



# Sensitivity Analysis in Observational Research: Introducing the E-Value

## Citation

VanderWeele, Tyler J., and Peng Ding. 2017. "Sensitivity Analysis in Observational Research: Introducing the E-Value." *Annals of Internal Medicine* (July 11). doi:10.7326/m16-2607.

## Published Version

doi:10.7326/M16-2607

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:36874927>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

## **Sensitivity analysis in observational research: introducing the E-value**

Tyler J. VanderWeele, Ph.D., Department of Epidemiology, Harvard T.H. Chan School of Public  
Health

and Peng Ding, Ph.D., Department of Statistics, University of California, Berkeley

**Abstract.** Sensitivity analysis can be useful in assessing how robust associations are to potential unmeasured or uncontrolled confounding. In this paper we introduce a new measure that we call the “E-value,” a measure related to the *evidence for causality* in observational studies, when they 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 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. We propose that in all observational studies intended to produce evidence for causality, the E-value be reported, or some other sensitivity analysis be used. We suggest calculating the E-value for both the observed association estimate (after adjustments for measured confounders) and for 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 be improved considerably, and ultimately, science would be strengthened.

Word Count: 3560

## Introduction

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, i.e., that some third factor related to both the treatment and the 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, i.e., how strong the unmeasured confounder would have to be associated with the treatment and outcome for the treatment-outcome association not to be causal. Sensitivity analysis makes clear quantitatively why Bradford Hill's criteria of "strength of association" (12) is important in establishing that a given association is in fact 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 the outcome to fully explain away a specific treatment-outcome association, conditional on the

measured covariates. To implement these sensitivity analysis techniques and to obtain E-values is relatively simple. If reporting E-values for sensitivity analysis were to become standard practice, our ability to assess evidence from observational studies would be improved and science would be strengthened.

### **Example**

As a motivating example, numerous 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. A population-based case-control study by Victora and colleagues (23) examined associations between breastfeeding and infant death by respiratory infection. After adjustment for age, birth weight, social status, maternal education, and family income, infants fed by formula only were 3.9 (95%CI:1.8,8.7) times more likely to die by respiratory infections than those who were exclusively breast-fed. The investigators controlled for markers of socioeconomic status but not for smoking, and smoking may be associated with less breastfeeding and also greater respiratory death.

### **Sensitivity analysis for unmeasured confounding**

Sensitivity analysis considers how strong an unmeasured confounder would have to be related to the treatment and the outcome to explain away the observed association. Numerous sensitivity analysis techniques have been developed for different statistical models (14-22,24-40). Often

these techniques involve specifying a number of parameters corresponding to the strength of the effect of the unmeasured confounder on the treatment and on the outcome and then using analytic formulae to determine what the true effect of the treatment on the outcome would be if an unmeasured confounder of specified strength were present. Such techniques are helpful in determining the strength of the evidence for causality.

These techniques are sometimes criticized for too much subjectivity – that regardless of the estimate that was in fact obtained, investigators will choose the sensitivity parameters so that the result looks robust to confounding. Other criticisms include 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 there is only one unmeasured confounder (24-28) or that there is no interaction between the effects of the unmeasured confounder and the treatment on the outcome (25-28). The criticisms then assert that these assumptions are needed to assess the impact of assumptions so that in fact the approach is not so useful after all.

These are not unreasonable criticisms. However, based on recent developments, it is now in fact possible to address them (37). Specifically, some techniques make no assumptions about the underlying structure of unmeasured confounders and still allow conclusions about the strength the unmeasured confounder(s) needed to explain away an observed association (37). We begin by describing one technique and then introduce the new E-value measure.

Suppose that an observational study controls for a number of covariates thought to be

confounders of the treatment-outcome association. After adjustment suppose that the estimated relative risk=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 could have controlled for one or more unmeasured confounders U, along with the observed covariates. The sensitivity analysis technique requires specifying two parameters. One corresponds to the strength of the association between the unmeasured confounder(s) U and the outcome D; the other corresponds to the strength of the association between the treatment or exposure E unmeasured confounder(s). Once these parameters are specified, we can calculate the most such a set of unmeasured confounders could alter the observed relative risk. We will let B denote the largest factor by which the observed relative risk could be altered by unmeasured confounder(s) of a particular strength.

In practice, we will not know the strengths of the unmeasured confounder associations, but we could in principle specify numerous different values and see how the estimate is affected by each setting. Let  $RR_{UD}$  denote the maximum risk ratio for the outcome comparing any two categories of the unmeasured confounder(s), within either treatment group, conditional on the observed covariates. Let  $RR_{EU}$  denote the maximum risk ratio for any specific level of the unmeasured confounder(s) comparing those with and without treatment, having already adjusted 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 non-breastfeeding mothers smoked as compared to 20% of breastfeeding mothers we would have  $RR_{EU}=2$ . The relationships are shown in Figure 1.

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

$$B = \frac{RR_{UD}RR_{EU}}{(RR_{UD} + RR_{EU} - 1)}$$

To obtain the most that this set of unmeasured confounders could alter an observed risk ratio RR one simply divides the observed risk ratio by this bias factor B (37). In fact, one can also divide the limits of the confidence interval by the bias factor B to obtain the maximum the unmeasured confounder(s) could move the confidence interval towards 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 from the Victora study (23), in which  $RR=3.9$  (95%CI:1.8,8.7) for those formula-fed rather than breast-fed. We might again be worried that this estimate is confounded by smoking status. Suppose that the maximum ratio by which smoking could increase respiratory death is  $RR_{UD}=4$ -fold and the maximum by which smoking differed by breastfeeding status was  $RR_{EU}=2$ -fold. 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 confidence interval by 1.6:  $RR=3.9/1.6=2.43$  (95%CI:1.1,5.4). Unmeasured confounding of this strength would not suffice to explain away the effect estimate.

One might object to a sensitivity analysis like this because of the assumptions of specifying the



strength of the confounding parameters,  $RR_{UD}$  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 would be to provide a large table with different values of  $RR_{UD}$  and  $RR_{EU}$ , including some which are large, to give readers and researchers a sense as to how sensitive the conclusions are to potential unmeasured confounders (37). One could also 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 Victora study. An alternative, and arguably simpler approach is to report what we will call the “E-value.”

