

ANNUAL
REVIEWS **Further**

Click [here](#) to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

Bias Analysis for Uncontrolled Confounding in the Health Sciences

Onyebuchi A. Arah

Department of Epidemiology, Fielding School of Public Health; UCLA Center for Health Policy Research; and California Center for Population Research, University of California, Los Angeles, California 90095; email: arah@ucla.edu

Annu. Rev. Public Health 2017. 38:23–38

First published online as a Review in Advance on January 6, 2017

The *Annual Review of Public Health* is online at publhealth.annualreviews.org

<https://doi.org/10.1146/annurev-publhealth-032315-021644>

Copyright © 2017 Annual Reviews. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 (CC-BY-SA) International License, which permits unrestricted use, distribution, and reproduction in any medium and any derivative work is made available under the same, similar, or a compatible license. See credit lines of images or other third-party material in this article for license information.



Keywords

bias modeling, causal inference, probabilistic bias analysis, Monte Carlo sensitivity analysis, quantitative methodology, unmeasured confounders

Abstract

Uncontrolled confounding due to unmeasured confounders biases causal inference in health science studies using observational and imperfect experimental designs. The adoption of methods for analysis of bias due to uncontrolled confounding has been slow, despite the increasing availability of such methods. Bias analysis for such uncontrolled confounding is most useful in big data studies and systematic reviews to gauge the extent to which extraneous preexposure variables that affect the exposure and the outcome can explain some or all of the reported exposure-outcome associations. We review methods that can be applied during or after data analysis to adjust for uncontrolled confounding for different outcomes, confounders, and study settings. We discuss relevant bias formulas and how to obtain the required information for applying them. Finally, we develop a new intuitive generalized bias analysis framework for simulating and adjusting for the amount of uncontrolled confounding due to not measuring and adjusting for one or more confounders.

INTRODUCTION

Observational studies play a central role in the health sciences (47–49). They are used for etiologic research, prediction research (e.g., to identify high-risk groups), prognostic research, and diagnostic research. Observational studies are becoming increasingly important for causal analysis given the practical and ethical costs of conducting randomized trials coupled with the increasing availability of secondary big data (1, 24, 25, 27). Both clinical and public health researchers rely heavily on observational studies. With the advent of big data, large observational studies are becoming common, and their sample size reduces the importance of sampling error or random variation to a secondary role relative to systematic error or bias. Uncontrolled confounding is one crucial source of systematic error or bias. It arises when variables that are not mediators of the effect under study, and that can explain part or all of the observed association between the study exposure and the outcome, are not measured and controlled for during study design or analysis (42, 49). For an unmeasured variable that is not a mediator (or not a consequence of the exposure more generally) to lead to uncontrolled confounding, once measured confounding variables are already adjusted for, the unmeasured variable must either: (a) be a cause of the outcome through a pathway other than the exposure and also be associated with the exposure, or (b) be a cause of the exposure and be associated with the outcome through a pathway other than the exposure. These criteria subsume a scenario in which the unmeasured variable is a common cause of the exposure and outcome (35, 42, 65).

Scope

This review focuses on uncontrolled confounding in studies that consider the total effect of one or more interventions (3, 29, 49, 63). The methods discussed here can also be used in mediation settings for exposure-outcome confounding only. Nonetheless, we refer the reader to the relevant literature for detailed considerations on bias analysis for uncontrolled mediator-outcome confounding (60–62). Bias analysis for unmeasured confounders under interaction analysis (64) is also not covered here. Specific methods for use in survival settings also exist and are the subject of ongoing research (26, 33, 61), but they are not discussed here. Also not covered here is the special case of uncontrolled confounding in multilevel or mixed model settings (10, 32, 39). Throughout, we focus on study settings in which the exposure-outcome association or effect is quantified using risk difference, mean difference, risk ratio (as in cohort studies), or odds ratio (as in case-control studies).

Definitions, Notation, and Assumptions

Let C , X , and Y stand respectively for measured confounder(s), exposure (or treatment), and outcome. Unless stated otherwise, we assume binary C , X , and Y , although the methods apply more generally. We use U to denote an unmeasured confounding variable (or a set of such variables). U may or may not be a known confounder; herein we only assume that it is unmeasured and uncontrolled in ensuing analysis. Capital and small letters are used to represent variables and their realized values, respectively. The reference levels of U , C , and X are denoted by u^* , c^* , and x^* . Let Y_x denote the potential value of the outcome when exposure X is set to x , perhaps contrary to fact.

Uncontrolled confounding is assumed to be present if Y_x is independent of X given both C and U but not given C only (41, 63, 66). That is, there is at least an open backdoor between X and Y that is not closed by conditioning on C but could be closed by conditioning on U had U been measured. See **Figures 1–3** for directed acyclic graphs (DAGs) that depict data-generating processes or causal structures whereby U and C confound the effect of X on Y . [There are now accessible introductions and detailed resources on DAGs (22, 42, 49).] **Figures 1–3** report information regarding U in relation to the observed X , Y , and possibly C ; omitting U leads to uncontrolled confounding and bias analysis using some of the methods described here. **Figure 4** represents an intractable scenario

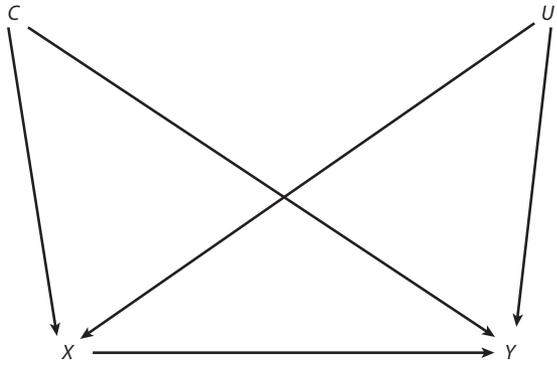


Figure 1

Directed acyclic graph (DAG) in which the exposure X and outcome Y share two common causes or confounders, namely a measured variable C and an unmeasured variable U .

in which measuring and controlling for U in addition to C does not allow us to estimate the effect of X on Y without bias. In **Figure 4**, conditioning on U would have controlled for confounding by U (via the path $X \leftarrow U \rightarrow Y$) but would have introduced a new bias (called collider-stratification bias) by opening up the colliding arrowheads at U ($X \leftarrow [U] \leftarrow Y$) (22, 42). Estimation of the total effect of X on Y in **Figure 4** requires additional measurements beyond C and U to fully control for confounding and eliminate any collider-stratification bias introduced by conditioning on U .

