

## Chapter 2

# A Guide to Implementing Quantitative Bias Analysis

### Introduction

Estimates of association from nonrandomized epidemiologic studies are susceptible to two types of error: random error and systematic error. Random error, or sampling error, is often called chance, and decreases toward zero as the sample size increases and the data are more efficiently distributed in the categories of the adjustment variables. The amount of random error in an estimate of association is measured by its precision. Systematic error, often called bias, does not decrease toward zero as the sample size increases or with more efficient distributions in the categories of the analytic variables. The amount of systematic error in an estimate of association is measured by its validity.

Conventional confidence intervals depict the random error about an estimate of association, but give no information about the amount of systematic error. The objective of quantitative bias analysis is to estimate quantitatively the systematic error that remains after implementing a study design and analysis. For comprehensive guidance on study design and analysis, the reader should consult an epidemiology methods textbook, such as *Modern Epidemiology* (Rothman et al., 2008b). This text not only assumes that the reader has applied established principles for study design and analysis but also recognizes that the systematic error remaining after implementing those principles merits quantification and presentation. The next sections briefly review principles of design and analysis that have presumably been applied. They are followed by sections on planning for quantitative bias analysis and a brief overview of the types of bias analyses described in this chapter.

### Reducing Error

The objective of analytic epidemiologic research is to obtain an accurate (precise and valid) estimate of the effect of an exposure on the occurrence of a disease. Epidemiologic studies should be designed and analyzed with this objective in mind, but epidemiologists should realize that this objective can never be completely achieved.

Every study has limited sample size, which means that every study contains some random error. Similarly, every study is susceptible to sources of systematic error. Even randomized epidemiologic studies are susceptible to selection bias from losses to follow-up and to misclassification of analytic variables. Since the research objective can never be achieved perfectly, epidemiologists should instead strive to reduce the impact of error as much as possible. These efforts are made in the design and analysis of the study.

## *Reducing Error by Design*

To reduce random error in a study's design, epidemiologists can increase the size of the study or improve the efficiency with which the data are distributed into the categories of the analytic variables. Increasing the size of the study requires enrolling more subjects and/or following the enrolled subjects for a longer period, and this additional information ought to reduce the estimate's standard error. A second strategy to improve an estimate's precision is to improve the efficiency with which the data are distributed into categories of the analytic variables. This strategy also reduces the standard error of the estimate of association, which improves its precision. Consider the standard error of the odds ratio, which equals the square root of the sum of inverses of the frequencies of the interior cells of a two-by-two table. As displayed in Table 2.1, the two-by-two table is the simplest contingency table relating exposure to disease.

**Table 2.1** The two-by-two contingency table relating exposure to disease

	Exposed	Unexposed
Disease	a	b
Undiseased	c	d

With this data arrangement, the odds ratio equals  $(a/c)/(b/d)$  and its standard error equals  $\sqrt{(1/a + 1/b + 1/c + 1/d)}$ . In a study with 100 subjects and each interior cell frequency equal to 25, the odds ratio equals its null value of 1.0 and the standard error of the odds ratio equals  $\sqrt{(1/25 + 1/25 + 1/25 + 1/25)} = 0.4$ . The 95% confidence interval about the null odds ratio equals 0.46 to 2.19. If only 40% of the cases were located, but the sample size remained constant by increasing the case to control ratio to 1 to 4, rather than 1 to 1, the odds ratio would remain null. The odds ratio's standard error would then equal  $\sqrt{(1/10 + 1/10 + 1/40 + 1/40)} = 0.5$  and the 95% confidence interval would equal 0.38 to 2.66. Although the sample size (100 subjects) did not change, the standard error and the width of the confidence interval (measured on the log scale) have both increased by 25% due only to the less efficient distribution of the subjects within the contingency table. This example illustrates how the efficiency of the distribution of data within the categories of analytic variables affects the study precision, given a fixed sample size, even when no bias is present.

Improving the efficiency with which the data are distributed requires an understanding of the distribution of the exposure and disease in the source population. If

one is interested in studying the relation between sunlight exposure and melanoma incidence, then a population in the northern United States might not have an efficient distribution of the exposure compared with a population in the southern United States where sunlight exposure is more common. If one is interested in studying the relation between tanning bed exposure and melanoma incidence, then a population in the northern United States might have a more efficient distribution of the exposure than a population in the southern United States. Careful selection of the source population is one strategy that investigators can use to improve the efficiency of the distribution of subjects within the categories of the analytic variables.

Matching is a second strategy to improve the efficiency of this distribution. Matching a predetermined number of controls to each case on potential confounders assures that the controls will appear in a constant ratio to cases within the categories of the confounder. For example, skin type (freckled vs unfreckled) might confound the relation between sunlight exposure and melanoma incidence. Cases of melanoma may be more likely to have freckled skin than the source population that gives rise to cases, and people with freckled skin might have different exposure to sunlight than people with unfreckled skin. Without matching, most of the cases will be in the category of the confounder denoting freckled skin, and most of the controls will be in the category of the confounder denoting unfreckled skin because it is more common in the source population. This disparity yields an inefficient analysis, and therefore a wider confidence interval. Matching controls to cases assures that controls appear most frequently in the stratum where cases appear most frequently (e.g., freckled skin), so the analysis is more efficient and the confidence interval narrower. Matching unexposed to exposed persons in cohort studies can achieve a similar gain in efficiency.

