Epidemiologic Reviews

Vol. 43, 2021
https://doi.org/10.1093/epirev/mxab012
Advance Access publication:
October 8, 2021

Review

Monte Carlo Simulation Approaches for Quantitative Bias Analysis: A Tutorial

Hailey R. Banack*, Eleanor Hayes-Larson, and Elizabeth Rose Mayeda

* Correspondence to Dr. Hailey R. Banack, Department of Epidemiology and Environmental Health, 270 Farber Hall, University at Buffalo, Buffalo, NY 14214 (e-mail: hrbanack@buffalo.edu).

Accepted for publication October 14, 2021.

Quantitative bias analysis can be used to empirically assess how far study estimates are from the truth (i.e., an estimate that is free of bias). These methods can be used to explore the potential impact of confounding bias, selection bias (collider stratification bias), and information bias. Quantitative bias analysis includes methods that can be used to check the robustness of study findings to multiple types of bias and methods that use simulation studies to generate data and understand the hypothetical impact of specific types of bias in a simulated data set. In this article, we review 2 strategies for quantitative bias analysis: 1) traditional probabilistic quantitative bias analysis and 2) quantitative bias analysis with generated data. An important difference between the 2 strategies relates to the type of data (real vs. generated data) used in the analysis. Monte Carlo simulations are used in both approaches, but the simulation process is used for different purposes in each. For both approaches, we outline and describe the steps required to carry out the quantitative bias analysis and also present a bias-analysis tutorial demonstrating how both approaches can be applied in the context of an analysis for selection bias. Our goal is to highlight the utility of quantitative bias analysis for practicing epidemiologists and increase the use of these methods in the epidemiologic literature.

bias analysis; confounding; measurement error; misclassification; Monte Carlo sampling; selection bias; simulation study

Abbreviations: CI, confidence interval; DAG, directed acyclic graph.

INTRODUCTION

Bias is a concept familiar to all epidemiologists and a frequently discussed concern in the epidemiologic literature. The definition of bias is deceptively simple: a systematic deviation of results or inferences from the truth (1). This definition requires an understanding of 2 key concepts. First, "systematic deviation" indicates a structural threat to validity of study inferences that arises from study design or analysis of study data, rather than from randomness or chance (due to sampling) (2–4). In other words, even if the study were repeated many times with new samples drawn from the same population, or the sample size was increased to reduce the influence of chance, the deviation from the truth would remain (5, 6).

This leads to the second key concept embedded in the definition of bias: the notion that research studies are designed to estimate some true value (e.g., prevalence of a disease in a descriptive study; the relationship between exposure and outcome in an etiologic study). It is important to recognize that results we obtain from our studies are not always equivalent to the "truth," and, in most instances, we never know the truth (7). This makes it challenging to quantify how far any individual study's results deviate from the truth. Nonetheless, this is an important theoretical concept to consider. Our goal as epidemiologists is to design studies to minimize bias and random error and obtain estimates as close to the truth as possible.

Threats to internal validity of study results are due to lack of exchangeability between exposed and unexposed groups; traditionally, they are categorized as confounding, selection bias, and information bias (4, 8, 10). Directed acyclic graphs (DAGs) are a useful tool for understanding the basic structure of each of these biases (11–16). Confounding bias occurs when the effect of exposure on outcome is mixed with the effect of a third variable that is a common cause of exposure and outcome (Figure 1A) (11, 17). The term "selection bias" is used to describe several

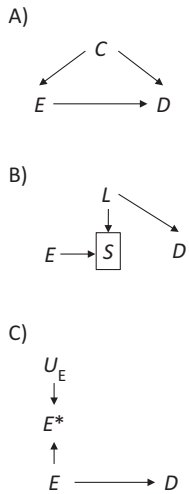

Figure 1. Directed acyclic graphs showing A) confounding bias, B) selection bias (collider stratification bias), C) information bias (nondifferential exposure misclassification). C, confounder; D, disease; E, exposure; E*, exposure as measured; L, common cause of collider (S) and disease; S, selection into sample; U_E, sources of error in measurement of exposure as measured.

concepts in epidemiology (18, 19). In this article, selection bias (Figure 1B) refers to the result of the processes by which individuals are selected into the analysis (e.g., participation in study; having complete data), in which the exposure–outcome association in a given study sample differs from the true causal exposure–outcome relationship in that sample (13, 20, 21). As Hernán et al. and others have described (13, 22, 23), this type of bias can be produced by conditioning on a variable known as a collider, a variable that is a common effect of the exposure (or a variable associated with exposure) and outcome (or a variable associated with the outcome). This is why some authors refer to selection bias as collider stratification bias or collider bias. Finally, information bias (Figure 1C) occurs when the estimated exposure–outcome relationship is distorted by mismeasurement or misclassification of exposure, outcome, or confounding variables (6, 24–26). When referring to mismeasurement of continuous variables, this type of bias is also called measurement error; when referring to incorrect classification of categorical variables, it is often called misclassification bias.

In many instances, careful study design can help minimize risk of confounding, selection, and information bias (4). Several authors have described best-practice recommendations to reduce potential for bias when designing epidemiologic studies (27–29). However, systematic sources of error often remain. Quantitative bias analysis is a class of analytic methods that can be used to assess how far study estimates may be from the "truth" (i.e., an estimate that is free of bias) (7, 30).

In this article, we review 2 strategies for quantitative bias analysis. The first approach, traditional probabilistic quantitative bias analysis, is used to examine the potential influence of systematic error on an effect estimate obtained from an analysis of real data (a check of robustness of study findings) (30). The second approach, quantitative bias analysis with generated data, uses simulation studies designed to emulate real-life data and study the magnitude of bias induced by specific scenarios created by the study investigator (31). Table 1 provides an overview comparing these approaches. In the following sections, we describe the steps for both approaches in detail and then present a bias analysis tutorial. Our goal is to make this article useful for practicing epidemiologists. In the tutorial, we discuss a worked example and provide sample code demonstrating how both approaches can be applied in the context of an analysis for selection bias.

APPROACH 1: TRADITIONAL PROBABILISTIC QUANTITATIVE BIAS ANALYSIS

The traditional approach to quantitative bias analysis involves correcting effect estimates obtained from analyses of real-world study data using bias correction equations (4, 7, 30). The parameters used in the bias-correction equations are called bias parameters; these parameters determine the direction and magnitude of bias adjustment (7, 32). The bias parameters required for the quantitative bias analysis depends on the type of bias under investigation. Bias-parameter estimates can come from 3 sources: 1) internal validation data, such as a validation subsample; 2) external validation data (e.g., population representative data); or 3) substantive knowledge or literature review (33, 34). In a traditional quantitative bias analysis, the investigator may choose to use a single estimate of each bias parameter to adjust study estimates, which is called simple or deterministic bias analysis; or to use a Monte Carlo sampling procedure to draw values of bias parameters from a probability distribution and use sampled values in the bias-correction equations, which is called probabilistic bias analysis (35, 36). In deterministic bias analysis, the investigator makes the implicit assumption that the bias parameter value(s) is known exactly. Probabilistic bias analysis incorporates uncertainty about the bias parameters and produces simulation intervals to quantify the magnitude of uncertainty in bias-adjusted effect estimates (6, 9, 36). For those interested in an introduction to deterministic bias analysis, we refer readers to the textbook by Lash and Fink (7). In this article, we focus on probabilistic approaches. Figure 2 provides a general framework for conducting probabilistic quantitative bias analysis.