CONSEQUENCES OF UNCONTROLLED CONFOUNDING

Before considering how to adjust for the bias left by an unmeasured confounder U , it can be instructive to see how not controlling for U leads to bias in more than just the effect of the exposure. Suppose we have study data on exposure X , outcome Y , and confounder C from the underlying causal structure in **Figure 1**. Were U known and had it been measured, it could have been used by an investigator to specify the following model and estimate the conditional risk difference for the effect of X on Y , $\alpha_X + \alpha_{XC}c$, when $C = c$:

$$E(Y | x, c, u) = \alpha_0 + \alpha_X x + \alpha_C c + \alpha_{XC} xc + \alpha_U u.$$

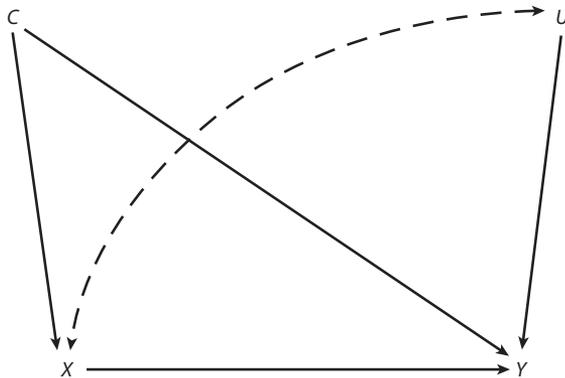


Figure 2

Directed acyclic graph (DAG) in which the exposure X and outcome Y share a measured common cause or confounder C and an unmeasured confounder U that is a cause of Y but is only associated with X through an unmeasured common cause depicted by a bidirectional dashed arrow.

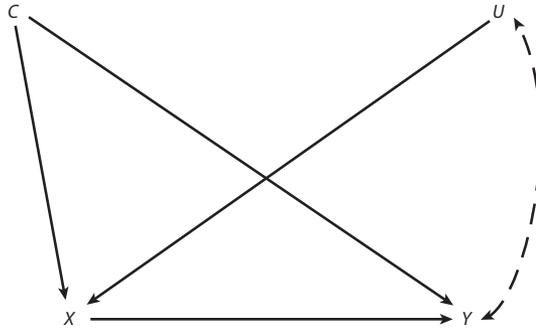


Figure 3

Directed acyclic graph (DAG) in which the exposure X and outcome Y share a measured common cause or confounder C and an unmeasured confounder U that is a cause of X but is only associated with Y through an unmeasured common cause depicted by a bidirectional dashed arrow.

In the absence of U , the investigator ends up fitting the following model in which ϕ_x is a biased estimate of the conditional risk difference for the conditional total effect of X on Y given C :

$$\begin{aligned}
 E(Y|x,c) &= \sum_u E(Y|x,c,u)P(u|x,c) \\
 &= \alpha_0 + \alpha_X x + \alpha_C c + \alpha_{XC}xc + \alpha_U(\lambda_0 + \lambda_X x + \lambda_C c + \lambda_{XC}xc) \\
 &= \alpha_0 + \alpha_U \lambda_0 + (\alpha_X + \alpha_U \lambda_X)x + (\alpha_C + \alpha_U \lambda_C)c + (\alpha_{XC} + \alpha_U \lambda_{XC})xc \\
 &= \phi_0 + \phi_X x + \phi_C c + \phi_{XC}xc
 \end{aligned}$$

Here, the model relating the unmeasured confounder U to the exposure X and measured confounder C is given by $E(U|x,c) = \lambda_0 + \lambda_X x + \lambda_C c + \lambda_{XC}xc$. In this model (assuming a binary U), $\lambda_X x + \lambda_{XC}xc$ is the risk difference in U due to one-unit change in the exposure X (that is, moving from $X = 0$ to $X = 1$ for binary X or moving from $X = x^*$ to $X = x$ for any X) conditional on $C = c$. From the model for Y given X and C only, it can be seen that ϕ_X —the coefficient of X or the estimate of the conditional association between X and Y —in the model adjusting for C only is biased for the true effect α_X : in this case, $\phi_X = \alpha_X + \alpha_U \lambda_X$. Similarly, the coefficient of

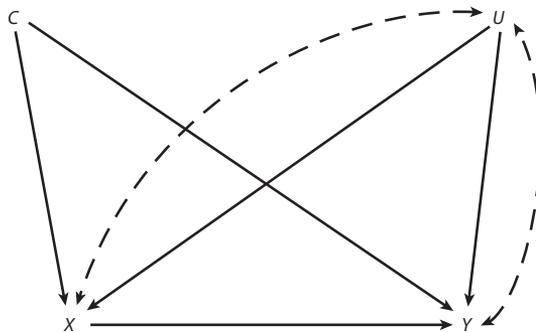


Figure 4

Directed acyclic graph (DAG) in which the exposure X and outcome Y share two common causes or confounders C (measured) and U (unmeasured), the effect of U on X is confounded by an unobserved common cause of U and X , and the effect of U on Y is confounded by an unobserved common cause of U and Y [each bidirectional dashed arrow represents unobserved common cause(s)].

Table 1 Contingency table showing hypothetical study data on measured binary confounder C , exposure X , and outcome Y (with a binary unmeasured confounder U)^a

	$C = 1$			$C = 0$		
	$X = 1$	$X = 0$	Total	$X = 1$	$X = 0$	Total
$Y = 1$	2,464	1,008	3,472	1,248	1,056	2,304
$Y = 0$	2,656	1,872	4,528	4,032	5,664	9,696
Total	5,120	2,880	8,000	5,280	6,720	12,000

^aTrue data generating process ($N = 20,000$):

$$P(U = 1) = 0.6;$$

$$P(C = 1) = 0.4;$$

$$P(X = 1 | c, u) = 0.35 + 0.2c + 0.15u;$$

$$P(Y = 1 | x, c, u) = 0.05 + 0.05x + 0.2c + 0.05xc + 0.2u.$$

C , namely ϕ_C , is also biased for α_C because $\phi_C = \alpha_C + \alpha_U \lambda_C$. It should be noted that α_C , the true or unbiased coefficient of C , need not have a causal interpretation in the first instance (63, 67). In **Figure 3**, for example, the association between C and Y conditional on X is not causal: The dashed bidirectional arc represents some unmeasured common cause of U and Y . It can also be seen that ϕ_{XC} , the coefficient of the product term XC , which captures the heterogeneity in the effect of X on Y across levels of C , is also affected by the lack of control for U . Not controlling for U has at least two consequences for the estimates from the model regressing Y on X and C only: It leads to (a) confounding of the $X \rightarrow Y$ effect and (b) collider-stratification bias in the C - Y association, or, where a causal interpretation is warranted, in the (direct) effect of C on Y not through X . The latter bias arises because, conditional on X that is a collider (a variable with two arrowheads on it) between C and U , C is now additionally associated with Y through the pathway $C \rightarrow [X] \leftarrow U \rightarrow Y$, where the square brackets indicate that the variable is conditioned on. **Table 1** presents an illustrative study data with an unmeasured U , and **Table 2** shows how the coefficients of X and C are biased by not controlling for U in conditional risk difference models. **Table 3** additionally shows the consequences of uncontrolled confounding for various marginal risk differences.

