_{XY})$ of $\exp\{\ln(1.422) \pm 1.96 \cdot 0.128\} = 1.11, 1.83$. Were there no bias, such results might be interpreted as providing an inference that OR_{XY} is above 1 but below 2. Nonetheless, experienced epidemiologists know better than to make this interpretation without cautions about common sources of bias: selection bias, uncontrolled confounding, and misclassification. As is traditional, the term "selection bias" will refer to non-random (differential or systematic) non-response and loss as well as non-random selection into the analysis.

Figure 19.1 shows the simplest representations of these three bias sources using causal directed acyclic graphs (causal diagrams), in which the arrows represent direct effects of one variable on another. With S denoting the selection indicator, the selection-bias problem (Fig. 19.1a) can be described by saying that we see the relation of X to Y conditional on $S = 1$ but we want to see that relation unconditionally (without regard to S). The confounding problem (Fig. 19.1b) can be described by saying that we see the unconditional relation of X to Y but we want to see that relation conditional on U. The misclassification problem (Fig. 19.1c) can be described by saying that we see the relation of X to Y but

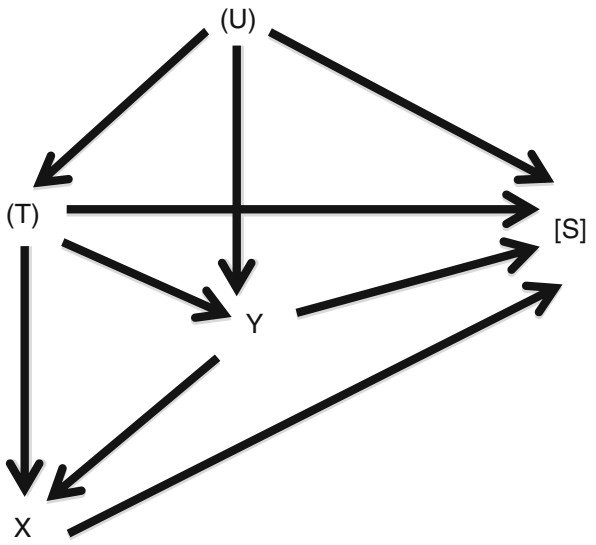

Fig. 19.2 Causal diagram combining selection, confounding, and measurement

Table 19.2 Imputed complete-population data table from SIDS study when S indicates selection into study, N_{xy} = population total with $X = x, Y = y$, and $\pi_{sxy} = \Pr(S = s|X = x, Y = y)$. Column totals form the target table

	$X = 1$		$X = 0$	
	$Y = 1$	$Y = 0$	$Y = 1$	$Y = 0$
$S = 1$	$173 = N_{11}\pi_{111}$	$134 = N_{10}\pi_{110}$	$602 = N_{01}\pi_{101}$	$663 = N_{00}\pi_{100}$
$S = 0$	$N_{11}\pi_{011}$	$N_{10}\pi_{010}$	$N_{01}\pi_{001}$	$N_{00}\pi_{000}$
Totals	$N_{11} = 173/\pi_{111}$	$N_{10} = 134/\pi_{110}$	$N_{01} = 602/\pi_{101}$	$N_{00} = 663/\pi_{100}$

we want the relation of T to Y unconditionally (without regard to X). More realistic situations involving combinations of these biases can also be represented, as in Fig. 19.2, which combines all three problems. Such diagrams are recommend as a preliminary to algebraic and statistical analyses, for they can reveal biases and assumption violations that might otherwise go undetected (Glymour and Greenland 2008; Pearl 2009; see also chapter ▸Directed Acyclic Graphs of this handbook).

Tables 19.2–19.4 display the corresponding data representations of these problems when the variables are binary, using count tables to show the target table (i.e., the unobserved structure of the target population). From these tables we can draw parallels and highlight differences among the three bias situations in Fig. 19.1:

(a) For selection bias (Fig. 19.1a and Table 19.2), we are interested in the unconditional relation of X to Y unconditional on selection S. But because we see only the relation of X to Y conditional on selection ($S = 1$), we must impute the unconditional relation of X to Y using probabilities of selection given what we do see (X and Y given $S = 1$).

(b) For confounding (Fig. 19.1b and Table 19.3), we are interested in the relation of X to Y conditional on U. But because we do not see U, we must impute its values using probabilities (bets) about the values of U given what we do see (again, X and Y).

(c) For misclassification (Fig. 19.1c and Table 19.4), we are interested in the unconditional relation of the true exposure T to the outcome Y so that if we saw T, we could discard X and just look at the TY association. But because we do not see T, we must impute its values using probabilities (bets) about the values of T given what we do see (again, X and Y).

In all three cases, we need a set of four probabilities to complete the three-way table of the three variables in the graph. These probabilities are denoted by π_{1xy} in Tables 19.2–19.4, where they correspond to probabilities of $S = 1$, $U = 1$, or $T = 1$ given the observed X and Y. We need only these four probabilities because the corresponding probabilities of $S = 0$, $U = 0$, or $T = 0$ are $\pi_{0xy} = 1 - \pi_{1xy}$. Using these probabilities, Table 19.2 displays the complete distribution of those unselected as well as those selected, while Tables 19.3 and 19.4 display the complete distribution of the observed subjects under confounding and misclassification.

Were these complete distributions known, we could compute our target parameter directly from them. But they are not known with certainty, and the best we can do

Table 19.3 Imputed complete-data table from SIDS study when U is an unmeasured confounder and $\pi_{uxy} = \Pr(U = u|X = x, Y = y)$. This table is the target table

	$X=1$		$X=0$	
	$Y=1$	$Y=0$	$Y=1$	$Y=0$
$U=1$	$A_{111} = 173\pi_{111}$	$A_{110} = 134\pi_{110}$	$A_{101} = 602\pi_{101}$	$A_{100} = 663\pi_{100}$
$U=0$	$A_{011} = 173\pi_{011}$ $= 173 - A_{111}$	$A_{010} = 134\pi_{010}$ $= 134 - A_{110}$	$A_{001} = 602\pi_{001}$ $= 602 - A_{101}$	$A_{000} = 663\pi_{000}$ $= 663 - A_{100}$
Totals	173	134	602	663

Table 19.4 Imputed complete-data and target tables from SIDS study when T indicates actual antibiotic use during pregnancy and $\pi_{txy} = \Pr(T = t|X = x, Y = y)$

(a) Complete data

	$X=1$		$X=0$	
	$Y=1$	$Y=0$	$Y=1$	$Y=0$
$T=1$	$A_{111} = 173\pi_{111}$	$A_{110} = 134\pi_{110}$	$A_{101} = 602\pi_{101}$	$A_{100} = 663\pi_{100}$
$T=0$	$A_{011} = 173\pi_{011}$ $= 173 - A_{111}$	$A_{010} = 134\pi_{010}$ $= 134 - A_{110}$	$A_{001} = 602\pi_{001}$ $= 602 - A_{101}$	$A_{000} = 663\pi_{000}$ $= 663 - A_{100}$
Totals	173	134	602	663

(b) Target table (complete data collapsed over X):

	$Y=1$	$Y=0$
$T=1$	$A_{1+1} = 173\pi_{111} + 602\pi_{101}$	$A_{1+0} = 134\pi_{110} + 663\pi_{100}$
$T=0$	$A_{0+1} = 173\pi_{011} + 602\pi_{001}$ $= 775 - A_{1+1}$	$A_{0+0} = 134\pi_{010} + 663\pi_{000}$ $= 797 - A_{1+0}$
Totals	$A_{++1} = 173 + 602 = 775$	$A_{++0} = 134 + 663 = 797$

is try to adjust our inferences using any relevant background information, such as information about the π_{1xy}, their combinations, or their components. This process is sometimes called *external* or *indirect* adjustment and is a core component of bias analysis.

