Sensitivity Analysis in Observational Research: Introducing the E-Value

Tyler J. VanderWeele, PhD, and Peng Ding, PhD

Sensitivity analysis is useful in assessing how robust an association is to potential unmeasured or uncontrolled confounding. This article introduces a new measure called the "E-value," which is related to the evidence for causality in observational studies that are potentially subject to confounding. The E-value is defined as the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and the outcome to fully explain away a specific treatment–outcome association, conditional on the measured covariates. A large E-value implies that considerable unmeasured confounding would be needed to explain away an effect estimate. A small E-value implies little unmeasured confounding would be needed to explain away an effect estimate. The authors propose that in all observational studies intended to produce evidence for causality, the E-value be reported or some other sensitivity analysis be used. They suggest calculating the E-value for both the observed association estimate (after adjustments for measured confounders) and the limit of the confidence interval closest to the null. If this were to become standard practice, the ability of the scientific community to assess evidence from observational studies would improve considerably, and ultimately, science would be strengthened.

Ann Intern Med. 2017;167:268-274. doi:10.7326/M16-2607 **Annals.org**
For author affiliations, see end of text.
This article was published at Annals.org on 11 July 2017.

Much empirical research is concerned with establishing causation. It is well-known, however, that with observational data, association (1–11) need not imply causation (12–22). A central concern with observational data is bias by unmeasured or uncontrolled confounding, that is, that some third factor related to both the treatment and outcome might explain their association, with no true causal effect (12–22). With observational data, we can never be certain that efforts to adjust for confounders or common causes are adequate.

An important approach to evaluating evidence for causation in the face of unmeasured confounding is "sensitivity analysis" (or "bias analysis") (14–22). Sensitivity analysis considers how strong unmeasured confounding would have to be to explain away the association, that is, how strongly the unmeasured confounder would have to be associated with the treatment and outcome for the treatment–outcome association not to be causal.

In this tutorial, we discuss a sensitivity analysis technique that makes minimal assumptions, and we propose that observational studies start reporting the "E-value," a new measure related to evidence for causality. The E-value represents the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome to fully explain away a specific treatment–outcome association, conditional on the measured covariates. Implementing these sensitivity analysis techniques and obtaining E-values are relatively simple. If reporting E-values for sensitivity analysis were standard practice, our ability to assess evidence from observational studies would improve and science would be strengthened.

See also:

Editorial comment 285

Web-Only
Supplement

Example

As a motivating example, several observational studies have reported associations between breastfeeding and various infant and maternal health outcomes. A common concern is that the effects of breastfeeding may be confounded by smoking behavior or by socioeconomic status. In a population-based case–control study, Victora and colleagues (23) examined associations between breastfeeding and infant death by respiratory infection. After adjusting for age, birthweight, social status, maternal education, and family income, the authors found that infants fed with formula only were 3.9 (95% CI, 1.8 to 8.7) times more likely to die of respiratory infections than those who were exclusively breastfed. The investigators controlled for markers of socioeconomic status but not for smoking, and smoking may be associated with less breastfeeding as well as greater risk for respiratory death.

Sensitivity Analysis for Unmeasured Confounding

Sensitivity analysis considers how strongly an unmeasured confounder would have to be related to the treatment and outcome to explain away the observed association. Several sensitivity analysis techniques have been developed for different statistical models (14–22, 24–41). Often, these techniques involve specifying several parameters corresponding to the strength of the effect of the unmeasured confounders on the treatment and on the outcome and then using analytic formulas to determine what the true effect of the treatment on the outcome would be if an unmeasured confounder of the specified strength were present. Such techniques are helpful in determining the strength of the evidence for causality.

These techniques sometimes are criticized for being too subjective—that is, regardless of the estimate obtained, investigators could choose the sensitivity parameters that make the result seem robust to confounding. Another criticism is that sensitivity analysis techniques themselves make simplifying assumptions about the unmeasured confounder. Such assumptions often stipulate that the unmeasured confounder is binary (22, 24, 26) or that only 1 unmeasured confounder exists (24–28), or that there is no interaction between the effects of the unmeasured confounder and of the treatment on the outcome (25–28). The criticisms then assert that these assumptions are needed to assess the effect of assumptions so that, in fact, the approach is not very useful after all.

These criticisms are not unreasonable; however, on the basis of recent developments, addressing them is now possible (37). Specifically, some techniques make no assumptions about the underlying structure of unmeasured confounders and still allow conclusions about the strength the unmeasured confounders must have to explain away an observed association (37). We begin by describing such a technique and then introduce the new E-value measure.