There is a third type of bias that could also result from not controlling for U : This is bias amplification of the uncontrolled confounding in the $X \rightarrow Y$ effect if the investigator adds to the model an instrumental variable—a preexposure variable that is only a cause of or is associated with the exposure X but not with Y except through X (40, 43). This third bias can be avoided by not adjusting for an instrumental variable as a covariate in the regression model.

Table 2 Conditional risk differences (95% confidence intervals) from linear binomial risk models relating exposure X to outcome Y , adjusted for confounding

	True (unbiased) model adjusting for C and U ^a	Biased model not controlling for confounding by U ^b
Coefficient of X when $C = 0$	0.050 (0.037–0.063)	0.079 (0.065–0.094)
Coefficient of X when $C = 1$	0.100 (0.078–0.121)	0.131 (0.109–0.153)
Coefficient of C (conditional on X)	0.200 (0.189–0.211)	0.193 (0.173–0.212)
Coefficient of U	0.200 (0.194–0.206)	Not applicable

^aTrue (unbiased) model: $P(Y = 1 | x, c, u) = 0.05 + 0.05x + 0.2c + 0.05xc + 0.2u$.

^bBiased model: $P(Y = 1 | x, c) = 0.1571 + 0.0792x + 0.1929c + 0.052xc$.

Table 3 Unbiased and biased conditional and marginal risk differences [95% confidence intervals (CI)] from linear binomial risk models for the effect of the exposure X on the outcome Y

	Unbiased risk difference (95% CI), controlling for C and U	Biased risk difference (95% CI), controlling for C only
Conditional risk difference for the conditional total effect at $C = 0$	0.050 (0.037–0.063)	0.079 (0.065–0.094)
Conditional risk difference RD_{YX} at $C = 1$	0.100 (0.078–0.121)	0.131 (0.109–0.153)
Marginal risk difference for the average treatment effect in the total population (ATE)	0.070 (0.058–0.083)	0.100 (0.088–0.112)
Marginal risk difference for the average treatment effect in the treated (ATT)	0.075 (0.062–0.087)	0.105 (0.093–0.117)
Marginal risk difference for the average treatment effect in the untreated (ATU)	0.065 (0.053–0.077)	0.095 (0.083–0.107)

To summarize, not controlling for a confounder like U in **Figures 1–3** biases the estimate of the effect of the exposure X as well as estimates of the associations between measured confounders (such as C) and the outcome Y .

BIAS FORMULAS FOR UNCONTROLLED CONFOUNDING

One of the oldest and most commonly used methods for adjusting an association estimate for an unmeasured confounder is the use of a bias formula to calculate a bias factor (3–11, 14, 17, 19, 20, 23, 32, 44, 53, 68). The calculated bias factor is then subtracted or removed from the biased (partially adjusted) estimate relating the exposure to the outcome (3, 29, 49, 50, 63) to obtain an externally adjusted estimate that could have been obtained if the assumptions about the bias parameters used in calculating the bias factor had held. In the example in **Figure 1**, uncontrolled confounding due to not measuring U is transmitted through the backdoor from X to Y , $X \leftarrow U \rightarrow Y$. A bias formula is a formula that is used to quantify the confounding via this backdoor.

Assuming that the set of variables C and U were sufficient to control for confounding when estimating the effect of X on Y , the relevant conditional risk differences ($RD_{YX(\text{target population})|c}$ conditional on C but standardized to U in different target populations, namely the total, exposed $X = x$, and unexposed $X = x^*$) and the marginal causal risk differences [$RD_{YX(\text{total})}$, $RD_{YX(x)}$, and $RD_{YX(x^*)}$, respectively, standardized to the joint distributions of C and U in the total, exposed and unexposed populations] would be given, without bias, by

$$\begin{aligned}
 RD_{YX(\text{total})|c} &= E(Y_x|c) - E(Y_{x^*}|c) \\
 &= \sum_u E(Y|x, c, u)P(u|c) - \sum_u E(Y|x^*, c, u)P(u|c); \\
 RD_{YX(x)|c} &= E(Y_x|x, c) - E(Y_{x^*}|x, c) \\
 &= E(Y|x, c) - \sum_u E(Y|x^*, c, u)P(u|c, x) \\
 &= \sum_u E(Y|x, c, u)P(u|c, x) - \sum_u E(Y|x^*, c, u)P(u|c, x);
 \end{aligned}$$

$$\begin{aligned}
RD_{YX(x^*)|c} &= E(Y_x | x^*, c) - E(Y_{x^*} | x^*, c) \\
&= \sum_u E(Y | x, c, u) P(u | c, x^*) - E(Y | x^*, c) \\
&= \sum_u E(Y | x, c, u) P(u | c, x^*) - \sum_u E(Y | x^*, c, u) P(u | c, x^*); \\
RD_{YX(\text{total})} &= E(Y_x) - E(Y_{x^*}) \\
&= \sum_{c,u} E(Y | x, c, u) P(u | c) P(c) - \sum_{c,u} E(Y | x^*, c, u) P(u | c) P(c); \\
RD_{YX(x)} &= E(Y_x | x) - E(Y_{x^*} | x) \\
&= E(Y | x) - \sum_{c,u} E(Y | x^*, c, u) P(u | c, x) P(c | x) \\
&= \sum_{c,u} E(Y | x, c, u) P(u | c, x) P(c | x) - \sum_{c,u} E(Y | x^*, c, u) P(u | c, x) P(c | x); \\
RD_{YX(x^*)} &= E(Y_x | x^*) - E(Y_{x^*} | x^*) \\
&= \sum_{c,u} E(Y | x, c, u) P(u | c, x^*) P(c | x^*) - E(Y | x^*) \\
&= \sum_{c,u} E(Y | x, c, u) P(u | c, x^*) P(c | x^*) - \sum_{c,u} E(Y | x^*, c, u) P(u | c, x^*) P(c | x^*).
\end{aligned}$$