19.2.1 Formulas for Selection Bias

Simplifications may arise depending on the available information, assumptions, and the target parameter. For example, in the selection-bias problem, the usual case-control target would be the population XY odds ratio OR_{XY} (the odds ratio unconditional on S), but the observed sample odds ratio, $\widehat{OR}_{XY|S=1}$, only estimates the odds ratio among those selected, $OR_{XY|S=1}$. The selection bias in $\widehat{OR}_{XY|S=1}$ relative to the population OR_{XY} is thus $\text{Bias}_{S=1} = \pi_{111}\pi_{100}/\pi_{110}\pi_{101} = OR_{XY|S=1}/OR_{XY}$. This bias equation suggests the following approximate estimating equation to adjust $\widehat{OR}_{XY|S=1}$ for selection bias:

$$\begin{aligned} OR_{XY} = N_{11}N_{00}/N_{10}N_{01} &\approx [(173/\pi_{111})(663/\pi_{100})]/[(134/\pi_{110})(602/\pi_{101})] \\ &= (173 \cdot 663)/(134 \cdot 602)/(\pi_{111}\pi_{100}/\pi_{110}\pi_{101}) \\ &= \widehat{OR}_{XY|S=1}/\text{Bias}_{S=1}. \end{aligned}$$

The approximation of chief statistical concern arises from substituting sample numbers for population numbers.

The equation shows that we need information on only one number, $\text{Bias}_{S=1}$, rather than all four selection probabilities π_{1xy}. $\text{Bias}_{S=1}$ might be estimated as the ratio of estimates without and with information from initial non-responders on whom data was later obtained ("call-back survey" information) from studies reporting such information (e.g., Hatch et al. 2000). Even with such information, however, concerns about generalization across studies must be addressed by considering how differences between the study populations may have affected response and selection.

19.2.2 Formulas for Confounding

Suppose there is no selection bias (so that $OR_{XY|S=1} = OR_{XY}$). Consider first the odds ratio $OR_{XY|U=1}$ in the $U = 1$ stratum (which is free of confounding by U). The bias in $\widehat{OR}_{XY}$ relative to $OR_{XY|U=1}$ is $\text{Bias}_{U=1} = \pi_{110}\pi_{101}/\pi_{111}\pi_{100} = OR_{XY}/OR_{XY|U=1}$. Table 19.3 then suggests the following approximate estimating equation to adjust $\widehat{OR}_{XY}$ for confounding by U:

$$\begin{aligned} OR_{XY|U=1} &\approx [A_{111}A_{100}]/[A_{110}A_{101}] \\ &= [(173\pi_{111})(663\pi_{100})]/[(134\pi_{110})(602\pi_{101})] \\ &= (173 \cdot 663)/(134 \cdot 602)/(\pi_{110}\pi_{101}/\pi_{111}\pi_{100}) \\ &= \widehat{OR}_{XY}/\text{Bias}_{U=1} \end{aligned}$$

A parallel formula can be written for $OR_{XY|U=0}$. To simplify analysis, however, it is usually assumed that the odds ratio is a constant $OR_{XY|U}$ across U ($OR_{XY|U=0} = OR_{XY|U=1} = OR_{XY|U}$, called the odds-ratio *homogeneity* assumption). In that case, $\text{Bias}_{U=1}$ is also the bias in $\widehat{OR}_{XY}$ relative to $OR_{XY|U=0}$, $\text{Bias}_{U=0} = \pi_{010}\pi_{001}/\pi_{011}\pi_{000} = OR_{XY}/OR_{XY|U=0}$, and both can be written as $\text{Bias}_U = OR_{XY}/OR_{XY|U} \approx \widehat{OR}_{XY}/OR_{XY|U}$. Thus, to adjust $\widehat{OR}_{XY}$ toward the constant U-specific odds ratio $OR_{XY|U}$, we need only information on one number, Bias_U, rather than all four confounder-prevalence probabilities π_{1xy}.

Suppose now that the target parameter what the effect of exposure would be on the entire population (often called the total-population or *marginal* effect of exposure). Evidence suggests that estimation of $OR_{XY|U}$ is not sensitive to heterogeneity (violations of homogeneity) in typical epidemiological settings in which the disease is uncommon and neither the XY effect nor the heterogeneity is large (Greenland and Maldonado 1994). Nonetheless, the assumption should be considered critically in any application, especially if the target parameter is the effect in a subpopulation (such as the exposed). Given the assumption, we still must estimate Bias_U. The relation $\text{Bias}_U = OR_{XY}/OR_{XY|U}$ suggests using the ratio of estimates without and with U adjustment from studies reporting such estimates; again however concerns about generalization across studies must be addressed.

The above formulation was derived for only one binary confounder U and an unadjusted $\widehat{OR}_{XY}$. Nonetheless, it extends directly to multiple confounders of any form and to situations involving a partially adjusted initial estimate (e.g., an initial estimate adjusted for age and sex but not smoking habits). This can be done by allowing $\widehat{OR}_{XY}$ to represent the partially adjusted estimate and Bias_U to represent the ratio $OR_{XY|partial}/OR_{XY|full}$ between a partially adjusted measure $OR_{XY|partial}$ and a fully adjusted measure $OR_{XY|full}$; this ratio can be estimated from reports or data providing estimates with both degrees of adjustment.

19.2.3 Sensitivity Analysis and Probabilistic Bias Analysis

The process of seeing how results change under different choices for a bias parameter (such as Bias_S or Bias_U) is often called a *sensitivity analysis*. When the focus is on the degree of bias implied by each parameter choice, the process is sometimes called *bias analysis*. The term "sensitivity analysis" is also used more broadly to include examination of the impact of any change in the statistical model, such as varying the model for the regression or random error; such model-sensitivity analysis is a large topic in econometrics and engineering (Leamer 1978, 1985; Saltelli et al. 2000) but is not covered here.

For selection bias or confounding, a sensitivity analysis might comprise nothing more than noting that, based on other data, the bias factor seems likely to fall in range, e.g., 0.80–1.33, and then displaying the adjusted point estimates for bias factors in this range. But it is not clear how to gauge or combine the uncertainty this range represents with the conventional statistics. One naïve approach would

expand the confidence limits of 1.11, 1.83 in the SIDS example using the extremes of the bias range, to produce limits of 1.11/1.33 = 0.83 and 1.83/0.80 = 2.29. As shown below, this approach tends to overstate the uncertainty because it combines only the extremes of bias and random error, ignoring that in the majority of possible combinations at least one of bias and random error is small or that in half the cases the two would be in opposite directions. For example, random error is just as likely to partially counterbalance bias as add to it.