Suppose that an observational study controls for several covariates thought to be confounders of the treatment–outcome association. After adjustment, suppose that the estimated relative risk equals *RR*. We may, however, still be concerned that this estimate is subject to unmeasured confounding. Suppose that all confounding would be removed if the study had controlled for 1 or more unmeasured confounders *U*, along with the observed covariates. The sensitivity analysis technique requires 2 parameters to be specified. One corresponds to the strength of the association between the unmeasured confounders *U* and the outcome *D*; the other corresponds to the strength of the association between the treatment or exposure *E* and the unmeasured confounders. Once these parameters are specified, we can calculate the extent to which such a set of unmeasured confounders could alter the observed relative risk. We let *B* denote the largest factor by which the observed relative risk could be altered by unmeasured confounders of a particular strength.

In practice, we do not know the strengths of the unmeasured confounder associations, but we could, in principle, specify many different values and determine how the estimate is affected by each setting. Let RR_{UD} denote the maximum risk ratio for the outcome comparing any 2 categories of the unmeasured confounders, within either treatment group, conditional on the observed covariates. Let RR_{EU} denote the maximum risk ratio for any specific level of the unmeasured confounders comparing those with and without treatment, with adjustment already made for the measured covariates. Thus, RR_{UD} captures how important the unmeasured confounder is for the outcome, and RR_{EU} captures how imbalanced the treatment groups are in the unmeasured confounder *U*. For example, if 40% of nonbreastfeeding mothers smoked, as compared with 20% of breastfeeding mothers, we would have $RR_{EU} = 2$. The relationships are shown in **Figure 1**.

Key Summary Points

Motivation: Observational studies that attempt to assess causality between a treatment and an outcome may be subject to unmeasured confounding.

Rationale: Sensitivity analysis can assess how strong an unmeasured confounder would have to be to explain away an observed treatment–outcome relationship. A sensitivity analysis technique that is easy to use, present, and interpret, and does not itself make strong assumptions, is desirable.

Definition of E-value: The E-value is the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome, conditional on the measured covariates, to fully explain away a specific treatment–outcome association.

Calculation: The E-value for an estimate, and for the limit of a 95% CI closest to the null, can be calculated in a straightforward way for risk ratios (**Table 1**) and for other measures (**Table 2**).

Conclusions: The E-value allows an investigator to make statements of the following form: "The observed risk ratio of 3.9 could be explained away by an unmeasured confounder that was associated with both the treatment and the outcome by a risk ratio of 7.2-fold each, above and beyond the measured confounders, but weaker confounding could not do so; the confidence interval could be moved to include the null by an unmeasured confounder that was associated with both the treatment and the outcome by a risk ratio of 3.0-fold each, above and beyond the measured confounders, but weaker confounding could not do so."

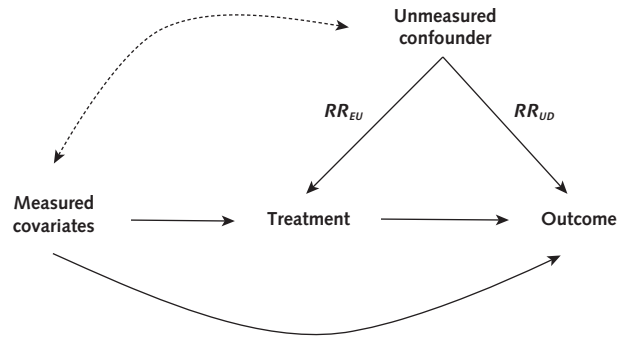
Once these variables are specified, the maximum relative amount by which such unmeasured confounding could reduce an observed risk ratio is given by the following formula (37):

$$B = RR_{UD}RR_{EU}/(RR_{UD} + RR_{EU} - 1).$$

To obtain the maximum amount this set of unmeasured confounders could alter an observed risk ratio *RR*, one simply divides the observed risk ratio by the bias factor *B* (37). In fact, one also may divide the limits of the CI by the bias factor *B* to obtain the maximum the unmeasured confounder could move the CI toward the null (37). The formula applies when the observed risk ratio *RR* is greater than 1. If the observed risk ratio is less than 1, then one multiplies by this bias factor rather than dividing by it.

We illustrate this approach with the association between maternal breastfeeding and respiratory death

Figure 1. Unmeasured confounder of the treatment–outcome relationship.



The maximum risk ratio for the outcome comparing any 2 categories of the unmeasured confounders, with adjustment already made for the measured covariates, is denoted in the diagram by RR_{UD}. The maximum risk ratio for any specific level of the unmeasured confounders comparing those with and without treatment, with adjustment already made for the measured covariates, is denoted in the diagram by RR_{EU}. The measured covariates are allowed to affect the unmeasured confounders, and vice versa.

from the study by Victora and colleagues (23), in which $RR = 3.9$ (CI, 1.8 to 8.7) for infants formula-fed rather than breastfed. Again, we might be worried that this estimate is confounded by smoking status. Suppose that the maximum ratio by which smoking could increase respiratory death is $RR_{UD} = 4$ and the maximum by which smoking differed by breastfeeding status was $RR_{EU} = 2$. Our bias factor is then $B = 4 \times 2/(4 + 2 - 1) = 1.6$. The most that unmeasured confounding could alter the effect estimate is obtained by dividing the observed risk ratio and its CI by 1.6: $RR = 3.9/1.6 = 2.43$ (CI, 1.1 to 5.4). Unmeasured confounding of this strength would not suffice to explain away the effect estimate.