Step 1. Estimating bias parameters

Different bias parameters are required for quantitative bias analysis, depending on the type of bias being considered. In this section, we describe how to calculate bias parameters using data from internal or external validation sources.

Quantitative bias analysis of an unmeasured confounder requires estimates of 1) the magnitude of the relationship between the confounder and exposure (φ) and 2) and between confounder and outcome (θ) (Figure 3). To quantify the confounder–exposure relationship, it is common to use information on the exposure-specific prevalence of the

Table 1. Comparing 2 Approaches for Quantitative Bias Analysis

Characteristic	Approach 1 Traditional Quantitative Bias Analysis	Approach 2 Bias Analysis With Generated Data
Overview of method	Use estimated bias parameters and bias correction equations to adjust naïve effect estimates from study data	Generate data to explore the magnitude of bias induced by a specific causal structure and/or analytic approach
Type of data analyzed	Real data collected from study participants	Data generated on the basis of parameters set by study investigators
Result	Bias-adjusted effect estimate and 95% simulation interval	Estimate of the magnitude of bias corresponding with the causal scenario under investigation
Comparison of interest	Bias-adjusted effect estimate vs. naïve effect estimate from study data	Estimate of exposure-outcome relationship from generated data vs. the known "truth" (i.e., true causal effect set by study investigator)
Includes Monte Carlo simulations?	Yes Bias-parameter values sampled from a probability distribution using Monte Carlo simulations	Yes Monte Carlo simulations used to generate the multiple data sets used in the bias analysis
Uses	Correct for confounding, selection bias, information bias	Examine the hypothetical empirical impact of confounding, selection bias, information bias on study results

confounder (i.e., prevalence in the exposed group and unexposed group) (7, 37). Bias parameters required for an analysis of an unmeasured confounding variable are most commonly drawn from a review of the literature or estimates from external data.

For selection bias, quantitative bias analysis parameters are estimates of sampling fractions, or selection probabilities, corresponding to the probability of selection into the study sample from the source population (32). The 4 sampling fractions (α, β, γ, δ) are calculated by dividing the number of individuals in the cells of the 2×2 table for the study sample by the number of individuals in the cells of a corresponding 2×2 table from the source population (Table 2).

To calculate bias parameters for quantitative bias analysis for misclassification requires information on the variable as classified in the analytic data set (e.g., E+, E−; D+, D−) as well as information about the true classification of the same variable on the same individuals, typically obtained from a referent or gold standard measure (7). It is then possible to create a table cross-classifying individuals according to their exposure or outcome status and calculate bias parameters (Table 3). Note how this type of 2×2 table, called a validation table, differs from a standard 2×2 table, which classifies individuals by exposure and outcome status. Bias parameters for an analysis of exposure or disease misclassification, either sensitivity and specificity (i.e., the proportion of true-positive exposures and -negative exposures identified correctly by the study measure, respectively) or positive and negative predictive value (i.e., the probability of being a true-positive or true-negative exposure, respectively, according to study measure), can be calculated from the cells of a validation 2×2 table (6). The decision to use sensitivity and specificity or positive and negative predictive values as bias parameters depends on available validation data (2, 33, 38). However, if the internal-validation substudy sample

Table 2. Calculating Sampling Fractions for Quantitative Bias Analysis of Selection Bias[a]

2×2 Table Corresponding to Data Collected From the Source Population		
	D^+	D^-
E^+	A	B
E^-	C	D
2×2 table corresponding to data collected from the study sample		
	D^+	D^-
E^+	A°	B°
E^-	C°	D°
Calculating sampling fractions		
	D^+	D^-
E^+	$\alpha = A°/A$	$\beta = B°/B$
E^-	$\gamma = C°/C$	$\delta = D°/D$

[a] E+, exposed group (E = 1); E−, unexposed group (E = 0); D+, individuals with disease (D = 1); D−, individuals without disease (D = 0).

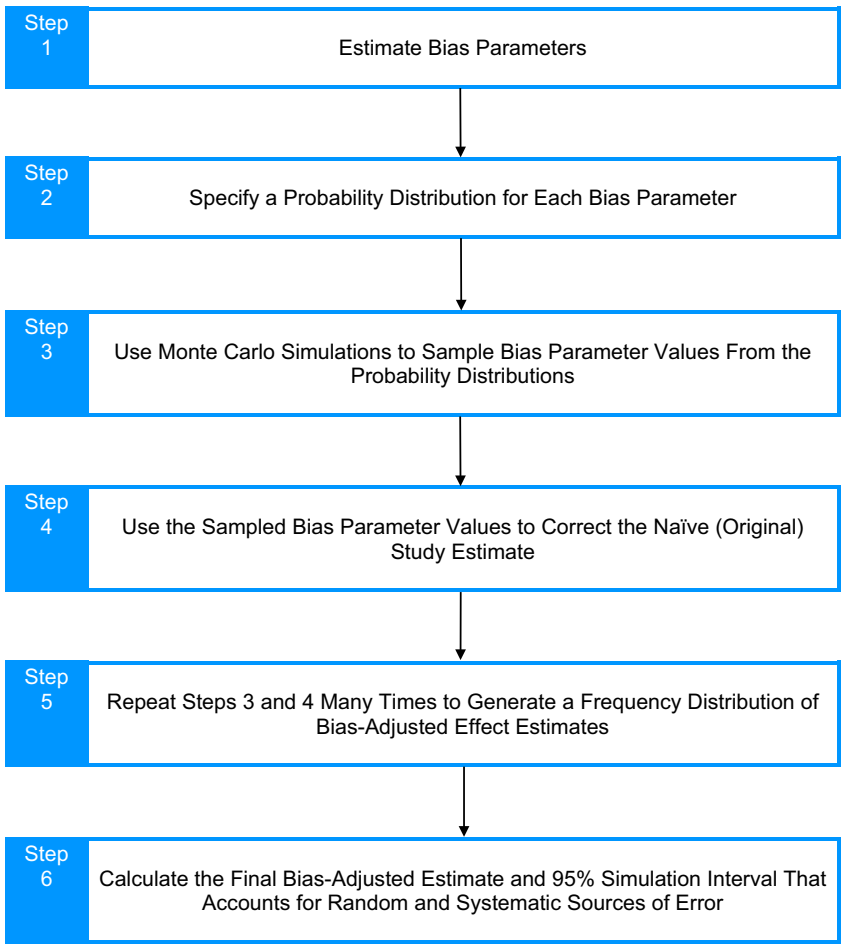

Figure 2. General framework for probabilistic quantitative bias analysis.

was selected at random from the study population, either sensitivity and specificity or positive and negative predictive value may be used in the quantitative bias analysis (36). Details on analogous procedures for continuous or polytomous variables that may be measured with error are beyond this scope of this tutorial; interested readers may consult the book by Lash et al. (7).

