# Multiple-bias Sensitivity Analysis Using Bounds

*Louisa H. Smith,*[a] *Maya B. Mathur,*[b] *and Tyler J. VanderWeele*[a,c]

**Abstract:** Confounding, selection bias, and measurement error are well-known sources of bias in epidemiologic research. Methods for assessing these biases have their own limitations. Many quantitative sensitivity analysis approaches consider each type of bias individually, although more complex approaches are harder to implement or require numerous assumptions. By failing to consider multiple biases at once, researchers can underestimate—or overestimate—their joint impact. We show that it is possible to bound the total composite bias owing to these three sources and to use that bound to assess the sensitivity of a risk ratio to any combination of these biases. We derive bounds for the total composite bias under a variety of scenarios, providing researchers with tools to assess their total potential impact. We apply this technique to a study where unmeasured confounding and selection bias are both concerns and to another study in which possible differential exposure misclassification and confounding are concerns. The approach we describe, though conservative, is easier to implement and makes simpler assumptions than quantitative bias analysis. We provide R functions to aid implementation.

**Keywords:** Bias analysis; Causal inference; Differential misclassification; Selection bias; Unmeasured confounding

(*Epidemiology* 2021;32: 625–634)



Assessing evidence for causation is fundamental to plan and target interventions and improve public health. However, many causal claims in epidemiologic studies are met with suspicion by both researchers and the general public, owing to the fact that such studies are well known to be subject to various biases. Although faults in these studies can sometimes be addressed directly—for example, through better sampling schemes, blinded outcome ascertainment, more extensive covariate measurements, and so on—other times confounding, selection bias, and measurement error are unavoidable. In such situations, our next best option is to assess the extent to which a given study's conclusions might be sensitive or robust to these biases and whether they threaten its conclusions. Often, however, this is limited to a few sentences in a discussion section qualitatively assessing the possibility of bias, sometimes appealing without quantitative justification to heuristics that may or may not hold true in a particular study.[1–3]

The weak uptake of quantitative bias analysis in epidemiology belies its long history. Over a half century ago, Cornfield and then Bross argued that the extent of possible bias was quantifiable based on observed data and possibly hypothetical quantities.[4–7] Attempts to generalize these results, as well as consider other biases, sometimes simultaneously with confounding, followed.[7–12] More recently, probabilistic bias analysis methods have been developed, allowing researchers to propose distributions for various bias parameters across multiple biases and to explore how various combinations of those parameters would affect their results.[13–18] Despite the availability of these methods in textbook and software form,[16,19,20] the actual uptake of such quantitative bias analysis in empirical research has been limited,[21] possibly because of the (at least perceived) computational complexity[22] or difficulty in proposing plausible distributions.

Even more recently, simpler approaches to sensitivity analysis have hearkened back to the early days of bias assessment, with the development of bounds for various biases that require limited assumptions and at most basic algebra.[23–25] However, a one-at-a-time approach is not sufficient for many studies subject to multiple sources of bias. In this article, we extend the simple sensitivity analysis framework to multiple biases, describing a bound for the total composite bias from confounding, selection, and differential exposure or outcome misclassification.

## THE PROBLEM OF MULTIPLE BIASES

We will describe a scenario in which all three types of bias are present, preventing the interpretation of an observed risk ratio as a causal risk ratio. Consider a binary exposure $A$, a binary outcome $Y$, and measured covariates $C$ ($C$ may also include unmeasured factors that are controlled by the study design). Let $S$ be an indicator of the subset of the population for which data have been collected, and let $A^*$ and $Y^*$ denote misclassified versions of the exposure and outcome, respectively.

***Editor's Note:*** Related articles are found on pp. 617 and 634.
Submitted May 6, 2020; accepted May 21, 2021
From the [a]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA; [b]Quantitative Sciences Unit, Stanford University, Palo Alto, CA; and [c]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA.
This work was supported by R01 CA222147 (T.J.V.W.).
The R package EValue can be used to implement the results, and the code to do so is given in the text.
The authors report no conflicts of interest.
SDC Supplmental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).
Correspondence: Louisa H. Smith, Department of Epidemiology, Harvard T. H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115. E-mail: Louisa_h_smith@g.harvard.edu.


ISSN: 1044-3983/21/325-625
DOI: 10.1097/EDE.0000000000001380

We use potential outcome notation to describe causal quantities: $Y_a$ is the outcome that would occur were exposure $A$ set to value $a$. We assume consistency, meaning that $Y_a = Y$ for observations for whom we observe $A = a$, and positivity, meaning that $0 < \Pr(A = 1 \mid \cdot) < 1$ within every stratum of the population.

We denote (conditional) independence between random variables with the symbol $\coprod$, such that $Y_a \coprod A \mid C$ implies conditional exchangeability; that is, potential outcomes are independent of exposure status conditional on $C$. However, when $C$ does not capture all of the exposure–outcome confounding, it is not true that $Y_a \coprod A \mid C$. We assume in that case that additionally adjusting for some unmeasured factor(s) $U_c$ would be sufficient to address confounding so that $Y_a \coprod A \mid C, U_c$. Here, $U_c$ may be a single random variable or a vector of variables, which may be continuous or take on any number of discrete values, or some combination. Similarly, we allow for selection bias, which we define as a lack of the conditional independence $Y_a \coprod A \mid C, U_c, S = 1$ when it is otherwise true that $Y_a \coprod A \mid C, U_c$. We likewise assume that the measurement of some variable(s) $U_s$, responsible for selection, would fully account for this bias, although the necessary conditions for it to do so will depend on whether we intend to make inferences about effects in the total population, or just the selected population. Finally, we allow for the possibility that the misclassification is differential, by which we mean that the sensitivity or specificity of the exposure measurement may differ depending on the value of the outcome, or that the sensitivity or specificity of the outcome measurement may depend on the exposure. In our notation, this means that it is not necessarily true that $A^* \coprod Y \mid A, C$ or that $Y^* \coprod A \mid Y, C$. In this work, we consider only misclassification of the exposure or the outcome but not both at once.