### **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 the outcome, conditional on the measured covariates, to explain away a treatment-outcome association. Rather than focusing on whether confounding of a specified strength would 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 is not choosing the parameters but merely reporting how much an unmeasured confounder would have to be related to the treatment and outcome to explain away an effect estimate; readers or other researchers can then assess whether the confounder associations of that magnitude are plausible.

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

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

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 above of RR=3.9 one can obtain the E-value as follows:

$$E\text{-value} = 3.9 + \sqrt{3.9 \times (3.9 - 1)} = 7.2.$$

From this E-value we could then conclude that “To explain away the observed risk ratio of RR=3.9, an unmeasured confounder that was associated with respiratory death and with breastfeeding by a risk ratio of 7.2-fold each, above and beyond the measured confounders, could explain it away, but weaker confounding could not” where the strength of an unmeasured confounder here is understood as 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 this observed treatment-outcome association.

The E-value is a continuous measure of how robust the association is to potential uncontrolled confounders. The lowest possible E-value is 1 (i.e. no unmeasured confounding is needed to explain away the observed association). The higher the E-value the stronger the confounder associations would have to be to explain away the effect. The E-value essentially sets the two parameters,  $RR_{UD}$  and  $RR_{EU}$ , equal to each other to see the minimum that they would both need

to be. The E-value for the Victora estimate corresponds to the point (7.26,7.26) on Figure 2. If one of the two parameters were smaller than the E-value, then the other would have to be larger.

In practice, of course, we care not simply about the estimate itself but also about the statistical uncertainty of the estimate, e.g., the confidence interval for the estimate. For this reason, good practice would be to report the E-value also for the limit of the confidence interval closest to the null. If the confidence interval includes the null of a risk ratio of 1, then the E-value for the confidence interval is simply 1 (since no confounding is needed to move the confidence interval to include 1). Otherwise, one simply calculates, using the formula above, the E-value for the limit of the confidence interval closest to the null. In the case of the respiratory death example, the E-value, for the lower limit of the confidence interval (1.8), is obtained by applying the formula above which produces an E-value for the confidence limit of 3.0.

An unmeasured confounder that was associated with respiratory death and with breastfeeding by risk ratios 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 as it would take substantial unmeasured confounding to reduce the observed association 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 Appendix.

The E-value interpretation does, however, depend on context, and in particular on the measured

covariates for which adjustment has been made. The E-value is the minimum strength of both of 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 two different studies of breastfeeding had the same E-value of 2.5, but one had controlled for multiple indicators of socioeconomic status (educational attainment, income, occupation, home-ownership, 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, since, in the former study, an unmeasured confounder would have to be associated with both breastfeeding and also the outcome by a risk ratio of 2.5-fold each, through pathways independent of multiple (rather than just one of the) socioeconomic markers.

The E-value results also do not guarantee that if there were a confounder with parameters of a particular strength 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 specifying the prevalence of, or making assumptions about the nature of, unmeasured confounders. But information about the distribution of the unmeasured confounder may be helpful in sensitivity analysis, and techniques are available to do so (21,25,29). These techniques do, however, make additional assumptions beyond the E-value approach.

The E-value should also 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 must also be carefully considered in

evaluating evidence. Other points of interpretation are discussed in Table 3. The E-value is a useful measure but needs to 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  $RR=0.80$  (95%CI:0.71,0.91). The  $p\text{-value}<0.001$  suggests strong evidence that breastfeeding and childhood leukemia are *associated*. But is this association causal? We can calculate the E-value by first taking the inverse of the risk ratio (since it is protective) and then applying the E-value formula which produces  $E=1.8$  for the estimate, and  $E=1.4$  for the upper confidence limit. In contrast with the respiratory death estimate, comparatively weaker confounder associations could explain away the observed association and even less could move the confidence interval to include a risk ratio of 1. An unmeasured confounder that was associated with childhood leukemia and with breastfeeding by a risk ratio of 1.4-fold does not seem implausible. Breastfeeding might have a protective effect on leukemia but the evidence for causality is not nearly as strong as it was for respiratory death.

As another example, a study by Moorman and colleagues (43) indicated that premenopausal women who breastfed 6-12 months were at a 0.5 (95%CI:0.3,0.8) lower odds of ovarian cancer than those who did not breastfeed; the analysis did not control for socioeconomic status. Here the E-value for the estimate is  $E=3.4$  and the E-value for the confidence interval is  $E=1.8$ . In this

case the estimate seems moderately robust, but substantial confounder associations, with breastfeeding and ovarian cancer, could potentially move the confidence interval to include 1. This constitutes perhaps some evidence for causality, but is intermediate between the E-values obtained for respiratory death and that for childhood leukemia.

Interestingly, for ovarian cancer, the p-value 0.006 calculated from the confidence interval, while still small, is not as extreme as it was for childhood leukemia as the outcome. But for ovarian cancer the E-value was more extreme than for childhood leukemia. Thus, for childhood leukemia, the evidence for *association* was stronger than for maternal ovarian cancer, but for maternal ovarian cancer, the evidence that the association is at least partially *causal* is arguably stronger than it is for childhood leukemia. As described in Table 3, the evidence provided by the p-value and the E-value are distinct; they are assessing different concepts; they can diverge; the p-value is more dependent on sample size than is the E-value; and both should arguably be routinely reported in observational studies.

The E-value can assess how robust an association is to potential unmeasured confounders and, in some cases, might provide strong evidence in support of causality. As with the p-value, however, the E-value cannot be used in an analogous way to definitively establish a null association. The E-value can be used to conclude that the evidence for causality from a study is weak; but the absence of evidence for an effect is not equivalent to evidence of no effect. However, as discussed in the Appendix, E-values can nevertheless still be used to assess how much confounding would be required to move a near-null association to levels that are clinically meaningful, or how much confounding would be needed even to reverse the direction of the

association. In fact, the E-value approach can be used to assess the minimum strength of association that an unmeasured confounder would need to have with both the treatment and the outcome to move the observed estimate to any other value. The E-value need not be used only for assessing overall evidence for causality but can be used simply to see how unmeasured confounders might change adjusted associations. The Appendix discusses the use of E-values for these other various purposes.