These preceding observations suggest that more modest combinations of bias and random error are more probable than extremes, an intuition that is captured mathematically in *probabilistic bias analysis*, and which is also treated under related topics such as uncertainty assessment and risk analysis (Eddy et al. 1992; Lash and Fink 2003; Phillips 2003; Steenland and Greenland 2004; Greenland 2001, 2003, 2005; Greenland and Lash 2008; Lash et al. 2009; Turner et al. 2009). As a simple example, one might specify a *prior probability* that the bias factor Bias_U for confounding by U falls in a stated range like 0.80, 1.33, typically 95%. The term "prior" signals that the assigned probability may be nothing more than an educated bet based on available relevant literature, although it could be based on detailed analysis of other studies. Regardless, one can use standard lognormal formulas to combine the bias uncertainty seen in this interval and the uncertainty from random error represented by the standard error or confidence interval.

Specifically, suppose the range 0.80, 1.33 were taken to represent the 2.5th and 97.5th percentiles of a lognormal prior-probability distribution ("prior") for Bias_U. Uncertainty about $\ln(\text{Bias}_U)$ would then follow a normal distribution, with mean m_U the average of the log limits: $m_U = [\ln(1.33) + \ln(0.80)]/2 = 0.0310$. We can subtract this mean bias m_U from the log odds-ratio estimate $\hat{\beta}_{XY} = \ln(\widehat{OR}_{XY})$ to get a bias-adjusted estimate of

$$\hat{\beta}_{XY} - m_U = \ln(1.422) - 0.0310 = 0.320.$$

The prior standard deviation would be the width of the log-bias interval $\ln(1.33) - \ln(0.80) = 0.508$ divided by the number of standard deviations in its width, which is $2 \cdot 1.96 = 3.92$ for a normal 95% interval. We then square this ratio to get the prior log-bias variance: $v_U = (0.508/3.92)^2 = 0.0168$. This bias variance is then added to the estimated random variance $v_R = 0.128^2 = 0.0164$ of $\ln(\widehat{OR}_{XY})$ to get a total variance $v_+ = v_R + v_U = 0.0164 + 0.0168 = 0.0332$. Finally, we compute an approximate bias-adjusted 95% *posterior probability* interval

$$\exp\{0.320 \mp 1.96(0.0332)^{1/2}\} = [0.96, 1.97].$$

This interval is more narrow than in the extreme-adjusted interval of [0.83, 2.29] but is still noticeably wider than the original confidence interval [1.11, 1.83] (as well as being shifted downward by a small amount).

To adjust for selection bias in addition to confounding, we could repeat the above process by specifying a selection-bias interval, converting that to a prior mean m_S

and variance v_S for $\ln(\text{Bias}_{S=1})$, then computing the posterior interval from the adjusted estimate $\hat{\beta}_{XY} - m_S - m_U$ and the total variance $v_R + v_S + v_U$. This extension assumes that the confounding and selection bias are independent, which would not be the case in the situation depicted in Fig. 19.2, where everything including U directly affects selection S.

The above log-linear approach to selection bias and confounding appears simple in concept but raises the question of how one can justify assigning a particular prior distribution to the bias factors. One way is to base the interval estimate for $\text{Bias}_U = OR_{XY}/OR_{XY|U}$ or $\text{Bias}_{S=1} = OR_{XY}/OR_{XY|S=1}$ seen upon adjustment in a past study, then use that interval for the current adjustment (Greenland and Mickey 1988; Greenland 2003). More often however such direct bias information is not readily available or is not considered generalizable from its source.

19.2.4 Confounding Parameters

The confounding bias factor Bias_U can be expressed as a function of the confounder distribution along with measures of the association of the confounder with exposure and disease (Kitagawa 1955; Cornfield et al. 1959; Bross 1966, 1967; Schlesselman 1978; Breslow and Day 1980, Sect. 3.4; Yanagawa 1984; Gail et al. 1988; Flanders and Khoury 1990; Schneeweiss 2006; VanderWeele and Arah 2011). For example, under homogeneity the odds-ratio bias Bias_U can be written

$$\text{Bias}_U = (1 + \theta \cdot OR_{UX|Y} \cdot OR_{UY|X})(1 + \theta)/(1 + \theta \cdot OR_{UX|Y})(1 + \theta \cdot OR_{UY|X})$$

where $\theta = \pi_{100}/\pi_{000} = \pi_{100}/(1 - \pi_{100})$ is the odds of $U = 1$ when $X = Y = 0$, $OR_{UX|Y}$ is the constant UX odds ratio given Y, and $OR_{UY|X}$ is the constant UY odds ratio given X (Yanagawa 1984). It can happen that there is some generalizable background information to place priors on these three bias parameters even though there is no direct information about Bias_U. For example, if the exposure ($X = 1$) and disease ($Y = 1$) are uncommon in the general population, then the prevalence π_{100} can be estimated from population survey data on U. If the exposure is rare or its effect on Y is "weak" ($0.5 < OR_{XY|U} < 2$), then we may expect $OR_{UY|X}$ to be similar to OR_{UY}, and OR_{UY} may be available from earlier studies.

The preceding approach assumes that U is a known confounder (e.g., a smoking indicator) that was unmeasured in the study in question but has been previously identified and subject to study in relation to disease if not exposure. If instead U represents an unspecified, unknown confounder, then the entire sensitivity exercise will remain far more speculative. Nonetheless, decomposition of the bias factor can still be successful in demonstrating that only implausibly strong confounder or selection effects can account for a strong observed association. Cornfield et al. (1959) is considered a landmark study in which such an approach was used to examine claims that the smoking-lung cancer relation might be attributable to confounding.

One approach to expanding intervals to account for uncertainties about multiple bias parameters is *Monte Carlo sensitivity analysis* (MCSA, also known as Monte Carlo risk analysis) which is based on taking random draws for each component parameter in the bias adjustment. There are many ways to proceed. One simple approach assigns prior distributions to $\text{Bias}_{S=1}$, θ, $OR_{UX|Y}$, and $OR_{UY|X}$ (which may be dependent) and a normal sampling distribution to the unadjusted log odds ratio $\hat{\beta}_{XY}$ with mean equal to the point estimate ln(1.422) and standard deviation equal to $v_R^{1/2} = 0.128$. We then cyclically generate thousands of draws from each of these distributions. Let $\text{Bias}^*_{S=1}, \theta^*, OR^*_{UX|Y}, OR^*_{UY|X}, \hat{\beta}^*_{XY}$ be the values drawn in one such cycle. At each draw we compute a Bias^*_U from θ^*, $OR^*_{UX|Y}$, and $OR^*_{UY|X}$ and then compute $\hat{\beta}^*_{adj} = \hat{\beta}^*_{XY} - \ln(\text{Bias}^*_{S=1}) - \ln(\text{Bias}^*_U)$. We may then create a histogram of the adjusted estimates from this simulation and take the 2.5th and 97.5th percentiles of the distribution of $\hat{\beta}^*_{adj}$ so generated as a simulated 95% posterior interval for the *XY* log odds ratio β_{XY}.

An alternative "bootstrap" approach to incorporating random error is to resample (with replacement) the entire data set at each cycle and recompute $\hat{\beta}^*_{XY}$ from the resampled data. This can work well with large cell counts but may slow computation considerably and may require complex corrections in small or sparse data sets in order to account for zero cells and other problems (Efron and Tibshirani 1994).