## MOTIVATING EXAMPLES

There is great interest in how exposures during pregnancy may affect offspring health. However important such questions are, they are difficult to answer with epidemiologic research. Ethics may limit inclusion of pregnant people in randomized trials, and many exposures of interest are not ethical or feasible to randomize to anyone. Case-control studies can efficiently capture rare childhood outcomes but recalling pregnancy exposures several years later can result in measurement error.[26] Prospective cohort studies can avoid this recall bias but are often subject to loss to follow-up when the duration between exposure and outcome assessment is long.[27] Observational studies of all types are threatened by uncontrolled confounding, and intergenerational confounders are particularly difficult to assess.[28] Importantly, studies like these are not affected by only one or another of these biases but may suffer from multiple threats to validity.

To demonstrate our sensitivity analysis approach, we will consider two questions about exposures during pregnancy and outcomes in children: whether HIV infection in utero causes wasting (low weight-for-length) and whether vitamin consumption during pregnancy protects against childhood leukemia.

Omoni et al[29] investigated the former hypothesis concerning HIV infection and wasting (among participants of a vitamin A supplementation trial in Zimbabwe) and found that, compared with children who were unexposed to HIV, those who had been infected with HIV in utero were more likely to be below a weight-for-length Z score of –2 as toddlers. The odds ratio comparing the two groups was 6.75 (95% CI, 2.79, 16.31) at 2 years. Although randomized trial data were used for the analysis, this was an observational study with respect to HIV infection, since infection is not randomized. The authors did not, however, adjust for any confounders. Furthermore, since enrollment occurred at delivery, after possible HIV exposure and transmission, the choice of whether to participate could have been affected by HIV status as well as other factors, leading to selection bias if those factors affect future child growth. We will consider the role that confounding and selection bias may play in this study.

As a second example, Ross et al[30] analyzed the relationship between vitamins and leukemia in a case-control study and found a decreased risk of acute lymphoblastic leukemia among children whose mothers consumed vitamin supplements during pregnancy. Their reported odds ratio, which, with a rare outcome, approximates a risk ratio of 0.51 (95% CI 0.30, 0.89), was conditional on maternal age, race, and a binary indicator of education. However, there may be other confounders that were not controlled, such as other indicators of a privileged or healthy lifestyle that are both associated with vitamin use and protection against leukemia. We also may be concerned about recall bias (differential exposure misclassification)—that mothers of children with a cancer diagnosis might be more likely to report *not* taking a vitamin even if they did so—so we consider how exposure misclassification and unmeasured confounding can be assessed simultaneously.

## THE MULTIPLE-BIAS BOUND

Two overarching types of bias analysis have been described: one that explores how biases of a given magnitude affect an estimate, which Phillips labeled "bias-level sensitivity analysis" and another that reduces the analysis to a summary of how much bias would be necessary for an observation to be compatible with a truly null effect (or some other specified nonnull effect), which he called "target-adjusted sensitivity analysis."[2] We focus on the former and address the latter in the eAppendix; http://links.lww.com/EDE/B824. Here, we present a multiple-bias bound, which allows researchers or consumers of research to explore the maximum factor by which unmeasured confounding, selection, and misclassification could bias a risk ratio.

We begin with outcome misclassification and then extend our results to exposure misclassification. We assume that the investigators have estimated

$\mathrm{RR}^{\mathrm{obs}}_{AY^*} = \dfrac{\Pr(Y^*=1 \mid A=1, S=1, c)}{\Pr(Y^*=1 \mid A=0, S=1, c)}$, the observed risk ratio conditional on some value $c$ of the covariates but wish to make inference about the causal conditional risk ratio $\mathrm{RR}^{\mathrm{true}}_{AY} = \dfrac{\Pr(Y_1=1 \mid c)}{\Pr(Y_0=1 \mid c)}$. We will assess bias on the relative scale so that we define the bias as $\mathrm{RR}^{\mathrm{obs}}_{AY^*} / \mathrm{RR}^{\mathrm{true}}_{AY}$.

Using bounds that have been previously described for misclassification, selection bias, and unmeasured confounding considered individually,[23–25, 31] we can bound $\mathrm{RR}^{\mathrm{obs}}_{AY^*}$ by factoring it into $\mathrm{RR}^{\mathrm{true}}_{AY}$ and components for each of the biases. The parameters that will be used to bound the biases are as follows:

$$\mathrm{RR}_{AY^*|y,S=1} = \max_y \frac{\Pr(Y^*=1 \mid Y=y, A=1, S=1, c)}{\Pr(Y^*=1 \mid Y=y, A=0, S=1, c)}$$

$$\mathrm{RR}_{U_sY|A=a} = \frac{\max_u \Pr(Y=1 \mid A=a, c, U_s=u)}{\min_u \Pr(Y=1 \mid A=a, c, U_s=u)} \text{ for } a = 0,1$$

$$\mathrm{RR}_{SU_s|A=a} = \max_u \frac{\Pr(U_s=u \mid A=a, S=a, c)}{\Pr(U_s=u \mid A=a, S=1-a, c)} \text{ for } a = 0,1$$

$$\mathrm{RR}_{U_cY} = \max_a \frac{\max_u \Pr(Y=1 \mid A=a, c, U_c=u)}{\min_u \Pr(Y=1 \mid A=a, c, U_c=u)}$$

$$\mathrm{RR}_{AU_c} = \max_u \frac{\Pr(U_c=u \mid A=1, c)}{\Pr(U_c=u \mid A=0, c)}.$$

These bias parameters have been described elsewhere, although separately.[23–25] Briefly, the bias parameter defining the misclassification portion of the bound ($\mathrm{RR}_{AY^*|y,S=1}$) describes the maximum of the false-positive probability ratio or sensitivity ratio *within* the selected population. The selection bias parameters ($\mathrm{RR}_{U_sY|A=a}$ and $\mathrm{RR}_{SU_s|A=a}$) describe the maximum factors by which the outcome risk differs by values of $U_s$, within strata of $A$, and the maximum factors by which some level of $U_s$ differs between the selected and nonselected groups, within strata of $A$. Finally, the unmeasured confounding parameters ($\mathrm{RR}_{U_cY}$ and $\mathrm{RR}_{AU_c}$) describe the maximum factor by which $U_c$ increases the outcome risk, conditional on $A$, and the maximum factor by which exposure is associated with some value of $U_c$. Each of the sensitivity parameters is conditional on the covariates adjusted for in the analysis and so describes the extent of bias above and beyond those factors.

To simplify notation, define the function $g(a,b) = \dfrac{a \times b}{a+b-1}$. Then, we have the following bound for the total composite bias.