## **Discussion**

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

Interpretative practices need to change. The p-value is sometimes used as the central measure of evidence for causality in randomized trials. While potentially subject to mis-use and misinterpretation (1,5-12), the p-value can 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 and thereby supplement 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 (e.g., 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 also going to require a large E-value as well?” We do not propose any threshold cut-off for the E-value. Enough mischief has been done by the arbitrary 0.05 p-value cut-off (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 like the p-value or the 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 confidence interval. If the association is not strong, the E-value will be quite small. The robustness to unmeasured confounding, and the evidence for causality, might thus be weak. But this result does not mean that there is no effect. And it, moreover, does not mean that the study should not be published. The study may be the best we can do with observational data, and it would then be important to know this fact. As noted above, a small E-value also does *not* mean that there is evidence for no effect; it only implies that the *evidence* for an effect is itself weak; but weak evidence for an effect does *not* imply evidence that the effect is absent. E-values can likewise be computed at the study design stage for hypothesized estimates when consideration is being made as 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 it is best not to



proceed with an observational study but to wait until resources are adequate to carry out a randomized trial.

It is important that evidence for causality be reported and accurately assessed. The E-value assists with that task. Observational research is sometimes criticized on the grounds that its results are constantly being overturned (44). This lack of reproducibility arises in part from too much reliance on the p-value and inadequate assessment of robustness to biases such as unmeasured or uncontrolled confounding. Again, the E-value would assist with this task. The introduction of the E-value may sometimes make publication more difficult, and may be subject to editorial abuse. But the end of science is not publication, but rather a collective attempt to come, as best as possible, to truth. Our hope is that the E-value will be of use in this regard. We believe its use should become routine.

## References

1. Ronald L. Wasserstein & Nicole A. Lazar (2016) The ASA's Statement on p-Values: Context, Process, and Purpose, *The American Statistician*, 70:2, 129-133.
2. Altman, D.G., Machin, D., Bryant, T.N., and Gardner, M.J. (eds.) (2000), *Statistics with Confidence* (2nd ed.), London: BMJ Books.
3. Rosner B. *Fundamentals of Biostatistics*. Cengage Learning; 8th edition, 2015.
4. Pagano, M. and Gavreau, K. *Principles of Biostatistics*. Brooks/Cole, 2000.
5. Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N. and Altman, D.G.: "Statistical Tests, P-values, Confidence Intervals, and Power: A Guide to Misinterpretations. *European Journal of Epidemiology*, 31(4):337-350.
6. Goodman, S. (2008), "A Dirty Dozen: Twelve P-Value Misconceptions," *Seminars in Hematology*, 45, 135–140.
7. Greenland, S. (2011), "Null Misinterpretation in Statistical Testing and its Impact on Health Risk Assessment," *Preventive Medicine*, 53, 225–228.

8. Greenland, S., and Poole, C. (2011), “Problems in Common Interpretations of Statistics in Scientific Articles, Expert Reports, and Testimony,” *Jurimetrics*, 51, 113–129.
9. Sterne, J. A. C., and Smith, G. D. (2001), “Sifting the Evidence—What’s Wrong with Significance Tests?” *British Medical Journal*, 322, 226–231.
10. Stang, A., Poole, C., and Kuss, O. (2010), “The Ongoing Tyranny of Statistical Significance Testing in Biomedical Research,” *European Journal of Epidemiology*, 25, 225–230.
11. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med*. 1999;130(12):995-1004.
12. Hill, A.B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*. 58(5):295–300.
13. Greenland, S. Randomization, statistics, and causal inference. *Epidemiology* 1990;1:421-429.
14. Imbens GW, Rubin DB. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015, Chapter 22.
15. Hernan M, Robins JM. *Causal Inference*. Chapman Hall, Expected Publication December 2017, Chapter 7. Available at: [https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2016/09/hernanrobins\\_v1.10.31.pdf](https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2016/09/hernanrobins_v1.10.31.pdf)

16. Rosenbaum PR. *Observational Studies*. Second Edition. Springer, 2002, Chapter 4.
17. Rosenbaum PR. *Design of Observational Studies*. Springer, 2010, Section 14.3.
18. Rosenbaum PR. (2010) Design sensitivity and efficiency in observational studies. *Journal of the American Statistical Association*, 105:490, 692-702.
19. Greenland, S. Multiple-bias modeling for analysis of observational data (with discussion). *Journal of the Royal Statistical Society Series A*, 2005; 168, 267-308.
20. Lash, T.L., Fox, M.P., and Fink, A.K. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York: Spring, 2009.
21. Rothman, K.J., Greenland, S., and Lash, T.L. *Modern Epidemiology*. Lippincott, 2008, Chapter 19.
22. 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.*, 22:173–203, 1959.
23. Victora C, Vaughan JP, Lombardi C, Fuchs SMC, Gigante LP, Smith PG, Nobre LC, Teixeira AMB, Moreira LB, Barros FC. Evidence for protection by breast-feeding against infant deaths from infectious diseases in Brazil. *Lancet*, 1987, Vol.330(8554):319-322.

24. Bross IDJ. Spurious effects from an extraneous variable. *J Chronic Dis.*, 19:637–647, 1966.
25. Schlesselman JJ. Assessing effects of confounding variables. *Am J Epidemiol.*, 108:3–8, 1978.
26. Rosenbaum PR and Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J R Stat Soc Series B*, 45:212–218, 1983.
27. Lin DY, Psaty BM, and Kronrnal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*, 54:948–963, 1998.
28. Imbens, G.W. Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 2003; 93, 126-132.
29. VanderWeele TJ and Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, 22(1):42–52, 2011.
30. Bross IDJ. Pertinency of an extraneous variable. *J Chronic Dis.*, 20:487–495, 1967.
31. Lee WC. Bounding the bias of unmeasured factors with confounding and effectmodifying potentials. *Stat Med.*, 30:1007–1017, 2011.