To reduce systematic error in a study's design, epidemiologists should focus on the fundamental criterion that must be satisfied to obtain a valid comparison of the disease incidence in the exposed group with the disease incidence in the unexposed group. That is, the unexposed group must have the disease incidence that the exposed group would have had, had they been unexposed (Greenland and Robins, 1986), within the strata of measured confounders. The ideal study would compare the disease occurrence in the exposed group (a factual, or observable, disease incidence) with the incidence they would have had, had they been unexposed (a counterfactual, or unobservable, disease incidence). Since the ideal comparison can never be realized, the disease incidence is measured in a surrogate group: a second group of subjects who are unexposed and whose disease experience we substitute for the counterfactual ideal. The validity of that substitution, which cannot be verified, directly impacts the validity of the estimate of association. The investigator must strive for the desired balance in the collapsed data, which is achievable within probability limits by randomization, or within strata of measured confounders.

With this criterion in mind, the design principles to enhance validity follow directly. The study population should be selected such that participation is not conditional on exposure status or disease status. When both exposure status and disease status affect the probability that a member of the source population participates in the study, the estimate of association will be susceptible to selection bias.

Enrolling subjects and/or documenting their exposure status before the disease occurs (i.e., prospectively) assure that disease status cannot be associated with initial participation.

Second, the study population should be selected such that the net effects of all other predictors of the outcome, aside from exposure itself, are in balance between the exposed and unexposed groups. This balance is commonly referred to as having no confounding. Randomization achieves this objective within limits that are statistically quantifiable. When exposure status cannot be assigned by randomization, which is usually the situation in studies of disease etiology, the investigator can limit confounding by restricting the study population to one level of the confounder or ensuring that data are collected on potential confounders so that their effects can be assessed in the analysis.

Finally, the data should be collected and converted to electronic form with as few classification errors as possible. Some errors in classification are, however, inevitable. Investigators often strive to assure that the rates of classification errors do not depend on the values of other variables (e.g., rates of exposure classification errors do not depend on disease status, which is called nondifferential exposure misclassification) or on the proper classification of other variables (e.g., errors in classification of exposure are as likely among those properly classified as diseased as among those improperly classified as diseased, which is called independent exposure misclassification). This second objective can be readily achieved by using different methods to collect information on disease status from those used to collect information on exposure status (as well as information on confounders). The data collection for disease status should be conducted so that the data collector is blinded to the information on exposure and confounder status. Nondifferential and independent errors in classification often yield the most predictable, and therefore most readily correctable, bias of the estimate of association. Nonetheless, one may choose to select a design expected to yield relatively small differential classification errors in preference to a design expected to yield relatively large nondifferential classification errors, since the former would yield less bias and uncertainty. Generalized advice always to balance information quality across compared categories (i.e., to strive for nondifferential classification errors) ignores the potential for this trade-off to favor small differential errors.

### ***Reducing Error in the Analysis***

Following implementation of a design that reduces random and systematic error to the extent practical, a well-designed analysis of the collected data can further reduce error. Data analysis should begin with a clearly specified definition of each of the analytic variables. The conversion algorithm and variable type should be defined for each analytic variable, after careful consideration of the variability in dose, duration, and induction period that will be characterized in the analysis.

After completing the definition and coding of analytic variables, the analysis proceeds to a descriptive characterization of the study population. The descriptive analysis

shows the demographic characteristics of the people in the study. For example, it might show the proportion of the population enrolled at each of the study centers, the distribution of age, and the proportion belonging to each sex. Descriptive analyses should include the proportion of the study population with a missing value assigned to each analytic variable. The proportion with missing data helps to identify analytic variables with problems in data collection, definition, or format conversion.

Examination of the bivariate relations between analytic variables is the third step in data analysis. Bivariate relations compare proportions, means, or medians for one study variable within categories of a second. These comparisons inform the analyst's understanding of the data distributions and can also identify data errors that would prompt an inspection of the data collection, variable definitions, or format conversions. The number of bivariate relations that must be examined grows exponentially as the number of analytic variables increases. If the number grows too large to be manageable, the analyst should restrict the examination to pairs that make sense *a priori*. However, whenever possible, all pairs ought to be examined because a surprising and important finding might easily arise from a pair that would be ignored *a priori*.

The comparisons of the proportions with the disease of interest within the categories of the analytic variables are a special subset of bivariate comparisons. These proportions can be explicitly compared with one another by difference or division, yielding estimates of association such as the risk difference, risk ratio, or a difference in means. When estimates of association are calculated as a part of the bivariate comparison, the analysis is also called a stratified analysis. Often one comparison is a focus of the stratified analysis, which is the comparison of the disease proportions in those exposed to the agent of interest with those unexposed to the agent of interest. This comparison relates directly to the original objective: a valid and precise estimate of the effect of an exposure on the occurrence of a disease. To continue the stratified analysis, the comparisons of disease proportions in exposed versus unexposed are expanded to comparisons within levels of other analytic variables. For example, the risk ratio comparing exposed with unexposed might be calculated within each of the three age groups. An average risk ratio can be calculated by standardization or pooling. Comparison of this average or summarized risk ratio with the crude or collapsed risk ratio (including all ages in one stratum) indicates whether age is an important confounder of the risk ratio. If the pooled risk ratio is substantially different from crude risk ratio, then the pooled risk ratio will provide an estimate of association that is unconfounded (by age) and is precision enhancing, in that its confidence interval will be narrower than those obtained from alternative methods for averaging the risk ratio across strata of age. Pooling reduces both the random error (by yielding a precision-enhancing estimate of association) and the systematic error (by yielding an estimate of association unconfounded by age). The correspondence between noncollapsibility and confounding holds also for the odds ratio, hazard, ratio, rate ratio, and rate difference, so long as the risk of disease is low (<10%) in every combination of the categories of exposure and the categories of controlled confounders. When the risk of disease is greater than 10%, these estimates of association may not be collapsible across strata of a control variable, even if that variable is not a confounder.