The posterior-interval interpretations given above are approximate and assume implicitly that there is no prior information on any parameter other than the bias parameters $\text{Bias}_{S=1}$, θ, $OR_{UX|Y}$, and $OR_{UY|X}$. This assumption is not realistic in practice because (apart from genes) few studies are the first to examine a given association. As a consequence, some authors call the resulting interval a "simulation interval" without specifying a statistical meaning for the interval (e.g., Lash et al. 2009). Nonetheless, a much more severe criticism applies to the conventional 95% confidence interval: It is routinely interpreted as a posterior interval for the target parameter $OR_{XY|U}$, yet this interpretation is based on the same assumption of no prior information about this target, along with the extremely unrealistic assumption that there is no bias at all (which would be represented by priors for $\text{Bias}_{S=1}$ and Bias_U that are point masses at 1).

If one wishes to employ informative priors for parameters other than the bias parameters, it will be necessary to switch to a penalized-likelihood or explicitly Bayesian analysis format in order to correctly interpret the resulting intervals as posterior probability intervals (Joseph et al. 1995; Graham 2000; Gustafson et al. 2001, 2010; Scharfstein et al. 2003; Steenland and Greenland 2004; Gustafson 2003, 2005; Chu et al. 2006; McCandless et al. 2007, 2012; Greenland 2005, 2009a, b; Molitor et al. 2009; Turner et al. 2009; Welton et al. 2009). The traditional alternative, ordinary sensitivity analysis, avoids this complication by simply displaying results for various choices of bias parameters (Copas and Li 1997; Rosenbaum 2002; Brumback et al. 2004), but can become unwieldy if the number of bias parameters is more than three (as will happen when misclassification is addressed).

For those seeking to avoid explicit prior distributions, the dimensionality problem may be addressed by estimating regions of compatibility between the data and the parameters of interest given the random-error model and sharp bounds for the bias parameters (Vansteelandt et al. 2006). These regions generalize confidence intervals to account for the bias parameters. Nonetheless, sharp bounds for the bias parameters are needed to construct the regions, and it is not clear why they provide a less arbitrary solution to the problem than do prior distributions, since the results are just as sensitive to choice of bounds as to choice of priors (Greenland 2009b).

19.3 Bias Formulas for Misclassification

Misclassification adjustment is usually much more difficult than others, in part because the target parameter is the association (such as the odds ratio OR_{TY}) between the completely unobserved true exposure indicator T and Y, not X and Y. Its complexity is sometimes overlooked, in part because it appears that the problem can be solved by simply dividing $\widehat{OR}_{XY}$ by the bias factor $\text{Bias}_M = OR_{XY}/OR_{TY}$ to adjust for the misclassification.

Unfortunately, unlike with selection bias and confounding, the bias-factor approach is not widely useful for misclassification adjustment. First, although the bias factor could be estimated as the ratio of estimates of OR_{XY} and OR_{TY} from studies reporting both estimates, such reports are very uncommon. Second, even when an estimate of Bias_M is available, there are exceptional concerns about generalization across studies: Bias_M can be highly sensitive to the prevalence of true exposure ($T = 1$), this prevalence may vary greatly across populations, and the extent of variation may be highly uncertain because exposure prevalence may have never been reported without misclassification.

19.3.1 Formulas Based on Sensitivity and Specificity

The aforementioned problems have led the misclassification literature to focus on quantities that are more often published and more transportable (generalizable) than bias factors. Let $\phi_{xty} = \Pr(X = x|T = t, Y = y)$. The *sensitivities* of X for T are then the ϕ_{11y} and the *specificities* are ϕ_{00y}, which are sometimes studied in themselves to inform later uses of X as a measurement of T. For example, patient recall of medical events or treatments (X) is often compared to medical record information (T). The latter information is of course imperfect and is quite subject to error regarding, for example, use of drugs but may be accurate enough for other exposures (e.g., implanted devices or drug *prescriptions*) to be cautiously treated as the truth or "gold standard" for those exposures. The non-differentiality assumption (independence of X and Y given T) then corresponds to $\phi_{xty} = \Pr(X = x|T = t, Y = y) = \Pr(X = x|T = t)$ so that sensitivities and specificities remain constant across Y ($\phi_{111} = \phi_{110}$ and $\phi_{001} = \phi_{000}$).

Sensitivities and specificities are often called "true positive rates" and "true negative rates," respectively. In parallel, the complements of sensitivities and specificities,

$$\Pr(X = 0|T = 1, Y = y) = \phi_{01y} = 1 - \phi_{11y} \text{ and}$$
$$\Pr(X = 1|T = 0, Y = y) = \phi_{10y} = 1 - \phi_{00y}$$

are often called the "false-negative rates" and "false-positive rates," respectively. Note that X being a perfect measurement of T ($X = T$) corresponds to perfect sensitivity and specificity ($\phi_{11y} = \phi_{00y} = 1$) as well as to perfect T prediction ($\pi_{11y} = \pi_{00y} = 1$).

Given a set of values for the ϕ_{xty}, we can estimate the four desired TY counts A_{t+y} from the observed XY counts A_{+xy} by solving a system of four equations (one for each X, Y combination) giving each observed XY count A_{+xy} as the sum of those with $T = 1$ and those with $T = 0$ who contribute to the count:

$$\begin{aligned}\text{observed count} = A_{+xy} &\approx \phi_{x1y}A_{1+y} + \phi_{x0y}A_{0+y} \\ &= \phi_{x1y}A_{1+y} + \phi_{x0y}(A_{++y} - A_{1+y}) \\ &= (\phi_{x1y} - \phi_{x0y})A_{1+y} + \phi_{x0y}A_{++y}\end{aligned}$$

provided all the solutions are positive. The solutions are $A_{1+y} = (A_{+xy} - \phi_{x0y}A_{++y})/(\phi_{x1y} - \phi_{x0y})$, which lead to an adjusted TY odds ratio collapsed (summed) over X:

$$\begin{aligned}\widehat{OR}_{TY} = &[(A_{+11} - \phi_{101}A_{++1})(A_{+00} - \phi_{010}A_{++0})]/ \\ &[(A_{+01} - \phi_{011}A_{++1})\ (A_{+10} - \phi_{100}A_{++0})].\end{aligned}$$

This formula does not simplify to a simple direct adjustment to $\widehat{OR}_{XY}$ and shows that without further assumptions, we need to know four parameters to adjust for the misclassification.

Background information on the four classification parameters may sometimes be found in *validation studies*, which examine the accuracy of X as a measure of T. If the validation study is a substudy of the study under analysis (internal validation), one can and should analyze its TXY data together with the main XY data using special methods. Those methods may be viewed as missing-data techniques because with a validation substudy T becomes a partially missing variable (Little and Rubin 2002; Carroll et al. 2006), and thus may be handled by methods as simple as multiple imputation (Cole et al. 2006). A key assumption for these techniques however is that T is missing at random (MAR), which is satisfied if subjects with T measurement are a random sample of subjects within each level of the observed variables.