32. Robins, J. M., Scharfstein, D., and Rotnitzky, A. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: *Statistical Models for Epidemiology, the Environment, and Clinical Trials*. Halloran, E. and Berry, D. (eds), 1-94. New York: Springer-Verlag.
33. McCandless, L. C., Gustafson, P., and Levy, A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in Medicine*, 2007; 26, 2331-2347.
34. Brumback, B. A., Hernán, M. A., Haneuse, S. J. P. A. and Robins J. M. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Statistics in Medicine*, 2004; 23, 749-767.
35. VanderWeele, T.J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Oxford University Press, Chapter 3.
36. VanderWeele, T.J. (2011). Sensitivity analysis for contagion effects in social networks. *Sociological Methods and Research*, 40:240-255.
37. Ding, P., and VanderWeele, T.J. (2016). Sensitivity analysis without assumptions. *Epidemiology*, 27;3:368-377.
38. Gilbert, P.B., Bosch, R., and Hudgens, M.G. (2003). Sensitivity analysis for the assessment

of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics*, 59:531-541.

39. Chiba, Y. and VanderWeele, T.J. (2011). A simple method for principal strata effects when the outcome has been truncated due to death. *American Journal of Epidemiology*, 173:745-751.

40. Hsuan T.-H., and Lee, W.-C. (2015). Bounding formulas for selection bias. *American Journal of Epidemiology*, 182:868-872.

41. Rosenbaum, P.R. (1991). Discussing hidden bias in observational studies. *Annals of Internal Medicine*, 115:901-905.

42. Ip S, Chung M, Raman G, Chew P, Magula N, DeVine D, Trikalinos T, Lau J. 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, MD: Agency for Healthcare Research and Quality. April 2007.

43. Moorman, X. Y., Calingaret, B., Palmieri, R. T., Iversen, E. S., Bentley, R. C., Halabi, S., Berchuck, A., and Schildkraut, J. M. (2008). Hormonal risk factors for ovarian cancer in premenopausal and postmenopausal women. *American Journal of Epidemiology*, 167:1059-1069.

44. Taubes G. Epidemiology faces its limits. *Science*. 1995 Jul 14;269(5221):164-169.
45. VanderWeele, T.J. On a square-root transformation of the odds ratio for a common outcome. *Epidemiology*, in press.
46. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Wiley, 2009, Chapter 7.
47. Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, 1995;117:167-178.
48. Oddy WH, Smith GJ, Jacoby P. A possible strategy for developing a model to account for attrition bias in a longitudinal cohort to investigate associations between exclusive breastfeeding and overweight and obesity at 20 years. *Annals of Nutrition and Metabolism* 2014, 65:234-235.
49. Reinisch, J., Sanders, S., Mortensen, E. and Rubin D.B. In-utero exposure to phenobarbital and intelligence deficits in adult men. *Journal of the American Medical Association*, 1995; 274, 1518-1525.
50. Huybrechts KF, Palmsten K, Avorn J, Cohen LS, Holmes LB, Franklin JM, Mogun H, Levin R, Kowal M, Setoguchi S, Hernández-Díaz S. Antidepressant use in pregnancy and the risk of cardiac defects. *N Engl J Med* 2014;370:2397-407.



51. Hammond, E. C. and Horn, D. Smoking and death rates: report on forty four months of follow-up of 187,783 men. *Journal of the American Medical Association*, 1958;166:1159–1172, 1294–1308.

Figure 1 Legend. Diagram illustrating an unmeasured confounder of the treatment-outcome relationship. The maximum risk ratio for the outcome comparing any two categories of the unmeasured confounder(s), having already adjusted for the measured covariates, is denoted in the diagram by  $RR_{UD}$ . The maximum risk ratio for any specific level of the unmeasured confounder(s) comparing those with and without treatment, having already adjusted for the measured covariates, is denoted in the diagram by  $RR_{EU}$ . The measured covariates are allowed to affect the unmeasured confounder(s) and/or vice versa.

Figure 2 Legend. Figure illustrating the value of the joint minimum strength of association on the risk ratio scale that an unmeasured confounder would need to have with the treatment and the outcome to fully explain away an observed treatment-outcome risk ratio of  $RR=3.9$  as in the Victora et al. study. The E-value essentially sets the two parameters,  $RR_{UD}$  and  $RR_{EU}$  equal to each other to see the minimum that they would both need to be. The E-value for the Victora estimate corresponds to the point (7.26,7.26).

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

Grant Support: The work was supported by NIH grant ES017876.

Mailing Addresses:

Tyler VanderWeele

Harvard T.H. Chan School of Public Health

677 Huntington Avenue

Boston MA 02115

Peng Ding

University of California Berkeley

425 Evans Hall

Berkeley, CA 94720-3860

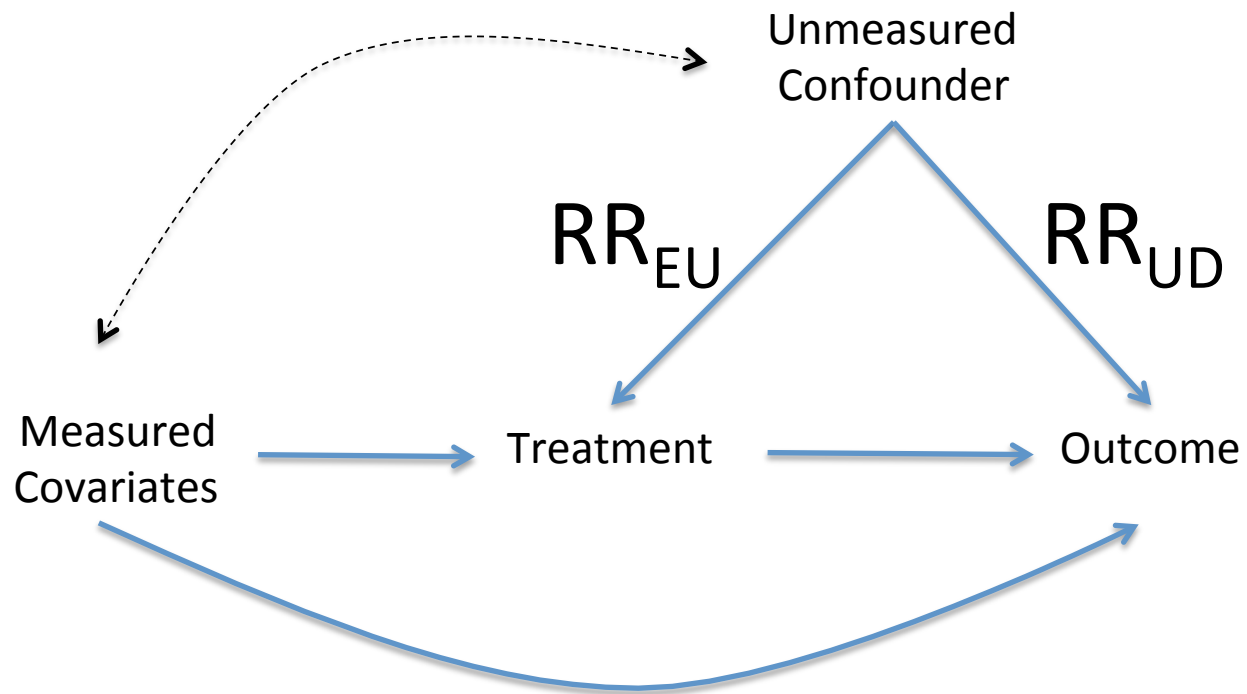
Address for Correspondence:

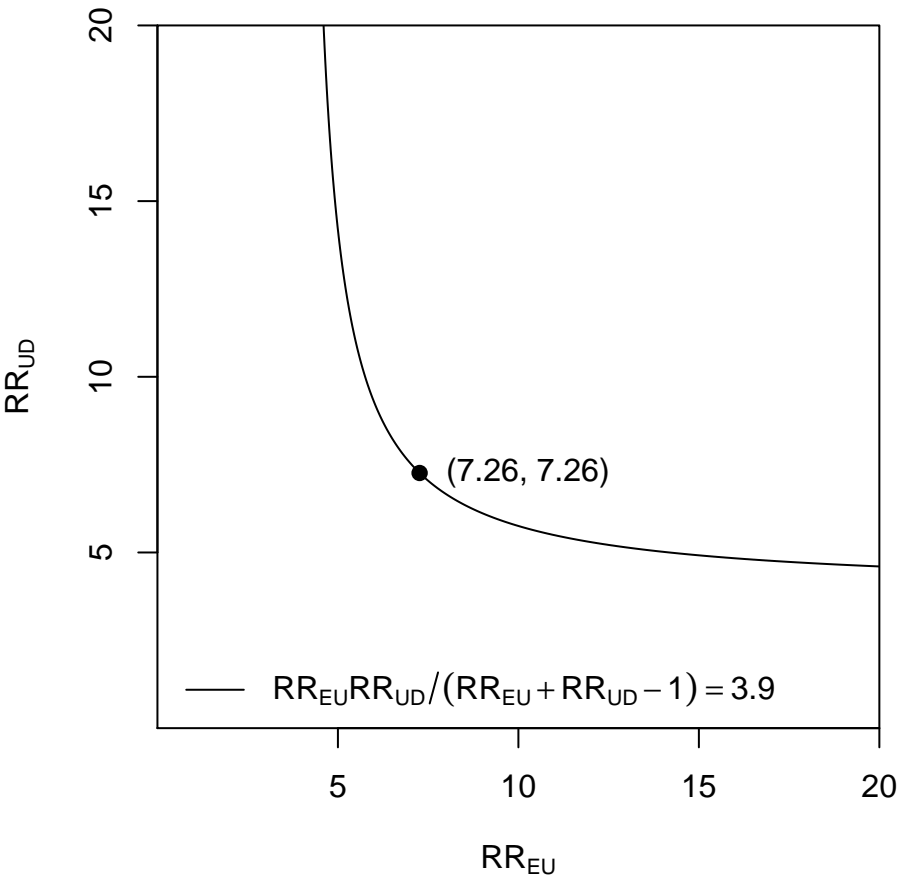
Tyler VanderWeele

Harvard T.H. Chan School of Public Health

677 Huntington Avenue

Boston MA 02115





## Key Points

<i>Motivation</i>	Observational studies that attempt to assess causality between a treatment and an outcome may be subject to unmeasured confounding.
<i>Rationale</i>	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, to present, and to interpret, and does not itself make strong assumptions, is desirable.
<i>Definition of E-value</i>	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 the outcome, conditional on the measured covariates, to fully explain away a specific treatment-outcome association.
<i>Calculation</i>	The E-value for an estimate, and for the limit of a 95% confidence interval closest to the null, can be calculated in a straightforward way for risk ratios as in Table 1, and for other measures as in Table 2.
<i>Conclusions</i>	The E-value allows an investigator to make statements of the form, “With an observed risk ratio of $RR=3.9$ , an unmeasured confounder that was associated with both the outcome and the treatment by a risk ratio of 7.2-fold each, above and beyond the measured confounders, could explain away the estimate, but weaker confounding could not; to move the confidence interval to include the null, an unmeasured confounder that was associated with the outcome and the treatment by a risk ratio of 3.0-fold each could do so, but weaker confounding could not.”

**Table 1. Calculating the E-Value for Risk Ratios**

Direction of Risk Ratio	Estimate or Confidence Interval	Computation of the E-Value
RR>1	Estimate	$E\text{-Value} = RR + \sqrt{RR \times (RR - 1)}$
	Confidence Interval	Let LL denote the lower limit of the confidence interval If $LL \leq 1$ then $E\text{-Value} = 1$ If $LL > 1$ then $E\text{-Value} = LL + \sqrt{LL \times (LL - 1)}$
RR<1	Estimate	Let $RR^* = 1/RR$ $E\text{-Value} = RR^* + \sqrt{RR^* \times (RR^* - 1)}$
	Confidence Interval	Let UL denote the upper limit of the confidence interval If $UL \geq 1$ then $E\text{-Value} = 1$ If $UL < 1$ then let $UL^* = 1/UL$ and $E\text{-Value} = UL^* + \sqrt{UL^* \times (UL^* - 1)}$