The analysis can proceed by further stratification on a second variable (e.g., sex groups) and pooling to simultaneously adjust for confounding by both age and sex. The number of strata increases geometrically as additional variables are analyzed, which can become confusing as the number of strata increases beyond what can be easily reviewed on a single page. In addition, the data quickly become too sparse for pooling as the frequencies in some cells fall below about five and may reach zero. A common solution to the problem engendered by this geometric progression is to use regression modeling rather than stratification. Regression models yield estimates of association that simultaneously adjust for multiple confounders and that are also precision-enhancing. Their advantage over stratification is that they do not become cumbersome or suffer from small numbers as easily as multiple stratification. However, regression modeling does not show the data distribution, so should not be used without first conducting the bivariate analysis and stratification on the critical confounders.

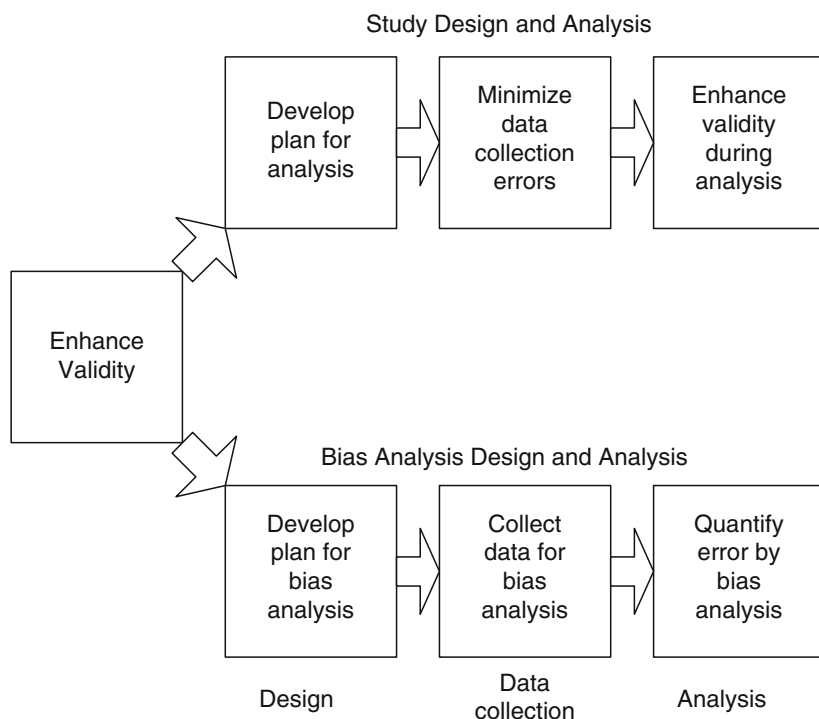
This analytic plan describes the conventional epidemiologic approach to data analysis. It yields a quantitative assessment of random error by producing confidence intervals about the crude or pooled estimates of association. It also adjusts the estimate of association for confounding variables included in the stratification or regression model. However, there is no adjustment for selection bias, measurement error, confounding by unmeasured confounders, or residual confounding by measured confounders that are poorly specified or poorly measured. Nor is there any quantification of uncertainty arising from these sources of bias. Quantitative bias analysis addresses these shortcomings in the conventional approach to epidemiologic data analysis.

## **Quantifying Error**

The goal of quality study design and analysis is to reduce the amount of error in an estimate of association. With that goal in mind, investigators have an obligation to quantify how far they are from this goal. Quantitative bias analysis achieves this objective. Conducting a study that will yield a measure of association with as little bias as practical requires careful planning and choices in the design of data collection and analysis. Similarly, quantifying the amount of residual bias requires choices in the design of data collection and analysis. Since conducting a high-quality bias analysis follows the same steps as conducting a high-quality epidemiologic study, plans for both should be integrated at each phase of the study, as depicted in Fig. 2.1.

## ***When Is Quantitative Bias Analysis Valuable?***

Before discussing the steps involved in planning and conducting a quantitative bias analysis, it is important to first consider when it makes the most sense to conduct a bias analysis. Quantitative bias analysis is most valuable when a study is likely to



**Fig. 2.1** Integration of planning for bias analysis with conventional study design and analysis

produce a precise estimate of association, such that the conventional confidence interval will be narrow. A narrow interval reflects a small amount of residual random error, tempting stakeholders to underestimate the true uncertainty and overstate their confidence that an association is truly causal and of the size estimated by the study. When a wider interval is obtained, inference from the study's results should be tenuous because of the substantial random error, regardless of whether systematic error has also been estimated quantitatively. Note that this formulation assumes that investigators use conventional confidence intervals as intended, that is, as a measure of error. Investigators who simply note whether the interval includes the null, a surrogate for statistical significance testing, will often be misled by statistically significant, but substantially imprecise estimates of association (Poole, 2001).

Quantitative bias analysis is also most valuable when a study is likely susceptible to a limited number of systematic errors. Studies susceptible to multiple substantial biases are not good candidates for quantitative bias analysis because the total error is too large to reliably quantify. These studies are similar to studies that yield wide conventional confidence intervals: the investigator or consumer should recognize that no inference will be reliable, so the effort of a quantitative bias analysis will

not be productive. Studies with wide conventional errors or that are susceptible to many large systematic errors might instead be useful for generating ideas for better-designed and larger subsequent studies. They should seldom however, provide a basis for inference or policy action, so the additional effort of quantitative bias analysis would not be an efficient use of resources.

Quantitative bias analysis is therefore most valuable when studies yield narrow conventional confidence intervals – so have little residual random error – and when these studies are susceptible to a limited number of systematic errors. Such studies often appear to be an adequate basis for inference or for policy action, even though only random error has been quantified by the conventional confidence interval. Quantification of the error due to the limited number of biases will safeguard against inference or policy action that takes account of only random error. Without a quantitative assessment of the second important source of error – systematic error – the inference or policy action would usually be premature.