If the validation data are from outside the study under analysis (external validation), uncertainty about generalizability must be considered, although the validation data can help one specify a prior distribution for the classification parameters (which

are bias parameters). Again, it appears that estimates of sensitivities and specificities (ϕ_{11y} and ϕ_{00y}) are more safely generalized than estimates of T-predictive probabilities (π_{1xy}) because the latter depend directly on the T distribution, which is unobserved in the analysis data and thus has unknown discrepancy from the T distribution in the validation data. An alternative approach is to use background information on sensitivities and specificities to help specify components of the π_{1xy}. This approach will be illustrated below.

19.3.2 Probabilistic Analysis Based on Sensitivity and Specificity Priors

One may parallel the earlier simulation analysis by setting priors on the sensitivities and specificities and sampling from these priors and the random errors to generate a simulation histogram of $\widehat{OR}_{TY}$ (Lash and Fink 2003; Lash et al. 2009; Fox et al. 2005; Greenland and Lash 2008). Unfortunately there are several complications that arise due to the more complex structure of misclassification adjustments. First, some combinations of ϕ_{1ty} allowed by the priors may lead to negative (and thus impossible) values for the imputed TXY counts. Such combinations are said to be *inadmissible*. If the priors chosen for the ϕ_{1ty} allow a non-negligible probability for inadmissible combinations, the resulting simulation intervals will no longer agree well with the correct Bayesian posterior intervals, whether or not the inadmissible values are discarded (Greenland 2005; MacLehose and Gustafson 2012).

Second, independent priors for the ϕ_{1ty} are grossly unrealistic. A large source of negative correlation between sensitivity and specificity is the stringency of criteria used to declare $X = 1$ (as an indicator of $T = 1$), with more stringency leading to lower sensitivity and higher specificity, and less stringency leading to higher sensitivity and lower specificity. A source of positive correlation is that use of more accurate measurement methods to create X (e.g., direct measurements) can increase both sensitivity and specificity compared to less accurate methods (e.g., self-reports). These sources of correlation need not balance to zero.

Of even greater concern, case and non-case sensitivity and specificity will be very highly positively correlated. If as usual the same method is used for measurement of cases and non-cases, the proper "null" reference point is not independence, but is instead its positive opposite, non-differentiality ($\phi_{xt1} = \phi_{xt0}$). Non-differentiality implies complete dependence (and thus 100% correlation) between case and non-case sensitivity and specificity (although complete dependence is not sufficient to produce non-differentiality). For the same reasons, realistic priors for the T-predictive probabilities π_{txy} will also be highly correlated.

It is possible to construct dependent priors to account for such correlation sources, although specifying the correlations can be difficult. See Fox et al. (2005) and Greenland and Lash (2008) for details and examples.

19.3.3 Probabilistic Analysis Based on Predictive-Value Components

The predictive approach is familiar in validation-study analysis (Lyles 2002) but is also a useful alternative to using direct priors for sensitivity and specificity, because it can use instead parameters that have no admissibility restrictions and for which prior independence may not be such a severely distortive assumption (Greenland 2009a). To this end, we recast the T-predictive probabilities π_{1xy} in terms of a logistic model:

$$\pi_{1xy} = \Pr(T = 1|X = x, Y = y) = \text{expit}(\beta_T + \beta_{TX}x + \beta_{TY}y + \beta_{TXY}xy)$$

where $\text{expit}(z) = e^z/(1 + e^z)$, which can be rewritten as the odds model:

$$\begin{aligned}\pi_{1xy}/\pi_{0xy} &= \Pr(T = 1|X = x, Y = y)/\Pr(T = 0|X = x, Y = y)\\ &= \exp(\beta_T + \beta_{TX}x + \beta_{TY}y + \beta_{TXY}xy).\end{aligned}$$

The odds ratio relating T and X when $Y = y$ is then seen to be

$$\begin{aligned}OR_{TX|Y=y} &= (\pi_{11y}\pi_{00y})/(\pi_{10y}\pi_{01y}) = (\pi_{11y}/\pi_{01y})/(\pi_{10y}/\pi_{00y})\\ &= \exp(\beta_{TX} + \beta_{TXY}y).\end{aligned}$$

Some algebra shows that (by the usual symmetry property of odds ratios), $OR_{TX|Y=y}$ also equals $\phi_{11y}\phi_{00y}/\phi_{10y}\phi_{01y}$, which is the product of sensitivity and specificity divided by the product of false-positive and false-negative rates. This ratio is the *receiver-operating characteristic* (ROC) odds ratio; it measures the quality of X as an indicator of T, being infinite when X is perfect and 1 when X is completely worthless.

Assuming a rare disease (so that the odds ratio among non-cases $OR_{TX|Y=0}$ approximates the odds ratio OR_{TX}), a prior for $\beta_{TX} = \ln(OR_{TX|Y=0})$ can be based on expectations for the population sensitivity and specificity via the relation $OR_{TX|Y=0} = \phi_{110}\phi_{000}/\phi_{100}\phi_{010}$. For X to be a high-quality indicator of T, this ROC odds ratio must be very large (Pepe et al. 2004). For example with sensitivity 0.6 and specificity 0.8 among non-cases, $OR_{TX0} = \exp(\beta_{TX}) = (0.6/0.4)/(0.2/0.8) = 6$; 0.6 and 0.9 yield $(0.6/0.4)/(0.1/0.9) = 13.5$; 0.8 and 0.8 yield $(0.8/0.2)/(0.2/0.8) = 16$; and 0.8 and 0.9 yield $(0.8/0.2)/(0.1/0.9) = 36$. Thus, even for a mediocre measurement (with sensitivity of only 0.6 and specificity of only 0.8) the TX odds ratio among non-cases, $OR_{TX0} = \exp(\beta_{TX})$ is 6 indicating that a prior for β_{TX} should be distributed well above zero. The preceding numeric examples would be encompassed by a lognormal prior for OR_{TX0} with median 13.5 (in the center of the examples) and 95% limits of 5 and 36, which translates to a normal prior with mean ln(13.5) and variance 0.25 for β_{TX}

Table 19.5 A set of independent normal prior distributions for the coefficients in the logistic regression of T on X and Y in the antibiotic-SIDS example

	Mean	Variance	95% prior limits for
β_T	logit(0.1)	0.16	$\pi_{100} = \text{expit}(\beta_T)$: 0.05, 0.20
β_{TX}	ln(13.5)	0.25	$\exp(\beta_{TX})$: 5, 36
β_{TY}	0	0.50	$\exp(\beta_{TY})$: 0.25, 4
β_{TXY}	0	0.125	$\exp(\beta_{TXY})$: 0.5, 2

The ratio of ROC odds ratios is $\text{ROR}_{TX} = OR_{TX|Y=1}/OR_{TX|Y=0} = \exp(\beta_{TXY})$, which is a summary measure of how much better (if over 1) or worse (if under 1) X is as a T indicator among cases compared to non-cases. Under non-differential misclassification, β_{TXY} is 0 and $\exp(\beta_{TY})$ equals the target parameter OR_{TY} (although $\beta_{TXY} = 0$ is not sufficient to produce non-differentiality or $\exp(\beta_{TY}) = OR_{TY}$). Thus, unless severe non-differentiality is expected, it is reasonable to use a prior for β_{TXY} that is concentrated near zero, for example, a normal prior with mean 0 and variance 0.125, which produces 95% prior limits for ROR_{TX} of 0.5 and 2. Likewise, it would be reasonable to use a prior for β_{TY} that is similar to but somewhat wider than what would be considered a reasonable prior for the target log odds ratio $\ln(\text{OR}_{TY})$, for example, a normal prior with mean zero and variance 0.5, which produces 95% prior limits for $\exp(\beta_{TY})$ of 0.25 and 4.