In the causal inference literature, $RD_{YX(\text{total})}$, $RD_{YX(x)}$, and $RD_{YX(x^*)}$ represent the risk differences for the average treatment effect in the total population (ATE), the average treatment effect among the treated (ATT), and the average treatment effect among the untreated (ATU), respectively (3, 63). Alternatively, these causal contrasts could have been defined as risk or odds ratios (2, 16, 29). For continuous U , integral signs and probability density functions replace the summation signs and probability mass functions, respectively, in these expressions. In the absence of U , the corresponding associational risk differences relating X to Y adjusted for C or standardized to the distribution of C are given by

$$\begin{aligned}
RD_{YX+(\text{total})|c} &= E(Y | x, c) - E(Y | x^*, c) \\
&= \sum_u E(Y | x, c, u) P(u | c, x) - \sum_u E(Y | x^*, c, u) P(u | c, x^*); \\
RD_{YX+(x)|c} &= E(Y | x, c) - E(Y | x^*, c) \\
&= \sum_u E(Y | x, c, u) P(u | c, x) - \sum_u E(Y | x^*, c, u) P(u | c, x^*); \\
RD_{YX+(x^*)|c} &= E(Y | x, c) - E(Y | x^*, c) \\
&= \sum_u E(Y | x, c, u) P(u | c, x) - \sum_u E(Y | x^*, c, u) P(u | c, x^*); \\
RD_{YX+(\text{total})} &= \sum_c E(Y | x, c) P(c) - \sum_c E(Y | x^*, c) P(c) \\
&= \sum_{c,u} E(Y | x, c, u) P(u | c, x) P(c) - \sum_{c,u} E(Y | x^*, c, u) P(u | x^*, c) P(c); \\
RD_{YX+(x)} &= E(Y | x) - \sum_c E(Y | x^*, c) P(c | x) \\
&= \sum_c E(Y | x, c) P(c | x) - \sum_c E(Y | x^*, c) P(c | x) \\
&= \sum_{c,u} E(Y | x, c, u) P(u | c, x) P(c | x) - \sum_{c,u} E(Y | x^*, c, u) P(u | c, x^*) P(c | x);
\end{aligned}$$

$$\begin{aligned}
RD_{YX+(x^*)} &= \sum_c E(Y | x, c) P(c | x^*) - E(Y | x^*) \\
&= \sum_c E(Y | x, c) P(c | x^*) - \sum_c E(Y | x^*, c) P(c | x^*) \\
&= \sum_{c,u} E(Y | x, c, u) P(u | c, x) P(c | x^*) - \sum_{c,u} E(Y | x^*, c, u) P(u | c, x^*) P(c | x^*).
\end{aligned}$$

The bias [$Bias_{RDYX(\text{target population})}$] due to not controlling for U in each of the risk differences is given in turn by the difference between the risk difference estimate not adjusted for U and the risk difference estimate adjusted for U :

$$\begin{aligned}
Bias_{RDYX(\text{total})|c} &= RD_{YX+(\text{total})|c} - RD_{YX(\text{total})|c} \\
&= \sum_u [E(Y | x, c, u) - E(Y | x, c, u^*)] [P(u | c, x) - P(u | c)] \\
&\quad - \sum_u [E(Y | x^*, c, u) - E(Y | x^*, c, u^*)] [P(u | c, x^*) - P(u | c)];
\end{aligned}$$

$$\begin{aligned}
Bias_{RDYX(x)|c} &= RD_{YX+(x)|c} - RD_{YX(x)|c} \\
&= \sum_u [E(Y | x^*, c, u) - E(Y | x^*, c, u^*)] [P(u | c, x) - P(u | c, x^*)];
\end{aligned}$$

$$\begin{aligned}
Bias_{RDYX(x^*)|c} &= RD_{YX+(x^*)|c} - RD_{YX(x^*)|c} \\
&= \sum_u [E(Y | x, c, u) - E(Y | x, c, u^*)] [P(u | c, x) - P(u | c, x^*)];
\end{aligned}$$

$$\begin{aligned}
Bias_{RDYX(\text{total})} &= RD_{YX+(\text{total})} - RD_{YX(\text{total})} \\
&= \sum_{c,u} [E(Y | x, c, u) - E(Y | x, c, u^*)] [P(u | c, x) - P(u | c)] P(c) \\
&\quad - \sum_{c,u} [E(Y | x^*, c, u) - E(Y | x^*, c, u^*)] [P(u | c, x^*) - P(u | c)] P(c);
\end{aligned}$$

$$\begin{aligned}
Bias_{RDYX(x)} &= RD_{YX+(x)} - RD_{YX(x)} \\
&= \sum_{c,u} [E(Y | x^*, c, u) - E(Y | x^*, c, u^*)] [P(u | c, x) - P(u | c, x^*)] P(c | x);
\end{aligned}$$

$$\begin{aligned}
Bias_{RDYX(x^*)} &= RD_{YX+(x^*)} - RD_{YX(x^*)} \\
&= \sum_{c,u} [E(Y | x, c, u) - E(Y | x, c, u^*)] [P(u | c, x) - P(u | c, x^*)] P(c | x^*).
\end{aligned}$$

The contrast on the right-hand side of each of these expressions is usually presented as a bias formula that is used to obtain the numerical value or bias factor on the left-hand side (3, 45, 63). Several techniques based on the idea of bias formulas have been in use for more than half a century (3–11, 14, 17, 19, 20, 23, 32, 44, 53, 68) but were generalized only recently (3, 60). One advantage of these bias formulas is that they are general and can be used for general outcomes, exposures, and confounders (63).

These bias formulas generally require specifying the following bias parameters: (a) the relation between U and Y conditional on C and X [for example, $E(Y | x, c, u) - E(Y | x, c, u^*)$]; (b) the distribution of U conditional on C and X [$P(u | c, x)$ and $P(u | c, x^*)$]; and sometimes (c) the distribution of

U conditional on C but not $X [P(u|c)]$. The second bias parameters $P(u|c, x)$ and $P(u|c, x^*)$ relate the exposure X to the unmeasured confounder U , conditional on the measured confounder(s) C . The first bias parameter $E(Y |x, c, u) - E(Y |x, c, u^*)$ relates U to Y conditional on C .

Typically, the investigator specifies the bias parameters and plugs them into the relevant bias formula to quantify the bias factor (for example, $Bias_{RDYX(\text{total})|c}$), which is then used to adjust the biased risk difference $RD_{YX+(\text{total})|c}$ to obtain the U -adjusted risk difference $RD_{YX(\text{total})|c}$:

$$RD_{YX(\text{total})|c} = RD_{YX+(\text{total})|c} - Bias_{RDYX(\text{total})|c}.$$

Parallel formulas for the risk ratio as well as approximate formulas for the odds ratio have also been derived and are reported elsewhere (3, 63).