One might object to a sensitivity analysis such as this because of the assumptions of specifying the strength of the confounding associations, RR_{UD} and RR_{EU}, and furthermore because an investigator could simply choose values of RR_{UD} and RR_{EU} that make the estimate seem robust. A potential remedy is to provide a large table with different values of RR_{UD} and RR_{EU}, including some that are large, to give readers and researchers a sense of how sensitive the conclusions are to potential unmeasured confounders (37). One also could plot all the values of RR_{UD} and RR_{EU} that suffice to explain away, or reverse, the association, as in Figure 2 for the estimate of $RR = 3.9$ from the study by Victora and colleagues (23). An alternative, and arguably simpler, approach is to report what we call the E-value, as described below.

THE E-VALUE FOR SENSITIVITY ANALYSIS

The E-value is the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome, conditional on the measured covariates, to explain away a treatment–outcome association. Rather than focusing on whether confounding of a specified strength would or would not suffice to explain away an effect estimate, as above, the E-value focuses on the magnitude of the confounder associations that could produce confounding bias equal to the observed treatment–outcome association. The investigator does not choose the variables but merely reports how strongly an unmeasured confounder must be related to the treatment and outcome to explain away an effect estimate; readers or other researchers may then assess whether the confounder associations of that magnitude are plausible.

E-value calculations are straightforward. For an observed risk ratio of RR:

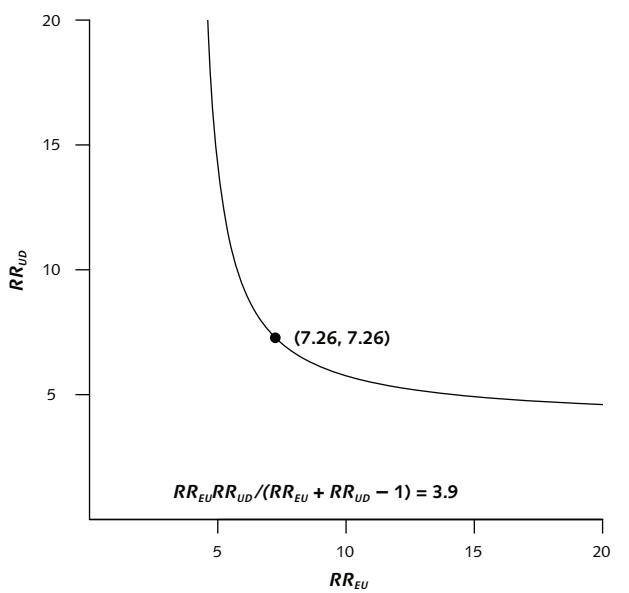
$$\text{E-value} = RR + \text{sqrt}\{RR \times (RR - 1)\}.$$

The proof appears elsewhere (37). The formula applies to a risk ratio greater than 1; for a risk ratio less than 1, one first takes the inverse of the observed risk ratio and then applies the formula. Thus, for the risk ratio $RR = 3.9$, one may obtain the E-value as follows:

$$\text{E-value} = 3.9 + \text{sqrt}\{3.9 \times (3.9 - 1)\} = 7.26.$$

From this E-value, we then may make statements such as "The observed risk ratio of 3.9 could be ex-

Figure 2. Value of the joint minimum strength of association on the risk ratio scale that an unmeasured confounder must have with the treatment and outcome to fully explain away an observed treatment–outcome risk ratio of $RR = 3.9$, as in the study by Victora and colleagues.



The E-value essentially sets the 2 parameters, RR_{UD} and RR_{EU}, equal to each other to determine the required minimum for both. The E-value for Victora and colleagues' (23) estimate corresponds to the point (7.26, 7.26). RR_{EU} = maximum risk ratio for any specific level of the unmeasured confounders comparing those with and without treatment, with adjustment already made for the measured covariates; RR_{UD} = maximum risk ratio for the outcome comparing any 2 categories of the unmeasured confounders, with adjustment already made for the measured covariates.

plained away by an unmeasured confounder that was associated with both the treatment and the outcome by a risk ratio of 7.2-fold each, above and beyond the measured confounders, but weaker confounding could not do so." The strength of an unmeasured confounder here is understood to be the maximum bias that could be generated in the bias formula for *B* given the confounder associations. Relatively strong confounding associations would be needed to completely explain away the observed treatment–outcome association of *RR* = 3.9.