### *Planning for Bias Analysis*

Quantitative bias analysis is best accomplished with foresight, just as with all aspects of epidemiologic research. The process of conducting a well-designed bias analysis goes beyond simply understanding the methods used for the analysis, but also includes a thorough planning phase to ensure that the information needed for quantification of bias is carefully collected. To facilitate this collection, investigators should consider the important threats to the validity of their research while designing their study. This consideration should immediately suggest the quantitative analyses that will explore these threats, and should thereby inform the data collection that will be required to complete the quantitative analyses.

For example, an investigator may design a retrospective case-control study of the relation between leisure exposure to sunlight and the occurrence of melanoma. Cases of melanoma and controls sampled from the source population will be interviewed by telephone regarding their exposures to sunlight and other risk factors for melanoma. The investigator should recognize the potential for selection bias to be an important threat to the study's validity: cases may be more likely than controls to agree to the interview, and those who spend substantial time in sunlight might also participate at a different rate than those who do not spend much time in the sun. To quantitatively address the potential selection bias (Chap. 4), the investigator will need to know the participation proportions in cases and controls, within groups of high and low exposure to sunlight. Case and control status will be known by design, but to characterize each eligible subject's sunlight exposure, the investigator will need to complete the interview. Sunlight exposure will not, therefore, be known for subjects who refuse to participate. However, in planning for a quantitative bias analysis, the investigator might ask even those who refuse to participate



whether they would be willing to answer a single question regarding their sunlight exposure. If the proportion of refusals who did agree to answer this one question was high, this alone would allow the investigator to crudely compare sunlight exposure history among cases and controls who refuse to participate, and to adjust the observed estimate of association for the selection bias.

To continue the example, the investigators might be concerned about the accuracy of subjects' self-report of history of leisure-time sunlight exposure. In particular, melanoma cases might recall or report their history of sunlight exposure differently than controls sampled from the source population. This threat to validity would be an example of measurement error (Chap. 6), which can also be addressed by quantitative bias analysis. To implement a bias analysis, the investigators would require estimates of the sensitivity and specificity of sunlight exposure classification among melanoma cases and among members of the source population. Classification error rates might be obtained by an internal validation study (e.g., comparing self-report of sunlight exposure history with a diary of sunlight exposure kept by subsets of the cases and controls) or by external validation studies (e.g., comparing self-report of sunlight exposure history with a diary of sunlight exposure kept by melanoma cases and noncases in a similar second population).

Finally, imagine that the investigator was concerned that the relation between leisure time exposure to sunlight and risk of melanoma was confounded by exposure to tanning beds. Subjects who use tanning beds might be more or less likely to have leisure time exposure to sunlight, and tanning bed use itself might be a risk factor for melanoma. If each subject's use of tanning beds was not queried in the interview, then tanning bed use would be an unmeasured confounder (Chap. 5). While tanning bed use would ideally have been assessed during the interview, it is possible that its relation to melanoma risk was only understood after the study began. To plan for bias analysis, the investigator might turn to published literature on similar populations to research the strength of association between tanning bed use and leisure time exposure to sunlight, the strength of association between tanning bed use and melanoma, and the prevalence of tanning bed use. In combination, these three factors would allow a quantitative bias analysis of the potential impact of the unmeasured confounder on the study's estimate of the association of leisure time exposure to sunlight on risk of melanoma.

In these examples, planning for quantitative bias analysis facilitates the actual analysis. Selection forces can be best quantified if the investigator plans to ask for sunlight information among those who refuse the full interview. Classification error can be best quantified if the investigator plans for an internal validation study or assures that the interview and population correspond well enough to the circumstances used for an external validation study. Unmeasured confounding can be best quantified if the investigator collects data from publications that studied similar populations to quantify the bias parameters. Table 2.2 outlines the topics to consider while planning for quantitative bias analysis. These topics are further explained in the sections that follow.

**Table 2.2** Planning for quantitative bias analysis

General tasks	Tasks for each type of bias		
	Misclassification	Selection bias	Confounding
Determine likely threats to validity	Ask whether misclassification or recall bias of any important analytic variables is a likely threat	Ask whether selection bias or loss-to-follow-up is a likely threat	Ask whether residual confounding or unmeasured confounding is likely
Determine data needed to conduct bias analysis	Internal validation study of classification rates or external validation data	Collect information on selection proportions	Collect information on prevalence of confounder and its associations with exposure and disease
Consider the population from which to collect data	Study population	Source population	Source population
Allocate resources	Develop databases that allow for recording of data, allocate time for data collection and analysis of data, write protocols for substudies, collect and understand software for analysis		
Set order of corrections	Usually first	Usually second	Usually third
Consider data level	Record-level data corrections or summary data		
Consider interrelations between biases	Important interrelations should be assessed with record-level data and multiple biases modeling		
Select a technique for bias analysis	Each method in 2.3 can be applied to each bias. Consider the number of biases to be analyzed, the interrelations between biases, the inferential question, and computational requirements		

### *Creating a Data Collection Plan for Bias Analysis*

As described in the preceding examples, planning for quantitative bias analysis during the study design will produce the most effective analyses. Investigators should examine the study design before data collection begins and ask, “What will likely be the major threats to validity once the data have been collected?” The answer will inform the plans for data collection necessary to conduct the quantitative bias analysis. If selection bias is a concern, then the investigator should collect the data required to calculate participation proportions within strata defined by the exposure, disease status, and important covariates. If classification errors are a concern, then the investigator should collect the data required to validate the study’s measurements. The validation data can be collected by an internal design or by applying validation data collected in a similar population (an external design). We provide further guidance on selecting and implementing an internal or external design in Chap. 3. If an important candidate confounder has not been measured, then the investigator should plan to use internal and external data (sometimes in combination) to estimate the impact of the unmeasured confounder.