**Table 2. E-Values for Other Effect Measures**

Effect Measure	
<i>Odds Ratio or Hazard Ratios for Rare Outcomes</i>	When the outcome is relatively rare (less than 15% say) by the end of follow-up, the E-value formula in Table 1 can be used (Ref 37).
<i>Rate Ratio for Count and Continuous Outcomes</i>	For ratio measures for count outcomes (or non-negative continuous outcomes), the E-value can be found by replacing the risk ratio with the rate ratio (or the ratio of expected values) in the E-value formula (Ref 37).
<i>Odds Ratio for Common Outcomes</i>	When the outcome is common (more than 15%), an approximate E-value can be obtained by replacing the risk ratio with the square root of the odds ratio (Ref 45), i.e. $RR \approx \sqrt{OR}$ , in the E-value formula in Table 1.
<i>Hazard Ratio for Common Outcomes</i>	When the outcome is common (more than 15%), an approximate E-value can be obtained (Ref 45) by applying the approximation $RR \approx (1 - 0.5^{\sqrt{HR}}) / (1 - 0.5^{\sqrt{1/HR}})$ in the E-value formula in Table 1.
<i>Difference in Continuous Outcomes</i>	With standardized effect sizes “d” (mean of the outcome variable divided by the standard deviation of the outcome) and a standard error for this effect size $s_d$ , an approximate E-value can be obtained (Ref 45-47) by applying the approximation $RR \approx \exp(0.91 \times d)$ in the E-value formula. An approximate confidence interval can be found 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 (Ref 27-29), but generally require more assumptions and the parameters do not necessarily have a corresponding E-value.
<i>Risk Difference</i>	If the adjusted risks for the treated and untreated are $p_1$ and $p_0$ , then the E-value can be obtained by replacing the risk ratio with $p_1/p_0$ in the E-value formula. The E-value for the confidence interval on a risk difference scale is more complex and software to do this is described in the Appendix. 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 above could be employed. Other sensitivity analysis techniques have been developed for this setting (Ref 27-29), but generally require more assumptions and do not provide a corresponding E-value.



**Table 3. Issues of Interpretation of the E-Value**

Issue	Interpretation
<i>Likely Effect Sizes</i>	The E-value should be interpreted in the context of the effect sizes an unmeasured confounder is likely to have with respect to the outcome and the treatment. In the context of biomedical and social sciences research, effect sizes of 2-fold or 3-fold or more do occasionally occur, but they are also not particularly common; a variable that affects both treatment and outcome each by 2-fold or 3-fold would likely be less common still. For purposes of comparison in any context, it may be helpful to calculate the analogous E-value for each of the measured covariates if they had been omitted.
<i>E-Values and Sensitivity Analysis</i>	The E-value for the respiratory death example was 7.2. In the formula for the bias factor B, it is possible that a confounder that was associated with the respiratory death by less than 7.2-fold could explain away the effect estimate but it would have to be associated with the treatment by a risk ratio more than 7.2-fold. Values of the sensitivity analysis parameters with a less extreme confounder-outcome association will require a more extreme treatment-confounder association, and vice versa.
<i>Sample Size, E-Values and P-Values</i>	A large study with a precisely estimated association will often have a very small p-value; the p-value can be made arbitrarily small by increasing the sample size. But 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 by simply increasing the sample size. The E-value for the confidence interval does depend on the sample size, but as the sample size gets larger the E-value for the confidence interval does not get arbitrary large; it is bounded by the strength of the association (the limit is sometimes referred to in other contexts as the “design sensitivity”, Ref 17-18). A large sample size may give a small p-value; a large effect size will give a large E-value.

## Appendix.

### Worked Example for an Odds Ratio with a Common Outcome

Oddy et al. (48) examined the relationship between cessation of breast-feeding before 6 months and the child later being overweight or obese at age 20. In the study being overweight or obese was defined as having body mass index above 25. The investigators assessed this association using logistic regression to adjust for the measured covariates of maternal age at conception, maternal BMI before pregnancy, maternal education before birth, family income at birth, and gender of the child. They did not adjust for smoking, which might be an unmeasured confounder of the relationship. Using logistic regression, they obtain an estimated odds ratio of  $OR=1.47$  (95% CI: 1.12, 1.93,  $p=0.005$ ). Unlike the outcomes considered in the text, here the outcome of being overweight or obese (i.e. body mass index above 25) is relatively common. Odds ratios will thus not approximate risk ratios but we can use the approximate approach with odds ratios given in Table 2.

We obtain an approximate risk ratio by taking the square root of the odds ratio (45)  
 $RR \approx \sqrt{OR} = \sqrt{1.47} = 1.21$  and we also take the square root of both limits of the 95% confidence interval to obtain an approximate 95% confidence interval for the risk ratio of (1.06, 1.39). If we apply the E-value formula in Table 1 to the risk ratio 1.21 we obtain

$$E\text{-value} = RR + \sqrt{RR \times (RR - 1)} = 1.21 + \sqrt{1.21 \times (1.21 - 1)} = 1.71$$

and if we apply the E-value formula in Table 1 to the lower limit of the confidence interval, 1.06, we obtain

$$E\text{-value} = RR + \sqrt{RR \times (RR - 1)} = 1.06 + \sqrt{1.06 \times (1.06 - 1)} = 1.31.$$

The approximate E-value for the estimate is thus 1.71 and the approximate E-value for the confidence interval is 1.31. Relatively modest unmeasured confounding could explain away the effect.

The E-value for an odds ratio for a common outcomes does, however, rely on an approximation. One obtains an approximate risk ratio from the odds ratio (45) using a square root conversion  $RR \approx \sqrt{OR}$ . This approximation works fairly well if the outcome probabilities are between 0.1 and 0.9, and it works very well if the outcomes probabilities are between 0.2 and 0.8. It is thus a reasonable approximation if the outcome is common (45).

### **Worked Example for a Difference in a Continuous Outcomes**

Reinisch et al. (49) examined the effect of in utero exposure to phenobarbital on intelligence in men using data from Copenhagen. The exposure group consisted of those who had been exposed in utero to phenobarbital and the control group of those who had not. Intelligence was measured by the Danish Military Board Intelligence Test taken by the participants in their early twenties. Adjustment was made for family socioeconomic status, breadwinner's education, sibling

position, whether the pregnancy was wanted, whether the mother attempted an abortion, maternal marital status, predisposing risk score indicating conditions were less than optimal for conception, mother's age, father's age, gestational length, birth weight, birth length, number of cigarettes per day in the third trimester, maternal weight gain divided by height cubed, and the maternal complaint score. The investigators did not have data on parental intelligence which they noted might be an unmeasured confounder. After adjustment the estimate of the effect of phenobarbital on intelligence was -4.77 (95% CI: -7.96,-1.58) points, with a standard error 1.63. The estimate and the standard error are for the scores themselves. A standard deviation for the test scores was 11.38. Dividing the estimate and the standard error by this standard deviation gives standardized effect sizes of  $d=-4.77/11.38=-0.42$  and  $s_d=1.63/11.38=0.14$ .

