Sensitivity Analysis Without Assumptions

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

Sensitivity Analysis Without Assumptions

Peng Ding[a] *and Tyler J. VanderWeele*[b]

Abstract: Unmeasured confounding may undermine the validity of causal inference with observational studies. Sensitivity analysis provides an attractive way to partially circumvent this issue by assessing the potential influence of unmeasured confounding on causal conclusions. However, previous sensitivity analysis approaches often make strong and untestable assumptions such as having an unmeasured confounder that is binary, or having no interaction between the effects of the exposure and the confounder on the outcome, or having only one unmeasured confounder. Without imposing any assumptions on the unmeasured confounder or confounders, we derive a bounding factor and a sharp inequality such that the sensitivity analysis parameters must satisfy the inequality if an unmeasured confounder is to explain away the observed effect estimate or reduce it to a particular level. Our approach is easy to implement and involves only two sensitivity parameters. Surprisingly, our bounding factor, which makes no simplifying assumptions, is no more conservative than a number of previous sensitivity analysis techniques that do make assumptions. Our new bounding factor implies not only the traditional Cornfield conditions that both the relative risk of the exposure on the confounder and that of the confounder on the outcome must satisfy but also a high threshold that the maximum of these relative risks must satisfy. Furthermore, this new bounding factor can be viewed as a measure of the strength of confounding between the exposure and the outcome induced by a confounder.

(*Epidemiology* 2016;27: 368–377)

Causal inference with observational studies is of great interest and importance in many scientific disciplines. Although unmeasured confounding between the exposure and the outcome may bias the estimation of the true causal effect, an approach often called "sensitivity analysis" or "bias analysis" over a range of sensitivity parameters sometimes allows researchers to make

Correspondence: Peng Ding, Department of Statistics, University of California, 425 Evans Hall, Berkeley, CA 94720. E-mail: pengdingpku@berkeley.edu.

causal inferences even without full control of the confounders of the relationship between the exposure and outcome.

Sensitivity analysis plays a central role in assessing the influence of the unmeasured confounding on the causal conclusions. However, many sensitivity analysis techniques often require additional untestable assumptions. For instance, some authors assume a single binary confounder.[1–6] Researchers also often assume a homogeneity assumption that there is no interaction between the effects of the exposure and the confounder on the outcome.[5–9] Some sensitivity analysis techniques only allow one to assess how strong an unmeasured confounder would have to be to completely explain away an effect[1–3,10,11] but do not allow one to assess what the effect estimate might be under weaker unmeasured confounding scenarios (i.e., do not allow one to do sensitivity analysis under alternative hypotheses). Performing sensitivity analysis under alternative hypotheses can be quite challenging due to more parameters needed in the sensitivity analysis. The Cornfield et al. early study[1] on sensitivity analysis for the cigarette smoking and lung cancer association, which helped initiate the entire field of sensitivity analysis, in fact made all three simplifying assumptions: a single binary confounder, no interaction, and only sensitivity analysis for the null hypothesis of no causal effect. Although some sensitivity analysis results exist for general confounders,[8,12] they are only easy to implement under some of the above simplifying assumptions.

In this article, we propose a new bounding factor and sensitivity analysis technique without any assumptions about the unmeasured confounder or confounders. None of the assumptions of the null hypothesis, a single binary confounder, or no interaction is required for using the bounding factor. Nonetheless, our new bounding factor, which makes no simplifying assumptions, is no more conservative than many previous sensitivity analysis techniques that do make assumptions and is furthermore easy to implement. Moreover, we show that the new bounding factor implies not only the classical Cornfield conditions[1] that both the relative risk of the exposure on the confounder and that of the confounder on the outcome must satisfy but also a stronger condition that the maximum of these relative risks must satisfy. The new bounding factor can be viewed as a measure of the strength of confounding between the exposure and the outcome resulting from the confounder. We begin by considering outcomes that are binary and extend our results further to time-to-event and non-negative count or continuous outcomes. We consider both ratio and difference scales.

The claim that our technique is "without assumptions" requires some clarification. As we will see below, we will, without any assumptions, be able to make statements of the form: "For an observed association to be due solely to unmeasured confounding, two sensitivity analysis parameters must satisfy [a specific inequality]." We will also, without assumptions, be able to make statements of the form: "For unmeasured confounding alone to be able to reduce an observed association [to a given level], two sensitivity analysis parameters must satisfy [another specific inequality]." We believe the ability to make statements of this form without imposing any specific structure on the nature of the unmeasured confounder or confounders constitutes a major advance in the literature.

However, if statements are made of the form, "If the sensitivity analysis parameter take [specified values], then such unmeasured confounding can reduce the observed estimate by no more than [a specific level]," then the specification of the sensitivity analysis parameters could itself of course be viewed as an assumption. Moreover, when placing the results within a counterfactual or potential outcomes framework, the assumptions implicit within that framework of course would be needed also to give a potential outcomes interpretation to the sensitivity analysis. Thus, certain types of statements concerning the sensitivity of conclusions to unmeasured confounding can be made "without assumptions," while other types of statements do require assumptions concerning the specification of the sensitivity analysis parameters themselves, or those implicit within the potential outcomes framework.

Our title perhaps merits one further qualification which is that what is called in this article "sensitivity analysis" is generally now referred to as "bias analysis" in the epidemiologic literature. Moreover, such "bias analysis" is relevant not only to problems of unmeasured confounding but also measurement error and selection bias, and our focus in this article only concerns unmeasured confounding. The term "sensitivity analysis" is, however, still employed in statistics, econometrics, and in many of the social sciences for issues of unmeasured confounding. We believe the technique presented in this article will be useful across this range of disciplines and have chosen to use the broader term, while acknowledging that terminology in epidemiology has shifted.

SENSITIVITY ANALYSIS: A NEW BOUNDING FACTOR

Let E denote the exposure, D denote a binary outcome, C denote the measured confounders, and U denote one or more unmeasured confounders. We will assume for what follows that the exposure E is binary, but all of the results on sensitivity analysis below are also applicable to a categorical or continuous exposure and could be applied comparing any two levels of E. For ease of notation, we assume that the unmeasured confounder U is categorical with levels $0, 1, \ldots, K-1$. But all the conclusions hold for U of general type (categorical,

continuous, or mixed; single or multiple confounders). We provide proofs and theoretical technical details for general U in the eAppendix (http://links.lww.com/EDE/B16).

Let $\mathrm{RR}^{\mathrm{obs}}_{ED|c} = P(D=1|E=1,C=c)/P(D=1|E=0,C=c)$ denote the observed relative risk of the exposure E on the outcome D within stratum of measured confounders $C = c$. Define $\mathrm{RR}_{EU,k|c} = P(U=k|E=1,C=c)/P(U=k|E=0,C=c)$ as the relative risk of exposure on category k of the unmeasured confounder within stratum of measured confounders $C = c$. We use $\mathrm{RR}_{EU|c} = \max_k \mathrm{RR}_{EU,k|c}$ to denote the maximum of these relative risks between E and U, which we will call the maximal relative risk of E on U within stratum $C = c$. Define

$$\mathrm{RR}_{UD|E=0,c} = \frac{\max_k P(D=1|E=0,C=c,U=k)}{\min_k P(D=1|E=0,C=c,U=k)}$$

as the maximum of the effect of U on D among the unexposed comparing any two categories of U (i.e., the ratio of the maximum and minimum of the probabilities of the outcome over strata of U without exposure and within stratum $C = c$); similarly, define

$$\mathrm{RR}_{UD|E=1,c} = \frac{\max_k P(D=1|E=1,C=c,U=k)}{\min_k P(D=1|E=1,C=c,U=k)}$$

as the maximum of the effect of U on D among the exposed comparing any two categories of U (i.e., the ratio of the maximum and minimum of the probabilities of the outcome over strata of U with exposure and within stratum $C = c$). We use $\mathrm{RR}_{UD|c} = \max(\mathrm{RR}_{UD|E=1,c}, \mathrm{RR}_{UD|E=0,c})$ to denote the maximum of the relative risks between U and D with and without exposure, defined as the maximal relative risk of U on D within stratum $C = c$. Note that if U is a vector that contains multiple unmeasured confounders, then $\mathrm{RR}_{EU|c}$ and $\mathrm{RR}_{UD|c}$ are defined as the maximum relative risk comparing any two categories of the vector U.