Finally, the investigator needs to consider the population targeted for collecting the validity data that will be used in the quantitative bias analysis. For example,

confounding arises at the level of the population, so data used to correct for an unmeasured confounder should arise from the same or a similar population, but should not necessarily be limited to the population sampled to participate in the study. The study sample is included in the source population, but is rarely the entire source population. Selection bias arises from disease and exposure-dependent participation. In assessing selection bias, the exposure and disease information are available for study participants, so the information required should be collected from nonparticipants. In contrast, information bias from classification error arises within the actual study population, so the data required for assessing information bias should be collected from a subset of participants (an internal validity study) or from a population similar to the participants (an external validity study). Careful consideration of the target population will lead to a more appropriate bias analysis.

Once the major threats to validity have been ascertained, and the population from which validity data will be collected has been identified, the investigator should devise a plan for collecting the validity data. If the validity data will be external, then the investigator should conduct a systematic review of the published literature to find applicable validity studies. For example, if the investigator of the sunlight-melanoma relation is concerned about errors in reporting of sunlight exposure, then she should collect all of the relevant literature on the accuracy of self-report of sunlight exposure. Studies that separate the accuracy of exposure by melanoma cases and noncases will be most relevant. From each of these studies, she should abstract the sensitivities and specificities (or predictive values) of self-report of sunlight exposure. Some estimates might be discarded if the population is not similar to the study population. Studies of the accuracy of self-report of sunlight exposure in teenagers would not provide good external validity information for a study of melanoma cases and controls, because there would be little overlap in the age range of the teenagers who participated in the validity study and the melanoma cases and controls who participated in the investigator's study. Even after discarding the poorly applicable validity data, there will often be a range of values reported in the literature, and the investigator should decide how to best use these ranges. An average value or a preferred value (e.g., the value from the external population most like the study population) can be used with simple bias analysis, or the range can be used with multidimensional bias analysis, probabilistic bias analysis, or multiple biases modeling.

If the validity data will be internal, then the investigator should allocate study resources to conduct the data collection required for the quantitative bias analysis. If nonparticipants will be crudely characterized with regard to basic demographic information such as age and sex, so that they can be compared to participants, then the data collection system and electronic database should allow for designation of nonparticipant status and for the data items that will be sought for nonparticipants. If a validity substudy will be implemented to characterize the sensitivity and specificity of exposure, then resources should be allocated to accomplish the substudy. A protocol should be written to sample cases and controls (usually at random) to participate in the diary verification of self-reported sunlight exposure. The substudy protocol might require additional informed consent, additional recruitment materials,

and will certainly require instructions for subjects on how to record sunlight exposure in the diary and a protocol for data entry.

These examples do not fully articulate the protocols required to plan and collect the data that will inform a quantitative bias analysis. The same principles for designing well-conducted epidemiologic studies apply to the design of well-conducted validity studies. The reader is again referred to texts on epidemiologic study design, such as *Modern Epidemiology* (Rothman et al., 2008b), to research the details of valid study design. The larger point, though, is that the data collection for a validity substudy should not be underestimated. The investigator should plan such studies at the outset, should allocate study resources to the data collection effort, and should assure that the validation substudy is completed with the same rigor as applied to the principal study.

### ***Creating an Analytic Plan for a Bias Analysis***

Valid epidemiologic data analysis should begin with an analytic strategy that includes plans for quantitative bias analysis at the outset. The plan for quantitative bias analysis should make the best use of the validation data collected per the design described above.

### **Order of Bias Analysis Corrections**

When multiple sources of systematic error are to be assessed in a single study, the order of corrections in the analysis can be important. In particular, adjustments for classification errors as a function of sensitivity and specificity do not reduce to a multiplicative bias factor. The place in the order in which an adjustment for classification error will be made can therefore affect the result of the bias analysis. In general, the investigator should reverse the order in which the errors arose. Errors in classification arise in the study population, as an inherent part of the data collection and analysis, so should ordinarily be corrected first. Selection bias arises from differences between the study participants and the source population, so should ordinarily be corrected second. Confounding exists at the level of the source population, so error arising from an unmeasured confounder should ordinarily be analyzed last.

While this order holds in general, exceptions may occur. For example, if internal validation data on classification errors are used to correct for information bias, and the internal validation data were collected after participants were selected into the study population, then one would correct first for classification error and then for selection bias. Were the internal validation data collected before participants were selected into the study population, then one would correct first for selection bias and then for classification error. In short, one should follow the study design in reverse to determine the appropriate order of bias analysis. See Chap. 9 on multiple bias analysis for a more complete discussion of the order of corrections.

## Type of Data: Record-Level Versus Summary

Many of the techniques for quantitative bias analysis described herein assume that the investigator has access to record-level data. That is, they assume that the original data set with information on each subject in the study is available for analysis. Record-level data, or original data, allow for a wider range of methods for quantitative bias analysis. With record-level data, corrections for classification errors can be made at the level of the individual subjects, which preserves correlations between the study variables and allows the analyst to adjust the corrected estimates of association for other confounders. Furthermore, when multiple sources of systematic error are assessed in the analysis, applying the bias analyses in the proper order to the record-level data can easily preserve the interactions of the biases.