The E-value is a continuous measure of an association's robustness to potential uncontrolled confounders. The lowest possible E-value is 1 (that is, no unmeasured confounding is needed to explain away the observed association). The higher the E-value, the stronger the confounder associations must be to explain away the effect. The E-value essentially sets the 2 parameters, RR_{UD} and RR_{EU}, equal to each other to determine the required minimum for both. The E-value for Victora and colleagues' estimate corresponds to the point (7.26, 7.26) in **Figure 2**. If 1 of the 2 parameters is smaller than the E-value, then the other must be larger. Sensitivity analysis makes clear quantitatively why Bradford Hill's criterion of "strength of association" (12) is important in establishing that a given association is, in fact, causal.

In practice, of course, we care not only about the estimate itself but also about the statistical uncertainty of the estimate, such as the CI for the estimate. For this reason, also reporting the E-value for the limit of the CI closest to the null is good practice. If the CI includes the null of a risk ratio of 1, then the E-value for the CI is simply 1 (because no confounding is needed to move the CI to include 1). Otherwise, one simply calculates, using the aforementioned formula, the E-value for the limit of the CI closest to the null. In the respiratory death example, the E-value for the lower limit of the CI (1.8) is obtained by applying the aforementioned formula, which produces an E-value for the confidence limit of 3.0.

An unmeasured confounder associated with respiratory death and breastfeeding by a risk ratio of 3.0-fold each could explain away the lower confidence limit, but weaker confounding could not. The evidence for causality from the E-value thus looks reasonably strong, because substantial unmeasured confounding would be needed to reduce the observed association or its CI to null.

Table 1 summarizes how to calculate E-values. For risk ratios, E-value calculations are straightforward. **Table 2** summarizes calculations for other effect measures. Some further worked examples for other effect measures are included in the **Supplement** (available at Annals.org).

The E-value interpretation, however, does depend on context, particularly on the measured covariates for which adjustment has been made. The E-value is the minimum strength of both the confounder associations that must be present, *above and beyond the measured covariates*, for an unmeasured confounder to explain away an association. Thus, for example, if 2 studies on breastfeeding had an E-value of 2.5 but 1 study had controlled for several indicators of socioeconomic status (educational attainment, income, occupation, homeownership, wealth) and the other had controlled for only a single binary marker of college education, then the former study would be more robust to unmeasured confounding, because in that study, an unmeasured confounder would have to be associated with both breastfeeding and the outcome by a risk ratio of 2.5-fold each, through pathways independent of several socioeconomic markers rather than just 1 of them.

Table 1. Calculating the E-Value for Risk Ratios

Estimate or CI, by Direction of Risk Ratio	Computation of the E-Value
***RR* >1**	
Estimate	E-value = $RR + \text{sqrt}\{RR \times (RR - 1)\}$
CI	If $LL \leq 1$, then E-value = 1 If $LL > 1$, then E-value = $LL + \text{sqrt}\{LL \times (LL - 1)\}$
***RR* <1**	
Estimate	Let $RR^* = 1/RR$ E-value = $RR^* + \text{sqrt}\{RR^* \times (RR^* - 1)\}$
CI	If $UL \geq 1$, then E-value = 1 If $UL < 1$, then let $UL^* = 1/UL$ and E-value = $UL^* + \text{sqrt}\{UL^* \times (UL^* - 1)\}$

LL = lower limit of the CI; *RR* = risk ratio; *RR** = inverse of *RR*; *UL* = upper limit of the CI; *UL** = inverse of *UL*.

The E-value results also do not guarantee that if a confounder with parameters of a particular strength existed, then it necessarily would explain away the effect, only that it is possible to construct scenarios in which it could (37). A rare unmeasured confounder would not bias an estimate as much. One of the strengths of the E-value approach is that it does not require one to specify the prevalence of unmeasured confounders or make assumptions about their nature. However, information about the distribution of the unmeasured confounder, when available, may be helpful in sensitivity analysis, and techniques are available to use such information (21, 25, 29). These techniques, however, make additional assumptions beyond the E-value approach.

The E-value also should be interpreted along with other strengths and weaknesses of the study and design. Unmeasured confounding is not the only source of potential bias in observational studies; measurement error, selection bias, and missing data also must be considered carefully in evaluating evidence. Other points of interpretation are discussed in **Table 3**. The E-value is a useful measure but must be interpreted in context.