If C and U suffice to control for confounding for the effect of E on D, the standardized relative risk

$$\mathrm{RR}^{\mathrm{true}}_{ED|c} = \frac{\sum_{k=0}^{K-1} P(D=1|E=1,C=c,U=k)P(U=k|C=c)}{\sum_{k=0}^{K-1} P(D=1|E=0,C=c,U=k)P(U=k|C=c)}$$

is the true causal relative risk of the exposure E on the outcome D within stratum $C = c$. In the main text, we focus the discussion on the whole population. We further show in eAppendix 2 (http://links.lww.com/EDE/B16) that all the conclusions also hold for exposed and unexposed subpopulations. We will for the next several sections assume all analyses are carried out within strata of C, and thus the condition $C = c$ is omitted and kept implicitly in all the conditional probabilities (e.g., $\mathrm{RR}^{\mathrm{obs}}_{ED|c}$ is replaced by $\mathrm{RR}^{\mathrm{obs}}_{ED}$ for notational simplicity). Later in the article, we will comment on how the results are applicable to estimation averaged over C, rather than conditional on C.

The relative risk pair $(\mathrm{RR}_{EU}, \mathrm{RR}_{UD})$ measures the strength of confounding between the exposure E and the outcome D induced by the confounder U. Our main result ties the ratio of the observed relative risk $\mathrm{RR}_{ED}^{\mathrm{obs}}$ adjusted only for measured confounders C and the true relative risk $\mathrm{RR}_{ED}^{\mathrm{true}}$ adjusted also for unmeasured confounders U, to the strength of confounding, $(\mathrm{RR}_{EU}, \mathrm{RR}_{UD})$. Without any assumptions, we have the following result:

Result 1:

$$\mathrm{RR}_{ED}^{\mathrm{true}} \geq \mathrm{RR}_{ED}^{\mathrm{obs}} \Big/ \frac{\mathrm{RR}_{EU} \times \mathrm{RR}_{UD}}{\mathrm{RR}_{EU} + \mathrm{RR}_{UD} - 1}. \tag{1}$$

Result 1 shows that even in the presence of unmeasured confounding, the true relative risk must be at least as large as $\mathrm{RR}_{ED}^{\mathrm{obs}} \Big/ \dfrac{\mathrm{RR}_{EU} \times \mathrm{RR}_{UD}}{\mathrm{RR}_{EU} + \mathrm{RR}_{UD} - 1}$. In eAppendix 2 (http://links.lww.com/EDE/B16), we provide a proof for Result 1 and also show that the inequality is sharp in the sense that we can always construct a model with a confounder U to attain the equality. The quantity $(\mathrm{RR}_{EU} \times \mathrm{RR}_{UD})/(\mathrm{RR}_{EU} + \mathrm{RR}_{UD} - 1)$ is a new joint bounding factor for the relative risk. Although quite simple, this bound using both RR_{EU} and RR_{UD} has several important implications.

First, the result essentially allows for sensitivity analysis without assumptions insofar as for an unmeasured confounder to reduce an observed estimated $\mathrm{RR}_{ED}^{\mathrm{obs}}$ to an actual relative risk of $\mathrm{RR}_{ED}^{\mathrm{true}}$ the sensitivity analysis parameters RR_{EU} and RR_{UD} must be sufficiently large to satisfy the inequality

$$\frac{\mathrm{RR}_{EU} \times \mathrm{RR}_{UD}}{\mathrm{RR}_{EU} + \mathrm{RR}_{UD} - 1} \geq \frac{\mathrm{RR}_{ED}^{\mathrm{obs}}}{\mathrm{RR}_{ED}^{\mathrm{true}}}.$$

This statement holds without any assumptions about the nature of the unmeasured confounder. One could plot those values of RR_{EU} and RR_{UD} that would be required to explain away the effect estimate (or the lower limit of a confidence interval). To conduct sensitivity analysis with prespecified strength of the unmeasured confounder, $(\mathrm{RR}_{EU}, \mathrm{RR}_{UD})$, we can divide the observed relative risk and its confidence limits by $(\mathrm{RR}_{EU} \times \mathrm{RR}_{UD})/(\mathrm{RR}_{EU} + \mathrm{RR}_{UD} - 1)$, to obtain a point estimate and confidence limits of the lower bound of the true causal effect of the exposure E on the outcome D. We will refer to the relative risk adjusted only for C, when divided by the bounding factor $(\mathrm{RR}_{EU} \times \mathrm{RR}_{UD})/(\mathrm{RR}_{EU} + \mathrm{RR}_{UD} - 1)$ as the corrected relative risk. It is "corrected" in the sense that an unmeasured confounder cannot reduce the relative risk any further than what is obtained by division by its bounding factor. As an example, suppose we have an observed relative risk of 2.1 with a 95% confidence interval $[1.4, 3.1]$. If we consider an unmeasured confounder with $(\mathrm{RR}_{EU}, \mathrm{RR}_{UD}) = (2,2)$, then the joint bounding factor is $2 \times 2/(2 + 2 - 1) = 1.33$,

and the corrected relative risk is $2.1/1.33 = 1.58$ with a 95% confidence interval $[1.4/1.33, 3.1/1.33] = [1.05, 2.33]$. Therefore, an unmeasured confounder with $(\mathrm{RR}_{EU}, \mathrm{RR}_{UD}) = (2, 2)$ cannot explain away the observed relative risk 2.1 or its lower confidence limit 1.4, i.e., it cannot reduce the point estimate and lower confidence limit of the relative risk to be smaller than 1. If we consider an unmeasured confounder with $(\mathrm{RR}_{EU}, \mathrm{RR}_{UD}) = (2.5, 3.5)$, then the joint bounding factor is $2.5 \times 3.5/(2.5 + 3.5 - 1) = 1.75$, and an estimate for the lower bound of the true causal relative risk is $2.1/1.75 = 1.20$ with a 95% confidence interval $[1.4/1.75, 3.1/1.75] = [0.8, 1.77]$. Although the confounder with $(\mathrm{RR}_{EU}, \mathrm{RR}_{UD}) = (2.5, 3.5)$ cannot explain away the observed relative risk of 2.1, it reduces the original lower confidence limit 1.4 to 0.8 (i.e., less than 1). Note that we are not merely assessing a binary confounder, and we are not imposing the no interaction assumption. Moreover, we are not restricted to only assessing how much confounding can explain away an effect nor are we even assuming that there is a single unmeasured confounder (since U can be a vector of unmeasured confounders). The corrected estimates and confidence intervals above are applicable irrespective of the underlying confounder (or confounders). We can apply the technique to obtain a range of values for the true causal effect under different specifications of RR_{EU} and RR_{UD}.

Table 1 shows the magnitudes of the joint bounding factor for different combinations of RR_{EU} and RR_{UD}. The entries in the table for the joint bounding factor are the largest observed relative risks that such an unmeasured confounder could explain away. We can see from the table that the joint bounding factor is always smaller than both of RR_{EU} and RR_{UD}, and much smaller than the maximum of them.