Some of the techniques described herein apply to summary data, or collapsed data. That is, they apply to data displayed as frequencies in summary contingency tables or as estimates of association and their accompanying conventional confidence intervals. Investigators or stakeholders with access to only summary data (e.g., a reader of a published epidemiology study) can use these techniques to conduct quantitative bias analysis. In addition, investigators with access to record-level data can generate these summary data and so might also use the techniques. However, these techniques do not necessarily preserve the interrelations between study variables and usually assume that multiple biases in an analysis are independent of one another (i.e., the biases do not interact). These assumptions are not usually testable and may often be incorrect. Investigators with access to record-level data should therefore use the analyses designed for record-level data in preference to the analyses designed for summary data.

## Selecting a Technique for Bias Analysis

Table 2.3 summarizes the analytic strategies available to accomplish a quantitative bias analysis. The first column provides the names of the bias analysis techniques. The second column explains how bias parameters are treated in the corresponding techniques. Bias parameters are the values that are required to complete the quantitative bias analysis. For example, to analyze bias due to classification errors, the sensitivity and specificity of the classification method (or its predictive values) are required. The sensitivity and specificity of classification are therefore the bias parameters of that bias analysis. The third column shows whether biases are analyzed individually or jointly for the corresponding techniques. The fourth column describes the output of the technique and the fifth column answers whether random error can be combined with the output to reflect the total error in the estimate of association. The last column depicts the computational difficulty of each technique. Note that different analytic techniques refer to a class of methods used to correct for biases, but do not refer to any particular bias. Each could be used to correct for selection bias, misclassification, or an unmeasured confounder.

**Table 2.3** Summary of quantitative bias analysis techniques

Analytic technique	Treatment of bias parameters	Number of biases analyzed	Output	Combines random error?	Computationally intensive?
Simple sensitivity analysis	One fixed value assigned to each bias parameter	One at a time	Single revised estimate of association	usually no	no
Multidimensional analysis	More than one value assigned to each bias parameter	One at a time	Range of revised estimates of association	no	no
Probabilistic analysis	Probability distributions assigned to bias parameters	One at a time	Frequency distribution of revised estimates of association	yes	yes
Multiple biases modeling	Probability distributions assigned to bias parameters	Multiple biases at once	Frequency distribution of revised estimates of association	yes	yes

Investigators designing or implementing a quantitative bias analysis should weigh three considerations as they choose the appropriate technique. When more than one bias will be analyzed, most of the analytic methods will treat them individually and/or independently. If more than one bias will be analyzed, investigators should consider using methods in the lower rows of the table. Only multiple bias modeling allows more than one bias to be analyzed simultaneously and allows the analyst to explicitly model the relationship between each of the biases. Ignoring these dependencies can produce different results than when they are taken into account. Second, investigators should consider the inferential goal, which relates most closely to the output in the foregoing table. The most common inferential goal is to adjust the estimate of association to take account of the bias. This goal can be accomplished with all of the analytic methods. Another common inferential goal is to adjust the confidence interval to reflect total error: the sum of the systematic error and the random error. This goal can be accomplished with only probabilistic bias analysis (when only one bias will be analyzed) or multiple bias modeling (when more than one bias will be analyzed). A last common inferential goal is to determine whether an estimate of association can be completely attributed to the bias. This goal requires examination of the bias from different combinations of the bias parameters, along with a determination of whether the combinations that yield a null result are reasonable. Because each combination is individually examined, and multiple combinations are required, this inferential goal is best accomplished by multidimensional bias analysis.

The third consideration is the computational difficulty of the quantitative bias analysis. Although each of the analyses can be accomplished using spreadsheets or

SAS code available on the web site (see Preface), the computational difficulty varies widely. As the computational difficulty grows, the researcher should expect to devote more time and effort to completing the analysis, and more time and presentation space to explaining and interpreting the method. In general, investigators should choose the computationally simplest technique that satisfies their inferential goal given the number of biases to be examined and whether multiple biases can be appropriately treated as independent of one another. When only one bias is to be examined, and only its impact on the estimate of association is central to the inference, then computationally straightforward simple bias analysis is sufficient. When more than one bias is to be examined, the biases are not likely independent, and an assessment of total error is required to satisfy the inferential goal, then the computationally most difficult and resource-intensive multiple bias modeling will be required.

The following paragraphs summarize each of the analytic techniques and illustrate the method with a brief example. The detailed chapters that follow show how to implement each technique and provide guidance for choosing from among the methods used to accomplish each of the techniques. That choice usually depends on the available bias parameters (e.g., the sensitivity and specificity of classification vs the positive and negative predictive values), the source of the bias parameters (i.e., internal or external validation data), and the data form (i.e., record-level or summary data).

## ***Bias Analysis Techniques***

### **Simple Bias Analysis**

With a simple bias analysis, the estimate of association obtained in the study is adjusted a single time to account for only one bias at a time. The output is a single revised estimate of association, which does not incorporate random error. For example, Marshall et al. (2003) investigated the association between little league injury claims and type of baseball used (safety baseball vs traditional baseball). They observed that safety baseballs were associated with a reduced risk of ball-related injury (rate ratio = 0.77; 95% CI 0.64, 0.93). They were concerned that injuries might be less likely to be reported when safety baseballs were used than when traditional baseballs were used, which would create a biased estimate of a protective effect. To conduct a simple bias analysis, they estimated that no more than 30% of injuries were unreported and that the difference in reporting rates was no more than 10% (the bias parameters). Their inferential goal was to adjust the estimate of association to take account of this differential underreporting. With this single set of bias parameters, the estimate of association would equal a rate ratio of 0.88. They concluded that a protective effect of the safety ball persisted after taking account of the potential for differential underreporting of injury, at least conditional on the accuracy of the values assigned to the bias parameters.