Step 2. Specifying probability distributions for bias parameters

In probabilistic bias analysis, uncertainty is incorporated in estimated bias-parameter values by specifying a probability distribution around each bias parameter. Consider an analysis of exposure misclassification. Using the process described previously, suppose the sensitivity was calculated as 75% and specificity as 85%. Instead of using these estimates in the bias analysis directly, a probability distribution is created for sensitivity and another for specificity. For example, it might be reasonable to use uniform distributions ranging from 65% to 85% for sensitivity and 75% to 95% for specificity. The range and type of probability distribution

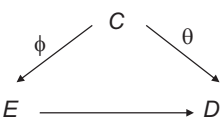

Figure 3. Calculating bias parameters for quantitative bias analysis of an unmeasured confounder. ϕ, effect of the confounder on exposure; θ, effect of the confounder on disease.

Table 3. Calculating Bias Parameters for Exposure Misclassification From a Validation 2 × 2 Table[a]

	Truly Exposed	Truly Unexposed
Classified as exposed	A	B
Classified as unexposed	C	D

[a] Sensitivity = A/(A + C); specificity = D/(B + D); positive predictive value (PPV) = A/(A + B); negative predictive value (NPV) = D/(C + D).

used in the analysis is determined by the study investigator; commonly used distributions include uniform, triangular, truncated normal, and β distributions (7, 35). Parameterization of the probability distribution should be based on substantive knowledge (27). For example, the investigator might choose a distribution with a range that encompasses sensitivity estimates from a literature review of validation studies for the measure of interest; a uniform distribution is preferred if the investigator is agnostic to the true value within the range, whereas normal or β distributions (truncated to a chosen range) would reflect an assumption about greater probability of certain values within the range.

Step 3. Sampling bias-parameter values from the probability distributions using Monte Carlo sampling

After specifying the probability distribution for a bias parameter, a specific value for the bias parameter is sampled from that probability distribution for use in bias-correction equations. Continuing the previous example, if the probability distribution for sensitivity is a uniform distribution ranging from 65% to 85%, any value within this range could be randomly chosen as the sensitivity in a particular iteration. The process of repeatedly drawing values from a probability distribution over multiple iterations is completed using Monte Carlo sampling. Monte Carlo sampling refers to a class of statistical techniques involving randomly sampling from a probability distribution (39–41). The ability to randomly sample from a probability distribution is a built-in feature of most common statistical software programs, such as Stata (e.g., *runiform[]*; SAS Institute Inc., Cary, North Carolina), R (e.g., *runif[]*; R Foundation for Statistical Computing, Vienna, Austria), and SAS (e.g., *RAND[]*; Statacorp, College Station, Texas). Regardless of software used, the underlying process remains the same: In each iteration, a value for the bias parameter is sampled from the probability distribution and then used in the bias correction equations in the next step.

Step 4. Using the sampled bias parameter values to correct the naïve (original) study estimate

In this step, the sampled bias-parameter value is used in a bias-correction equation, which incorporates information obtained from the original effect estimate with the bias parameters, to adjust the original study estimates (7, 30, 32), and produces a revised (bias-adjusted) estimate. Bias correction formulae for simple scenarios of unmeasured confounding, selection bias, and information bias are listed in Web Figure 1 (available at https://doi.org/10.1093/aje/mxab012). It is helpful to examine these formulae to gain intuition about how bias parameters are used to update original study estimates and produce bias-adjusted estimates. For example, the selection-bias formula involves correcting the estimate by multiplying by probabilities of selection into the study on the basis of exposure and disease status. Correction formulae for additional scenarios are available (4, 7); highly complex scenarios may be better suited to quantitative bias analysis with data generation (see approach 2 in this tutorial).

Step 5. Repeat steps 3 and 4 many times to generate a frequency distribution of bias-adjusted effect estimates

To obtain a distribution of bias-adjusted effect estimates, the process of sampling bias parameters from a probability distribution and correcting the original study estimates using the sampled bias parameters (steps 3 and 4) is repeated many times. Each repetition (i.e., iteration) returns a bias-corrected estimate; estimates vary due to the use of Monte Carlo sampling to draw a bias-parameter value from its probability distribution. Estimates from the multiple iterations yield a frequency distribution of bias-adjusted effect estimates, with the number of estimates matching the number of iterations. A figure showing a sample distribution is provided in the worked example in the section Quantitative Bias Analysis Tutorial.

The number of iterations required depends on the specific bias scenario being considered. There is no single correct answer for the number of iterations; investigators must balance the need to obtain a distribution of stable estimates with pragmatic concerns about computational time. In the epidemiologic literature, there are several examples of quantitative bias analysis with 100,000 iterations (38, 42) but other examples that range from 10,000 (43) to 30,000 (2) to 80,000 iterations (36).

Step 6. Final bias-adjusted estimates and 95% simulation intervals to account for random and systematic error

An important advantage of probabilistic bias analysis is that it allows for quantifying uncertainty around a final bias-adjusted effect estimate with a simulation interval (typically a 95% simulation interval), defined as the interval that contains 95% of bias-adjusted effect estimates. A simulation interval is conceptually similar to a confidence interval (CI) around a standard effect estimate but does not assume the absence of bias and, therefore, typically has better coverage of the true effect than a conventional CI (4). Estimating a 95% simulation interval that accounts for random and systematic error is a 2-step process that includes: 1) multiplying the conventional standard error (SE) for the original (uncorrected) effect estimate by a random number drawn from a standard normal distribution and 2) subtracting this product from each of the bias-adjusted effect estimates (7). This process produces an updated distribution of bias-adjusted effect estimates that incorporates random and systematic error. The 50th percentile of this distribution is the final bias-adjusted effect estimate and the 2.5th and 97.5th percentiles of the distribution are the lower and upper bounds of the 95% simulation interval.