In settings where recall bias is expected (differences in case and non-case recall), it is important to not be deceived into thinking such bias alone implies that ROR_{TX} is above or below 1 (β_{TXY} above or below 0). In the example in Table 19.1, the traumatic event of SIDS may trigger both true and false recall, increasing sensitivity (which increases ROR_{TX}) as well as decreasing specificity (which decreases ROR_{TX}). The net result is that uncertainty about β_{TXY} should be more symmetrically distributed around 0 than naïve reasoning might suggest.

Finally, note that $\text{expit}(\beta_T) = \pi_{100} = \Pr(T = 1|X = 0, Y = 0)$ is the probability that a "test negative" ($X = 0$) in a non-case is erroneous. For an exposure with an expected prevalence well below 50% (as was antibiotic use in unselected pregnancies during the study period) and a reasonably sensitive measure X of T, we should concentrate our prior distribution for π_{100} well below 0.5, which forces the prior for $\beta_T = \text{logit}(\pi_{100})$ to be well below 0. An example would be a normal prior with mean for β_T with mean $\text{logit}(0.1) = -2.20$ and variance 0.16, which produces 95% prior limits for π_{100} of 0.05 and 0.20.

Table 19.5 summarizes the independent normal priors suggested above for the logistic predictive-value coefficients β_T, β_{TX}, β_{TY}, β_{TXY}, which are also consistent with the limited background information about the antibiotic-SIDS relation at the time (Kraus et al. 1989). These can be used in a Monte Carlo sensitivity analysis in which at each cycle:

1. All counts are resampled from Table 19.1.
2. β_T^*, β_{TX}^*, β_{TY}^*, and β_{TXY}^* are drawn from their priors and used to compute the π_{txy}^*.
3. These π_{txy}^* are multiplied against the resampled counts to get a simulated TXY table, as in Table 19.4a.

4. The TXY table is collapsed over X, and $\widehat{OR}^*_{TY}$ is computed from the resulting TY table (Table 19.4b).

After 250,000 repetitions of sampling cycle 1–4, the 2.5th and 97.5th percentiles of the resulting $\widehat{OR}^*_{TY}$ distribution are 0.37 and 3.4, with a median of 1.2. Various Bayesian calculations with non-informative priors on all parameters besides β_T, β_{TX}, β_{TY}, and β_{TXY} yield similar results (Greenland 2009a, b). These results show that the precision of the conventional 95% confidence limits of 1.1, 1.8, if interpreted as an interval for OR_{TY}, is due to its assumption that X is a perfect measurement of T ($X = T$, i.e., $\pi_{11y} = \pi_{00y} = 1$ or $\phi_{11y} = \phi_{00y} = 1$). The expansion of the 95% interval for OR_{TY} to [0.37, 3.4] further suggests that the data add little information about OR_{TY} beyond that conveyed by the priors.

The simple simulation analysis just described can be extended to allow for selection bias and confounder adjustment (both measured and unmeasured) as well as for misclassification, although it can become quite unwieldy due the number of bias parameters and data cells involved (Greenland 2005). Unfortunately, it does not extend easily to situations in which all parameters (not just bias parameters) are given priors, whereas Bayesian computation has no such difficulty.

19.4 Pointwise Versus Probabilistic Analysis

While basic sensitivity or bias analyses are widely accepted and even promoted, there are difficulties if not objections to arbitrary elements in the priors are needed for probabilistic extensions (such as prior shape, e.g., normal vs. beta vs. trapezoidal). These difficulties are understandable given the lack of neutral and accepted conventions for these elements.

To address these elements, the analysis may be repeated using different identifying distributions and the results from different choices tabulated and contrasted. This process is called *prior sensitivity analysis*. Sensitivity analyses that instead directly vary bias parameters (*pointwise analyses* or deterministic analyses) are the special case of prior sensitivity analysis in which the varied priors are limited to point-mass priors (which assign 100% probability to just one value for the parameter).

From this viewpoint, pointwise analyses are of limited value compared to full probabilistic sensitivity analysis. Their main limitation is that they provide no explicit accounting for uncertainty about the bias parameters (Greenland 1998, 2005; Gustafson et al. 2001). The point-mass priors implicit in these analyses have no scientific justification given the uncertainty inherent in the parameters and do not escape the arbitrariness problem because they are based on a range of bias-parameter variation that is itself arbitrary. Too broad a range will lead to inclusion of values that would be recognized by subject experts as misleadingly absurd (e.g., a smoking prevalence of 90% in a general population), while too narrow a range will exclude important possibilities that cannot be ruled out. It is thus worthwhile to use genuine background information to create credible or plausible priors, or at least to specify the values used in a pointwise analysis.

In accounting for background information, there is a need to avoid giving that information more weight than it is worth, in recognition of the generalization problems touched on earlier. In particular, prior distributions should not be too narrow (incautious), lest our results remain overconfident. Of special note is that confidence intervals from other studies do not account for generalization problems and thus are too narrow to be used directly as prior intervals; at the very least, their probability percentage would need to be reduced below the original confidence percentage. For example, a 95% confidence interval for an effect from one study (even a perfect trial) should not be accorded 95% probability of containing the effect in another study, since generalization should reduce our confidence that the interval contains the latter effect.

Even with attention to such details, what appear to be cautious priors to one analyst may appear incautious to another. An ideal solution would allow each analyst to plug in their own priors and repeat the procedure, thus making the audience an active participant in sensitivity analysis. Macros and spreadsheets allowing such a solution have been offered (Steenland and Greenland 2004; Fox et al. 2005; Orsini et al. 2008; Lash et al. 2009); in principle these do not require audience access to the data, so confidentiality can be assured. Similar extensions to model-sensitivity analysis and influence analysis are also possible (e.g., by re-specifying external constraints on inclusion and exclusion criteria).

19.5 Some Theoretical Considerations for General Bias Analysis

This section concerns general formulations that come into play in analyzing more complex data and in multiple-bias modeling (Greenland 2005, 2009b; Vansteelandt et al. 2006; Molitor et al. 2009). It presumes some familiarity with mathematical statistics and may be skipped by less theoretical readers.

For the purposes of basic bias analysis of a single data set, the following definitions can be useful. A parameter in a data-generating model is *fully identified* by a system of estimating equations for the model parameters if the parameter is constant over the set of solutions to the system. More generally, a function of the model parameters is *fully identified* by the equations for the model parameters if the function maps the solution set to a unique value. At the opposite extreme, the function is *not identified* by the equations if it maps the solutions onto its entire range. Thus, in the confounding example (Table 19.3), the target parameter is not identified by its estimating equation $OR_{XY|U} \approx \exp(\widehat{\beta}_{XY})/\text{Bias}_U$, because given XY data with a finite $\widehat{\beta}_{XY}$ (and no further constraint), there is a solution for every possible value r of $OR_{XY|U}$; one merely chooses the bias parameters to make $\text{Bias}_U = \exp(\widehat{\beta}_{XY})/r$.