Cain et al. (2006, 2007) conducted a simple bias analysis with the inferential goal of determining whether their estimate of association could be completely



attributed to bias. Their study objective was to estimate the association between highly active antiretroviral therapy (HAART) and multiple acquired immunodeficiency syndrome (AIDS)-defining illnesses. Averaging over multiple AIDS-defining illnesses, the hazard of an AIDS-defining illness in the HAART calendar period was 0.34 (95% CI 0.25, 0.45) relative to the reference calendar period. The authors were concerned that differential loss-to-follow-up might account for the observed protective effect. They conducted a “worst-case” simple bias analysis by assuming that the 68 men lost-to-follow-up in the HAART calendar period had an AIDS-defining illness on the date of their last follow-up, and that the 16 men lost-to-follow-up in the calendar periods before HAART was introduced did not have an AIDS-defining illness by the end of follow-up. With these bounding assumptions, the estimated effect of HAART equaled a hazard ratio of 0.52. The inference is that differential loss-to-follow-up could not account for all of the observed protective effect of HAART against multiple AIDS-defining illnesses, presuming that this analysis did in fact reflect the worst case influence of this bias.

Note that in both examples, the estimate of association was adjusted for only one source of error, that the adjustment was not reflected in an accompanying interval (only a point estimate was given), and that random error was not simultaneously incorporated to reflect total error. These are hallmarks of simple bias analysis.

## Multidimensional Bias Analysis

Multidimensional bias analysis is an extension of simple bias analysis in which the analyst examines multiple values or combinations of values of the bias parameters, rather than single values. For example, Sundararajan et al. (2002) investigated the effectiveness of 5-fluorouracil adjuvant chemotherapy in treating elderly colorectal cancer patients. Patients who received 5-fluorouracil therapy had a lower rate of colorectal cancer mortality than those who did not (hazard ratio 0.66; 95% CI 0.60, 0.73). The investigators were concerned about bias from confounding by indication because the therapy assignment was not randomized. To assess the potential impact of this unmeasured confounder, they made assumptions about the range of (1) the prevalence of an unknown binary confounder, (2) the association between the confounder and colorectal mortality, and (3) the association between the confounder and receipt of 5-fluorouracil therapy (these are the bias parameters). The inferential goal was to determine whether confounding by indication could completely explain the observed protective effect. Most combinations of the bias parameters also yielded a protective estimate of association; only extreme scenarios resulted in near-null estimates of association. The range of revised estimates of association, which does not incorporate random error, is the output of the multidimensional bias analysis. The authors wrote, “Confounding could have accounted for this association only if an unmeasured confounder were extremely unequally distributed between the treated and untreated groups or increased mortality by at least 50%.” They therefore concluded that the entire protective effect could not be reasonably attributed to confounding by indication, which answered their inferential goal, at least conditional



on the accuracy of the ranges assigned as values to the bias parameters. While multidimensional bias analysis provides more information than simple bias analysis in that it provides a set of corrected estimates, it does not yield a frequency distribution of adjusted estimates of association. Each adjusted estimate of association stands alone, so the analyst or reader gains no sense of the most likely adjusted estimate of association (i.e., there is no central tendency) and no sense of the width of the distribution of the adjusted estimate of association (i.e., there is no frequency distribution of corrected estimates). Multidimensional bias analysis also addresses only one bias at a time and does not simultaneously incorporate random error, disadvantages that it shares with simple bias analysis.

### **Probabilistic Bias Analysis**

Probabilistic bias analysis is an extension of simple bias analysis in which the analyst assigns probability distributions to the bias parameters, rather than single values (as with simple bias analysis) or a series of discrete values within a range (as with multidimensional bias analysis). By repeatedly sampling from the probability distributions and correcting for the bias, the result is a frequency distribution of revised estimates of association, which can be presented and interpreted similarly to a conventional point estimate and frequentist confidence interval. Like the earlier methods, only one bias at a time is examined. For example, in a study of the association between periconceptional vitamin use and preeclamptic risk, Bodnar et al. (2006) were concerned about confounding by fruit and vegetable intake, which had not been measured. The odds ratio associating regular periconceptional use of multivitamins with preeclampsia equaled 0.55 (95% CI 0.32, 0.95). High intake of fruits and vegetables is more common among vitamin users than nonusers and also reduces the risk of preeclampsia. Bodnar et al. created a distribution of the potential relative risk due to confounding using external information about the strength of association between fruit and vegetable consumption and vitamin use, strength of association between fruit and vegetable consumption and preeclamptic risk, and prevalence of high intake of fruits and vegetables. They used Monte Carlo methods to integrate the conventional odds ratio, distribution of the relative risk due to confounding, and the random error to generate output that reflects both an adjusted point estimate and uncertainty intervals. As expected, this probabilistic bias analysis suggested that the conventional results were biased away from the null. The conventional OR (0.55) was attenuated to 0.63 (95% simulation interval: 0.56, 0.72 after accounting for only systematic error; 0.36 and 1.12 after accounting for both systematic and random error, respectively). Unlike with the previous methods, there is now a sense of the central tendency of the corrected estimate of association (0.63) and the amount of uncertainty in that estimate (as portrayed by the simulation intervals), and random error is integrated with systematic error. The bias analysis suggests that vitamin use is associated with a reduced risk of preeclampsia, even after taking account of the unmeasured confounding by fruit and vegetable intake and random error, which satisfies the inferential goal. The original

analysis somewhat overestimated the protective effect and the conventional interval somewhat underestimated the total error, at least conditional on the accuracy of the distributions assigned to the bias parameters.