APPROACH 2: BIAS ANALYSIS APPROACHES INVOLVING DATA GENERATION

A second approach to quantitative bias analysis involves using Monte Carlo simulations to quantify bias under different scenarios in hypothetical study data generated by the study investigator (31). In contrast to the first approach, this approach focuses on understanding the magnitude of

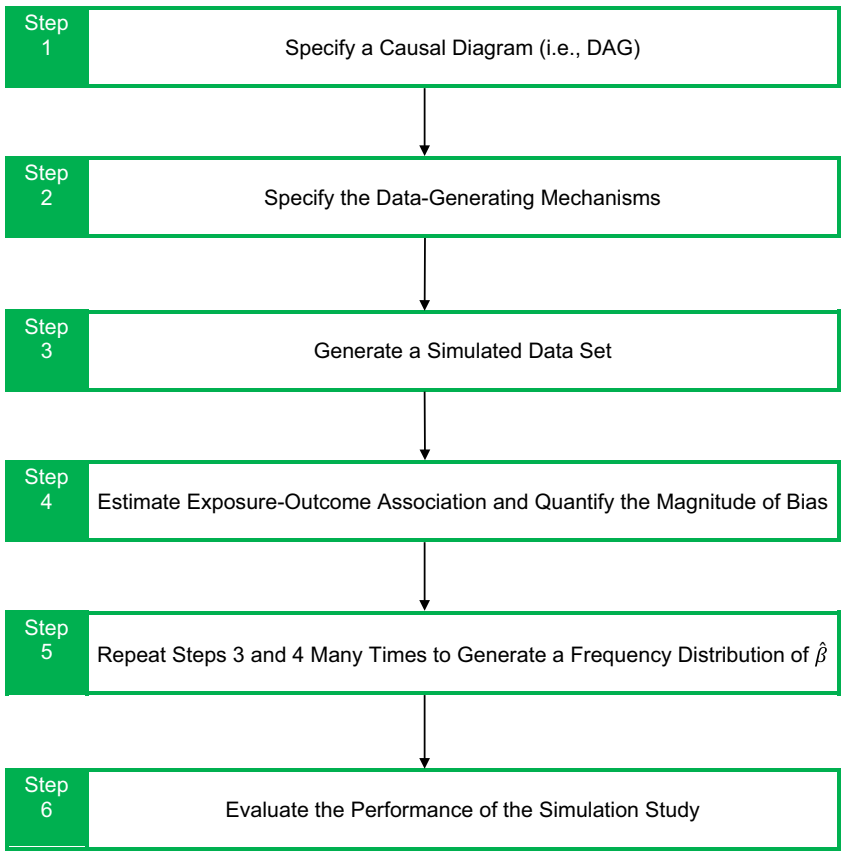

Figure 4. Framework for quantitative bias analysis from a simulated (user generated) data set with data generation under specific causal scenarios. DAG, directed acyclic graph.

bias induced by specific hypothetical causal structures and is driven by substantive knowledge of the exposure, covariate, and outcome relationships. An advantage of this approach is that it can be used to evaluate the plausibility of specific hypotheses related to bias that are difficult, or impossible, to investigate using real data (44).

Figure 4 describes the process of using Monte Carlo simulations for quantitative bias analysis in a user-generated data set, which can be applied to questions related to confounding, selection bias, and/or information bias. In the following paragraphs, we describe each step in more detail.

Step 1. Specify a causal diagram for the bias scenario of interest

The first step in designing the simulation study is creating a DAG corresponding with the causal structure of the scenario under investigation. The DAG, which encodes assumptions about variables involved in producing the outcome and the causal relationships between them, is specific to the research question of interest and source of bias of interest (a sample is provided in the worked example). The DAG is an essential precursor for generating appropriate data and must specify the exposure and outcome for the causal estimand of interest (45).

Step 2. Specify the data-generating mechanism

In the second step, we specify data-generating mechanisms (i.e., rules) corresponding with the DAG from step 1 (46). Because all the data are being generated, the investigator must specify the causal effect of exposure on outcome (the estimand of interest β) and the effects relating all the other variables in the DAG to each other and to the exposure and outcome. It is helpful to remember that these data-generating rules are intended to create a data set that mimics real data. When variables are measured in a real-life study population, the relationships between variables of interest already exist, but when variables are created in a generated data set, these relationships and expected frequencies of variables must be specified using subject matter expertise.

Data-generating rules for exogenous variables (i.e., those without any parents in the DAG (15)) are specified first as marginal means or probabilities with a distribution. Often, continuous variables are specified as being normally or uniformly distributed, and dichotomous variables as binomially distributed. Data-generating rules for endogenous variables specify conditional means or probabilities that depend on the values of their parent variables and also include a distribution (e.g., random error term or binomial

draw based on conditional probability) to ensure they are not deterministic. To mimic real data, the variable forms, frequencies, and effect sizes in the data-generating rules should be a realistic representation of those values in real data, and often the marginal and conditional means and distributions are taken directly from real data. However, investigators may still need to ensure that data-generating rules produce values consistent with data (e.g., use logistic transformations to ensure an endogenous variable is between 0 and 1; use an exponential distribution if simulating time-to-event outcomes). Importantly, in this step, the investigator sets the true magnitude of β, the estimand interest, as part of the data-generating rule for the outcome.

Step 3. Generate a simulated data set

The third step is to use a statistical software program to generate data according to the specified rules and for a specified sample size. Most standard programs, including SAS (SAS Institute, Cary, North Carolina), Stata (StataCorp, College Station, Texas), and R (R Foundation for Statistical Computing, Vienna, Austria), include functions that will facilitate the generation of simulated data set (43, 47). Although the form of the simulated data set varies, a simple data set might include 1 row per simulated person in the sample with columns for the exposure, outcome, and other variables in the DAG. For each simulated person, values are assigned using data-generating mechanisms described in step 2, first for exogenous variables and subsequently for endogenous variables. At this stage, it is important to check that the generated data (e.g., variable means and distributions) are consistent with the data-generating rules specified in step 2 to ensure the simulations are operating as expected; these checks often also reassure the investigator that the generated data set adequately mimics real data, increasing confidence that bias in the simulation should operate similarly to bias in real data. The choice of sample size may be driven by mimicking a particular empirical study or by balancing between precision and computational feasibility; simulating more individuals in the data set reduces random error and increases precision (and may reduce the number of iterations needed in step 5) but is also more computationally intensive.

Step 4. Estimate exposure–outcome association and quantify the magnitude of bias

In the next step, the generated data set is used to estimate the exposure–outcome relationship ($\hat{\beta}$) in the simulated world created with the generated data. The simulated data are analyzed as a real data set would be, although the investigator can analyze the data in multiple ways (e.g., assuming a simulated confounder was unmeasured and could not be adjusted for; assuming this confounder was measured). The $\hat{\beta}$ estimated under each set of analytic conditions can be compared with β specified by the investigator to determine the amount of error observed in the analysis of the generated data set.