As a second important consequence of our main Result 1, we also show in eAppendix 2 (http://links.lww.com/EDE/B16) that once we specify one of the unmeasured confounding measures, for example RR_{EU}, then to be able to reduce an observed relative risk of $\mathrm{RR}_{ED}^{\mathrm{obs}}$ to a true causal relative risk of $\mathrm{RR}_{ED}^{\mathrm{true}}$ the other confounding measure RR_{UD} must be at least of the magnitude

$$\mathrm{RR}_{UD} \geq \frac{\mathrm{RR}_{EU} \times \mathrm{RR}_{ED}^{\mathrm{obs}} - \mathrm{RR}_{ED}^{\mathrm{obs}}}{\mathrm{RR}_{EU} \times \mathrm{RR}_{ED}^{\mathrm{true}} - \mathrm{RR}_{ED}^{\mathrm{obs}}}.$$

For an unmeasured confounder to completely explain away the relative risk, i.e., reduce $\mathrm{RR}_{ED}^{\mathrm{obs}}$ to $\mathrm{RR}_{ED}^{\mathrm{true}} = 1$, once we specify RR_{EU} the other unmeasured confounding measure much be at least of the magnitude

$$\mathrm{RR}_{UD} \geq \frac{\mathrm{RR}_{EU} \times \mathrm{RR}_{ED}^{\mathrm{obs}} - \mathrm{RR}_{ED}^{\mathrm{obs}}}{\mathrm{RR}_{EU} - \mathrm{RR}_{ED}^{\mathrm{obs}}}.$$

For example, if we have an observed relative risk $\mathrm{RR}_{ED}^{\mathrm{obs}} = 2.5$, and we specify the exposure–confounder

TABLE 1.   Magnitudes of the Joint Bounding Factor for Different Combinations of the Exposure–Confounder Association RR_{EU} and the Confounder–Outcome Association RR_{UD}

Bounding Factor	RR_{UD}											
	1.3	**1.5**	**1.8**	**2**	**2.5**	**3**	**3.5**	**4**	**5**	**6**	**8**	**10**
RR_{EU}												
1.3	1.06	1.08	1.11	1.13	1.16	1.18	1.20	1.21	1.23	1.24	1.25	1.26
1.5	1.08	1.12	1.17	1.20	1.25	1.29	1.31	1.33	1.36	1.38	1.41	1.43
1.8	1.11	1.17	1.25	1.29	1.36	1.42	1.47	1.50	1.55	1.59	1.64	1.67
2	1.13	1.20	1.29	1.33	1.43	1.50	1.56	1.60	1.67	1.71	1.78	1.82
2.5	1.16	1.25	1.36	1.43	1.56	1.67	1.75	1.82	1.92	2.00	2.11	2.17
3	1.18	1.29	1.42	1.50	1.67	1.80	1.91	2.00	2.14	2.25	2.40	2.50
3.5	1.20	1.31	1.47	1.56	1.75	1.91	2.04	2.15	2.33	2.47	2.67	2.80
4	1.21	1.33	1.50	1.60	1.82	2.00	2.15	2.29	2.50	2.67	2.91	3.08
5	1.23	1.36	1.55	1.67	1.92	2.14	2.33	2.50	2.78	3.00	3.33	3.57
6	1.24	1.38	1.59	1.71	2.00	2.25	2.47	2.67	3.00	3.27	3.69	4.00
8	1.25	1.41	1.64	1.78	2.11	2.40	2.67	2.91	3.33	3.69	4.27	4.71
10	1.26	1.43	1.67	1.82	2.17	2.50	2.80	3.08	3.57	4.00	4.71	5.26

association $RR_{EU} = 3$. Then to reduce the observed relative risk to a true causal relative risk $RR_{ED}^{true} = 1.5$, the confounder–outcome association must be at least as large as $(3 \times 2.5 - 2.5)/(3 \times 1.5 - 2.5) = 2.5$; to completely explain away the observed relative risk (i.e., to reduce the observed relative risk to $RR_{ED}^{true} = 1$), the confounder–outcome association must be at least as large as $(3 \times 2.5 - 2.5)/(3 - 2.5) = 10$. The symmetry of Result 1 implies that a similar result also holds for RR_{EU} with prespecified RR_{UD}.

Third, we show in eAppendix 2 (http://links.lww.com/EDE/B16) that if both the generalized relative risks RR_{EU} and RR_{UD} have the same magnitude, for an unmeasured confounder to reduce an observed relative risk of RR_{ED}^{obs} to a true causal relative risk of RR_{ED}^{true} both of the confounding relative risks must thus be at least as large as

$$RR_{EU} = RR_{UD} \geq \left\{ RR_{ED}^{obs} + \sqrt{RR_{ED}^{obs}(RR_{ED}^{obs} - RR_{ED}^{true})} \right\} \Big/ RR_{ED}^{true}.$$

For an unmeasured confounder to completely explain away an observed relative risk of RR_{ED}^{obs} (i.e., to reduce RR_{ED}^{obs} to a true causal relative risk of $RR_{ED}^{true} = 1$), both RR_{EU} and RR_{UD} must be at least as large as

$$RR_{EU} = RR_{UD} \geq RR_{ED}^{obs} + \sqrt{RR_{ED}^{obs}(RR_{ED}^{obs} - 1)}.$$

If one of the confounding relative risks is smaller than the lower bound above, we then know that the other one must be larger. Thus even if RR_{EU} and RR_{UD} are not of the same magnitude, the maximum of RR_{EU} and RR_{UD} must satisfy the inequality above. We then have the following "high threshold" condition:

$$\max(RR_{EU}, RR_{UD})$$
$$\geq \left\{ RR_{ED}^{obs} + \sqrt{RR_{ED}^{obs}(RR_{ED}^{obs} - RR_{ED}^{true})} \right\} \Big/ RR_{ED}^{true}.$$

For example, to reduce an observed relative risk of $RR_{ED}^{obs} = 2.5$ to a true causal relative risk of $RR_{ED}^{true} = 1.5$, the high threshold is $(2.5 + \sqrt{2.5 \times 1})/1.5 = 2.72$; at least one of RR_{EU} and RR_{UD} must be of magnitude 2.72 or above. To completely explain away an observed relative risk of $RR_{ED}^{obs} = 2.5$, the high threshold is $2.5 + \sqrt{2.5 \times 1.5} = 4.44$; at least one of RR_{EU} and RR_{UD} must be of magnitude 4.44 or higher to completely explain away the effect.

Fourth, the bias formula in (1) is relevant for an apparently causative exposure, which allows researchers to get lower bounds of the true causal relative risk given prespecified sensitivity parameters RR_{EU} and RR_{UD}. If the exposure E is apparently preventive with $RR_{ED}^{obs} < 1$, we can use the following formula to conduct sensitivity analysis:

$$RR_{ED}^{true} \leq RR_{ED}^{obs} \times \frac{RR_{EU} \times RR_{UD}}{RR_{EU} + RR_{UD} - 1}, \qquad (2)$$

where we modify the definition of RR_{EU} as $\max_k RR_{EU,k}^{-1}$ (i.e., the maximum of the inverse relative risks relating E and U), or equivalently the inverse of the minimum of the relative risks relating E and U. For an apparently preventive exposure, (2) allows researchers to obtain an upper bound of the causal relative risk RR_{ED}^{true} by multiplying the observed relative risk RR_{ED}^{obs} by the joint bounding factor $RR_{EU} \times RR_{UD} \big/ (RR_{EU} + RR_{UD} - 1)$. We present the proof in eAppendix 2 (http://links.lww.com/EDE/B16), and omit analogous discussion based on (2).