We then use the conversion in Table 2 to get an approximate  $RR \approx \exp(0.91 \times d) = 1.46$  with approximate 95% confidence interval  $(\exp\{0.91 \times d - 1.78 \times s_d\}, \exp\{0.91 \times d + 1.78 \times s_d\}) = (1.14, 1.88)$ . If we apply the E-value formula in Table 1 to the risk ratio 1.46 we obtain

$$E\text{-value} = RR + \sqrt{RR \times (RR - 1)} = 1.46 + \sqrt{1.46 \times (1.46 - 1)} = 2.28$$

and if we apply the E-value formula in Table 1 to the lower limit of the confidence interval, 1.14, we obtain

$$E\text{-value} = RR + \sqrt{RR \times (RR - 1)} = 1.14 + \sqrt{1.14 \times (1.14 - 1)} = 1.54.$$

The approximate E-value for the estimate is thus 2.28 and the approximate E-value for the confidence interval is 1.54.

The E-value for a difference in continuous outcomes does, however, rely on two approximations. First an approximate conversion between standardized effect sizes and odds ratios is often used in meta-analysis which assumes that a binary variable is in fact based on an underlying continuous variable with a particular cut-off. One can obtain an approximate OR from a standardized effect size (46,47) by  $OR \approx \exp\{1.81 \times d\}$  and an approximate confidence interval by  $[\exp\{1.81 \times d - 3.55 \times s_d\}, \exp\{1.81 \times d + 3.55 \times s_d\}]$ . The approximation works fairly well if the data follow a normal or logistic model and if the binary outcome does not have very low or very high probability. Second, one then can obtain an approximate risk ratio from the odds ratio (45) using a square root conversion  $RR \approx \sqrt{OR}$ . The application of both of these transformations gives rise to the approximation in Table 2 of  $RR \approx \exp(0.91 \times d)$  with approximate confidence interval  $(\exp\{0.91 \times d - 1.78 \times s_d\}, \exp\{0.91 \times d + 1.78 \times s_d\})$ .

### **E-Values for Non-Null Hypotheses**

In the text we discussed how to calculate E-values so as to assess the minimum strength of the association an unmeasured confounder would need to have with both the treatment and outcome to move the point estimate, or one limit of the confidence interval, to the null. However, a similar procedure can be used to assess the minimum magnitude of both confounder associations that would be needed to move an estimate to some other value of the risk ratio. If we have an observed risk ratio of RR and want to assess the minimum strength of both associations that

would be needed to shift the estimate to some other value  $RR^T$  then we first take the ratio of the two values,  $RR/RR^T$ , and then apply the E-value formulas in Table 1 to this quantity.

For example, the estimate in the AHRQ report (42) for the effect of breastfeeding on childhood leukemia was 0.80 (95%CI:0.71,0.91). We calculated the E-value as 1.8 for the estimate and the 1.4 for the confidence interval. We might refer to these as the “null E-values” i.e. 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, conditional on the measured covariates, to shift the estimate (or confidence interval) to the null. Suppose we wanted to see how large both unmeasured confounding associations would need to be to shift the estimate from  $RR=0.80$  to a risk ratio of  $RR^T=0.90$ . We would first take the ratio  $RR/RR^T=0.80/0.90=0.89$  and then we would apply the procedure in Table 1, to obtain an E-value of 1.5, which is the magnitude of the associations an unmeasured confounder would need to have with breastfeeding and childhood leukemia to move the observed risk ratio of  $RR=0.80$  to an adjusted risk ratio of  $RR^T=0.90$ . We might refer to this as a “non-null E-value”, or an E-value to shift to an adjusted risk ratio of  $RR^T=0.90$ . The interpretation of this non-null E-value is that for an unmeasured confounder to shift the observed estimate of  $RR=0.80$  to an estimate of  $RR^T=0.90$ , an unmeasured confounder that was associated with both breastfeeding and childhood leukemia by a risk ratio of 1.5-fold each could do so, but weaker confounding could not. We can also calculate E-values for the values of the risk ratio on the other side of the null hypothesis. Thus if we wanted to assess the minimum strength of both confounder associations that would be needed to move the risk ratio estimate of  $RR=0.80$  to a risk ratio estimate of  $RR^T=1.20$ , we could again take the ratio  $RR/RR^T=0.80/1.20=0.67$  and then apply the formula in Table 1 to 0.67 to obtain the non-null E-value 2.4

to shift the adjusted risk ratio to  $RR^T=1.20$ ; we could apply a similar procedure for the upper limit of the confidence interval so that we use the formulas in Table 1 with the value  $0.91/1.20=0.76$  to obtain 1.95 as the non-null E-value to shift to an adjusted risk ratio of 1.2 for the upper limit of the confidence interval. The interpretation of these non-null E-values would then be that for an unmeasured confounder to shift the observed risk ratio estimate of  $RR=0.80$  to a risk ratio of 1.2, an unmeasured confounder that was associated with both breastfeeding and childhood leukemia by a risk ratio of 2.4-fold each could do so, but weaker confounder could not; to shift the upper confidence interval of 0.91 to 1.2, an unmeasured confounder that was associated with both breastfeeding and childhood leukemia by a risk ratio of 1.95-fold each could do so, but weaker confounding could not.