Step 5. Repeat steps 3 and 4 many times to generate frequency distribution of $\hat{\beta}$

Steps 3 and 4 are repeated many times (e.g., 1,000). These iterations of sample generation produce multiple simulated data sets and, correspondingly, multiple exposure–outcome association estimates ($\hat{\beta}$) from each data set. As previously described, there is no single correct number of iterations required to produce valid estimates from a simulation study. There are formulas that can be used to guide this decision, analogous to sample size calculations used when collecting real data (40). A greater number of simulations will reduce the amount of random error in the simulation study, but this decision needs to be balanced against pragmatic concerns such as increased computational time.

After iterating the data generation and estimation (steps 3 and 4), the final magnitude of bias induced by the causal structure specified in step 1 is estimated by comparing the average of the $\hat{\beta}$ estimates from the N simulated data sets ($\bar{\hat{\beta}} = \sum_{k=1}^{N} = \frac{\hat{\beta}_k}{N}$) with β, the true causal effect of exposure on outcome (31). Bias can be assessed by measuring the absolute difference between the average estimated value ($\hat{\beta}$) from the multiple iterations of sample generation and the truth (β):

$$Bias = \bar{\hat{\beta}} - \beta. \quad (1)$$

Bias can also be represented as a percentage of the true value or as a percentage of the SE of estimated values, SE($\hat{\beta}$):

$$Percent\ bias = \left(\frac{\bar{\hat{\beta}}-\beta}{\beta}\right) \times 100 \quad (2)$$

$$Standardized\ bias = \left(\frac{\bar{\hat{\beta}}-\beta}{\mathrm{SE}\left(\hat{\beta}\right)}\right) \times 100. \quad (3)$$

The result of this simulation procedure is an estimate of $\bar{\hat{\beta}}$ and bias corresponding to the causal scenario (DAG) and data-generating rules specified in steps 1 and 2. One key advantage of using simulated data to examine bias is the flexibility it provides; it allows investigators to explore many sources of bias under different scenarios. By varying the causal structure (e.g., incorporating multiple types of bias) and data-generating rules, it is possible to compare the magnitude of bias arising from various causal structures and parameterizations. This facilitates comparison of multiple estimates of $\bar{\hat{\beta}}$ with each other, as well as with β, the true causal effect.

Step 6. Evaluate performance of the simulation study

After completing a simulation study to empirically examine sources of bias, there are several metrics that can be used to compare the simulated results with the truth, including measures of bias, variability, accuracy, and coverage (39,

40). It is good practice to examine more than 1 of these performance criteria.

Variability in the estimates from the N repeated iterations of sample generation can be assessed using the empirical SE, the standard deviation of the $\hat{\beta}$ estimates across N generated data sets (39, 46):

$$\textit{Empirical standard error} = \sqrt{\frac{1}{(N-1)}\sum_{k=1}^{N}\left(\hat{\beta}_k - \bar{\hat{\beta}}\right)^2}. \quad (4)$$

When comparing 2 simulation scenarios (e.g., scenario A: collider stratification vs. scenario B: collider stratification with exposure–collider interaction) to determine whether one induces a greater degree of bias than the other, it may also be informative to examine relative precision as a measure of variability (39). Relative precision is calculated using the estimated empirical SE of scenario A ($\widehat{EmpSE}_A$) and scenario B ($\widehat{EmpSE}_B$), as follows:

$$\textit{Relative precision}\,(\%) = \left[\left(\frac{\widehat{EmpSE}_A}{\widehat{EmpSE}_B}\right)^2 - 1\right] \times 100 \quad (5)$$

Measures of accuracy incorporate both bias and variability to assess the results of the simulation. The mean squared error (MSE in equation 6) is a commonly used measure that is calculated from the sum of the squared bias and variance of $\bar{\hat{\beta}}$ (40):

$$MSE = \left(\bar{\hat{\beta}} - \beta\right)^2 + \left(\mathrm{SE}\left(\hat{\beta}\right)\right)^2 \quad (6)$$

Coverage assessment measures the proportion of times the CI for $\hat{\beta}$ contains the true causal effect, β (40). Coverage depends on both bias and precision; greater bias and more precise confidence limits yield reduced coverage, whereas imprecision in confidence limits leads to higher coverage. For a 95% CI around an unbiased estimate, coverage should be approximately 95%.

QUANTITATIVE BIAS ANALYSIS TUTORIAL

In the previous sections, we described the steps for carrying out a probabilistic quantitative-bias analysis and quantitative-bias analysis with generated data. We now present an example focused on selection bias to illustrate both of these approaches. Sample code for these analyses is available on Github (https://github.com/Mayeda-Research-Group/MC-QBA-example).

Selection-bias example: osteoporosis and dementia risk

Recall, selection bias can occur when the process of selecting study participants leads to a statistical association in the study sample that is a biased estimate of the causal effect in the population of interest (11, 48). Numerous selection processes can give rise to selection bias, including who is invited to participate in the study, who agrees to participate, who remains in the study over time, and who has complete data (22, 45, 49, 50). These selection processes can threaten external validity, wherein study results do not generalize to a specific population of interest, and/or threaten internal validity, when study results are spurious associations that do not represent causal effects for any population, including the individuals in the study sample.

Case study:

> *A research team decides to study the relationship between osteoporosis and dementia incidence in a cohort of older adults (>65 years of age). They are interested in the following causal question: "Does osteoporosis influence dementia incidence?"*

To answer this question, the team collected data from 1,600 older adults from a local orthopedic follow-up clinic of patients recently discharged from the hospital after a hip fracture and who agreed to participate in this research study. Given their interest in studying osteoporosis, recruiting from an orthopedic clinic was a sensible choice. The exposure (A) in this analysis is a binary exposure variable with 2 levels: diagnosed osteoporosis (A = 1) and no osteoporosis (A = 0). The outcome is also a binary variable: incident dementia (Y = 1) and no dementia (Y = 0). Web Figure 2 summarizes the data from this study in a 2 × 2 table. The crude odds ratio (OR) for the exposure-outcome relationship is (120 × 632)/(677 × 171) = 0.66 (95% CI: 0.50, 0.85). On the basis of this result, it appears as though individuals with osteoporosis have 33% lower odds of incident dementia than those without osteoporosis.

All of the older adults in this study were patients of a network of local primary care physicians. As a follow-up to the first analysis, the investigators decided to examine the same relationship between osteoporosis and incident dementia in a community sample of older adults (n = 5,000). For this study, they recruited participants regardless of history of hip fracture. The crude OR for the exposure–outcome relationship from this larger study is (147 × 3,160) / (1,353 × 340) = 1.01 (95% CI: 0.81, 1.24), demonstrating a null association between these 2 variables (Web Figure 2).