Finally, all the results above are within strata of the observed covariates C as would be obtained from a log-binomial regression model or a logistic regression model with rare outcome. If averaged relative risk over the observed covariates C is of interest, the true causal relative risk must be at least as large as the minimum of $RR_{ED|c}^{obs} \Big/ \dfrac{RR_{EU|c} \times RR_{UD|c}}{RR_{EU|c} + RR_{UD|c} - 1}$ over c.

If we assume a common causal relative risk among the levels of C as in the usual log-linear or logistic regression with rare outcomes, then the true causal relative risk must be at least as large as the maximum of $\mathrm{RR}_{ED|c}^{\mathrm{obs}} \Big/ \dfrac{\mathrm{RR}_{EU|c} \times \mathrm{RR}_{UD|c}}{\mathrm{RR}_{EU|c} + \mathrm{RR}_{UD|c} - 1}$ over c. See eAppendix 2 (http://links.lww.com/EDE/B16) for further discussion.

RELATION WITH CORNFIELD CONDITIONS

Under the assumptions of a binary confounder U and the conditional independence between the exposure E and the outcome D given the confounder U, Cornfield et al.[1] showed that the exposure–confounder relative risk must be at least as large as the observed exposure–outcome relative risk:

$$\mathrm{RR}_{EU} \geq \mathrm{RR}_{ED}^{\mathrm{obs}}. \qquad (3)$$

Schlesselman[7] further showed that the confounder–outcome relative risk must also be at least as large as the observed exposure–outcome relative risk:

$$\mathrm{RR}_{UD} \geq \mathrm{RR}_{ED}^{\mathrm{obs}}. \qquad (4)$$

We show in eAppendix 2 (http://links.lww.com/EDE/B16) that the classical Cornfield conditions (3) and (4) are just special cases of our result by letting one of RR_{EU} or RR_{UD} go to infinity in (1). Moreover, our results apply to general confounders not just binary confounders, and our results also apply to other possible values of the true causal relative risk of the exposure on the outcome. We are not restricted to only assessing how strong the unmeasured confounder would have to be to completely explain away the effect. Thus, for example, for confounding to reduce the observed relative risk $\mathrm{RR}_{ED}^{\mathrm{obs}}$ to a true causal relative risk of $\mathrm{RR}_{ED}^{\mathrm{true}}$, the unmeasured confounding measures have to satisfy

$$\mathrm{RR}_{EU} \geq \mathrm{RR}_{ED}^{\mathrm{obs}}/\mathrm{RR}_{ED}^{\mathrm{true}} \text{ and } \mathrm{RR}_{UD} \geq \mathrm{RR}_{ED}^{\mathrm{obs}}/\mathrm{RR}_{ED}^{\mathrm{true}}. \qquad (5)$$

Perhaps even more importantly with regard to Cornfield-like conditions, our main Result 1 not only leads to the conditions in (5) that both RR_{EU} and RR_{UD} must satisfy but also implies the following condition that the maximum of RR_{EU} and RR_{UD} must satisfy:

$$\begin{aligned} &\max(\mathrm{RR}_{EU}, \mathrm{RR}_{UD}) \\ &\geq \left\{\mathrm{RR}_{ED}^{\mathrm{obs}} + \sqrt{\mathrm{RR}_{ED}^{\mathrm{obs}}(\mathrm{RR}_{ED}^{\mathrm{obs}} - \mathrm{RR}_{ED}^{\mathrm{true}})}\right\}\Big/\mathrm{RR}_{ED}^{\mathrm{true}}, \end{aligned} \qquad (6)$$

to reduce an observed relative risk $\mathrm{RR}_{ED}^{\mathrm{obs}}$ to a true causal relative risk $\mathrm{RR}_{ED}^{\mathrm{true}}$. We show this in eAppendix 2 (http://links. lww.com/EDE/B16). As a special case, for the unmeasured confounder to completely explain away the observed relative risk (i.e., $\mathrm{RR}_{ED}^{\mathrm{true}} = 1$), it is necessary that

$$\max(\mathrm{RR}_{EU}, \mathrm{RR}_{UD}) \geq \mathrm{RR}_{ED}^{\mathrm{obs}} + \sqrt{\mathrm{RR}_{ED}^{\mathrm{obs}}(\mathrm{RR}_{ED}^{\mathrm{obs}} - 1)}.$$

Once again the results do not require a binary unmeasured confounder. They are applicable to any unmeasured confounder. Similar low and high threshold Cornfield conditions that the minimum and maximum of the confounding measures must satisfy to completely explain away an effect were derived on an odds ratio scale of exposure–confounder association by Flanders and Khoury[12] and Lee,[10] and we comment and extend these results in eAppendix 2 (http://links. lww.com/EDE/B16).

The classical Cornfield conditions and the high threshold generalization are useful to answer the question about the magnitude of the association between the exposure and the confounder and that between the confounder and the outcome, to explain away the observed exposure–outcome association or with our new results, to reduce it to a prespecified magnitude. The Cornfield conditions in (5) and (6) are especially useful, when we want to specify only one of the marginal associations RR_{EU} or RR_{UD} as well as their relative magnitudes. However, they are inferior to the main Result 1, which is essentially the condition that the joint values of $(\mathrm{RR}_{EU}, \mathrm{RR}_{UD})$ must satisfy. As will be seen below, although the high threshold conclusions are a useful heuristic, they are weaker than the use of our new joint bounding factor in Result 1 insofar as there are scenarios which the joint bounding factor in Result 1 can rule out an estimate as being due to unmeasured confounding but the high threshold conditions cannot. For example, when we have an observed exposure–outcome relative risk of $\mathrm{RR}_{ED}^{\mathrm{obs}} = 3$, the low threshold (i.e., the classical Cornfield condition) is given by

$$\min(\mathrm{RR}_{EU}, \mathrm{RR}_{UD}) \geq 3,$$

the high threshold is given by

$$\max(\mathrm{RR}_{EU}, \mathrm{RR}_{UD}) \geq 3 + \sqrt{3 \times 2} = 5.45,$$

and the joint threshold condition is given by

$$\frac{\mathrm{RR}_{EU} \times \mathrm{RR}_{UD}}{\mathrm{RR}_{EU} + \mathrm{RR}_{UD} - 1} \geq 3.$$

Thus, the low Cornfield threshold is 3, and so we know that we must have that both RR_{EU} and RR_{UD} be greater than 3 to explain away the effect. The high Cornfield threshold is 5.45, and so at least one of RR_{EU} and RR_{UD} must be larger than 5.45 to explain away the effect. Consider an unmeasured confounder with $(\mathrm{RR}_{EU} = 5.5, \mathrm{RR}_{UD} = 3.1)$, they would exceed both the low Cornfield threshold (since $\mathrm{RR}_{EU} > 3, \mathrm{RR}_{UD} > 3$) and the high threshold (since $\mathrm{RR}_{EU} > 5.45$), and we might thus think it can explain away the observed exposure–outcome relative risk. However, using our joint threshold condition in (1), an unmeasured confounder with $(\mathrm{RR}_{EU} = 5.5, \mathrm{RR}_{UD} = 3.1)$ has a bounding factor $5.5 \times 3.1 / (5.5 + 3.1 - 1) = 2.24 < 3$ and thus such confounding could not explain away an observed relative risk of 3. We can see this from our result in (1), but

we cannot see this from the classical Cornfield conditions and even the new high threshold Cornfield condition. The Cornfield conditions, both low and high thresholds, although a useful heuristic, are not as useful for sensitivity analysis as our bounding factor. This is because there are scenarios, such as the one above, in which our bounding factor can rule out an estimate as due to unmeasured confounding, while the low and high threshold Cornfield conditions cannot.