Between the extremes of full and no identification, a function is *partially identified* by the equations if it maps solutions to a proper subset of its range but not to a unique value. In bias settings this usually means the solutions are mapped into a subset of the range of reduced but positive dimension, although it may also or

instead mean that the image of the solutions is sharply bounded inside the function range. The entire model is *non-identified* if some function of its parameters is only partially identified. An *identification problem* exists if the target parameter is not fully identified, for that means multiple possible values for that parameter arise from solutions to the estimating equations.

These are data-dependent definitions and so oriented toward likelihood and Bayesian analysis; they are modified for asymptotic sampling theory since non-identification whose probability goes to zero with increasing sample size can be ignored in that theory. Regardless, identification problems are traditionally solved by introducing artificial, arbitrary constraints (such as no bias for the target parameter) that at most have justification in a very narrow range of real circumstances (e.g., randomized trials with perfect compliance and no loss) and may be generalized only in very limited ways (e.g., to generalized linear semi-parametric models like Cox proportional hazards regression).

A common example is the assumption of *no residual selection bias or confounding* (often referred to as "no uncontrolled selection bias or confounding" although the terminology varies considerably). This condition corresponds to an ignorability condition that selection for analysis and treatment assignment are randomized conditional on the regression covariates, making selection and assignment independent of the potential outcomes that compose the target effect parameter (the ignorability being *strong* or *weak* according to whether it refers to joint or component-wise independence of potential outcomes and treatment assignment). Coupled with the assumption that the regression model is correct, the regression model for Y then coincides with the structural equation for Y (the Y-potential-outcome function), and further confounding concerns are obviated; unbiased conditional-effect estimation then reduces to the familiar problem of unbiased estimation of the regression of Y on treatment and the covariates. Alternatively, assuming instead correct models for selection and treatment-assignment probabilities, the fitted models for these probabilities can be used to estimate marginal effects (as in inverse-probability-weighting methods).

A problem with this traditional approach (of forcing identification through no-residual bias assumptions) is that it generates highly overconfident inferences: Confidence intervals so derived fail to cover the target near the nominal rate, and posterior intervals are far too narrow given the actual amount of external information available. This overconfidence follows from forcing unknown bias parameters to equal unique values, which takes no account of the huge uncertainty surrounding those parameters.

To reflect this uncertainty properly, Leamer (1974, 1978) and many others recommended instead one begin with non-identified models, and then examine the impact of augmenting the estimating functions with contextually based parameter constraints (Graham 2000; Gustafson et al. 2001; Greenland 2003, 2005, 2009a, b; Gustafson 2003, 2005). Those constraints are in turn derived from a *prior distribution* or *penalty function* for some or all of the parameters. If one allows for the fallibility of these contextual inputs, bias analysis then serves chiefly as an antidote or diagnostic to counterbalance the overconfident results from the

conditionally randomized model. The conditionally randomized model remains a simple if doubtful reference point, arguably a necessary convention given the enormous number of prior distributions that would identify the target parameter.

Randomization assumptions correspond to prior distributions that assign positive probability to only a proper subspace of the non-identified parameters (making the priors degenerate with respect to the bias-parameter space). In the selection-bias example, random sampling corresponds to no association of X with S, which occurs in the subspace defined by $\pi_{s1y} = \pi_{s0y}$; in the confounding example, randomization induces no association of U with X, which occurs in the subspace defined by $\pi_{u1y} = \pi_{u0y}$. More generally, no net bias ($\text{Bias}_{S=1} * \text{Bias}_U * \text{Bias}_M = 1$ in the example) constrains all bias parameters to a manifold of much lower dimension than the entire bias-parameter space. Realistic priors rarely entail any such dimension reduction, even when they concentrate near lower-dimensional manifolds, and thus reveal the source of overconfidence from traditional identifying assumptions.

19.6 Conclusions

The present chapter has addressed settings in which a "correct" model for the data generation can never be known or approximated, and hence the data cannot tell us whether we are close to or far from the target, even probabilistically. In these settings, we cannot guarantee that inferences from a posterior distribution will be superior to inferences from our prior alone (Neath and Samaniego 1997). Thus, the importance of specific models and priors is de-emphasized in favor of providing a framework for sensitivity analysis across plausible models and priors. This framework need not be all-encompassing, because often just a few plausible specifications can usefully illustrate the illusory nature of an apparently conclusive conventional analysis.

When inference is mandated in these settings (as in pooled analyses to advise policy), we must admit we can only propose models that incorporate or are at least consistent with facts as we know them and that all inferences are completely dependent on these modeling choices (including non-parametric or semi-parametric inferences). There will be an infinite number of such models, and they will not all yield similar inferences, necessitating some sort of sensitivity analysis to provide a picture of the problem unless the effects in question are so large as to be obvious (as is typically the case in disease outbreaks).

In this task statistical modeling provides only inferential possibilities rather than inferences. Any analysis should thus be viewed as part of a sensitivity analysis which depends on external plausibility considerations to reach conclusions (Greenland 2005, 2009b; Vansteelandt et al. 2006). Results from single models are merely examples of what might be plausibly inferred, although just one plausible inference may suffice to demonstrate inherent limitations of the data.

Whatever their value for summarization, conventional models treat unknown parameters as if known and thus do not satisfy plausibility considerations. These unknowns include many bias parameters that can be forced to their null by

successful design strategies but are probably not null in most observational settings. The use of prior distributions can provide more plausible inferences by allowing expansion of conventional models to include bias parameters as unknowns.

In any given observational context, progress beyond prior-based analyses can be made only by obtaining data from a study design that pins down a previously unknown bias parameter. Examples of such designs include studies with successful randomization, highly accurate measurement, or that have isolated a controversial effect in a setting where the effect is so large as to exceed plausible bias magnitudes. Because such designs are often infeasible (e.g., in the electromagnetic field-childhood cancer controversy), proper construction and use of prior distributions should become a component of statistical training for observational research.

Despite these considerations, there is no basis for mandating a bias analysis of every study or even most studies. For example, bias analysis is superfluous when conventional intervals show that no useful conclusion could be drawn from the study even if it were perfect apart from random error. More generally, rather than providing a bias analysis, a study may provide greater service by refraining from inference; instead it can focus on carefully reporting its design, conduct, and data in great detail to facilitate pooling and meta-analysis (Greenland et al. 2004). Inferences are best based on a more complete account of evidence than can be provided in a single study report, and thus the effort of bias analysis is more justifiable in research synthesis (Turner et al. 2009; Welton et al. 2009). Even there, bias analysis becomes essential only when doing risk assessment or when authors claim to offer near-definitive conclusions.