Result 1:

If $Y_a \coprod A \mid C, U_c$ and $Y \coprod S \mid A, C, U_s$

then: $\mathrm{RR}^{\mathrm{obs}}_{AY^*} / \mathrm{RR}^{\mathrm{true}}_{AY} \le \mathrm{BF}_m \times \mathrm{BF}_s \times \mathrm{BF}_c$

where $\mathrm{BF}_m = \mathrm{RR}_{AY^*|y,S=1}$, $\mathrm{BF}_s = g\left(\mathrm{RR}_{U_sY|A=1}, \mathrm{RR}_{SU_s|A=1}\right) \times g\left(\mathrm{RR}_{U_sY|A=0}, \mathrm{RR}_{SU_s|A=0}\right)$, and $\mathrm{BF}_c = g\left(\mathrm{RR}_{AU_c}, \mathrm{RR}_{U_cY}\right)$. The derivation of this and the results that follow are given in the eAppendix; http://links.lww.com/EDE/B824.

Result 1 can be used to quantify the maximum amount of bias that could be produced by parameters of a given value. Values for the sensitivity parameters may be taken from validation studies, previous literature, or expert knowledge or proposed as hypotheticals. Because the sensitivity parameters are maxima, they are always greater than or equal to 1, and the composite bound will thus be greater than or equal to 1. For an apparently causative observed exposure–outcome risk ratio (>1), one could divide the estimate and its confidence interval by the bound to obtain the maximum that the specified biases could shift the estimate and its confidence interval. For a preventive observed exposure–outcome risk ratio (<1), one could multiply the estimate and its confidence interval by the bound to obtain the maximum that the specified biases could shift the estimate and its confidence interval, or equivalently reverse the coding of the exposure to obtain a risk ratio >1. By applying the bound to the confidence interval closest to the null, we can make statements such as: "In 95% of repeated samples with the same sources of bias, adjusting the confidence interval in this way would result in a lower bound that is less than the true causal risk ratio, provided the proposed parameter values adequately bound (i.e., are as large as or larger than) the true parameter values."

Although the bound allows for terms for all three biases, if any of them is judged not to threaten a given study or to bias toward the null, that factor can be omitted. Furthermore, the selection bias term can be simplified under certain assumptions;[24] we illustrate in the first example below.

Result 1 is summarized in the first row of Table 1. The assumptions required for the bound to hold are listed under the biases to which they pertain, and the bound itself is in the final column. Note that the factorization of the bound implies an ordering of the biases: the misclassification parameters are defined within the stratum $S=1$. Intuitively, this corresponds to a study in which outcome measurement is done after people have been selected into the study and so requires considering the strength of differential misclassification only within that group. In general, we can think of biases as layers that we must peel off sequentially and the order in which we do so is the reverse of the order in which they occurred in the data.[32,33] Confounding is generally thought of as a property of nature within the population of interest, so occurs first (although if parameters describing the strength of confounding are derived based on misclassified exposure or outcome, that may not be the case[32]), but the order in which selection and

**TABLE.** Multiple Bias Bounds for Various Combinations of Biases

| Biases and Associated Assumptions | | | Bound Under the Stated Assumptions |
|---|---|---|---|
| **Bias 1** | **Bias 2** | **Bias 3** | |
| Unmeasured confounding $Y_a \perp\!\!\!\perp A \mid C, U_c$ | General selection bias $Y \perp\!\!\!\perp S \mid A, C, U_s$ | Outcome misclassification | $g\left(\mathrm{RR}_{AU_c}, \mathrm{RR}_{U_cY}\right) \times g\left(\mathrm{RR}_{U_sY\mid A=1}, \mathrm{RR}_{SU_s\mid A=1}\right) \times g\left(\mathrm{RR}_{U_sY\mid A=0}\mathrm{RR}_{SU_s\mid A=0}\right) \times \mathrm{RR}_{AY^*\mid y, S=1}$ |
| Unmeasured confounding $Y_a \perp\!\!\!\perp A \mid C, U_c$ | Outcome misclassification | General selection bias $Y^* \perp\!\!\!\perp S \mid A, C, U_s$ | $g\left(\mathrm{RR}_{AU_c}, \mathrm{RR}_{U_cY}\right) \times \mathrm{RR}_{AY^*\mid y} \times g\left(\mathrm{RR}_{U_sY^*\mid A=1}, \mathrm{RR}_{SU_s\mid A=1}\right) \times g\left(\mathrm{RR}_{U_sY^*\mid A=0}\mathrm{RR}_{SU_s\mid A=0}\right)$ |
| Unmeasured confounding $Y_a \perp\!\!\!\perp A \mid S=1, C, U_c, U_s$ | Selected population | Outcome misclassification | $g\left(\mathrm{RR}_{AU_{sc}\mid S=1}, \mathrm{RR}_{U_{sc}Y\mid S=1}\right) \times \mathrm{RR}_{AY^*\mid y}$ |
| Unmeasured confounding $Y_a \perp\!\!\!\perp A \mid C, U_c$ | General selection bias $Y \perp\!\!\!\perp S \mid A, C, U_s$ | Exposure misclassification $\Pr(Y=0 \mid a, c, S=1) \approx 1$ | $g\left(\mathrm{RR}_{AU_c}, \mathrm{RR}_{U_cY}\right) \times g\left(\mathrm{RR}_{U_sY\mid A=1}, \mathrm{RR}_{SU_s\mid A=1}\right) \times g\left(\mathrm{RR}_{U_sY\mid A=0}\mathrm{RR}_{SU_s\mid A=0}\right) \times \mathrm{OR}_{YA^*\mid a, S=1}$ |
| Unmeasured confounding $Y_a \perp\!\!\!\perp A \mid C, U_c$ | Exposure misclassification $\Pr(Y=0 \mid a, c) \approx 1$ | General selection bias $Y \perp\!\!\!\perp S \mid A^*, C, U_s$ | $g\left(\mathrm{RR}_{AU_c}, \mathrm{RR}_{U_cY}\right) \times \mathrm{OR}_{YA^*\mid a} \times g\left(\mathrm{RR}_{U_sY\mid A^*=1}, \mathrm{RR}_{SU_s\mid A^*=1}\right) \times g\left(\mathrm{RR}_{U_sY\mid A^*=0}\mathrm{RR}_{SU_s\mid A^*=0}\right)$ |
| Unmeasured confounding $Y_a \perp\!\!\!\perp A \mid S=1, C, U_c, U_s$ | Selected population | Exposure misclassification $\Pr(Y=0 \mid a, c, S=1) \approx 1$ | $g\left(\mathrm{RR}_{AU_{sc}\mid S=1}, \mathrm{RR}_{U_{sc}Y\mid S=1}\right) \times \mathrm{OR}_{YA^*\mid a}$ |
| Unmeasured confounding $Y_a \perp\!\!\!\perp A \mid C, U_c$ | General selection bias $Y \perp\!\!\!\perp S \mid A, C, U_s$ | Exposure misclassification $\Pr(Y=0 \mid a, c, S=1) \approx 1$ $\Pr(A=0 \mid y, c, S=1) \approx 1$ | $g\left(\mathrm{RR}_{AU_c}, \mathrm{RR}_{U_cY}\right) \times g\left(\mathrm{RR}_{U_sY\mid A=1}, \mathrm{RR}_{SU_s\mid A=1}\right) \times g\left(\mathrm{RR}_{U_sY\mid A=0}\mathrm{RR}_{SU_s\mid A=0}\right) \times \mathrm{RR}_{YA^*\mid a, S=1}$ |
| Unmeasured confounding $Y_a \perp\!\!\!\perp A \mid C, U_c$ | Exposure misclassification $\Pr(Y=0 \mid a, c) \approx 1$ $\Pr(A=0 \mid y, c) \approx 1$ | General selection bias $Y \perp\!\!\!\perp S \mid A^*, C, U_s$ | $g\left(\mathrm{RR}_{AU_c}, \mathrm{RR}_{U_cY}\right) \times \mathrm{RR}_{YA^*\mid a} \times g\left(\mathrm{RR}_{U_sY\mid A^*=1}, \mathrm{RR}_{SU_s\mid A^*=1}\right) \times g\left(\mathrm{RR}_{U_sY\mid A^*=0}\mathrm{RR}_{SU_s\mid A^*=0}\right)$ |
| Unmeasured confounding $Y_a \perp\!\!\!\perp A \mid S=1, C, U_c, U_s$ | Selected population | Exposure misclassification $\Pr(Y=0 \mid a, c, S=1) \approx 1$ $\Pr(A=0 \mid y, c, S=1) \approx 1$ | $g\left(\mathrm{RR}_{AU_{sc}\mid S=1}, \mathrm{RR}_{U_{sc}Y\mid S=1}\right) \times \mathrm{RR}_{YA^*\mid a}$ |