ILLUSTRATION

Consider the historical study conducted by Hammond and Horn,[13] in which the point estimate of the observed relative risk of cigarette smoking on lung cancer was $\text{RR}_{ED}^{\text{obs}} = 10.73$ with 95% confidence interval $[8.02, 14.36]$. Fisher[14] suggested that the observed relative risk of the exposure E on the outcome D might be completely due to the existence of a common genetic confounder. The work of Cornfield et al.[1] showed that for a binary unmeasured confounder to completely explain away the observed relative risk, both the exposure–confounder relative risk and the confounder–outcome relative risk would have to be at least 10.73. Let us now assume then that both the exposure–confounder relative risk and the confounder–outcome relative risk have the magnitude 10.73. The joint bounding factor is

$$\frac{\text{RR}_{EU} \times \text{RR}_{UD}}{\text{RR}_{EU} + \text{RR}_{UD} - 1} = \frac{10.73 \times 10.73}{10.73 + 10.73 - 1} = 5.63.$$

Even if we assume such a strong confounder, the point estimate of the causal relative risk of cigarette smoking and lung cancer must still be at least as large as $\text{RR}_{ED}^{\text{true}} \geq \text{RR}_{ED}^{\text{obs}} / 5.63 = 10.73 / 5.63 = 1.91 > 1$, and the 95% confidence interval is $[8.02 / 5.63, 14.36 / 5.63] = [1.42, 2.55]$ with the lower confidence limit still larger than 1. Thus in fact, not even exposure–confounder and confounder–outcome relative risks of 10.73 suffice to explain away the effect nor the lower confidence limit. In fact, to explain away the point estimate of the observed relative risk 10.73, the magnitude of RR_{EU} and RR_{UD} (if $\text{RR}_{EU} = \text{RR}_{UD}$) should be at least as large as $10.73 + \sqrt{10.73 \times 9.73} = 20.95$. And to explain away the lower confidence limit 8.02, these two confounding relative risks should be at least as large as $8.02 + \sqrt{8.02 \times 7.02} = 15.52$. More generally, we can plot those values of RR_{EU} and RR_{UD} that would be required to explain away the effect estimate or the lower limit of the confidence interval. This is given in the Figure. To explain away the point estimate, the two parameters would have to lie on or above the solid line. To explain away the lower confidence limit, the two parameters would have to lie on or above the dotted line. These results hold without any assumptions on the structure of the unmeasured confounding. The numerical results above show that, by using the new joint bounding factor, it is even more implausible than using the Cornfield conditions that a genetic confounder explains away the relative risk between cigarette smoking and lung cancer.
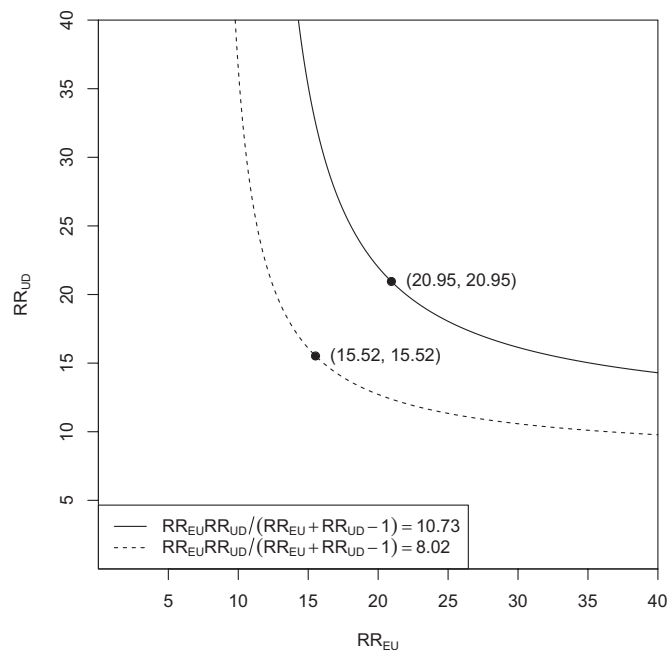

FIGURE. The areas above the two lines are the joint values of the exposure–confounder association RR_{EU} and the confounder–outcome association RR_{UD} that can would be required to explain away the effect estimate 10.73 and the lower confidence limit 8.02.

More generally, we could consider corrected estimates and confidence intervals for the effect over a range of different values of the sensitivity analysis parameters, RR_{EU} and RR_{UD}, as in Table 2. The columns of Table 2 correspond to RR_{UD} and the rows to RR_{EU}. The entries are the corrected estimates and confidence intervals for the effect under the different confounding scenarios. In general, a table like this one is most informative for sensitivity analysis. SAS code to carry out such a sensitivity analysis and to provide such a table is given in eAppendix 9 (http://links.lww.com/EDE/B16).

DISCUSSION

A crucial task in causal inference with observational studies is to assess the sensitivity of causal conclusions with respect to unmeasured confounding. In sensitivity analysis, because one is assessing the sensitivity of conclusions to the assumption of no unmeasured confounding, additional untestable assumptions may often seem undesirable and suspect to researchers. We have introduced a new joint bounding factor that allows researchers to conduct sensitivity analysis without assumptions, that is, we provide an inequality, which is applicable without any assumptions, such that the sensitivity analysis parameters must satisfy the inequality if an unmeasured confounder is to explain away the observed effect estimate or reduce it to a particular level. We can obtain a conservative estimate of the true causal effect by dividing the observed relative risk by the bounding factor; the method does not assume a single binary confounder or no exposure–confounder interaction on the outcome.

TABLE 2. Bounds on Corrected Estimates, Lower Confidence Limits, and Upper Confidence Limits for Unmeasured Confounding (Each Cell Contains Bounds on Point Estimate, Lower and Upper Confidence Limits; Columns Correspond to Increasing Strength of the Risk Ratio of U on the Outcome; Rows Correspond to Increasing Strength of Risk Ratio Relating the Exposure and U)