The investigators were perplexed by these findings. How could osteoporosis look strongly protective among individuals who had recently experienced a hip fracture but completely null in a larger sample of older adults from the same practice group? They discussed several potential mechanistic, biological, and physiological explanations, but the epidemiologist on the team also suggested the possibility of a methodological explanation: selection bias.

The epidemiologist explained that their study design could have (inadvertently) created selection bias. She drew out a DAG to help them understand this better (Figure 5). By recruiting 1,600 study participants from an orthopedic clinic and including only those with a recent hip fracture, the team conditioned the results on a variable (hip fracture) that is a consequence of the exposure, osteoporosis, and shares common causes with the outcome, dementia. Individuals

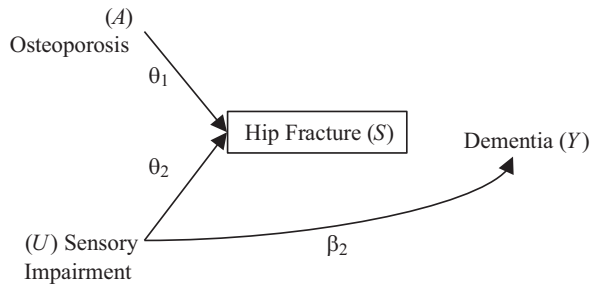

Figure 5. Directed acyclic graph representing the causal structure for the relationship between osteoporosis and dementia.

with osteoporosis are at a substantially increased risk of hip fracture (51). One potential common cause of hip fracture and dementia is sensory impairment, defined as hearing and/or vision loss, which has been associated with both hip fracture and dementia risk (52, 53). Thus, by recruiting patients with a recent hip fracture, the investigators unintentionally conditioned on a collider and induced selection bias, which did not exist in the primary care practice sample. To thoroughly explore the impact of this design decision on their results and attempt to correct their findings for selection bias, the investigators decided to undertake a probabilistic quantitative bias analysis.

APPROACH 1: PROBABILISTIC QUANTITATIVE BIAS ANALYSIS

The investigators first estimated the bias parameters required for this analysis, sampling fractions, by dividing the number of individuals in the cells of the 2 × 2 table in the selected population (those with hip fracture; Web Figure 2) by the number of individuals in the cells of the 2 × 2 table in the source population (those recruited from the primary care practice group; Web Figure 2). The resulting sampling fractions, also reported in Web Figure 2, were $\alpha = 0.82$, $\beta = 0.50$, $\gamma = 0.50$, and $\delta = 0.20$. Rather than using these sampling fractions directly in the correction equation for selection bias, the investigators created a uniform probability distribution for each of the bias parameter values. The probability distributions ranged from 10% lower to 10% higher than the calculated bias parameters: the distribution for α ranged from 0.72 to 0.92, the distributions for β and γ each ranged from 0.40 to 0.60, and the distribution for δ ranged from 0.10 to 0.30. The investigators used Monte Carlo sampling to draw a value for each of the bias parameters from these uniform distributions and used the sampled values for α, β, γ, and δ to adjust the "crude" OR estimate of 0.66, using the formula in Web Figure 1 (32).

The process of sampling bias-parameter values and adjusting the crude OR was repeated 10,000 times. The result of this process was 10,000 bias-adjusted ORs, represented, via histogram, as a distribution of bias-adjusted ORs (OR = 0.99; 95% CI: 0.56, 2.11; Web Figure 3). Finally, the investigators accounted for random and systematic sources of error and calculated a 95% simulation interval. Using the data from the cells of the 2 × 2 table from the original analysis (Web Figure 2), they calculated the SE for the original estimate:

$$SE\left(\log\left(\mathrm{OR}_{\mathrm{crude}}\right)\right) = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} = \sqrt{\frac{1}{120}+\frac{1}{677}+\frac{1}{171}+\frac{1}{632}} = 0.13 \quad (7)$$

They simulated total error ($\mathrm{OR}_{\mathrm{total\ error}}$), including both random and systematic sources of error, by randomly drawing a standard normal deviate ($\mathrm{random}_{0,1}$), multiplying it by the SE estimate of 0.13, and then subtracting the product from the 10,000 bias-adjusted OR estimates:

$$\mathrm{OR}_{\mathrm{total\ error}} = \left(\mathrm{OR}_{\mathrm{adj}}\right) - \left(\mathrm{SE}\left(\log\left(\mathrm{OR}_{\mathrm{crude}}\right)\right) \times \mathrm{random}_{0,1}\right) \quad (8)$$

The results of this quantitative bias analysis are illustrated in Web Figure 3. The final, bias-adjusted effect estimate from this simulation process is 1.01 (95% SI: 0.53, 2.21), which is nearly identical to the estimated OR value obtained in the larger study population.

Thus, the original OR value obtained in the study sample of 1,600 older adults, an OR of 0.66 (95% CI: 0.50, 0.85), a result of collider stratification, was not valid. This is clear when the results are compared with either the results from the larger study sample (n = 5,000) from the source population (1.01; 95% CI: 0.81, 1.24) or the results from the probabilistic quantitative bias analysis (1.01; 95% SI: 0.53, 2.21).

APPROACH 2: QUANTITATIVE BIAS ANALYSIS WITH GENERATED DATA

To further explore the influence of selection-bias mechanism on their study results, the investigators decided to also conduct a simulation study. The data generated in the simulation study was intended to mimic the real data they collected in their study, based on the DAG in Figure 5. Given their results, demonstrating a null effect of osteoporosis on dementia risk in the unselected sample, they simulated data under the sharp null hypothesis, a situation in which there is no causal effect of osteoporosis on incident dementia risk, and no arrow from osteoporosis to dementia, as shown in Figure 5.

The investigators specified the data-generating model corresponding to the DAG in Figure 5. When generating the simulated data, each "person" was assigned an exposure (0 if no osteoporosis, 1 if diagnosed with osteoporosis) from a Bernoulli distribution with marginal probability $P = 0.2$; continuous sensory impairment scores (U) were assigned from a standard normal distribution ($U \sim N(0,1)$). Data for hip fracture, a binary variable, were generated in a 2-step process, first calculating the conditional

Table 4. Simulated Sample Data from 1 Iteration of Data Generation Using Data-Generating Rules Described in the Case Study for Approach 2: Quantitative Bias Analysis With Generated Data

Patient ID No.	Osteoporosis	Sensory Impairment Score	Hip Fracture	Dementia
1	0	−0.630	0	0
2	1	−1.018	0	0
3	0	−0.147	0	0
...				
5,000	1	−1.420	0	0

Abbreviations: ID, identification.

probability of hip fracture using the expit transformation (expit(x) = $e^x/[1 + e^x]$) to ensure values stayed between 0 and 1:

$$\begin{aligned} &\mathrm{P}(\text{hip fracture} \mid \text{osteoporosis, sensory impairment}) \\ &= expit(\gamma_0 + \gamma_1 osteoporosis + \gamma_2 sensory\ impairment \\ &\quad + \gamma_3 osteoporosis \times sensory\ impairment.) \end{aligned}$$

Hip fracture (1 or 0) was determined by a draw from a Bernoulli distribution with probability of success equal to P(hip fracture | osteoporosis, sensory impairment).