The first three columns show different combinations and ordering of the biases, as well as the implied assumptions. The fourth column contains the expression that bounds the bias when those assumptions hold. The definitions of the parameters are given in the main text.

misclassification occur may depend on the study design. We could alternatively derive a bound that depends on a parameter describing the extent to which the outcome is misclassified in the total population, and on others describing how selection is associated with the misclassified outcome. These parameters may be more intuitive in a study with case-control sampling. Additionally, another ordering of the biases may be preferable if data exist to justify estimates of the alternative parameters. We define those alternative parameters and derive that bound in the eAppendix; http://links.lww.com/EDE/B824. The second row of Table 1 summarizes those results; there, the assumption for selection bias is an assumption about the misclassified outcome, and the bound in the final column is defined in terms of parameters that reflect that ordering.

## Example

We illustrate the use of the multiple-bias bound to assess possible bias in the study by Omoni and colleagues regarding the effect of HIV status on wasting.[29] Wasting is defined by weight-for-length Z score of –2 or below and is a rare outcome, so we can interpret the reported OR of 6.75 (95% CI, 2.79, 16.31) as an approximate risk ratio. Since we have no reason to believe that misclassification of wasting was differential by exposure status (i.e., child or mother HIV status) status, and nondifferential outcome misclassification would on average bias toward the null in this situation,[9] we will focus on unmeasured confounding and selection bias in this example.

The choice of whether to participate in the trial, and therefore in the analysis in question, may have been influenced by prior maternal HIV status. For example, people with HIV infection may be hesitant to enroll owing to stigma regarding infection or fear of confirming their status. Other factors may affect enrollment as well: parents with food insecurity may be more likely to enroll in a vitamin-supplementation trial than those without if they think it will improve their children's nutrition. This benefit could outweigh the hesitancy for some, resulting in selection bias: if a mother in the study is living with HIV, it is likely that her family is also food insecure, making her child more at risk of wasting. Participation in the trial is therefore a collider in a directed acyclic graph describing these relationships, as shown in Figure 1. Similarly, there are factors that are associated with HIV status that may also affect wasting; if these are not on the causal pathway, we may be worried about unmeasured confounding. The authors did not adjust for parity or marital status, although they report that primiparous women were less likely to have HIV, as were married women. We may be concerned that children in single-parent households and those with more siblings are at higher risk of wasting. To demonstrate interpretation of the bound, we will propose values for the parameters describing the strength of these relationships based on the data presented in the original article as well as our background knowledge.

Suppose that the most vulnerable in the population were more likely to participate in the trial, and thus that wasting

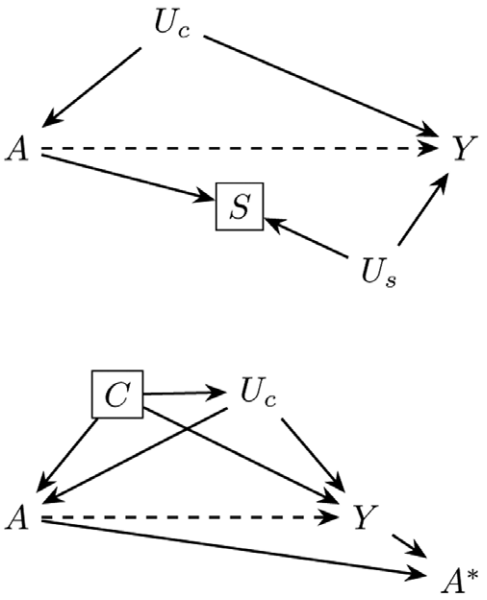


**FIGURE** 1. DAGs depicting the examples described in the text. A, This DAG depicts unmeasured confounding (due to $U_c$) and selection bias (due to $U_s$). In the graph, the assumptions $Y_a \coprod A \mid U_s, S = 1$ and $Y_a \coprod A \mid U_c$ are met. This corresponds to the first example in the text, where $A$ indicates HIV infection, $U_c$ family factors including parity and marital status, $S$ participation in the trial, $U_s$ food insecurity, and $Y$ wasting. B, This DAG depicts unmeasured confounding (due to $U_c$) and differential misclassification of the exposure (due to the $Y \rightarrow A^*$ edge). In the graph, the assumption $Y_a \coprod A \mid C, U_c$ is met. This corresponds to the second example in the text, where $A$ indicates vitamin consumption during pregnancy, $A^*$ reported vitamin consumption, $U_c$ breastfeeding, $C$ maternal age, race, and education, and $Y$ child leukemia. DAGs indicates directed acyclic graphs.