FURTHER EXAMPLES

To illustrate the E-value's usefulness and interpretation, we consider the potential effects of breastfeeding on other childhood and maternal outcomes. A study by the Agency for Healthcare Research and Quality (42) reported the association between breastfeeding and childhood leukemia as *RR* = 0.80 (CI, 0.71 to 0.91). The *P* value less than 0.001 suggests strong evidence that breastfeeding and childhood leukemia are associated. However, is this association causal? We can calculate the E-value by first taking the inverse of the risk

Table 2. E-Values for Other Effect Measures

Effect Measure	Computation of Approximate E-Value
OR or HR for rare outcomes	When the outcome is relatively rare (e.g., <15%) by the end of follow-up, the E-value formula in **Table 1** may be used (37). In a case–control study, the outcome only needs to be rare in the underlying population, not in the case–control study.
Rate ratio for count and continuous outcomes	For ratio measures for count outcomes (or nonnegative continuous outcomes), the E-value may be found by replacing the risk ratio with the rate ratio (or the ratio of expected values) in the E-value formula in **Table 1** (37).
OR for common outcomes	When the outcome is common (>15% at the end of follow-up), an approximate E-value may be obtained by replacing the risk ratio with the square root of the OR (45), i.e., $RR \approx \text{sqrt(OR)}$, in the E-value formula in **Table 1**.
HR for common outcomes	When the outcome is common (>15% at the end of follow-up), an approximate E-value may be obtained (45) by applying the approximation $RR \approx (1 - 0.5^{\text{sqrt(HR)}})/(1 - 0.5^{\text{sqrt(1/HR)}})$ in the E-value formula in **Table 1**.
Difference in continuous outcomes	With standardized effect sizes d (mean of the outcome variable divided by the SD of the outcome) and an SE for this standardized effect size s_d, an approximate E-value may be obtained (45–47) by applying the approximation $RR \approx \exp(0.91 \times d)$ in the E-value formula. An approximate CI for the risk ratio may be found by using the approximation $(\exp\{0.91 \times d - 1.78 \times s_d\}, \exp\{0.91 \times d + 1.78 \times s_d\})$. This approach relies on additional assumptions and approximations. Other sensitivity analysis techniques have been developed for this setting (27–29), but they generally require additional assumptions, and the variables do not necessarily have a corresponding E-value.
Risk difference	If the adjusted risks for the treated and untreated are p_1 and p_0, then the E-value may be obtained by replacing the risk ratio with p_1/p_0 in the E-value formula. The E-value for the CI on a risk difference scale is more complex, and software to obtain this is described in the **Supplement** (available at Annals.org). Alternatively, if the outcome probabilities p_1 and p_0 are not very small or very large (e.g., if they are between 0.2 and 0.8), then the approximate approach for differences in continuous outcomes given previously may be used. Other sensitivity analysis techniques have been developed for this setting (27–29) but generally require additional assumptions and do not provide a corresponding E-value.

HR = hazard ratio; OR = odds ratio; *RR* = risk ratio.

ratio (because it is protective; see **Table 1**) and then applying the E-value formula that produces $E = 1.8$ for the estimate and $E = 1.4$ for the upper confidence limit. In contrast to the respiratory death estimate, comparatively weaker confounder associations could explain away the observed association and even weaker associations could move the CI to include a risk ratio of 1. An unmeasured confounder associated with childhood leukemia and breastfeeding by a risk ratio of 1.4-fold does not seem implausible. Breastfeeding might protect against leukemia, but the evidence for causality from the estimate is not nearly as strong as it was for respiratory death.

As another example, a study by Moorman and colleagues (43) indicated that in premenopausal women who breastfed for 6 to 12 months, the odds of developing ovarian cancer were 0.5 (CI, 0.3 to 0.8) times lower than in women who did not breastfeed; the analysis did not control for socioeconomic status. Here, the E-value is $E = 3.4$ for the estimate and $E = 1.8$ for the CI. In this case, the estimate seems moderately robust, but substantial confounder associations with breastfeeding and ovarian cancer could potentially move the CI to include 1. This perhaps constitutes some evidence for causality but is intermediate between the E-value obtained for respiratory death and the one for childhood leukemia.

Of interest, for ovarian cancer, the *P* value of 0.006 calculated from the CI, although still small, is not as extreme as it was for childhood leukemia as the outcome. However, for ovarian cancer, the E-value was more extreme than for childhood leukemia. Thus, for childhood leukemia, the evidence for association was stronger than it was for maternal ovarian cancer, but for maternal ovarian cancer, the evidence that the association was at least partially causal arguably was stronger than it was for childhood leukemia. As described in **Table 3**, the evidence provided by the *P* value and that provided by the E-value are distinct; the measures relate to different concepts; the measures may diverge; the *P* value is more dependent than the E-value on