The bias formulas for uncontrolled confounding can be simplified further in some cases if we are willing to make additional (usually parametric) assumptions, such as assuming homogeneity of the bias parameters across levels of the exposure and measured confounders (3, 63).

To make the use of these formulas more concrete, consider the data in **Table 1**, in which U is assumed unobserved. The conditional linear risk model for estimating the effect of X on Y conditional on C and U is

$$\begin{aligned} E(Y |x, c, u) \\ &= \alpha_0 + \alpha_X x + \alpha_C c + \alpha_{XC} xc + \alpha_U u \\ &= 0.05 + 0.05x + 0.2c + 0.05xc + 0.2u. \end{aligned}$$

The true unbiased conditional X - Y risk difference would have been 0.05 at $C = 0$. In the absence of U , the following is obtained by regressing Y on X and C only:

$$\begin{aligned} E(Y |x, c) \\ &= \sum_u E(Y |x, c, u) P(u|x, c) \\ &= \phi_0 + \phi_X x + \phi_C c + \phi_{XC} xc \\ &= 0.157 + 0.079x + 0.192c + 0.052xc. \end{aligned}$$

At $C = 0$, the estimated conditional risk difference $RD_{YX+(x)|c}$ for the association between X and Y is 0.079 and is biased for the true conditional risk difference of 0.05. The appropriate formula for a conditional linear risk model can be used to estimate the bias factor that can be subtracted from the biased estimate 0.079 to obtain the U -adjusted risk difference estimate. The formula for $Bias_{RDYX(x)|c}$ can be used in which x is 1, x^* is 0, u is 1, and u^* is 0. The bias formula requires the following bias parameters: (a) the risk difference $E(Y |x^*, c, u) - E(Y |x^*, c, u^*)$ relating U to Y conditional on the exposure X and measured confounder C rewritten as $E(Y |X = 0, c, U = 1) - E(Y |X = 0, c, U = 0)$ at $U = 1$ but as $E(Y |X = 0, c, U = 0) - E(Y |X = 0, c, U = 0) = 0$ at $U = 0$, using $U = 0$ as reference; and (b) the prevalence of each level of U among $X = x$ and $C = c$, which can be expressed generally as $P(U = 1|c, x) = \lambda_0 + \lambda_X x + \lambda_C c + \lambda_{XC} xc$. In this case, we secretly know that $P(U = 1|c, x) = \lambda_0 + \lambda_X x + \lambda_C c + \lambda_{XC} xc = 0.536 + 0.146x - 0.036c + 0.010xc$. In real applications, this bias parameter model will not be known and must be obtained from an external source, as discussed in the next section. To apply the bias formula in this illustration, recall that, for the binary U , X , and Y used in this illustration,

$$\begin{aligned} E(Y |x^*, c, u) - E(Y |x^*, c, u^*) \\ &= (\alpha_0 + \alpha_X \cdot x^* + \alpha_C c + \alpha_{XC} \cdot x^* \cdot c + \alpha_U u) - (\alpha_0 + \alpha_X \cdot x^* + \alpha_C c + \alpha_{XC} \cdot x^* \cdot c + \alpha_U \cdot u^*) \\ &= \alpha_U. \end{aligned}$$

Therefore,

$$\begin{aligned}
Bias_{RD_{YX(x)|c}} &= \sum_u [E(Y|x^*, c, u) - E(Y|x^*, c, u^*)][P(u|c, x) - P(u|c, x^*)] \\
&= [E(Y|X=0, c, U=1) - E(Y|X=0, c, U=0)][P(U=1|c, X=1) \\
&\quad - P(U=1|c, X=0)] + [E(Y|X=0, c, U=0) \\
&\quad - E(Y|X=0, c, U=0)][P(U=0|c, X=1) - P(U=0|c, X=0)] \\
&= \alpha_U \cdot [P(U=1|c, X=1) - P(U=1|c, X=0)] + 0 \cdot [P(U=0|c, X=1) \\
&\quad - P(U=0|c, X=0)] \\
&= \alpha_U \cdot [P(U=1|c, X=1) - P(U=1|c, X=0)] \\
&= \alpha_U [(\lambda_0 + \lambda_X \cdot 1 + \lambda_{cC} + \lambda_{XC} \cdot 1 \cdot c) - (\lambda_0 + \lambda_X \cdot 0 + \lambda_{cC} + \lambda_{XC} \cdot 0 \cdot c)] \\
&= \alpha_U (\lambda_X + \lambda_{XC} \cdot c) \\
&= \alpha_U (\lambda_X + \lambda_{XC} \cdot 0) \\
&= \alpha_U \lambda_X \\
&= 0.2 \times 0.146 \\
&= 0.029.
\end{aligned}$$

The bias-adjusted X - Y risk difference $RD_{YX(x)|c}$ is then obtained by applying the formula

$$\begin{aligned}
RD_{YX(x)|c} &= RD_{YX+(x)|c} - Bias_{RD_{YX(x)|c}} \\
&= 0.079 - 0.029 \\
&= 0.05.
\end{aligned}$$

OBTAINING THE VALUES OF THE BIAS PARAMETERS

Specifying the values for bias parameters can be formidable without deep prior knowledge or external data. In particular, the bias parameter values for $P(u|c, x)$ and $P(u|c, x^*)$ can be hard to determine, being usually less intuitive than the association between U and Y conditional on X and C , namely $[E(Y|x, c, u) - E(Y|x, c, u^*)]$. Therefore, a particular challenge in applying bias formulas and related methods is how to obtain the magnitude and direction of the bias parameters needed for relating U to X and C as well as relating U to Y conditional on X and C (3, 29, 49, 55, 57, 63). Several sources can be used to obtain the bias parameters for use in the bias formulas. Beyond the investigator's background knowledge, a validation (sub)study that is internal or external to the primary study can be a source of bias parameters (12, 20, 30, 38, 52, 54, 56). An internal validation substudy is specific to the primary study and can be especially useful if it can spend more resources improving and expanding measurements that can inform the larger primary study. The validation study collects invaluable data that can be used to address selection bias (especially due to selective nonresponse), measurement error in variables, and uncontrolled confounding due to confounders not measured in the larger primary study (20, 29, 30, 56, 57). An external validation study is not a substudy of the primary study and can be used similarly where appropriate. Examples of external validation data can come from other published study data, systematic reviews, and meta-analyses. Although it can supply bias parameter values for one or more unmeasured confounders more readily than an investigator's background knowledge or intelligent guesses, a validation (sub)study can still be prone to similar sources of bias as the primary study. Therefore, it is important not to be overly optimistic about the value of the validation (sub)study, and the investigator should allow for such uncertainty in using bias parameters from the validation (sub)study (20, 30, 49).