References

Breslow NE, Day NE (1980) Statistical methods in cancer research. Vol I: the analysis of case-control data. IARC, Lyon

Bross IDJ (1966) Spurious effects from an extraneous variable. J Chronic Dis 19:637–647

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

Brumback BA, Hernan MA, Haneuse S, Robins JM (2004) Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. Stat Med 23: 749–767

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu C (2006) Measurement error in nonlinear models, 2nd edn. Chapman and Hall, Boca Raton

Cole SR, Chu H, Greenland S (2006) Multiple-imputation for measurement-error correction (with comment). Int J Epidemiol 35:1074–1082

Copas JB, Li HG (1997) Inference for non-random samples (with discussion). J R Stat Soc Ser B 59:55–77

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

Chu H, Wang Z, Cole SR, Greenland S (2006) Sensitivity analysis of misclassification: a graphical and a Bayesian approach. Ann Epidemiol 16:834–841

Eddy DM, Hasselblad V, Schachter R (1992) Meta-analysis by the confidence profile method. Academic press, New York

Efron B, Tibshirani RJ (1994) An introduction to the bootstrap. Chapman and Hall, New York

Flanders WD, Khoury MJ (1990) Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. Epidemiology 1:199–246

Fox MP, Lash TL, Greenland S (2005) A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int J Epidemiol 34:1370–1376

Gail MH, Wacholder S, Lubin JH (1988) Indirect corrections for confounding under multiplicative and additive risk models. Am J Ind Med 13:119–130

Glymour MM, Greenland S (2008) Causal diagrams. Chapter 12. In: Rothman KJ, Greenland S, Lash TL (eds) Modern epidemiology, 3rd edn. Lippincott-Williams-Wilkins, Philadelphia, pp 183–209

Graham P (2000) Bayesian inference for a generalized population attributable fraction. Stat Med 19:937–956

Greenland S (1998) The sensitivity of a sensitivity analysis. In: 1997 Proceedings of the biometrics section. American Statistical Association, Alexandria, pp 19–21

Greenland S (2001) Sensitivity analysis, Monte-Carlo risk analysis, and Bayesian uncertainty assessment. Risk Anal 21:579–583

Greenland S (2003) The impact of prior distributions for uncontrolled confounding and response bias. J Am Stat Assoc 98:47–54

Greenland S (2005) Multiple-bias modeling for observational studies (with discussion). J R Stat Soc Ser A 168:267–308

Greenland S (2008) Introduction to Bayesian statistics. Chapter 18. In: Rothman KJ, Greenland S, Lash TL (eds) Modern epidemiology, 3rd edn. Lippincott-Williams-Wilkins, Philadelphia, pp 328–344

Greenland S (2009a) Bayesian perspectives for epidemiologic research. III. Bias analysis via missing-data methods. Int J Epidemiol 38:1662–1673. corrigendum (2010) Int J Epidemiol 39:1116

Greenland S (2009b) Relaxation penalties and priors for plausible modeling of nonidentified bias sources. Stat Sci 24:195–210

Greenland S, Lash TL (2008) Bias analysis. Chapter 19. In: Rothman KJ, Greenland S, Lash TL (eds) Modern epidemiology, 3rd edn. Lippincott-Williams-Wilkins, Philadelphia, pp 345–380

Greenland S, Maldonado G (1994) The interpretation of multiplicative model parameters as standardized parameters. Stat Med 13:989–999

Greenland S, Mickey RM (1988) Closed form and dually consistent methods for inference on strict collapsibility in 2x2xK and 2xJxK tables. Appl Stat 37:335–343

Greenland S, Gago-Domiguez M, Castellao JE (2004) The value of risk-factor ("black-box") epidemiology (with discussion). Epidemiology 15:519–535

Gustafson P (2003) Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments. Chapman and Hall/CRC, Boca Raton

Gustafson P (2005) On model expansion, model contraction, identifiability, and prior information (with discussion). Stat Sci 20:111–140

Gustafson P, Le ND, Saskin R (2001) Case–control analysis with partial knowledge of exposure misclassification probabilities. Biometrics 57:598– 609

Gustafson P, McCandless LC, Levy AR, Richardson SR (2010) Simplified Bayesian sensitivity analysis for mismeasured and unobserved confounders. Biometrics 66:1129–1137

Hatch EE, Kleinerman RA, Linet MS, Tarone RE, Kaune WT, Auvinen A, Baris D, Robison LL, Wacholder S (2000) Do confounding or selection factors of residential wire codes and magnetic fields distort findings of electromagnetic fields studies? Epidemiology 11:189–198

Joseph L, Gyorkos TW, Coupal L (1995) Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am J Epidemiol 141: 263–272

Kitagawa EM (1955) Components of a difference between two rates. J Am Stat Assoc 50: 1168–1194

Kraus JF, Greenland S, Bulterys MG (1989) Risk factors for sudden infant death syndrome in the U.S. Collaborative Perinatal Project. Int J Epidemiol 18:113–120

Lash TL, Fink AK (2003) Semi-automated sensitivity analysis to assess systematic errors in observational epidemiologic data. Epidemiology 14:451–458
Lash TL, Fox MP, Fink AK (2009) Applying quantitative bias analysis to epidemiologic data. Springer, New York
Leamer EE (1974) False models and post-data model construction. J Am Stat Assoc 69:122–131
Leamer EE (1978) Specification searches. Wiley, New York
Leamer EE (1985) Sensitivity analyses would help. Am Econ Rev 75:308–313
Little RJA, Rubin DB (2002) Statistical analysis with missing data, 2nd edn. Wiley, New York
Lyles RH (2002) A note on estimating crude odds ratios in case-control studies with differentially misclassified exposure. Biometrics 58:1034–1037
MacLehose RF, Gustafson P (2012) Is probabilistic bias analysis approximately Bayesian? Epidemiology 23:151–158
McCandless LC, Gustafson P, Levy AR (2007) Bayesian sensitivity analysis for unmeasured confounding in observational studies. Stat Med 26:2331–2347
McCandless LC, Gustafson P, Levy AR, Richardson SR (2012) Hierarchical priors for bias parameters in Bayesian sensitivity analysis for unmeasured confounding. Stat Med 31: 383–396
Molitor N-T, Best N, Jackson C, Richardson S (2009) Using Bayesian graphical models to model biases in observational studies and to combine multiple sources of data: application to low birth weight and water disinfection by-products. J R Stat Soc Ser A 172:615–638
Neath AA, Samaniego FJ (1997) On the efficacy of Bayesian inference for nonidentifiable models. Am Stat 51:225–232
Orsini N, Bellocco R, Bottai M, Wolk A, Greenland S (2008) A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. Stata J 8:29–48
Pearl J (2009) Causality, 2nd edn. Cambridge University Press, New York
Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P (2004) Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 159:882–890
Phillips CV (2003) Quantifying and reporting uncertainty from systematic errors. Epidemiology 14:459–466
Rosenbaum PR (2002) Observational studies, 2nd edn. Springer, New York
Saltelli A, Chan K, Scott EM (eds) (2000) Sensitivity analysis. Wiley, New York
Scharfstein DO, Daniels MJ, Robins JM (2003) Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. Biostatistics 4:495–512
Schlesselman JJ (1978) Assessing effects of confounding variables. Am J Epidemiol 108:3–8
Schneeweiss S (2006) Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf 15:291–303
Steenland K, Greenland S (2004) Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. Am J Epidemiol 160:384–392
Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG (2009) Bias modeling in evidence Synthesis. J R Stat Soc Ser A 172:21–47
VanderWeele TJ, Arah OA (2011) Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments and confounders. Epidemiology 22:42–52
Vansteelandt S, Goetghebeur E, Kenward MG, Molenberghs G (2006) Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. Stat Sin 16:953–980
Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC (2009) Models for potentially biased evidence in meta-analysis using empirically based priors. J R Stat Soc Ser A 172:119–136
Yanagawa T (1984) Case-control studies: assessing the effect of a confounding factor. Biometrika 71:191–194