Tyler J. VanderWeele*, Peng Ding, and Maya Mathur

Technical Considerations in the Use of the E-Value

Abstract: The E-value is defined as the minimum strength of association on the risk ratio scale that an unmeasured confounder would have to have with both the exposure and the outcome, conditional on the measured covariates, to explain away the observed exposure-outcome association. We have elsewhere proposed that the reporting of E-values for estimates and for the limit of the confidence interval closest to the null become routine whenever causal effects are of interest. A number of questions have arisen about the use of E-value including questions concerning the interpretation of the relevant confounding association parameters, the nature of the transformation from the risk ratio scale to the E-value scale, inference for and using E-values, and the relation to Rosenbaum's notion of design sensitivity. Here we bring these various questions together and provide responses that we hope will assist in the interpretation of E-values and will further encourage their use.

Keywords: Bias Analysis, Causal Inference, Covariate Adjustment, Design Sensitivity, Sensitivity Analysis, Treatment Effects, Unmeasured Confounding

1 Introduction

In 2017, we introduced the E-value metric to help assess sensitivity of results to potential unmeasured confounding [1]. The E-value was defined as the minimum strength of association on the risk ratio scale that an unmeasured confounder would have to have with both the exposure and the outcome, conditional on the measured covariates, to explain away the observed exposure-outcome association [1]. Formulas for computing E-values or approximate E-values in a variety of settings were provided. Software and also an online calculator for E-values have since been provided [2]. Since its introduction a number of, often more technical, questions have been posed concerning the use and interpretation of E-values. The purpose of this paper is to document and address some of the more common questions that have arisen.

2 Calculation of E-values and interpretation of the parameters

The formal derivation of the E-value relies on two parameters [3]. Let E denote an exposure of interest, D the outcome, C the measured covariates, and U one or more unmeasured confounders. The observed exposure-outcome association on the risk ratio scale, conditional on covariates C, is given by

$$RR_{obs} = \frac{P(D = 1|E = 1, c)}{P(D = 1|E = 0, c)}$$

*Corresponding author: Tyler J. VanderWeele, Harvard University, Departments of Epidemiology and Biostatistics, Boston, Massachusetts, United States, e-mail: tvanderw@hsph.harvard.edu
Peng Ding, University of California Berkeley, Department of Statistics, 425 Evans Hall, Berkeley, California, United States, e-mail: pengdingpku@berkeley.edu
Maya Mathur, Harvard University, Department of Biostatistics, Boston, Massachusetts, United States, e-mail: maya.z.mathur@gmail.com

The association, conditional on C, but adjusted also for U would be:

$$RR_{true} = \frac{\sum_u P(D = 1|E = 1, c, u)P(u|c)}{\sum_u P(D = 1|E = 0, c, u)P(u|c)}$$

If covariates (C,U) suffice to control for confounding of the effect of E on D, then the latter expression RR_{true} can be interpreted as the causal risk ratio of E on D conditional on C. More formally, let D_e denote the counterfactual outcome if E is set to e, and let $X \parallel Y \mid Z$ be used to denote that X is independent of Y given Z. The effect of E on D is said to be unconfounded conditional on C if $D_e \parallel E \mid C$ for all e. We have that if the effect of E on D is unconfounded given (C,U) then

$$\frac{P(D_1|c)}{P(D_0|c)} = \frac{\sum_u P(D = 1|E = 1, c, u)P(u|c)}{\sum_u P(D = 1|E = 0, c, u)P(u|c)}$$

and hence we denote the expression $\frac{\sum_u P(D=1|E=1,c,u)P(u|c)}{\sum_u P(D=1|E=0,c,u)P(u|c)}$ as RR_{true}.

Consider now the following two sensitivity analysis parameters [1, 3]:

$$RR_{UD} = max \left\{ \frac{max_u P(D = 1|E = 1, c, u)}{min_u P(D = 1|E = 1, c, u)}, \frac{max_u P(D = 1|E = 0, c, u)}{min_u P(D = 1|E = 0, c, u)} \right\}$$

$$RR_{EU} = max_u \frac{P(U = u|E = 1, c)}{P(U = u|E = 0, c)}$$

Essentially, RR_{UD} is the maximum effect that U can have on D, conditional on C = c, comparing any two categories of U, for either the exposed or unexposed; and RR_{EU} is the maximum risk ratio relating the exposure to any particular level of U, conditional on C = c. We showed that [3]:

$$\frac{RR_{obs}}{RR_{true}} \leq \frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1}$$

so that $\frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1}$ was the maximum bias (comparing the ratio of the observed association adjusted for C, to the true association adjusted also for U) that could be generated by such an unmeasured confounder. We then further derived that for the unmeasured confounder(s) to shift the observed risk ratio to the null of 1, if one wanted both RR_{UD} and RR_{EU} to be as small as possible, then the minimum they could both be (which was what we called the E-value) was [1, 3]:

$$E\text{-}value := min_{RR_{UD}, RR_{EU}: \frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1} \geq RR_{obs}} max(RR_{UD}, RR_{EU})$$

$$= RR_{obs} + \sqrt{RR_{obs}(RR_{obs} - 1)}$$

The E-value is thus straightforward to calculate from the observed risk ratio. We had noted previously that what constitutes a large E-value is context dependent. It is relative to the outcome, to the exposure, and to the measured covariates for which adjustment has been made. For example, an E-value of 2 when all-cause mortality is the outcome may provide more evidence for robustness to confounding than would an E-value of 2 when the outcome is suicide, because risk ratios of 2 are much less common for all-cause mortality in empirical analyses than they are for suicide, for which there are a variety of known risk factors with risk ratios of 5- or even 10-fold. The actual evidence for causality is dependent on the context including the nature of the exposure, the outcome, the measured covariates, and other potential sources of bias. However, within a given context, the E-value, properly interpreted, can help assess the robustness of an estimate to unmeasured confounding, and this consideration is relevant in assessing the overall evidence for causality.

Questions have been raised with regard to the interpretation of the E-value for a continuous exposure. In that context, depending on the magnitude of the exposure change examined, the magnitude of the corresponding risk ratio will differ and thus the E-value will differ as well. One will often be able to make the E-value larger simply by specifying a larger change in the two exposure levels being compared. However, that the E-value may get larger for a larger change in the exposure levels being compared makes sense, both

because it is more plausible that a large exposure change has a causal effect (a difference in body weights comparing 300 vs. 80 pounds is more likely to have a causal effect on various outcomes than a difference in body weights of 170 vs. 169 pounds), but also because it is more likely, for a larger exposure change, that the two exposure groups differ more on the unmeasured confounder(s) U, so a larger E-value is needed to indicate genuine evidence of robustness.

Questions have come up concerning the consequences of having a potentially continuous or many-valued unmeasured confounder U. In such cases, because many pairwise comparisons of categories of U are possible, it may be more plausible than it is with a binary U that the maxima of these numerous pairwise comparisons produce RR_{UD} and RR_{EU} exceeding the E-value. Hence, it might be the case that a large E-value still in fact does not contribute all that much evidence for a causal effect. This is a reasonable concern. Several interpretative points here, however, are important.

First, the confounding associations RR_{UD} and RR_{EU} are both conditional on the measured covariates C so that the confounding associations RR_{UD} and RR_{EU} reflect residual confounding not captured by the measured covariates C. It is the association between U and both D and E, *independent of C*, that is relevant here. We have, in our paper, referred to these conditional associations as the unmeasured confounding "above and beyond the measured confounders." [1] In many cases, control for pre-exposure covariates C will reduce the amount of bias due to confounding. For example, if income is an unmeasured variable, but control has been made in the covariates C for education, occupation, and home-ownership, then income may itself, conditional on these other socio-economic markers, not generate all that much bias. There is an entire graphical calculus on when covariate conditioning suffices to eliminate bias and when conditioning on a covariate can introduce bias that would have been otherwise absent [4]. Within that graphical models literature, two now-classic examples of when conditioning on a pre-exposure covariate can introduce additional bias include the conditioning on a pre-exposure variable that is a "collider", a common effect of two variables, one of which is associated with the exposure and the other of which is associated with the outcome [5–7]. Another example when conditioning on a baseline covariate can increase bias is when there is an unmeasured common cause U of E and D, then conditioning on a covariate C that is a cause of only the exposure but not the outcome except through the exposure, i. e. for an instrument of the effect of the exposure on the outcome, can likewise increase bias, though researchers will often not be certain if a particular covariate is in fact an instrument [8–10]. While conditioning on such a covariate C, i. e. an instrument, may not increase the sensitivity parameter, RR_{UD}, it is the case that conditioning on an instrument can increase the sensitivity parameter RR_{EU}. One must thus be careful with regard to believing that controlling for measured covariates always necessarily reduces confounding.

Second, the inequality holds for any U and thus the results are relevant for any set of covariates U such that the effect of E on D is unconfounded conditional on (C,U). One could thus define the parameters RR_{UD} and RR_{EU} for each possible U such that (C,U) suffice to control for confounding and then take the minimum over U of the resulting bias $\frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1}$. One would then have

$$\frac{RR_{obs}}{RR_{true}} \leq min_{U:D_e \parallel E|(C,U)} \left\{ \frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1} \right\}$$

The E-value calculated as $RR_{obs} + \sqrt{RR_{obs}(RR_{obs} - 1)}$ is then the minimum strength of association on the risk ratio scale that *any and every* unmeasured confounder, that suffices along with C to control for confounding, would have to have with both the exposure and the outcome, conditional on the measured covariates, to explain away the observed exposure-outcome association.