FIXED VERSUS PROBABILISTIC BIAS ANALYSIS

After obtaining the bias parameters, the investigator can use them in the bias formulas for bias analysis in several ways. First, simple fixed analysis involving a fixed (one-time) value assignment to

the bias parameters can be used to obtain single bias-adjusted estimates of the exposure–outcome association. This does not account for random error or even uncertainty in the values of the bias parameters. This is sometimes referred to as simple sensitivity analysis (29, 30, 49). Second, the investigator can repeat the simple bias analysis for several different fixed values of the bias parameters and report several bias-adjusted exposure–outcome association estimates. As before, this so-called multidimensional bias analysis does not account for random error or for uncertainty in each fixed bias parameter value. Third, to overcome the shortcomings of the preceding approaches, probability distributions, rather than fixed values, can be assigned to the bias parameters in what is called probabilistic bias analysis to obtain a distribution of bias-adjusted exposure–outcome association estimates, while accounting for study random error and uncertainty in the choice of bias parameters (3, 29, 30, 49).

OTHER BIAS ANALYSIS METHODS

Scholars have developed methods other than the bias formulas for adjusting for uncontrolled confounding. These include: (a) the direct specification of the bias factor (that is, the numerical value from the bias formula without specifying the underlying bias parameters relating U to X given C and relating U to Y given X and C) (21) or related methods (47, 48); (b) the simulation or imputation of U using external information (subsumed under missing data methods) (12, 20, 28); (c) propensity calibration using validation data (51, 56, 57); (d) intensity scores (9, 46); (e) the use of negative controls (34); and (f) the use of bounding techniques (13, 18, 31, 36), among others. Some of these techniques are still evolving and have the potential to become routine in bias analysis for uncontrolled confounding.

MULTIPLE UNMEASURED CONFOUNDERS

With the exception of a few cases, such as propensity calibration and Bayesian bias analysis, many existing methods are not easily amenable to multiple unmeasured confounders (51, 54, 57, 59). Nonetheless, it is possible to view the bias formulas discussed in this article as being extensible to multiple unmeasured confounder settings by seeing U as a set of variables and adapting the formulas to reflect the implied joint distribution of multiple U s. This could substantially increase the number of bias parameters needed. More work is needed in this area.

GENERALIZED BIAS ANALYSIS FRAMEWORK

To overcome some of the challenges facing the existing methods described above, we propose a novel generalized framework for bias analysis that simulates the amount of uncontrolled confounding due to one or more unmeasured confounders under one or more scenarios in which the exposure X has no effect or some effect on the outcome Y given U and C ; **Figure 5** provides an example.

This new generalized bias analysis using simulated confounding is intuitive because it allows the investigator to reason in the direction of the arrows in the DAG or the information flow in the assumed data-generating process to quantify the amount of uncontrolled confounding due to the specified values of assumed bias parameters. For example, instead of reasoning about bias parameters backward from X and C to U , as seen in the bias formula approach, in the simulated confounding approach, one reasons from U and C to X to obtain a new simulated exposure X_{sim} from $P(x_{\text{sim}}|c, u)$, from which $P(u|c, x_{\text{sim}})$ can be estimated by regressing U on C . The new X_{sim} can also be used in a bias formula if so desired (although this last part is not required). Similarly,

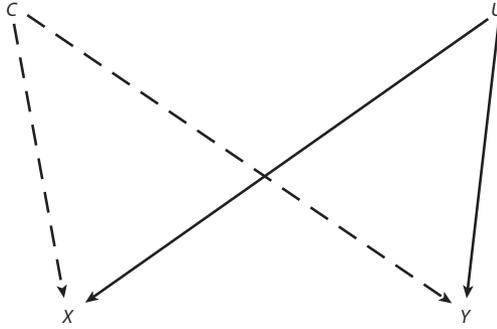


Figure 5

Directed acyclic graph (DAG) in which the exposure X and outcome Y share two common causes C (measured) and U (unmeasured), but X is assumed to have no effect on Y .

U , C , and X_{sim} are used to generate a new outcome Y_{sim} , assuming a null $X_{\text{sim}}-Y_{\text{sim}}$ association. Regressing Y_{sim} on X_{sim} and C , but not U , yields a non-null association between exposure X_{sim} and outcome Y_{sim} that is due to uncontrolled confounding by U . Overall, this simulated confounding framework entails the following algorithm:

- Step 1: Simulate the unmeasured confounder U_{sim} from its marginal distribution $P(u_{\text{sim}})$; for example, simulate a binary U from $P(U_{\text{sim}} = 1) = \hat{\mu}_U$ where $0 < \hat{\mu}_U < 1$.
- Step 2: Using the observed study data, obtain the parameters of the conditional distribution $P(x|c)$; for example, regress X on C to obtain $P(X = 1|c) = \hat{\delta}_0 + \hat{\delta}_C c$.
- Step 3: Simulate X_{sim} from $P(x_{\text{sim}}|c, u_{\text{sim}})$ using the parameter $\hat{\mu}_U$ and U_{sim} from step 1, parameters $\hat{\delta}_0$ and $\hat{\delta}_C$ from step 2, and externally obtained parameter $\hat{\delta}_U$ for the assumed $U-X$ associational risk difference given C ; for example, simulate a binary X_{sim} from a Bernoulli trial using $P(X_{\text{sim}} = 1|c, u_{\text{sim}}) = \hat{\delta}_0 + \hat{\delta}_C c + \hat{\delta}_U u_{\text{sim}} - \hat{\delta}_U \hat{\mu}_U$, where the constant term $\hat{\delta}_U \hat{\mu}_U$ offsets the intercept to account for marginalizing over the unobserved U in the observed data used to obtain the parameters in step 2.
- Step 4: Use the observed study data to obtain the parameters of the conditional expression $P(y|x, c)$; for example, regress Y on X and C to obtain $P(Y = 1|x, c) = \hat{\phi}_0 + \hat{\phi}_X x + \hat{\phi}_C c$.
- Step 5: Simulate Y_{sim} from $P(y_{\text{sim}}|x_{\text{sim}}, c, u_{\text{sim}})$ using U_{sim} from step 1, X_{sim} from step 3, and externally obtained parameter (risk difference) $\hat{\alpha}_U$ for the assumed conditional $U-Y$ associational risk difference given X and C in the unobserved model for $E(Y|x, c, u) = \alpha_0 + \alpha_X x + \alpha_C c + \alpha_{XC} xc + \alpha_U u$. For example, simulate a binary Y_{sim} from a Bernoulli trial using $P(Y_{\text{sim}} = 1|x_{\text{sim}}, c, u_{\text{sim}}) = \hat{\phi}_0 + 0 \cdot x_{\text{sim}} + \hat{\phi}_C c + \hat{\alpha}_U u_{\text{sim}}$.
- Step 6: Regress Y_{sim} on X_{sim} and C to obtain $P(Y_{\text{sim}} = 1|x_{\text{sim}}, c) = \hat{\gamma}_0 + \hat{\gamma}_{X_{\text{sim}}} x_{\text{sim}} + \hat{\gamma}_C c$, and read off the coefficient $\hat{\gamma}_{X_{\text{sim}}}$ of X_{sim} as the amount of confounding due to omitting U_{sim} . Note that $\hat{\gamma}_{X_{\text{sim}}}$ is also an estimate of the bias factor for the conditional association model.
- Step 7: Repeat step 4 and use the simulated confounding estimate or bias factor $\hat{\gamma}_{X_{\text{sim}}}$ to offset the U -biased coefficient of X in the model for $P(Y = 1|x, c)$ that is based only on the observed data, and thus to obtain U -bias-adjusted X - Y association.