Table 3. Issues of Interpretation of the E-Value

Issue	Interpretation
Likely effect sizes	The E-value should be interpreted in the context of the effect sizes that an unmeasured confounder is likely to have with respect to the outcome and treatment. In the context of biomedical and social sciences research, effect sizes ≥2- or 3-fold occasionally occur but are not particularly common; a variable that affects both treatment and outcome each by 2- or 3-fold would likely be even less common. For purposes of comparison, calculating the analogous E-value for each of the measured covariates if they had been omitted may be helpful.
E-values and sensitivity analysis	The E-value for the respiratory death example was 7.2. In the formula for the bias factor B, a confounder that was associated with the respiratory death by less than 7.2-fold might explain away the effect estimate but would have to be associated with the treatment by a risk ratio more than 7.2-fold. Values of the sensitivity analysis variables with a less extreme confounder–outcome association will require a more extreme treatment–confounder association, and vice versa.
Sample size, E-values, and *P* values	A large study with a precisely estimated association often has a very small *P* value; the *P* value may be made arbitrarily small by increasing the sample size. However, if the effect size is small, then the E-value will be small. The E-value depends on the magnitude of the association; it cannot be made arbitrarily large simply by increasing the sample size. The E-value for the CI does depend on the sample size. However, as the sample size increases, the E-value for the CI does not get arbitrarily large; it is bounded by the strength of the association (the limit sometimes is referred to in other contexts as the "design sensitivity" [17, 18]). A large sample size may give a small *P* value; a large effect size will give a large E-value.

sample size; and both measures arguably should be reported routinely in observational studies.

The E-value can assess the robustness of an association to potential unmeasured confounders and, in some cases, might provide strong evidence to support causality. As with the *P* value, however, the E-value cannot be used analogously to establish a null association definitively. The E-value may be used to conclude that the evidence for causality from a study is weak, but the absence of evidence for an effect is not the same as evidence of no effect. Nevertheless, as discussed in the **Supplement**, E-values still may be used to assess how much confounding would be required to move a near-null association to clinically meaningful levels, or even to reverse the direction of the association. In fact, the E-value approach may be used to assess the minimum strength of association that an unmeasured confounder would need to have with both the treatment and outcome to move the observed estimate to any other value. The E-value need not be used only to assess overall evidence for causality but also may be used simply to determine how unmeasured confounders might change adjusted associations. The **Supplement** discusses the use of E-values for these other purposes.

DISCUSSION

We propose that all observational studies that assess causality (that is, are not strictly about description or predictive or prognostic modeling) report the E-value for the estimate and the CI or use some other sensitivity analysis technique. Journals should strongly encourage such reporting. An investigator may use text such as we have outlined in our examples (see the Key Summary Points) as succinct but highly informative statements about evidence of causality. E-values also may be reported for any estimate discussed in a systematic review.

Interpretative practices must change. The *P* value sometimes is used as the central measure of evidence for causality in randomized trials. Although potentially subject to misuse and misinterpretation (1, 5–12), the *P* value may be informative as a continuous measure of evidence as to whether an association is present. With observational studies, however, association does not imply causation, and relying on the *P* value is wholly inadequate. Unmeasured confounding is often the central challenge in assessing evidence for causality in observational research, and E-values assess robustness to such unmeasured confounding, thereby supplementing *P* values.

The E-value, of course, does not address all issues of bias. It does not assess measurement error or selection bias, nor does it address selective reporting of results (such as when multiple tests are done, and only significant ones are reported).

A possible objection might be, "It was already difficult enough to achieve a *P* value below some threshold; are you now going to require a large E-value as well?" We do not propose any threshold cutoff for the E-value. Enough mischief has been done by the arbitrary 0.05 *P*-value cutoff (1, 5–12). The E-value, like the *P* value, is a continuous measure. Publication decisions should never rest simply on the magnitude of measures such as the *P* value or E-value.

Moreover, it is possible to obtain a very precise estimate of a relatively small effect from a large, well-designed observational study that has extensive covariate control with a very narrow CI. If the association is not strong, the E-value will be quite small. The robustness to unmeasured confounding, as well as the evidence for causality, thus might be weak. However, this result does not mean that no effect exists. Moreover, it does not mean that the study should not be published. The study may be the best we can do with observational data; therefore, knowing this fact would be important and worth publishing. As noted earlier, a small E-value also does not mean that there is evidence for no effect; it implies only that the evidence for an effect is itself weak. However, weak evidence for an effect does not imply evidence that the effect is absent.

E-values likewise may be computed at the study design stage for hypothesized estimates when consideration is given to whether, and to what extent, covariates will be available to adequately control for confounding. A small E-value for a hypothesized estimate may indicate that one should not proceed with an observational study but should wait until resources are adequate to carry out a randomized trial.

Reporting and accurate assessment of evidence for causality are important; the E-value assists with these tasks. Observational research is sometimes criticized on the grounds that its results constantly are being overturned (44). This state of affairs arises in part from overreliance on the *P* value and inadequate assessment of robustness to biases, such as unmeasured or uncontrolled confounding. Again, the E-value would help with this task. Introduction of the E-value may sometimes make publication more difficult and may be subject to editorial abuse. However, the end of science is not publication but rather a collective attempt to arrive, as best as possible, at the truth. Our hope is that the E-value will be of use in this regard. We believe its use should become routine.