Examining non-null E-values can also be helpful in examining the confounder associations that would be needed to move an estimate that is close to the null to an estimate that might be clinically meaningful. For example, Huybrechts et al. (50) report an adjusted risk ratio estimate for the use of anti-depressants during pregnancy on infant cardiac defects of 1.06 (95%CI:0.93,1.22). The confidence interval includes the null and so the null-E-value described in the text for the confidence interval is simply 1 – no further unmeasured confounding is needed to shift the confidence interval to include 1. However, we could also examine the confounder associations that would be needed to shift the estimate to 1.20. As before, we take the ratio  $1.06/1.20=0.88$ , and we apply the formula in Table 1 to this value to obtain a non-null E-value of 1.5 to shift the risk ratio to 1.2. Thus, an unmeasured confounder that was associated with both anti-depressant use during pregnancy and infant cardiac defects by a risk ratio of 1.5-fold each could move the risk ratio to 1.20, but weaker confounding could not. We could also consider the

confounder associations that would be needed to shift the confidence interval so as not to include the null e.g. to bring the lower confidence limit of 0.93 to a lower limit of 1.01. We would again take the ratio  $0.93/1.01=0.92$  and we could again apply the formulas in Table 1 to this value to obtain an E-value of 1.4. Thus, an unmeasured confounder that was associated with both anti-depressant use during pregnancy and infant cardiac defects by a risk ratio of 1.4-fold each could shift the confidence interval to exclude the null, but weaker confounding could not. As noted in the text, the use of E-values cannot be used to definitely establish a null association but the use of E-values can still be helpful in assessing the confounding associations that would be needed to move a null association to something that might be clinically meaningful.

### **E-Values for Null and Non-Null Risk Differences**

It is also possible to calculate the exact E-value for a risk difference for a binary outcome. Unfortunately, the E-value for the confidence interval of a risk difference cannot be easily obtained by hand calculation. We have, however, provided an R function that will calculate the E-value for a risk difference and its confidence interval. The function and a read-me file are downloadable from the journal website. We have provided similar code in SAS elsewhere (Ref 37, Section 10 of the eAppendix). To use the function the user must input the total number of subjects in the study,  $N$ , the proportion of the population with the exposure or treatment,  $f$ , the adjusted risk for the exposed or treated,  $p_1$ , and its estimated standard error,  $se.p_1$ , the adjusted risk for the unexposed or untreated,  $p_0$ , and its estimated standard error,  $se.p_0$ , and as well the null or non-null value of the risk difference,  $RD.true$ , for which one wants to calculate E-values. If the E-value for the null hypothesis of no treatment effect is desired then the user simply sets



RD.true=0; otherwise if the E-value for a non-null risk difference is desired this can be specified.

After running the R code, the user is required just to put these values in one function:

```
evaluesRD(N, f, p1, p0, se.p1, se.p0, RD.true)
```

The function will then output the E-value corresponding to the minimum amount of confounding on the risk ratio scale that could move the risk difference estimate or the confidence interval to the value of RD.true that was specified. While the estimate and confidence interval themselves are given on the risk difference scale, the E-values still corresponding to the strength of the confounding relationships between exposure and the unmeasured confounder and between the unmeasured confounder and the outcome are on the risk ratio scale.

As an example, Hammond and Horn (51) report associations between smoking and lung cancer deaths from a cohort study of 187,783 men, of which 42% were classified as having a history of regular cigarette smoking (exposed) versus others (no smoking or only occasional smoking). The estimated risk among the regular smokers was 0.00503 with standard error 0.000252. The estimated risk among those who were not regular smokers was 0.000469 with standard error 0.0000656. The risk difference is 0.00456 (95% CI: 0.00405, 0.00507). To calculate the E-value to reduce the estimate or confidence interval to the null of no risk difference we input:

N=187,783, f=0.42, p1=0.00503, p0=0.000469, se.p1=0.000252, se.p0=0.0000656, RD.true=0.

If we put these into the R function we have:

```
evaluesRD(187783, 0.42, 0.00503, 0.000469, 0.000252, 0.0000656, RD.true=0)
```

and we would obtain output from the R function of the E-values of 20.9 for the estimate and 15.9 for the confidence interval. We could thus say, “With an observed risk difference of  $RD=0.00456$ , an unmeasured confounder that was associated with both regular smoking and lung cancer death by a risk ratio of 20.9-fold each, above and beyond the measured confounders, could explain away the estimate, but weaker confounding could not; to move the confidence interval to include the null, an unmeasured confounder that was associated with both regular smoking and lung cancer death by a risk ratio of 15.9-fold each could do so, but weaker confounding could not.” These would be the E-values for the null of no effect.

We can also calculate non-null E-values. For example, if we were to see the strength of the unmeasured confounding associations that could suffice to move the observed risk difference of  $RD=0.00456$  to the value 0.001 we could input into the R-function:

```
evaluesRD(187783, 0.42, 0.00503, 0.000469, 0.000252, 0.0000656, RD.true=0.001)
```

and we would obtain output from the R function of the non-null E-values of 12.5 for the estimate and 9.5 for the confidence interval. We could then say, “For an unmeasured confounder to shift the observed risk difference estimate of  $RD=0.00456$  to a risk difference of  $RD=0.001$ , an unmeasured confounder that was associated with both regular smoking and lung cancer death by a risk ratio of 12.5-fold each could do so, but weaker confounder could not; to shift the lower confidence limit of  $RD=0.00398$  to  $RD=0.001$ , an unmeasured confounder that was associated

with both regular smoking and lung cancer death by a risk ratio of 9.5-fold each could do so, but weaker confounding could not.”

### **E-Values and the Bias Formula**

As in the text, if  $RR_{UD}$  is the maximum risk ratio for the outcome comparing any two categories of the unmeasured confounder(s) and if  $RR_{EU}$  is the maximum risk ratio for any specific level of the unmeasured confounder(s) comparing those with and without treatment, then the maximum relative amount such unmeasured confounding could reduce an observed risk ratio is given by:

$$B = \frac{RR_{UD} \cdot RR_{EU}}{RR_{UD} + RR_{EU} - 1}$$

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 the outcome to explain away a treatment-outcome association. We can obtain the E-value formula by setting the two parameters equal to each other  $RR_{UD} = RR_{EU}$  (which we then call the value E) and then examining how large these would have to be so that the bias factor B is equal to the observed risk ratio RR. Using the bias factor formula we thus have:

$$RR = \frac{E \cdot E}{E + E - 1}$$

$$RR = \frac{E^2}{2E - 1}$$

$$RR \cdot (2E - 1) = E^2$$

$$0 = E^2 - 2RR \cdot E + RR$$

The positive solution to this quadratic equation is  $E = RR + \sqrt{RR \times (RR - 1)}$  which gives the formula for the E-value.