The programming of steps 3 and 5 can be simplified when there are many measured covariates C by omitting the coefficient(s) of set C while maintaining the coefficients of U and X at the levels they would have attained under the omitted coefficient set. This simulated confounding is quite general and can be used for difference and ratio measures, general outcomes, exposures and confounders, and multiple unmeasured confounders, and it can incorporate different functional forms into the models for the exposure and outcome. The new algorithm only simulates U , X , and Y using

parameters taken from the observed study data, covariates from the observed data (optionally), and externally obtained bias parameters (not unlike the bias formula technique, although this method is more intuitive). The algorithm can be repeated in combination with bootstrapping or can be programmed into a more extensive probabilistic sensitivity analysis. This new algorithm is different from the semi-automated sensitivity analysis of Lash and Fink, which simulates U as a function of the observed X and Y (28). Their method, therefore, also involves the challenge of reasoning backward from X and Y to U as was seen in the bias formulas discussed earlier in this article. The new generalized bias analysis method introduced here avoids that challenge by appealing to the intuition encoded in the assumed DAG and reasoning forward from U to X and Y .

CONCLUSION

Bias analysis for uncontrolled confounding is crucial for causal inference studies that rely on observational data or less-than-perfect randomized controlled trials, as the ones seen in the health sciences. Concerns about uncontrolled confounding should accompany any covariate selection issue for confounding control in empirical quantitative analysis (2, 4, 15, 17, 23, 42, 58). Methodological development of bias analysis for uncontrolled confounding has been ongoing for more than half a century, and general methods and software have become more readily available in the last decade (3, 20, 29, 30, 37, 49, 55, 56, 63). Although its adoption and applications have been slow, bias analysis needs and applications are likely to rise with the growth of big data, computational platforms, causal inference tools, needs for replication, data sharing, and journal peer-review and reporting requirements. This article has provided a broad overview of the key approaches and introduced a new framework that hopefully contributes to faster adoption and further methodological development and refinement.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was partly supported by European Commission FP7 grant 241822, NIDDK grant R01DK095668-02, and NICHD grant R01HD072296-01A1. I thank Amy Chai, Vahe Khachadourian, Roch Nianogo, and the anonymous reviewers for all the feedback that helped improve this article.

LITERATURE CITED

1. Angus DC. 2015. Fusing randomized trials with big data: the key to self-learning health care systems? *JAMA* 314(8):767–68
2. Arah OA. 2008. The role of causal reasoning in understanding Simpson’s paradox, Lord’s paradox, and the suppression effect: covariate selection in the analysis of observational studies. *Emerg. Themes Epidemiol.* 5:5
3. Arah OA, Chiba Y, Greenland S. 2008. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann. Epidemiol.* 18(8):637–46
4. Arah OA, Sudan M, Olsen J, Kheifets L. 2013. Marginal structural models, doubly robust estimation, and bias analysis in perinatal and paediatric epidemiology. *Paediatr. Perinat. Epidemiol.* 27(3):263–65

5. Axelson O, Steenland K. 1988. Indirect methods of assessing the effects of tobacco use in occupational studies. *Am. J. Ind. Med.* 13(1):105–18
6. Breslow NE, Day NE. 1980. *Statistical Methods in Cancer Research*, Vol. 1: *The Analysis of Case-Control Studies*. Int. Agency Res. Cancer Sci. Publ. 32. Lyon, Fr.: Int. Agency Res. Cancer
7. Bross IDJ. 1966. Spurious effects from an extraneous variable. *J. Chronic Dis.* 19(6):637–47
8. Bross IDJ. 1967. Pertinency of an extraneous variable. *J. Chronic Dis.* 20(7):487–95
9. Brumback B, Greenland S, Redman M, Kiviat N, Diehr P. 2003. The intensity-score approach to adjusting for confounding. *Biometrics* 59(2):274–85
10. Cai Z, Brumback BA. 2015. Model-based standardization to adjust for unmeasured cluster-level confounders with complex survey data. *Stat. Med.* 34(15):2368–80
11. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. 1959. Smoking and lung cancer: recent evidence and a discussion of some questions. *J. Natl. Cancer Inst.* 22:173–203
12. Faries D, Peng X, Pawaskar M, Price K, Stamey JD, Seaman JW. Evaluating the impact of unmeasured confounding with internal validation data: an example cost evaluation in type 2 diabetes. *Value Health* 16(2):259–66
13. Flanders WD, Khoury MJ. 1990. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology* 1(3):239–46
14. Gail MH, Wacholder S, Lubin JH. 1988. Indirect corrections for confounding under multiplicative and additive risk models. *Am. J. Ind. Med.* 13(1):119–30
15. Goto A, Arah OA, Goto M, Terauchi Y, Noda M. 2013. Severe hypoglycaemia and cardiovascular disease: systematic review and meta-analysis with bias analysis. *BMJ* 347:f4533
16. Greenland S. 1996. Basic methods for sensitivity analysis of biases. *Int. J. Epidemiol.* 25(6):1107–16
17. Greenland S. 2003. The impact of prior distributions for uncontrolled confounding and response bias. *J. Am. Stat. Assoc.* 98(461):47–54
18. Greenland S. 2004. Bounding analysis as an inadequately specified methodology. *Risk Anal.* 24(5):1085–92
19. Greenland S. 2005. Multiple-bias modelling for analysis of observational data (with discussion). *J. R. Stat. Soc. Ser. A* 168(2):267–306
20. Greenland S. 2009. Bayesian perspectives for epidemiologic research. III: Bias analysis via missing-data methods. *Int. J. Epidemiol.* 38(6):1662–73
21. Greenland S. 2014. Sensitivity analysis and bias analysis. In *Handbook of Epidemiology*, ed. W Ahrens, I Pigeot, pp. 685–706. New York: Springer. 2nd ed.
22. Greenland S, Pearl J, Robins JM. 1999. Causal diagrams for epidemiologic research. *Epidemiology* 10(1):37–48
23. Helmich E, Boerebach BCM, Arah OA, Lingard L. 2015. Beyond limitations: improving how we handle uncertainty in health professions education research. *Med. Teach.* 37(11):1–8
24. Jain SH, Rosenblatt M, Duke J. 2014. Is big data the new frontier for academic-industry collaboration? *JAMA* 311(21):2171
25. Kaufmann SHE, Fletcher HA, Guzman CA, Ottenhoff THM. 2015. Big data in vaccinology: introduction and section summaries. *Vaccine* 33(40):5237–40
26. Klungsoyr O, Sexton J, Sandanger I, Nygård JF. 2009. Sensitivity analysis for unmeasured confounding in a marginal structural Cox proportional hazards model. *Lifetime Data Anal.* 15(2):278–94
27. Larson EB. 2013. Building trust in the power of “big data” research to serve the public good. *JAMA* 309(23):2443–44
28. Lash TL, Fink AK. 2003. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 14(4):451–58
29. Lash TL, Fox MP, Fink AK. 2011. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York: Springer
30. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. 2014. Good practices for quantitative bias analysis. *Int. J. Epidemiol.* 43(6):1969–85
31. Lee W-C. 2011. Bounding the bias of unmeasured factors with confounding and effect-modifying potentials. *Stat. Med.* 30(9):1007–17
32. Li L, Brumback BA, Weppelmann TA, Morris JG, Ali A. 2016. Adjusting for unmeasured confounding due to either of two crossed factors with a logistic regression model. *Stat. Med.* 35(18):3179–88