is more likely in children of participants than of nonparticipants, both among those with HIV as well as those without. The assumption that the outcome is more likely in the selected population of both exposure groups allows us to simplify the selection bias component of the bound, so that the bounding factor only relies on two selection terms, as described by Smith and VanderWeele.[24] Suppose now that children of the most food-insecure mothers are 3 times as likely to have extremely low weight-for-length scores than the least likely group so that $\mathrm{RR}_{U_s Y \mid A=1} = 3$ and that the mothers with HIV infection in the study compared with those not in the study are twice as likely to be food insecure so that $\mathrm{RR}_{SU_s \mid A=1} = 2$.

Although the odds ratios from this study were not adjusted for parity and marital status, the authors reported proportions of these characteristics stratified by exposure,[29] which can aid in coming up with a reasonable value for $\mathrm{RR}_{AU_c}$. For example, suppose we estimate that 3% of the women whose infants are infected with HIV are multiparous and unmarried, but that this is true of 7% of the women without HIV. If this is

the family situation with the largest disparity between exposure groups, then we can specify $\mathrm{RR}_{AU_c} = 2.3$. Now suppose that children in these most precarious families have 2.5 times the risk of wasting than those in the least precarious, so that $\mathrm{RR}_{U_cY} = 2.5$.

Then, we can calculate the bound as $\frac{3\times 2}{3+2-1}\times\frac{2.3\times 2.5}{2.3+2.5-1} = 2.27$. If those are the only sources of selection bias and unmeasured confounding, and there is no measurement error, then this amount of bias cannot fully explain the approximate observed $\mathrm{RR}^{\mathrm{obs}}_{AY}$ of 6.75, since 6.75/2.27 = 2.97. Of course, this observed value is subject to statistical uncertainty, so we can also consider the lower limit of the confidence interval, 2.79. If the proposed parameter values hold, then even in the worst-case scenario, $\mathrm{RR}^{\mathrm{true}}_{AY}$ is still consistent with $2.79/2.27 = 1.23$, an increase of about 23% in risk of wasting at 2 years of age owing to HIV infection. However, the parameter values, we used represent only a single reasonable choice. We might proceed by exploring a range of values, as we demonstrate below when we introduce software.

## EXPOSURE MISCLASSIFICATION

When differential exposure misclassification is a concern, we can derive a similar bound under similar assumptions. However, unlike the bound for outcome misclassification, the bound for exposure misclassification that is used applies to the odds ratio, not the risk ratio, and the sensitivity parameters are also not themselves risk ratios.[25] We therefore cannot factor the observed risk ratio as in the previous section. However, for a sufficiently rare outcome, odds ratios approximate risk ratios, which allows for some progress.

In this section, $\mathrm{RR}^{\mathrm{obs}}_{A^*Y} = \frac{\Pr(Y=1\mid A^*=1,S=1,c)}{\Pr(Y=1\mid A^*=0,S=1,c)}$ refers to the observed (approximate) risk ratio under exposure misclassification, when the outcome is rare in the selected population. Denote with $\mathrm{OR}_{A^*Y|y,S=1}$ the largest out of the false-positive odds ratio $\{f_1'/f_0'\}/\{(1-f_1')/(1-f_0')\}$, the sensitivity odds ratio $\{s_1'/s_0'\}/\{(1-s_1')/(1-s_0')\}$, the correct classification ratio $\{s_1'/s_0'\}/\{(1-f_1')/(1-f_0')\}$, and incorrect classification ratio $\{f_1'/f_0'\}/\{(1-s_1')/(1-s_0')\}$, where $f_y' = \Pr(A^*=1\mid Y=y,A=0,S=1,c)$ and $s_y' = \Pr(A^*=1\mid Y=y,A=1,S=1,c)$. Then the following bound holds approximately; that is, to the extent that the odds ratio approximates the risk ratio.

**Result 2:**

If $\Pr(Y=0\mid A^*=a,S=1,c)\approx 1$

and $\Pr(Y=0\mid A=a,S=1,c)\approx 1$

then:

$$\mathrm{RR}^{\mathrm{obs}}_{A^*Y}/\mathrm{RR}^{\mathrm{true}}_{AY} \le \mathrm{BF}_m\times\mathrm{BF}_s\times\mathrm{BF}_c$$

where $\mathrm{BF}_m = \mathrm{OR}_{A^*Y|y,S=1}$ and $\mathrm{BF}_s$ and $\mathrm{BF}_c$ are as previously defined.

This result is summarized in the fourth row of Table 1, as are extensions involving exposure misclassification.

## Example

We can jointly assess the magnitude of bias owing to differential recall of vitamin use and unmeasured confounding in the study of leukemia risk by Ross and colleagues, in which $\mathrm{RR}^{\mathrm{obs}}_{A^*Y} = 0.51$ (95% CI 0.30, 0.89), by proposing realistic values for the bias parameters. A probabilistic bias analysis for misclassification was previously done in relation to this study, in which Jurek et al conducted a literature search for validation studies of multivitamin use during the periconceptional period.[34] They found no pertinent articles and instead used expert knowledge and bounds from the data (e.g., by assuming correct classification is better than chance) to propose distributions for false negative and false-positive probabilities for the cases and controls, which we can use to inform our choice of parameters. Because we think the case-control differential in false negatives is stronger than that for false positives, we might choose that $\Pr(A^*=0\mid Y=1,A=1)=0.15$ and $\Pr(A^*=0\mid Y=0,A=1)=0.1$ to compute $\mathrm{BF}_m'$. Since we are dealing with a possibly protective factor, however, and the bound is greater than 1 by definition, we reverse the coding of the exposure to reflect that the original estimate of $\mathrm{RR}^{\mathrm{obs}}_{A^*Y}=0.51$ represents a $1/0.51 = 1.96$-fold increase in risk associated with *not* taking vitamins. Therefore, $f_1' = 0.15$ and $f_0' = 0.10$, and $\mathrm{BF}_m' = 1.59$.