## Multiple Biases Modeling

Multiple biases modeling is also an extension of simple bias analysis in which the analyst assigns probability distributions to the bias parameters, rather than single values or ranges, but now the analyst examines the impact of more than one bias at a time. For example, we conducted a case-control study of the effect of pregnancy termination (induced and spontaneous) on breast cancer risk among parous residents of Massachusetts ages 25–55 years at breast cancer diagnosis (Lash and Fink, 2004). The study included all Massachusetts breast cancer cases reported to the Massachusetts cancer registry between 1988 and 2000 arising from the population of women who gave birth in Massachusetts between 1987 and 1999. The conditional adjusted odds ratio estimate of the risk ratio of breast cancer, comparing women who had any history of pregnancy termination with women who had no history of pregnancy termination, equaled 0.91 (95% CI 0.79, 1.0). Information on history of pregnancy termination and potential confounders was recorded on birth certificates before the breast cancer diagnosis, so errors in recall or reporting of this history should have been nondifferentially and independently related to breast cancer status (Rothman et al., 2008d). It may be that the observed null result derives from nondifferential, independent misclassification of history of termination, thereby masking a truly nonnull result. In addition, the study may have been subject to a selection bias if women who migrated from Massachusetts between the time they gave birth and the time they developed breast cancer differed from those who did not migrate with respect to pregnancy terminations. The inferential goal was to adjust the estimate of association and its interval to account for these biases. We first implemented a probabilistic bias analysis with the following bias parameters: (1) a triangular distribution of sensitivity of termination classification ranging from 69% to 94% with a mode of 85%, (2) a triangular distribution of specificity of termination classification ranging from 95% to 100% with a mode of 99%, and (3) a prevalence of termination in the source population ranging from 20% to 30% with a mode of 25% (Holt et al., 1989; Werler et al., 1989; Wilcox and Horney, 1984). To allow for small deviations from perfectly nondifferential misclassification, we allowed the sensitivity and specificity of termination classification in cases, versus controls, to vary independently of one another between 0.9-fold and 1.1-fold (e.g., if the sensitivity in cases was chosen to be 85%, then the sensitivity in the controls could be no less than 76.5% and no greater than 93.5%). These were the bias parameters used to address misclassification. The probabilistic bias analysis yielded a median odds ratio estimate of 0.90 (95% simulation interval 0.62, 1.2). Conditional on the accuracy of the distributions assigned to the bias parameters, this probabilistic bias analysis (which only accounts for one source of bias) supports the notion that the result is unlikely to arise from a bias toward the null induced

by nondifferential nondependent misclassification of the dichotomous exposure variable, because the central tendency remained near null and the interval remained narrow.

The expectation was that the estimates of association should be immune to initial selection bias because study eligibility did not require active participation. However, loss-to-follow-up by migration out of Massachusetts after giving birth may have been differentially related to history of pregnancy termination and breast cancer incidence. In a previous investigation of similar design, Tang et al. (2000) used state-specific migration data to estimate that the loss-to-follow-up may have led to a 5% underestimate of the relative effect. To implement a multiple bias model, we combined the probabilistic misclassification bias analysis results above with a triangular bias parameter distribution ranging from 1 to 1.1 with mode 1.05 to account for the potential selection bias induced by migration. This multiple bias model, which accounts for the selection bias and the misclassification as well as random error, yielded a median estimate of 0.95 (95% simulation interval 0.65, 1.3; with random error incorporated). While the multiple bias analysis median estimate and the conventional point estimate are nearly identical, the width of the multiple bias model's simulation interval on the log scale is more than twice the width of the conventional 95% confidence interval, which conveys the additional uncertainty arising from the systematic errors. Taken together and conditional on the accuracy of the distributions assigned to the bias parameters, the bias analysis shows that the null result is unlikely to have arisen from misclassification of termination status or from selection bias arising from differential migration of subjects between the date of giving birth and the date of cancer diagnosis record, and that the total uncertainty is larger than reflected in the conventional confidence interval, but still not large enough to infer a nonnull result with any confidence.

### *A Note on Inference*

In the inference segment of each of the preceding examples, the inference was always said to be conditional on the accuracy of the values or distributions assigned to the bias parameters. It is, of course, impossible to know the accuracy of these assignments. Nonetheless, the analyst should think that the assignments are more accurate than the inherent assignments made to these bias parameters in a conventional data analysis (e.g., no unmeasured confounding and perfect classification). If stakeholders other than analyst support a different set of values, the bias analysis can and should be repeated with the alternate set of values to see whether the results of the bias analysis and the inference change substantially.

As will be described in Chap. 3, the assignment of values and distributions to bias parameters is equal parts art, educated guess, and science. Were the values known with certainty, then a bias analysis would not be necessary because alternate empirical methods would be superior. Circumstances such as this are rare. It is imperative, therefore, that in any bias analysis the values assigned to the bias

parameters are explicitly given, the basis for the assignment explicitly provided, and any inference resting on the results of the bias analysis explicitly conditioned on the accuracy of the assignments.

## **Conclusion**

Well-conducted epidemiologic research begins with a sound design, including valid methods of data collection and assurance that the collected data will yield a sufficiently precise and valid estimate of association. Nonetheless, even the estimate of association obtained from a well-designed study will inevitably be susceptible to residual error. Analysts conventionally calculate a confidence interval to quantify residual random error; quantitative bias analysis similarly quantifies residual systematic error. Just as investigators plan for their conventional analyses as they design their study, so too should they plan for quantitative bias analysis as they design their study. By incorporating quantitative bias analysis into all elements of study design and analysis, investigators will be best able to achieve the overarching objective of obtaining a valid and precise estimate of the effect of an exposure on the occurrence of disease.

Applying Quantitative Bias Analysis to Epidemiologic  
Data

Lash, T.L.; Fox, M.P.; Fink, A.K.

2009, XII, 192 p., Hardcover

ISBN: 978-0-387-87960-4