33. Lin DY, Psaty BM, Kronmal RA. 1998. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 54(3):948
34. Lipsitch M, Tchetgen ET, Cohen T. 2010. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 21(3):383–88
35. Luna X De, Waernbaum I, Richardson TS. 2011. Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika* 98(4):861–75
36. MacLehose RF, Kaufman S, Kaufman JS, Poole C. 2005. Bounding causal effects under uncontrolled confounding using counterfactuals. *Epidemiology* 16(4):548–55
37. McCandless LC, Gustafson P, Levy A. 2007. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat. Med.* 26(11):2331–47
38. McCandless LC, Richardson S, Best N. 2012. Adjustment for missing confounders using external validation data and propensity scores. *J. Am. Stat. Assoc.* 107(497):40–51
39. McCulloch CE, Searle SR, Neuhaus JM. 2009. *Generalized, Linear, and Mixed Models*. Hoboken, NJ: John Wiley & Sons
40. Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, et al. 2011. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am. J. Epidemiol.* 174(11):1213–22
41. Pearl J. 2009. Causal inference in statistics: an overview. *Stat. Surv.* 3:96–146
42. Pearl J. 2009. *Causality: Models, Reasoning and Inference*. New York: Cambridge Univ. Press. 2nd ed.
43. Pearl J. 2011. Invited commentary: understanding bias amplification. *Am. J. Epidemiol.* 174(11):1223–27
44. Phillips CV. 2003. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 14(4):459–66
45. Porta M, ed. 2014. *A Dictionary of Epidemiology*. New York: Oxford Univ. Press. 6th ed.
46. Robins JM, Rotnitzky A, Scharfstein DO. 2000. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, ed. ME Halloran, D Berry, pp. 1–94. New York: Springer
47. Rosenbaum PR. 2002. *Observational Studies*. New York: Springer. 2nd ed.
48. Rosenbaum PR. 2010. *Design of Observational Studies*. New York: Springer
49. Rothman KJ, Greenland S, Lash TL. 2008. *Modern Epidemiology*. Philadelphia: Lippincott Williams & Wilkins. 3rd ed.
50. Schlesselman JJ. 1978. Assessing effects of confounding variables. *Am. J. Epidemiol.* 108(1):3–8
51. Schneeweiss S. 2006. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol. Drug Saf.* 15(5):291–303
52. Stamey JD, Beavers DP, Faries D, Price KL, Seaman JW. 2014. Bayesian modeling of cost-effectiveness studies with unmeasured confounding: a simulation study. *Pharm. Stat.* 13(1):94–100
53. Steenland K. 2004. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am. J. Epidemiol.* 160(4):384–92
54. Stürmer T, Glynn RJ, Rothman KJ, Avorn J, Schneeweiss S. 2007. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. *Med. Care* 45(10 Suppl. 2):S158–65
55. Stürmer T, Rothman KJ, Avorn J, Glynn RJ. 2010. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution—a simulation study. *Am. J. Epidemiol.* 172(7):843–54
56. Stürmer T, Schneeweiss S, Avorn J, Glynn RJ. 2005. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am. J. Epidemiol.* 162(3):279–89
57. Stürmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. 2007. Performance of propensity score calibration—a simulation study. *Am. J. Epidemiol.* 165(10):1110–18
58. Sudan M, Kheifets L, Arah OA, Olsen J. 2013. Cell phone exposures and hearing loss in children in the Danish national birth cohort. *Paediatr. Perinat. Epidemiol.* 27(3):247–57
59. Uddin MJ, Groenwold RHH, Ali MS, de Boer A, Roes KCB, et al. 2016. Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. *Int. J. Clin. Pharm.* 38(3):714–23
60. VanderWeele TJ. 2010. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology* 21(4):540–51

61. VanderWeele TJ. 2013. Unmeasured confounding and hazard scales: sensitivity analysis for total, direct, and indirect effects. *Eur. J. Epidemiol.* 28(2):113–17
62. VanderWeele TJ. 2016. Mediation analysis: a practitioner's guide. *Annu. Rev. Public Health* 37:17–32
63. Vanderweele TJ, Arah OA. 2011. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 22(1):42–52
64. Vanderweele TJ, Mukherjee B, Chen J. 2012. Sensitivity analysis for interactions under unmeasured confounding. *Stat. Med.* 31(22):2552–64
65. Vanderweele TJ, Shpitser I. 2011. A new criterion for confounder selection. *Biometrics* 67(4):1406–13
66. VanderWeele TJ, Shpitser I. 2013. On the definition of a confounder. *Ann. Stat.* 41(1):196–220
67. Westreich D, Greenland S. 2013. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am. J. Epidemiol.* 177(4):292–98
68. Yanagawa T. 1984. Case-control studies: assessing the effect of a confounding factor. *Biometrika* 71(1):191–94