Jurek et al's probabilistic bias analysis used the crude 2-by-2 table from the original article, so did not take into account even the few measured confounders.[34] However, even those measured confounders would likely not be sufficient to control for confounding by healthy lifestyle, as there is evidence that other healthy behaviors are associated with leukemia. For example, a recent meta-analysis found that not breastfeeding compared with breastfeeding for at least 6 months was associated with an increase in acute lymphoblastic leukemia risk by a factor of 1.22.[35] Using breastfeeding as a proxy for healthy lifestyle, for the unmeasured confounding parameters, we will take $\mathrm{RR}_{U_cY} = 1.22$ and $\mathrm{RR}_{AU_c} = 2$, suggesting that children who were not breastfed are 1.22 times as likely to get leukemia, and that mothers who take multivitamins are twice as likely to breastfeed than those who do not. A directed acyclic graph depicting this example is shown in Figure 1.

Using these values, we find that $1.59\times\frac{1.22\times 2}{1.22+2-1} = 1.75$, indicating that the observed risk ratio may be biased by a factor of 1.75 if the differential misclassification and unmeasured confounding were of the strengths we proposed. Since we are dealing with a possible protective factor, we multiply

the observed estimate of 0.51 and its confidence interval (95% CI 0.30, 0.89) by the bound (or equivalently divide the reverse-coded estimate of 1.96 by the bound), resulting in a bias-adjusted estimate and confidence interval of 0.89 (95% CI 0.52, 1.56). Unlike the Jurek et al sensitivity analysis,[34] which found that results were largely unchanged by exposure misclassification, we have focused specifically on a situation in which misclassification is differential by outcome, and have additionally taken both measured and unmeasured confounding into account. Doing so indicates that the results may be sensitive to misclassification and uncontrolled confounding, as can be seen if the chosen parameter values are thought to be reasonable.

## INFERENCE IN THE SELECTED POPULATION

Results 1 and 2 are derived with respect to the true causal effect in the total population, despite possible selection bias. In other situations, we may only be interested in the existence and magnitude of a causal effect in the selected population. In this case, our estimand of interest is $\mathrm{RR}^{\mathrm{true}}_{AY|S=1} = \frac{\Pr(Y_1 \mid S=1, c)}{\Pr(Y_0 \mid S=1, c)}$. If only selection bias is present, one can derive a bound under the assumption that $Y_a \coprod A \mid S=1, c, U_s$.[24] In the present context, we additionally accommodate unmeasured confounding and measurement error. Consider unmeasured confounding by $U_c$ such that it is only the case that $Y_a \coprod A \mid S=1, c, U_s, U_c$. Therefore, we must consider the vector of factors causing selection bias and unmeasured confounding $U_{sc} = (U_s, U_c)$. Define the sensitivity parameters $\mathrm{RR}_{U_{sc}Y} = \max_a \frac{\max_u \Pr(Y=1 \mid A=a, c, U_{sc}=u)}{\min_u \Pr(Y=1 \mid A=a, c, U_{sc}=u)}$ and $\mathrm{RR}_{AU_{sc}} = \max_u \frac{\Pr(U_{sc}=u \mid A=1, c)}{\Pr(U_{sc}=u \mid A=0, c)}$. Then under outcome misclassification, we have the following bound.

**Result 3:**

If $Y_a \coprod A \mid S=1, c, U_c, U_s$
then:

$$\mathrm{RR}^{\mathrm{obs}}_{AY^*} / \mathrm{RR}^{\mathrm{true}}_{AY|S=1} \leq \mathrm{BF}_m \times \mathrm{BF}_{sc}$$

where $\mathrm{BF}_m$ is defined as in Result 1, and $\mathrm{BF}_{sc} = g\left(\mathrm{RR}_{U_{sc}Y}, \mathrm{RR}_{AU_{sc}}\right)$. These latter parameters now refer to the maximum risk ratio for the outcome among the selected comparing any two levels of any of $U_s$ and $U_c$, and the maximum ratio for any joint level of $U_s$ and $U_c$ comparing exposed to unexposed, among the selected. This bound holds under exposure misclassification with a rare outcome in the selected population as well, with $\mathrm{BF}'_m = \mathrm{OR}_{A^*Y|y,S=1}$.

This result is summarized in the third row of Table 1.

## SOFTWARE

The R package EValue[36] allows for easy calculation of the multiple-bias bounds for various combinations of biases and assumptions, including all those presented in Table 1, as well as the possible simplifications to the selection bias bound, as in the first example. The function *multi_bias()* creates a set of biases according to the user's specifications. The user can then input this object along with a proposed set of parameter values to the *multi_bound()* function to calculate a bound.

For example, the biases in the HIV example can be set with *HIV_biases <-multi_bias(confounding(),selection("general", "increased risk"))*. The command to calculate the bound is then *multi_bound(biases = HIV_biases, RRAUc = 2.3, RRUcY = 2.5, RRUsYA1 = 3, RRSUsA1 = 2)*. Similarly, for the vitamins-leukemia example, the biases are set with *leuk_biases <- multi_bias(confounding(), misclassification("exposure", rare_outcome = TRUE, rare_exposure = FALSE))* and the bound command is *multi_bound(biases = leuk_biases, RRAUc = 2, RRUcY = 1.22, ORYAa = 1.59)*.

These functions can be used to prepare a table or figure of bounded bias-adjusted estimates across a range of proposed parameter values. For example, Figure 2 shows the upper bound for the estimate of the protective effect of multivitamin use on leukemia across various values for $\mathrm{RR}_{U_cY}$ and $\mathrm{RR}_{AU_c}$ (in the columns and rows), and for two values of the misclassification ratio $\mathrm{OR}_{A^*Y|y}$ above and below the diagonal. We can use this table to describe multiple scenarios under which it would be possible for the true effect to be null. We can also see that even if, for example, the prevalence of the unmeasured confounder, or set of confounders, differs greatly between consumers and nonconsumers of multivitamins (e.g., $\mathrm{RR}_{AU_c} = 3$), a relatively small association between the unmeasured confounder and leukemia (e.g., $\mathrm{RR}_{U_cY} = 1.25$) and between vitamin use and misclassifcation (e.g., $\mathrm{OR}_{A^*Y|y} = 1.25$) would at most lead to a bias-adjusted estimate of 0.74.