A similar process was used to generate the binary outcome variable, incident dementia, conditional on only sensory impairment (because of the assumed sharp null relationship between osteoporosis and dementia):

$$\begin{aligned} &\mathrm{P}(\text{dementia} \mid \text{sensory impairment}) \\ &= expit\,(\beta_0 + \beta_1 sensory\ impairment) \end{aligned}$$

Dementia (1 or 0) was generated by a draw from a Bernoulli distribution with probability of success equal to p(dementia | sensory impairment).

With these data-generating rules, the team created a simulated data set with 5,000 observations and 5 variables: participant identification, osteoporosis, sensory impairment score, hip fracture (0/1), and dementia (0,1). A snapshot of the simulated data is presented in Table 4. Once the data were created, the investigators verified that the data were generated as intended by checking the means and distributions of the newly created variables for consistency with the data-generating rules they specified. The mean (standard deviation) values were as follows: osteoporosis: 0.20 (0.40); sensory impairment: 0.012 (1.02); hip fracture: 0.22 (0.42); dementia: 0.09 (0.30).

Finally, the team calculated the exposure–outcome OR in the generated data using simple logistic regression. Using the full (unselected) sample from the generated data, they estimated the OR for the effect of osteoporosis on risk of incident dementia as 1.04 (95% CI: 0.83, 1.31). They then examined the exposure–outcome relationship only among individuals who experienced hip fracture (hip fracture =1), and observed an OR of 0.69 (95% CI: 0.51, 0.94).

Satisfied that the data generating process and analysis code were correct, they repeated this process of sample generation and calculating effect estimates 1,000 times, resulting in 1,000 OR estimates from 1,000 simulated

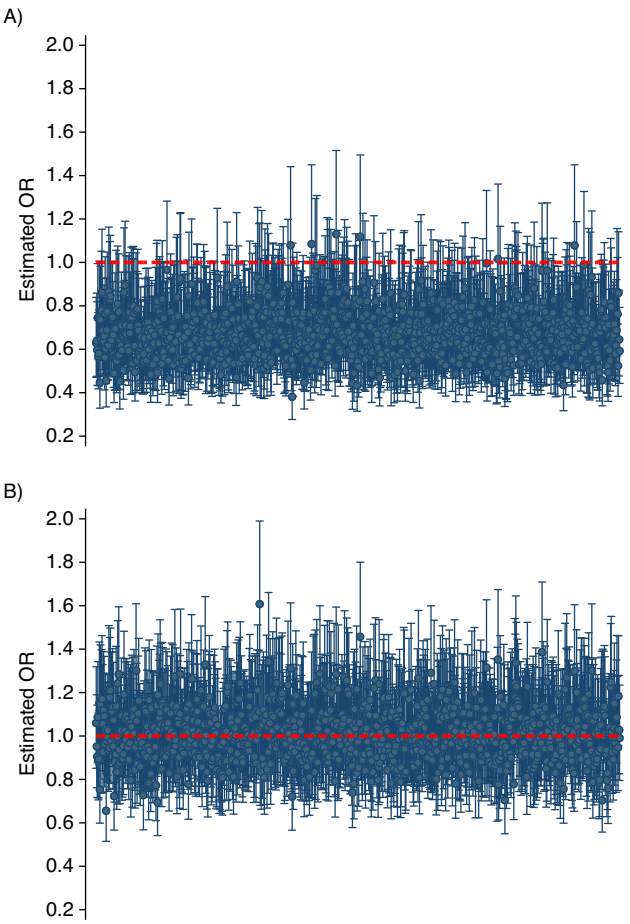

Figure 6. Estimated odds ratios (ORs) and 95% confidence intervals from (A) 1,000 simulated samples in the total population (mean OR = 1.00) and (B) among those with hip fracture (mean OR = 0.66). Dashed line indicates null value (OR = 1.00).

data sets, each with 5,000 simulated participants. Results from these iterations are plotted in Figure 6. In the total population, the estimated mean OR ($\bar{\hat{\beta}}$) was 1.00 and the 95% CI coverage was 0.94, so 94% of the 95% CIs include the true causal effect (exp(β)=1.00). Among those who had recently experienced a hip fracture, the estimated mean OR was 0.66 and the CI coverage was 0.24, so only 24% of the 95% CIs include the true causal effect (exp(β) = 1.00). These results are consistent with the data shared by the research team from their real-world data: There was no effect of osteoporosis on dementia in the community sample, but a spurious protective association due to collider bias was observed among individuals who recently experienced a hip fracture.

CONCLUSIONS AND PRACTICAL TIPS

We discussed 2 approaches for quantitative bias analysis, one using bias parameters to adjust effect estimates in real data and the other using generated data to explore the potential magnitude of bias induced under specific hypothetical scenarios. The step-by-step instructions and tutorial with sample code are intended to provide epidemiologists with a sufficient level of detail to apply these strategies to their own work. As Lash et al. (27) discuss, a key barrier to implementation and uptake of quantitative bias analysis is lack of training in the practice of bias analysis.

One of our objectives for this review paper was to provide specific how-to guidance to encourage epidemiologists to use these strategies more widely. To accomplish this objective, we focused on simple scenarios correcting for 1 form of bias at a time. However, bias is rarely unidimensional, and the approaches we discussed can be adapted to more complex scenarios with multiple types and sources of bias (4, 7). These approaches can become complex quickly, and perhaps this is one reason why more investigators do not routinely use these approaches. In our experience, it is often helpful to start with a simple scenario and gradually add more complexity (e.g., multiple bias types; variable interactions). This stepwise approach may help clarify the independent contributions of multiple bias sources and reduce errors in implementation. As with any type of analysis, learning how to implement quantitative bias analysis takes practice. However, we believe it is a worthwhile investment of time and effort. An increase in the number of epidemiologists using these strategies would result in important methodological advancements in our field.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology and Environmental Health, School of Public Health and Health Professions, University at Buffalo, Buffalo, New York, United States (Hailey R. Banack); and Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, California, United States (Eleanor Hayes-Larson, Elizabeth Rose Mayeda).

This work was supported by the National Institutes of Health, National Institute on Aging (grant R00AG053410).

Conflict of interest: none declared.