From Harvard T.H. Chan School of Public Health, Boston, Massachusetts, and University of California, Berkeley, Berkeley, California.

Acknowledgment: The authors thank Sander Greenland, James Robins, 2 reviewers, and the editors for helpful comments on an earlier draft of this paper.

Grant Support: By National Institutes of Health grant ES017876.

Disclosures: Authors have disclosed no conflicts of interest. Forms can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M16-2607.

Requests for Single Reprints: Tyler J. VanderWeele, PhD, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115; e-mail, tvanderw@hsph.harvard.edu.

Current author addresses and author contributions are available at Annals.org.

References

1. **Wasserstein RL, Lazar NL.** The ASA's statement on p-values: context, process, and purpose. Am Stat. 2016;70:129-33.
2. **Altman DG, Machin D, Bryant TN, Gardner MJ, eds.** Statistics with Confidence. 2nd ed. London: BMJ Books; 2000.
3. **Rosner B.** Fundamentals of Biostatistics. 8th ed. Boston: Cengage Learning; 2015.
4. **Pagano M, Gavreau K.** Principles of Biostatistics. Belmont, CA: Brooks/Cole; 2000.
5. **Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al.** Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31:337-50. [PMID: 27209009] doi:10.1007/s10654-016-0149-3
6. **Goodman S.** A dirty dozen: twelve p-value misconceptions. Semin Hematol. 2008;45:135-40. [PMID: 18582619] doi:10.1053/j.seminhematol.2008.04.003
7. **Greenland S.** Null misinterpretation in statistical testing and its impact on health risk assessment. Prev Med. 2011;53:225-8. [PMID: 21871481] doi:10.1016/j.ypmed.2011.08.010
8. **Greenland S, Poole C.** Problems in common interpretations of statistics in scientific articles, expert reports, and testimony. Jurimetrics. 2011;51:11329.
9. **Sterne JA, Davey Smith G.** Sifting the evidence–what's wrong with significance tests? BMJ. 2001;322:226-31. [PMID: 11159626]
10. **Stang A, Poole C, Kuss O.** The ongoing tyranny of statistical significance testing in biomedical research. Eur J Epidemiol. 2010;25:225-30. [PMID: 20339903] doi:10.1007/s10654-010-9440-x
11. **Goodman SN.** Toward evidence-based medical statistics. 1: The *P* value fallacy. Ann Intern Med. 1999;130:995-1004. [PMID: 10383371]
12. **Hill AB.** The environment and disease: association or causation? Proc R Soc Med. 1965;58:295-300. [PMID: 14283879]
13. **Greenland S.** Randomization, statistics, and causal inference. Epidemiology. 1990;1:421-9. [PMID: 2090279]
14. **Imbens GW, Rubin DB.** Sensitivity analysis and bounds. In: Causal Inference for Statistics, Social, and Biomedical Sciences. New York: Cambridge Univ Pr; 2015:496-512.
15. **Hernan MA, Robins JR.** Confounding. In: Causal Inference. 11 September 2016. Accessed at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2016/09/hernanrobins_v1.10.31.pdf on 22 May 2017.
16. **Rosenbaum PR.** Sensitivity to hidden bias. In: Observational Studies. 2nd ed. New York: Springer; 2002:105-70.
17. **Rosenbaum PR.** Design sensitivity. In: Design of Observational Studies. New York: Springer; 2010:269-71.
18. **Rosenbaum PR.** Design sensitivity and efficiency in observational studies. J Am Stat Assoc. 2010;105:692-702.
19. **Greenland S.** Multiple-bias modeling for analysis of observational data. J R Stat Soc Ser A. 2005;168:267-308.
20. **Lash TL, Fox MP, Fink AK.** Applying Quantitative Bias Analysis to Epidemiologic Data. New York: Springer; 2009.
21. **Greenland S, Lash TL.** Bias analysis. In: Rothman KJ, Greenland S, Lash TL, eds. Modern Epidemiology. Philadelphia: Lippincott Williams & Wilkins; 2008:345-80.
22. **Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL.** Smoking and lung cancer: recent evidence and a discussion of some questions. J Natl Cancer Inst. 1959;22:173-203. [PMID: 13621204]
23. **Victora CG, Smith PG, Vaughan JP, Nobre LC, Lombardi C, Teixeira AM, et al.** Evidence for protection by breast-feeding against infant deaths from infectious diseases in Brazil. Lancet. 1987;2:319-22. [PMID: 2886775]
24. **Bross ID.** Spurious effects from an extraneous variable. J Chronic Dis. 1966;19:637-47. [PMID: 5966011]
25. **Schlesselman JJ.** Assessing effects of confounding variables. Am J Epidemiol. 1978;108:3-8. [PMID: 685974]