More examples are available in the eAppendix; http://links.lww.com/EDE/B824, and the package documentation is available online.

## DISCUSSION

We have described an approach to sensitivity analysis that we hope can help bridge the gap between complex methods that require specifying many parameters and making restrictive assumptions, and simpler methods that allow for assessment of only one type of bias at a time. The multiple-bias bound can be used to simultaneously consider the possible effects of biases that are of different strengths. Researchers can propose values for the parameters based on background knowledge, validation studies, or simply hypothetical situations, and assess the minimum possible true risk ratio that would be compatible if the observed value were affected by

| $RR_{AU_c}$ \ $RR_{U_cY}$ | 1.25 | 1.5 | 1.75 | 2 | 2.25 | 2.5 | 2.75 | 3 |
|---|---|---|---|---|---|---|---|---|
| 1.25 | 0.80<br>0.66 | 0.82 | 0.84 | 0.85 | 0.86 | 0.87 | 0.88 | 0.88 |
| 1.5 | 0.68 | 0.86<br>0.72 | 0.89 | 0.92 | 0.94 | 0.96 | 0.97 | 0.98 |
| 1.75 | 0.70 | 0.74 | 0.94<br>0.78 | 0.97 | 1.00 | 1.03 | 1.05 | 1.07 |
| 2 | 0.71 | 0.76 | 0.81 | 1.02<br>0.85 | 1.06 | 1.09 | 1.12 | 1.15 |
| 2.25 | 0.72 | 0.78 | 0.84 | 0.88 | 1.11<br>0.92 | 1.15 | 1.18 | 1.22 |
| 2.5 | 0.72 | 0.80 | 0.86 | 0.91 | 0.96 | 1.20<br>1.00 | 1.24 | 1.27 |
| 2.75 | 0.73 | 0.81 | 0.88 | 0.94 | 0.99 | 1.03 | 1.29<br>1.07 | 1.33 |
| 3 | 0.74 | 0.82 | 0.89 | 0.96 | 1.01 | 1.06 | 1.11 | 1.38<br>1.15 |

**FIGURE** 2. Corrected estimates for the effect of multivitamin use in pregnancy on childhood leukemia, taking into account unmeasured confounding and recall bias. The original estimate was 0.51. Corrected estimates are arranged in rows and columns by the parameters defining the unmeasured confounding, $RR_{AUc}$ and $RR_{UcY}$. The two parameters are interchangeable with respect to the bound, so a table of estimates corrected only for unmeasured confounding would be symmetric. However, the estimates in the upper and lower triangles have been corrected by misclassification parameters of different magnitudes. Below the diagonal, the misclassification ratio is assumed to be 1.25; above the diagonal, it is assumed to be 1.5.

biases of that magnitude. When planning for future research, the bound can be used to compare the effects of biases within a given situation and prioritize more extensive confounder assessment, a more valid sampling or inclusion scheme, and better measurement techniques if resource constraints or data collection options forced one to choose among them. It may also show that certain improvements to study design are futile; if the amount of an unavoidable bias greatly attenuates the anticipated risk ratio estimate, investing resources into reducing another type of bias may not be worth it.[37,38]

There are a number of caveats and limitations to this approach. Although the calculations involved in our approach are simple, the entire process of assessing bias should not be. Importantly, it should be specific to the study design, the available data, and the research question; values for the sensitivity parameters are meaningless without a frame of reference. Indeed, critiques of the E value for unmeasured confounding have emphasized the importance of clearly specifying the confounder, or set of confounders, that have not been measured.[39–41] The same should be true for factors potentially causing selection bias or the reason behind possible differential misclassification. Unmeasured confounders could be anything from a single missing risk factor to the "ultimate covariate,"[42] the variable encoding an individual's causal type. Misclassification may be negligible or close to nondifferential, or as bad as chance in one or another group; it is up to researchers and readers to assess the plausibility of these situations with respect to a given study and what was conditioned on in the analysis and then assess how much bias they would create. Like any tool, the multiple-bias bound can be misused; we encourage researchers to not be careless, owing to its apparent simplicity, but rather to be thoughtful in its use.

Additionally, in avoiding certain assumptions, we have necessarily invoked others. In particular, the bounds we propose describe a "worst-case scenario" for the bias; in almost all settings, the actual bias will be smaller than the bound. For example, for the actual bias to obtain the bound assumes that the unmeasured confounder has the distribution that maximizes confounding, given the two parameters defining it.[43] The same is true of selection bias and misclassification; for example, the general selection bias bound implies that outcomes and exposures in the nonselected group are distributed to result in the most possible bias. This is of course necessary for a bound to be a bound, but many realistic conditions would not result in as much bias, and the bound should be interpreted as the bias that *could* result from parameters of a given magnitude not that necessarily *would* result. The few assumptions that are required for the bound to hold may not be reasonable in all settings; for example, the general selection bias assumption is unlikely to hold in case-control studies. In addition, the interpretation of the bias-adjusted confidence interval pertains to the application of the adjustment to repeated samples with the same sets of biases; if the biases were truly resolved in the design or analysis, the bounds would differ.

Finally, although we have suggested two possible orderings for factoring the bias, others that take into account, for example, misclassification that is also differential by an unmeasured confounder, are possible. We have presented results for risk ratios, which can in many cases be extended to odds ratios. However, our bound for exposure misclassification relies on a rare outcome assumption that limits its use and results in an approximate bound. Because we are never sure of the true values of the parameters that make up the bound, and the bound represents a worst-case scenario not likely to hold anyway, this approximation is not likely to meaningfully affect interpretation. Further work could be done to extend this approach to risk differences or mean differences, which may be especially challenging because the bounds are more frequently noninformative.[44] Other approaches exist to quantify as simply as possible unmeasured confounding in linear or probit models,[45–48] but to our knowledge, they have not yet been extended to multiple biases.

There is no single solution to the problem of bias in epidemiologic research. Some biases can be corrected at the design phase, others in the main analysis, but the assessment of what bias may remain should be a regular component of any study that attempts to quantify causal effects. The multiple-bias bound can make it simpler to do so, and we hope to encourage thoughtful consideration of multiple sources of bias in epidemiologic research.