REFERENCES

1. Miquel P. *A Dictionary of Epidemiology*. 6th ed. New York, New York: Oxford University Press; 2014.
2. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology*. 2003;14(4):451–458.
3. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology*. 2003;14(4):459–466.
4. Rothman KJ, Greenland S, Lash TL. Chapter 9. Validity in epidemiologic studies. In: *Modern Epidemiology*. Philadelphia, PA: Lippincott, Williams & Wilkins; 2008: 128–147.
5. Greenland S. Multiple-bias modelling for analysis of observational data. *J R Stat Soc A Stat Soc*. 2005;168(2): 267–306.
6. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2005;34(6):1370–1376.
7. Lash T, Fox M, Fink A. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York, NY: Springer; 2009.
8. Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004;58(8):635–641.
9. Rothman KJ, Greenland S, Lash T. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
10. Edwards JK, Cole SR, Westreich D. All your data are always missing: incorporating bias due to measurement error into the potential outcomes framework. *Int J Epidemiol*. 2015;44(4): 1452–1459.
11. Hernán MA, Robins JM. *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC; 2019.
12. Hernán MA, Cole SR. Invited commentary: causal diagrams and measurement bias. *Am J Epidemiol*. 2009;170(8): 959–962.
13. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5): 615–625.
14. Hernández-Díaz S, Schisterman EF, Hernán MA. The birth weight "paradox" uncovered? *Am J Epidemiol*. 2006;164(11): 1115–1120.
15. Glymour MM. Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS, eds. *Methods in Social Epidemiology*. Hoboken, NJ: Jossey-Bass/Wiley; 2006:393–428.
16. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37–48.
17. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci*. 1999;14(1):29–46.
18. Hernán MA. Invited commentary: selection bias without colliders. *Am J Epidemiol*. 2017;185(11):1048–1050.
19. Schwartz S, Campbell UB, Gatto NM, et al. Toward a clarification of the taxonomy of "bias" in epidemiology textbooks. *Epidemiology*. 2015;26(2):216–222.
20. Glymour MM, Mayeda ER, Selby VN. Selection bias in clinical epidemiology causal thinking to guide

patient-centered research. *Epidemiology*. 2016;27(4):466–468.
21. Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol*. 2010;39(2):417–420.
22. Flanders DW, Eldridge RC, McClellan W. A nearly unavoidable mechanism for collider bias with index-event studies. *Epidemiology*. 2014;25(5):762–764.
23. Luque-Fernandez MA, Schomaker M, Redondo-Sanchez D, et al. Educational note: paradoxical collider effect in the analysis of non-communicable disease epidemiological data: a reproducible illustration and web application. *Int J Epidemiol*. 2018;48(2):640–653.
24. Wacholder S. When measurement errors correlate with truth: surprising effects of nondifferential misclassification. *Epidemiology*. 1995;6(2):157–161.
25. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol*. 1991;134(10):1233–1246.
26. Greenland S. Statistical uncertainty due to misclassification: implications for validation substudies. *J Clin Epidemiol*. 1988;41(12):1167–1174.
27. Lash TL, Fox MP, MacLehose RF, et al. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969–1985.
28. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidimiology*. 2008;19(6):766–779.
29. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758–764.
30. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol*. 1996;25(6):1107–1116.
31. Mayeda ER, Tchetgen Tchetgen EJ, Power MC, et al. A simulation platform for quantifying survival bias: an application to research on determinants of cognitive decline. *Am J Epidemiol*. 2016;184(5):378–387.
32. Kleinbaum DG, Morgenstern H, Kupper LL. Selection bias in epidemiologic studies. *Am J Epidemiol*. 1981;113(4):452–463.
33. Fox MP, Lash TL, Bodnar LM. Common misconceptions about validation studies. *Int J Epidemiol*. 2020;49(4):1392–1396.
34. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *Am J Epidemiol*. 1993;137(11):1251–1258.
35. Orsini N, Bellocco R, Bottai M, et al. A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. *Stata J*. 2008;8(1):29–48.
36. Banack HR, Stokes A, Fox MP, et al. Stratified probabilistic bias analysis for body mass index-related exposure misclassification in postmenopausal women. *Epidemiology*. 2018;29(5):604–613.
37. Arah OA. Bias analysis for uncontrolled confounding in the health sciences. *Annu Rev Public Health*. 2017;38:23–38.
38. Ahrens K, Lash TL, Louik C, et al. Correcting for exposure misclassification using survival analysis with a time-varying exposure. *Ann Epidemiol*. 2012;22(11):799–806.
39. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074–2102.
40. Burton A, Altman DG, Royston P, et al. The design of simulation studies in medical statistics. *Stat Med*. 2006;25(24):4279–4292.
41. Metropolis N, Ulam S. The Monte Carlo method. *J Am Stat Assoc*. 1949;44(247):335–341.
42. Radin RG, Rothman KJ, Hatch EE, et al. Maternal recall error in retrospectively reported time-to-pregnancy: an assessment and bias analysis. *Paediatr Perinat Epidemiol*. 2015;29(6):576–588.
43. Jurek AM, Greenland S, Maldonado G, et al. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol*. 2005;34(3):680–687.
44. Mayeda ER, Binder A, Kobayashi LC. Life course epidemiology and analysis. In: Roger D, Martin G, Quarraisha Abdool K, et al., eds. *Oxford Textbook of Global Public Health*. Oxford, UK: Oxford University Press; 2015.
45. Mayeda ER, Banack HR, Bibbins-Domingo K, et al. Can survival bias explain the age attenuation of racial inequalities in stroke incidence?: a simulation study. *Epidemiology*. 2018;29(4):525–532.
46. Mayeda ER, Filshtein TJ, Tripodis Y, et al. Does selective survival before study enrolment attenuate estimated effects of education on rate of cognitive decline in older adults? A simulation approach for quantifying survival bias in life course epidemiology. *Int J Epidemiol*. 2018;47(5):1507–1517.
47. Lukácsy K. Generating random samples from user-defined distributions. *Stata J*. 2011;11(2):299–304.
48. Westreich D, Edwards JK, Lesko CR, et al. Target validity and the hierarchy of study designs. *Am J Epidemiol*. 2018;188(2):438–443.
49. Banack HR, Kaufman JS, Wactawski-Wende J, et al. Investigating and remediating selection bias in geriatrics research: the selection bias toolkit. *J Am Geriatr Soc*. 2019;67:1970–1976.
50. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. *Epidemiology*. 2012;23(1):119–128.
51. Johnell O, Kanis J. Epidemiology of osteoporotic fractures. *Osteoporos Int*. 2005;16(2):S3–S7.
52. Hwang PH, Longstreth WT Jr, Brenowitz WD, et al. Dual sensory impairment in older adults and risk of dementia from the GEM study. *Alzheimers Dement (Amst)*. 2020;12(1):e12054.
53. Grue EV, Kirkevold M, Mowinchel P, et al. Sensory impairment in hip-fracture patients 65 years or older and effects of hearing/vision interventions on fall frequency. *J Multidiscip Healthc*. 2008;2:1–11.