26. **Rosenbaum PR, Rubin DB.** Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. J R Stat Soc Series B Stat Methodol. 1983;45:212-8.
27. **Lin DY, Psaty BM, Kronmal RA.** Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics. 1998;54:948-63. [PMID: 9750244]
28. **Imbens GW.** Sensitivity to exogeneity assumptions in program evaluation. Am Econ Rev. 2003;93:126-32.
29. **Vanderweele TJ, Arah OA.** Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology. 2011;22:42-52. [PMID: 21052008] doi:10.1097/EDE.0b013e3181f74493
30. **Bross ID.** Pertinency of an extraneous variable. J Chronic Dis. 1967;20:487-95. [PMID: 6028268]
31. **Lee WC.** Bounding the bias of unmeasured factors with confounding and effect modifying potentials. Stat Med. 2011;30:1007-17.
32. **Robins J M, Scharfstein D, Rotnitzky A.** Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran E, Berry D, eds. Statistical Models for Epidemiology, the Environment, and Clinical Trials. New York: Springer-Verlag; 2000:1-94.
33. **McCandless LC, Gustafson P, Levy A.** Bayesian sensitivity analysis for unmeasured confounding in observational studies. Stat Med. 2007;26:2331-47. [PMID: 16998821]
34. **Brumback BA, Hernán MA, Haneuse SJ, Robins JM.** Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. Stat Med. 2004;23:749-67. [PMID: 14981673]
35. **VanderWeele TJ.** Sensitivity analysis for mediation. In: Explanation in Causal Inference: Methods for Mediation and Interaction. New York: Oxford Univ Pr; 2015:66-97.
36. **VanderWeele TJ.** Sensitivity analysis for contagion effects in social networks. Sociol Methods Res. 2011;40:240-55. [PMID: 25580037]
37. **Ding P, VanderWeele TJ.** Sensitivity analysis without assumptions. Epidemiology. 2016;27:368-77. [PMID: 26841057] doi:10.1097/EDE.0000000000000457
38. **Gilbert PB, Bosch RJ, Hudgens MG.** Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. Biometrics. 2003;59:531-41. [PMID: 14601754]
39. **Chiba Y, VanderWeele TJ.** A simple method for principal strata effects when the outcome has been truncated due to death. Am J Epidemiol. 2011;173:745-51. [PMID: 21354986] doi:10.1093/aje/kwq418
40. **Huang TH, Lee WC.** Bounding formulas for selection bias. Am J Epidemiol. 2015;182:868-72. [PMID: 26519426] doi:10.1093/aje/kwv130
41. **Rosenbaum PR.** Discussing hidden bias in observational studies. Ann Intern Med. 1991;115:901-5. [PMID: 1952480]
42. **Ip S, Chung M, Raman G, Chew P, Magula N, DeVine D, et al.** Breastfeeding and Maternal and Infant Health Outcomes in Developed Countries. Evidence Report/Technology Assessment no. 153. (Prepared by Tufts-New England Medical Center Evidence-based Practice Center under contract no. 290-02-0022.) AHRQ publication no. 07-E007. Rockville: Agency for Healthcare Research and Quality; April 2007.
43. **Moorman PG, Calingaert B, Palmieri RT, Iversen ES, Bentley RC, Halabi S, et al.** Hormonal risk factors for ovarian cancer in premenopausal and postmenopausal women. Am J Epidemiol. 2008;167:1059-69. [PMID: 18303003] doi:10.1093/aje/kwn006
44. **Taubes G.** Epidemiology faces its limits. Science. 1995;269:164-9. [PMID: 7618077]
45. **VanderWeele TJ.** On a square-root transformation of the odds ratio for a common outcome. Epidemiology. 2017. [Forthcoming].
46. **Borenstein M, Hedges LV, Higgins JPT, Rothstein HR.** Converting among effect sizes. In: Introduction to Meta-Analysis. Hoboken, NJ: Wiley; 2009:45-51.
47. **Hasselblad V, Hedges LV.** Meta-analysis of screening and diagnostic tests. Psychol Bull. 1995;117:167-78. [PMID: 7870860]

Current Author Addresses: Dr. VanderWeele: Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115.
Dr. Ding: University of California, Berkeley, 425 Evans Hall, Berkeley, CA 94720.

Author Contributions: Conception and design: T.J. VanderWeele, P. Ding.
Analysis and interpretation of the data: T.J. VanderWeele, P. Ding.
Drafting of the article: T.J. VanderWeele.
Critical revision for important intellectual content: T.J. VanderWeele.
Final approval of the article: T.J. VanderWeele, P. Ding.
Provision of study materials or patients: T.J. VanderWeele.
Statistical expertise: T.J. VanderWeele, P. Ding.
Obtaining of funding: T.J. VanderWeele.
Administrative, technical, or logistic support: T.J. VanderWeele.
Collection and assembly of data: T.J. VanderWeele, P. Ding.