	1.2	1.3	1.5	1.8	2.0	2.5	3.0	5.0	8.0	10.0
1.2	10.43 (7.80,13.96)	10.32 (7.71,13.81)	10.13 (7.57,13.56)	9.94 (7.43,13.30)	9.84 (7.35,13.16)	9.66 (7.22,12.92)	9.54 (7.13,12.76)	9.30 (6.95,12.45)	9.17 (6.85,12.27)	9.12 (6.82,12.21)
1.3	10.32 (7.71,13.81)	10.16 (7.59,13.60)	9.90 (7.40,13.26)	9.63 (7.20,12.89)	9.49 (7.09,12.70)	9.24 (6.91,12.37)	9.08 (6.79,12.15)	8.75 (6.54,11.71)	8.56 (6.40,11.46)	8.50 (6.35,11.38)
1.5	10.13 (7.57,13.56)	9.90 (7.40,13.26)	9.54 (7.13,12.76)	9.14 (6.83,12.23)	8.94 (6.68,11.97)	8.58 (6.42,11.49)	8.35 (6.24,11.17)	7.87 (5.88,10.53)	7.60 (5.68,10.17)	7.51 (5.61,10.05)
1.8	9.94 (7.43,13.30)	9.63 (7.20,12.89)	9.14 (6.83,12.23)	8.61 (6.44,11.52)	8.35 (6.24,11.17)	7.87 (5.88,10.53)	7.55 (5.64,10.11)	6.91 (5.17,9.25)	6.56 (4.90,8.78)	6.44 (4.81,8.62)
2.0	9.84 (7.35,13.16)	9.49 (7.09,12.70)	8.94 (6.68,11.97)	8.35 (6.24,11.17)	8.05 (6.01,10.77)	7.51 (5.61,10.05)	7.15 (5.35,9.57)	6.44 (4.81,8.62)	6.04 (4.51,8.08)	5.90 (4.41,7.90)
2.5	9.66 (7.22,12.92)	9.24 (6.91,12.37)	8.58 (6.42,11.49)	7.87 (5.88,10.53)	7.51 (5.61,10.05)	6.87 (5.13,9.19)	6.44 (4.81,8.62)	5.58 (4.17,7.47)	5.10 (3.81,6.82)	4.94 (3.69,6.61)
3.0	9.54 (7.13,12.76)	9.08 (6.79,12.15)	8.35 (6.24,11.17)	7.55 (5.64,10.11)	7.15 (5.35,9.57)	6.44 (4.81,8.62)	5.96 (4.46,7.98)	5.01 (3.74,6.70)	4.47 (3.34,5.98)	4.29 (3.21,5.74)
5.0	9.30 (6.95,12.45)	8.75 (6.54,11.71)	7.87 (5.88,10.53)	6.91 (5.17,9.25)	6.44 (4.81,8.62)	5.58 (4.17,7.47)	5.01 (3.74,6.70)	3.86 (2.89,5.17)	3.22 (2.41,4.31)	3.00 (2.25,4.02)
8.0	9.17 (6.85,12.27)	8.56 (6.40,11.46)	7.60 (5.68,10.17)	6.56 (4.90,8.78)	6.04 (4.51,8.08)	5.10 (3.81,6.82)	4.47 (3.34,5.98)	3.22 (2.41,4.31)	2.51 (1.88,3.37)	2.28 (1.70,3.05)
10.0	9.12 (6.82,12.21)	8.50 (6.35,11.38)	7.51 (5.61,10.05)	6.44 (4.81,8.62)	5.90 (4.41,7.90)	4.94 (3.69,6.61)	4.29 (3.21,5.74)	3.00 (2.25,4.02)	2.28 (1.70,3.05)	2.04 (1.52,2.73)

Previous sensitivity analysis approaches in the literature often relied on the assumption of a single binary confounder and no-interaction between the effects of the exposure and the confounder on the outcome.[5,8,9] For example, Schlesselman[7] assumed a binary confounder, a common relative risk, γ, of the confounder on the outcome for both with and without exposure (i.e., a no interaction assumption). Under these assumptions, he obtained the bias factor $\mathrm{RR}_{ED}^{obs}/\mathrm{RR}_{ED}^{true} = \{1+(\gamma-1)\,P(U=1|E=1)\}/\{1+(\gamma-1)\,P(U=1|E=0)\}$ for sensitivity analysis requiring specifications of γ, $P(U=1|E=1)$ and $P(U=1|E=0)$. Our result requires fewer assumptions and fewer sensitivity parameters (two rather than three). We further discuss in eAppendix 4 (http://links.lww.com/EDE/B16) that, under Schlesselman's formula, if $P(U=1|E=1)/P(U=1|E=0)$ is constrained to be no larger than some limit RR_{EU}, then the maximum bias factor that can be obtained from Schlesselman's formula is $\mathrm{RR}_{EU} \times \gamma/(\mathrm{RR}_{EU} + \gamma-1)$, which is the same as our bounding factor. Thus, in this setting Schlesselman's no interaction assumption does not strengthen the bounds; the no interaction assumption is unnecessary. Without the no interaction assumption, Flanders and Khoury[12] and VanderWeele and Arah,[8] derived general formulas for sensitivity analysis. However, unless the confounder is binary, these formulas require specifying a very large number of parameters. They also require specifying the prevalence of each confounder level. Flanders and Khoury[12] derive bounds for the true causal relative risk for the exposed population which are potentially applicable without specifying the prevalence of the unmeasured confounder. However, without specifying the prevalence, their formula only leads to a low threshold Cornfield condition, and these bounds are thus much weaker than those in this article. We discuss further the relation between their results and ours in eAppendix 4 (http://links.lww.com/EDE/B16).

The relative risk scale is widely used for sensitivity analysis in epidemiology and elsewhere, but the risk difference scale is also often of interest and importance.[11,15] We show, in Appendix 1, that similar conditions for sensitivity analysis also hold for the risk difference. If we use similar sensitivity parameters on the *relative risk scale* for the risk difference estimate, then we can derive similar lower bounds on the effects and determine how much confounding is required to explain away an effect or reduce it to a specific level. SAS code for this approach is also given in eAppendix 10 (http://links.lww.com/EDE/B16). We can also do sensitivity analysis for the risk difference using sensitivity parameters on the *risk difference scale*. Unfortunately, however, these conditions for the risk difference using risk difference sensitivity parameters then depend on the number of categories of the unmeasured confounder, and become weaker for confounders with more categories. This is not the case for sensitivity analysis of the risk difference (or the relative risk) if the sensitivity

parameters themselves are expressed on the relative risk scale, in which case the bounding factor is applicable and is the same regardless of the number of categories. Due to this property, it is perhaps more suitable to conduct sensitivity analysis for the risk difference using sensitivity parameters on the relative risk scale. See Appendix 2 for further discussion.

The hazard ratio is widely used for analyzing data with time-to-event outcomes. In eAppendix 7 (http://links.lww.com/EDE/B16), we show that under the assumption of having a rare outcome at the end of follow-up, the same bounding factor also applies to the hazard ratio with the confounder–outcome relative risk replaced by the confounder–outcome hazard ratio. Likewise similar results also apply to non-negative outcomes (e.g., counts or positive continuous outcomes) by replacing the confounder–outcome relative risk by the maximum ratio by which the confounder may increase the expected outcome comparing any two confounder categories.

The new joint bounding factor $(\mathrm{RR}_{EU} \times \mathrm{RR}_{UD}) / (\mathrm{RR}_{EU} + \mathrm{RR}_{UD} - 1)$ plays a central role in our sensitivity analysis approach, which, in turn, gives us a new measure of the strength of unmeasured confounding induced by a confounder U. Our approach has the advantage of making no assumptions about the structure of the unmeasured confounder or confounders, and of delivering conclusions much stronger than the original Cornfield conditions.

In general, a table with many different possible sensitivity analysis parameters including values that are quite extreme, such as Table 2, will be most informative. However, at the very least, in any observational study, researchers should report how much confounding would be needed to reduce the estimate, and how much confounding would be needed to reduce the confidence interval, to include the null. We believe that if this were always done in observational studies, the evidence for causality could much more easily be assessed and science would be better served.