## REFERENCES


1. Ioannidis JP. Limitations are not properly acknowledged in the scientific literature. *J Clin Epidemiol*. 2007;60:324–329.
2. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology*. 2003;14:459–466.
3. Lash TL. Heuristic thinking and inference from observational epidemiology. *Epidemiology*. 2007;18:67–72.
4. Cornfield J, Haenszel W, Hammond EC, et al. Smoking and lung cancer: Recent evidence and a discussion of some questions. *J Natl Cancer Inst*. 1959;22:173–203.
5. Bross ID. Spurious effects from an extraneous variable. *J Chronic Dis*. 1966;19:637–647.
6. Bross ID. Pertinency of an extraneous variable. *J Chronic Dis*. 1967;20:487–495.
7. Schlesselman JJ. Assessing effects of confounding variables. *Am J Epidemiol*. 1978;108:3–8.
8. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
9. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol*. 1977;105:488–495.
10. Barron BA. The effects of misclassification on the estimation of relative risk. *Biometrics*. 1977;33:414–418.
11. Greenland S, Neutra R. An analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. *J Chronic Dis*. 1981;34:433–438.
12. Greenland S, Kleinbaum DG. Correcting for misclassification in two-way tables and matched-pair studies. *Int J Epidemiol*. 1983;12:93–97.
13. Lash TL, Silliman RA. A sensitivity analysis to separate bias due to confounding from bias due to predicting misclassification by a variable that does both. *Epidemiology*. 2000;11:544–549.
14. Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc*. 2003;98:47–54.
15. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology*. 2003;14:451–458.
16. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol*. 2005;34:1370–1376.
17. Greenland S. Multiple-bias modelling for analysis of observational data (with discussion). *J R Stat Soc Ser A Stat Soc*. 2005;168:267–306.
18. Lash TL, Schmidt M, Jensen AØ, Engebjerg MC. Methods to apply probabilistic bias analysis to summary estimates of association. *Pharmacoepidemiol Drug Saf*. 2010;19:638–644.
19. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. Dordrecht; New York: Springer; 2009.
20. Orsini N, Bellocco R, Bottai M, et al. A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. *Stata J*. 2008;8:29–48.
21. Hunnicutt JN, Ulbricht CM, Chrysanthopoulou SA, Lapane KL. Probabilistic bias analysis in pharmacoepidemiology and comparative effectiveness research: a systematic review. *Pharmacoepidemiol Drug Saf*. 2016;25:1343–1353.
22. Lash TL, Abrams B, Bodnar LM. Comparison of bias analysis strategies applied to a large data set. *Epidemiology*. 2014;25:576–582.
23. Ding P, VanderWeele TJ. Sensitivity analysis without assumptions. *Epidemiology*. 2016;27:368–377.
24. Smith LH, VanderWeele TJ. Bounding bias due to selection. *Epidemiology*. 2019;30:509–516.
25. VanderWeele TJ, Li Y. Simple sensitivity analysis for differential measurement error. *Am J Epidemiol*. 2019;188:1823–1829.
26. Chin HB, Baird DD, McConnaughey DR, Weinberg CR, Wilcox AJ, Jukic AM. Long-term recall of pregnancy-related events. *Epidemiology*. 2017;28:575–579.
27. Greene N, Greenland S, Olsen J, Nohr EA. Estimating bias from loss to follow-up in the Danish National Birth Cohort. *Epidemiology*. 2011;22:815–822.
28. Mumford SL, Yeung EH. Intergenerational effects-causation or confounding? *Fertil Steril*. 2018;110:52–53.
29. Omoni AO, Ntozini R, Evans C, et al. Child growth according to maternal and child HIV status in Zimbabwe. *Pediatr Infect Dis J*. 2017;36:869–876.
30. Ross JA, Blair CK, Olshan AF, et al. Periconceptional vitamin useand leukemia risk in children with down syndrome: a Children's Oncology Group study. *Cancer*. 2005;104:405–410.
31. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med*. 2017;167:268–274.

32. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol*. 1996;25:1107–1116.
33. Maclure M, Schneeweiss S. Causation of bias: the episcope. *Epidemiology*. 2001;12:114–122.
34. Jurek AM, Maldonado G, Spector LG, Ross JA. Periconceptional maternal vitamin supplementation and childhood leukaemia: an uncertainty analysis. *J Epidemiol Community Health*. 2009;63:168–172.
35. Amitay EL, Keinan-Boker L. Breastfeeding and childhood leukemia incidence: a meta-analysis and systematic review. *JAMA Pediatr*. 2015;169:1–9.
36. Mathur MB, Ding P, Riddell CA, VanderWeele TJ. Web site and R package for computing E-values. *Epidemiology*. 2018;29:e45–e47.
37. Lash TL, Ahern TP. Bias analysis to guide new data collection. *Int J Biostat*. 2012;8:/j/ijb.2012.8.issue–/j/ijb.2012.82/1557.
38. Fox MP, Lash TL. Quantitative bias analysis for study and grant planning. *Ann Epidemiol*. 2020;43:32–36.
39. Blum MR, Tan YJ, Ioannidis JPA. Use of E-values for addressing confounding in observational studies-an empirical assessment of the literature. *Int J Epidemiol*. 2020;49:1482–1494.
40. Fox MP, Arah OA, Stuart EA. Commentary: the value of E-values and why they are not enough. *Int J Epidemiol*. 2020;49:1505–1506.
41. VanderWeele TJ, Mathur MB. Commentary: developing best-practice guidelines for the reporting of E-values. *Int J Epidemiol*. 2020;49:1495–1497.
42. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol*. 1986;15:413–419.
43. VanderWeele TJ, Ding P, Mathur M. Technical considerations in the use of the E-value. *J Causal Inference*. 2019;7:1–11.
44. Ding P, Vanderweele TJ. Generalized Cornfield conditions for the risk difference. *Biometrika*. 2014;101:971–977.
45. Frank KA. Impact of a confounding variable on a regression coefficient. *Sociol Methods Res*. 2000;29:147–194.
46. Altonji JG, Elder TE, Taber CR. Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools. *J Polit Econ*. 2005;113:151–184.
47. Oster E. Unobservable selection and coefficient stability: theory and evidence. *J Bus Econ Stat*. 2019;37:187–204.
48. Cinelli C, Hazlett C. Making sense of sensitivity: extending omitted variable bias. *J R Stat Soc Ser B Stat Methodol*. 2019;82:39–67.