REFERENCES

1. Cornfield J, Haenszel W, and Hammond EC et al. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst*. 1959;22:173–203.
2. Bross IDJ. Spurious effects from an extraneous variable. *J Chronic Dis*. 1966;19:637–647.
3. Bross IDJ. Pertinency of an extraneous variable. *J Chronic Dis*. 1967;20:487–495.
4. Yanagawa T. Case-control studies: assessing the effect of a confounding factor. *Biometrika*. 1984;71:191–194.
5. Rosenbaum PR, Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J R Stat Soc Series B*. 1983;45:212–218.
6. Imbens GW. Sensitivity to exogeneity assumptions in program evaluation. *Am Econ Rev*. 2003;93:126–132.
7. Schlesselman JJ. Assessing effects of confounding variables. *Am J Epidemiol*. 1978;108:3–8.
8. VanderWeele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*. 2011;22:42–52.
9. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*. 1998;54:948–963.
10. Lee WC. Bounding the bias of unmeasured factors with confounding and effect-modifying potentials. *Stat Med*. 2011;30:1007–1017.
11. Ding P, VanderWeele TJ. Generalized Cornfield conditions for the risk difference. *Biometrika*. 2014;101:971–977.
12. Flanders WD, Khoury MJ. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology*. 1990;1:239–246.
13. Hammond EC, Horn D. Smoking and death rates: report on forty four months of follow-up of 187,783 men. *J Am Med Assoc*. 1958;166:1159–1172, 1294–1308.
14. Fisher RA. Dangers of cigarette smoking [letter]. *Brit Med J*. 1957;2:297–298.
15. Poole C. On the origin of risk relativism. *Epidemiology*. 2010;21:3–9.
16. Gastwirth JL, Krieger AM, and Rosenbaum PR. Cornfield's inequality. In: Armitage P, Colton T, eds. *Encyclopedia of Biostatistics*. New York, NY: Wiley; 1998:952–955.
17. Wang L, Krieger AM. Causal conclusions are most sensitive to unobserved binary covariates. *Stat Med*. 2006;25:2257–2271.
18. Ichino A, Mealli F, Nannicini T. From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *J Appl Econom*. 2008;23:305–327.

APPENDIX 1

Conditions for the Risk Difference Using Sensitivity Parameters on the Relative Risk Scale

As in the text, we assume analysis is conducted, and all probabilities below are, conditional on, or within strata of the measured covariates C. Define the bounding factor as $\mathrm{BF}_U = \mathrm{RR}_{EU} \times \mathrm{RR}_{UD} / (\mathrm{RR}_{EU} + \mathrm{RR}_{UD} - 1)$, the prevalence of the exposure as $f = P(E = 1)$, and the probabilities of the outcome with and without exposure as $p_1 = P(D = 1 | E = 1)$ and $p_0 = P(D = 1 | E = 0)$. The causal risk differences for the exposed and unexposed populations are

$$\mathrm{RD}_{ED+}^{\mathrm{true}} = p_1 - \sum_{k=0}^{K-1} P(D = 1 | E = 0, U = k) P(U = k | E = 1),$$

$$\mathrm{RD}_{ED-}^{\mathrm{true}} = \sum_{k=0}^{K-1} P(D = 1 | E = 1, U = k) P(U = k | E = 0) - p_0,$$

and the causal risk difference for the whole population is

$$\mathrm{RD}_{ED}^{\mathrm{true}} = \sum_{k=0}^{K-1} \{P(D = 1 | E = 1, U = k) - P(D = 1 | E = 0,$$

$$U = k)\} P(U = k)$$

$$= f \mathrm{RD}_{ED+}^{\mathrm{true}} + (1 - f) \mathrm{RD}_{ED-}^{\mathrm{true}}.$$

We show in eAppendix 5 (http://links.lww.com/EDE/B16) that the lower bounds for the causal risk differences are

$$\mathrm{RD}^{\mathrm{true}}_{ED+} \geq p_1 - p_0 \times \mathrm{BF}_U,$$

$$\mathrm{RD}^{\mathrm{true}}_{ED-} \geq p_1/\mathrm{BF}_U - p_0,$$

$$\mathrm{RD}^{\mathrm{true}}_{ED} \geq (p_1 - p_0 \times \mathrm{BF}_U) \times \left\{ f + (1-f)/\mathrm{BF}_U \right\}$$
$$= (p_1/\mathrm{BF}_U - p_0) \times \left\{ f \times \mathrm{BF}_U + (1-f) \right\}.$$

Note that even without knowing f, we can use the inequality $\mathrm{RD}^{\mathrm{true}}_{ED} \geq \min(\mathrm{RD}^{\mathrm{true}}_{ED+}, \mathrm{RD}^{\mathrm{true}}_{ED-})$ to obtain a lower bound for $\mathrm{RD}^{\mathrm{true}}_{ED}$.

As an example, suppose the probabilities of the outcome with and without exposure are $p_1 = 0.25, p_0 = 0.1$, and therefore the observed risk difference is $\mathrm{RD}^{\mathrm{obs}}_{ED} = p_1 - p_0 = 0.15$. If we assume that the unmeasured confounding measures are $(\mathrm{RR}_{EU}, \mathrm{RR}_{UD}) = (2,2)$ with the joint bounding factor of $2 \times 2/(2+2-1) = 1.33$, then the true risk difference for the exposed is at least as large as $0.25 - 0.1 \times 1.33 = 0.12$, the true risk difference for the unexposed is at least as large as $0.25/1.33 - 0.1 = 0.09$, and the true risk difference for the whole population is at least as large as $\min(0.12, 0.09) = 0.09$. If we further know that the prevalence of the exposure is $f = 0.2$, the true risk difference for the whole population is at least as large as $0.12 \times 0.2 + 0.09 \times 0.8 = 0.10$.

The above results imply that, for an unmeasured confounder to reduce the observed risk difference to be $\mathrm{RD}^{\mathrm{true}}_{ED+}, \mathrm{RD}^{\mathrm{true}}_{ED-}$ and $\mathrm{RD}^{\mathrm{true}}_{ED}$, respectively, the Cornfield conditions for the joint bounding factor for the exposed, the unexposed, and the whole population, respectively, are

$$\mathrm{BF}_U \geq (p_1 - \mathrm{RD}^{\mathrm{true}}_{ED+})/p_0,$$

$$\mathrm{BF}_U \geq p_1/(p_0 + \mathrm{RD}^{\mathrm{true}}_{ED-}),$$

$$\mathrm{BF}_U \geq \left(\frac{1}{2} p_0 f\right) \left[\sqrt{ \begin{array}{l} \{\mathrm{RD}^{\mathrm{true}}_{ED} + p_0(1-f) - p_1 f\}^2 \\ + 4 p_1 p_0 f(1-f) \end{array}} - \{\mathrm{RD}^{\mathrm{true}}_{ED} + p_0(1-f) - p_1 f\} \right].$$

Note that if the true causal risk difference is $\mathrm{RD}^{\mathrm{true}}_{ED} = 0$, the above conditions all reduce to $\mathrm{BF}_U \geq \mathrm{RR}^{\mathrm{obs}}_{ED}$. Suppose, again, the probabilities of the observed outcome with and without exposure are $p_1 = 0.25, p_0 = 0.1$, and the prevalence of the exposure is $f = 0.2$. For an unmeasured confounder U to reduce the observed risk difference of $\mathrm{RD}^{\mathrm{obs}}_{ED} = 0.15$ to a true risk difference of $\mathrm{RD}^{\mathrm{true}}_{ED} = 0.05$, the joint bounding factor resulting from the confounder must be at least as large as

$$\mathrm{BF}_U \geq \left(\frac{1}{2 \times 0.1 \times 0.2}\right) \left[\sqrt{ \begin{array}{l} (0.05 + 0.1 \times 0.8 - 0.25 \times 0.2)^2 + 4 \\ \times 0.25 \times 0.1 \times 0.2 \times 0.8 \\ -(0.05 + 0.1 \times 0.8 - 0.25 \times 0.2) \end{array}} \right] = 1.74.$$

Therefore, as in the text both of the confounding measures RR_{EU} and RR_{UD} must be at least as large as 1.74, and the maximum of them must be at least as large as $1.74 + \sqrt{1.74 \times 0.74} = 2.88$.

The above results are useful for apparently causative exposures with $\mathrm{RD}^{\mathrm{obs}}_{ED} > 0$, which give (possibly positive) lower bounds for the causal risk differences. However, for apparently preventive exposure with $\mathrm{RD}^{\mathrm{obs}}_{ED} < 0$, we need to modify the definition of RR_{EU} as $\mathrm{RR}_{EU} = \max_u \mathrm{RR}^{-1}_{EU}(u)$. And we have the following analogous results on the upper bounds of the causal risk differences:

$$\mathrm{RD}^{\mathrm{true}}_{ED+} \leq p_1 \times \mathrm{BF}_U - p_0,$$

$$\mathrm{RD}^{\mathrm{true}}_{ED-} \leq p_1 - p_0/\mathrm{BF}_U,$$

$$\mathrm{RD}^{\mathrm{true}}_{ED} \leq (p_1 \times \mathrm{BF}_U - p_0) \times \left\{ f + (1-f)/\mathrm{BF}_U \right\}$$
$$= (p_1 - p_0/\mathrm{BF}_U) \times \left\{ f \times \mathrm{BF}_U + (1-f) \right\}.$$

The results above are conditional on measured covariates C. Due to the linearity of the risk difference, we can also obtain the lower bound of the marginal risk differences averaged over the observed covariates C using

$$\mathrm{RD}^{\mathrm{true}}_{ED+} = \sum_c \mathrm{RD}^{\mathrm{true}}_{ED|c+} P(C = c | E = 1),$$

$$\mathrm{RD}^{\mathrm{true}}_{ED-} = \sum_c \mathrm{RD}^{\mathrm{true}}_{ED|c-} P(C = c | E = 0),$$

$$\mathrm{RD}^{\mathrm{true}}_{ED} = \sum_c \mathrm{RD}^{\mathrm{true}}_{ED|c} P(C = c).$$

In eAppendix 5 (http://links.lww.com/EDE/B16), we provide details and proofs for the results above, discuss statistical inference for the causal risk difference bounds under finite samples, and give formulas for how large the bounding factor would have to be to reduce an estimate or a confidence interval to 0 or to some other specified quantity. In the eAppendix (http://links.lww.com/EDE/B16), we also provide software code to implement this sensitivity analysis approach for the risk difference.

APPENDIX 2

Conditions for the Risk Difference Using Sensitivity Parameters on the Risk Difference Scale

In Appendix 1, we considered sensitivity analysis for the risk difference with sensitivity analysis parameters on the relative risk scale. In this Appendix, we consider sensitivity analysis for the risk difference with parameters defined on the risk difference scale. Unfortunately, for the reasons described below, the results for the risk difference with parameters defined on the difference scale are not as practically useful as when the parameters are defined on the relative risk scale.

Let $\mathrm{RD}^{\mathrm{obs}}_{ED} = P(D = 1 | E = 1) - P(D = 1 | E = 0)$ denote the observed risk difference, and

$$\mathrm{RD}_{ED}^{\mathrm{true}} = \sum_{k=0}^{K-1} \{P(D=1|E=1, U=k) - P(D=1|E=0,$$
$$U=k)\} P(U=k)$$

denote the standardized risk difference.

Define $\alpha_k = P(U=k|E=1) - P(U=k|E=0)$ as the difference in the probability that the confounder U takes a particular value k comparing exposed and unexposed. We use $\mathrm{RD}_{EU} = \max_{k\geq 1}|\alpha_k|$, the maximum of these absolute differences, to measure the exposure–confounder association on the risk difference scale, defined as the maximal risk difference of the exposure E on the confounder U. Define $\beta_k^* = P(D=1|E=1, U=k) - P(D=1|E=1, U=0)$ and $\beta_k = P(D=1|E=0, U=k) - P(D=1|E=0, U=0)$ as the difference in the probability of the outcome comparing the category k and 0 of the confounder U with and without exposure. Define $\mathrm{RD}_{UD|E=1} = \max_{k\geq 1}|\beta_k^*|$ and $\mathrm{RD}_{UD|E=0} = \max_{k\geq 1}|\beta_k|$ as the maximums of these differences with and without exposure, respectively. We use $\mathrm{RD}_{UD} = \max(\mathrm{RD}_{UD|E=1}, \mathrm{RD}_{UD|E=0})$ to measure the confounder–outcome association in the risk difference scale, defined as the maximal risk difference of the confounder U on the outcome D.

We first consider a *binary* unmeasured confounder. For binary confounder U, the maximal risk difference RD_{EU} becomes the ordinary risk difference RD_{EU}, and the maximal risk difference becomes the maximum of two conditional risk difference $\mathrm{RD}_{UD} = \max(\mathrm{RD}_{UD|E=1}, \mathrm{RD}_{UD|E=0})$. We have that

$$\mathrm{RD}_{EU} \times \mathrm{RD}_{UD} \geq \mathrm{RD}_{ED}^{\mathrm{obs}} - \mathrm{RD}_{ED}^{\mathrm{true}},$$

which further leads to the following low and high thresholds:

$$\min(\mathrm{RD}_{EU}, \mathrm{RD}_{UD}) \geq \mathrm{RD}_{ED}^{\mathrm{obs}} - \mathrm{RD}_{ED}^{\mathrm{true}},$$
$$\max(\mathrm{RD}_{EU}, \mathrm{RD}_{UD}) \geq \sqrt{\mathrm{RD}_{ED}^{\mathrm{obs}} - \mathrm{RD}_{ED}^{\mathrm{true}}},$$

which generalize previous results under the null of zero causal effect of E on D.[11,15,16]

For categorical confounder U, no simple form of the bounding factor is available, but we can still show that RD_{EU} and RD_{UD} must satisfy the following conditions:

$$\mathrm{RD}_{EU} \geq (\mathrm{RD}_{ED}^{\mathrm{obs}} - \mathrm{RD}_{ED}^{\mathrm{true}})/(K-1),$$

$$\mathrm{RD}_{UD} \geq (\mathrm{RD}_{ED}^{\mathrm{obs}} - \mathrm{RD}_{ED}^{\mathrm{true}})/2,$$

$$\max(\mathrm{RD}_{EU}, \mathrm{RD}_{UD})$$
$$\geq \max\left\{\sqrt{(\mathrm{RD}_{ED}^{\mathrm{obs}} - \mathrm{RD}_{ED}^{\mathrm{true}})/(K-1)}, (\mathrm{RD}_{ED}^{\mathrm{obs}} - \mathrm{RD}_{ED}^{\mathrm{true}})/2\right\}.$$

When $K=3$ such as a three-level genetic confounder, these conditions reduce to

$$\min(\mathrm{RD}_{EU}, \mathrm{RD}_{UD}) \geq (\mathrm{RD}_{ED}^{\mathrm{obs}} - \mathrm{RD}_{ED}^{\mathrm{true}})/2,$$
$$\max(\mathrm{RD}_{EU}, \mathrm{RD}_{UD}) \geq \sqrt{(\mathrm{RD}_{ED}^{\mathrm{obs}} - \mathrm{RD}_{ED}^{\mathrm{true}})/2}.$$

The results above generalize previous results[11] from the null hypothesis of no effect ($\mathrm{RD}_{ED}^{\mathrm{true}} = 0$) to alternative hypotheses ($\mathrm{RD}_{ED}^{\mathrm{true}}$ arbitrary). We show the proofs and extensions for the above results in eAppendix 6 (http://links.lww.com/EDE/B16).

We can see from above that the generalized Cornfield conditions for the risk difference under alternative hypotheses depend on the number of categories of U, and become less informative as the number of categories increases. Therefore, a binary confounder is not the most conservative case for sensitivity analysis with parameters expressed the risk difference scale. However, the Cornfield conditions for the relative risk do not suffer from this problem. Therefore, it seems that it is more appropriate to conduct sensitivity analysis with parameters expressed on the risk ratio scale, and a binary confounder is the most conservative case for sensitivity analysis with parameters expressed on the risk ratio scale.[17,18]