Third, and perhaps most importantly when combined with the second observation above, the reality of our estimates and attempts at confounding control are at best approximate. Often we would be content, and indeed very pleased, if our estimates were only a few percent away from the truth. Let S denote the set of all possible covariates U such that adjustment for (C,U) would bring the observed association between E and D, conditional on C and adjusted for U, within a factor of say 1.03 (i. e. 3 %) of the actual causal effect i. e.

$$S = \left\{ U: \frac{1}{1.03} \leq \frac{\frac{\sum_u P(D=1|E=1,c,u)P(u|c)}{\sum_u P(D=1|E=0,c,u)P(u|c)}}{\frac{P(D_1|c)}{P(D_0|c)}} \leq 1.03 \right\}$$

One could then define the parameters RR_{UD} and RR_{EU} for each possible U in the set S such that (C,U) would suffice to bring the bias within 3 % of the causal effect. One could then further take the minimum over U of the resulting bias $\frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1}$. One would then have

$$\frac{RR_{obs}}{P(D_1|c)/P(D_0|c)} \le (1.03) \times min_{U \in S} \left\{ \frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1} \right\}$$

Once we allow for up to 3 % bias say, a considerably coarsened version of the relevant unmeasured confounders may well suffice. For such coarsened versions of the unmeasured confounders, the relevant confounding association parameters RR_{UD} and RR_{EU} may be considerably smaller for the coarsened unmeasured confounder than for the original underlying unmeasured confounder. Once we further take the minimum over all possible unmeasured confounders and all possible coarsenings that would result in at most 3 % bias, the confounding association parameters RR_{UD} and RR_{EU} will be yet smaller still.

The E-value calculated as $RR_{obs} + \sqrt{RR_{obs}(RR_{obs} - 1)}$ could then be interpreted as the minimum strength of association on the risk ratio scale that *any and every* unmeasured confounder or coarsening thereof, that suffices along with the observed covariates to bring the observed association within 3 % of the true causal effect, would have to have with both the exposure and the outcome, conditional on the measured covariates, to explain away the observed exposure-outcome association. Thus, even if an unmeasured confounder that completely eliminated bias had very large confounding association parameters RR_{UD} and RR_{EU}, the E-value may arguably still be a useful metric for robustness to unmeasured confounding as it can be applied, in an approximate sense, as above, to coarsenings and approximate confounding control as well. We give a worked example of this in the Appendix A. Cochran [11] also noted that in many settings a coarsening to five or six strata is often sufficient to remove at least 90 % of the bias due to that covariate. Informally then, if one is content with being within 10 % of the unconfounded estimate, a consideration of the sensitivity analysis parameters when comparing e. g. the top to the bottom quintile of the unmeasured confounder may be a reasonable way to think about the magnitude of the confounding parameters for a continuous unmeasured covariate.

A somewhat related issue that pertains to the definition of the confounding parameters concerns the possibility of multiple unmeasured confounders being needed to eliminate confounding. The bias analysis and E-value calculations above are in fact applicable to the setting of multiple unmeasured confounders [3]. The confounding parameters RR_{UD} is then simply interpreted as the maximum effect that U can have on D, conditional on C = c, comparing any two categories of the entire vector of unmeasured confounders U, for either the exposed or unexposed; and RR_{EU}, is the maximum risk ratio relating the exposure to any particular level of the entire vector U, conditional on C = c. In such settings large values of RR_{UD} and RR_{EU} may not be particularly implausible. While an E-value of 5, say, for all-cause mortality as the outcome, may seem, when considering a single confounder, to require very substantial confounding associations and it is perhaps unlikely a single unmeasured confounder could increase the probability of the outcome by 5-fold conditional on the measured covariates, an increase of that magnitude may not be quite as implausible if one is considering a whole group of potential unmeasured confounders. The effect comparing the most favorable values of a set of confounders U to the least favorable values of that set U might plausibly increase the probability of the outcome by 5-fold, perhaps even conditional on the measured covariates. For example, if the unmeasured confounders were age, income, baseline health, and country, then a risk ratio of 5 for all-cause mortality might be quite plausible comparing someone young, rich, in excellent health, and in a country with good safety and medical care, versus someone who is old, poor, exceedingly frail, and in a country with poor medical care and in which civil war has begun. However if there are in fact multiple important unmeasured confounders, one should perhaps question whether the data available are in fact adequate to get a reasonable estimate of the causal effect at all. If it is known in advance that there are not just one, but numerous known unmeasured confounders, strongly associated with the outcome and exposure and independent of the measured covariates, then arguably this is not a good study setting in which to attempt to draw conclusions. If it is thought plausible that a 5-fold increase in the probability of the outcome could be generated by the unmeasured confounders conditional on the measured covariates, then it is perhaps time to leave that study data alone and pursue other more

adequate data sources. A large E-value can only contribute strong evidence for a true causal effect if the set of measured covariates adjusted for plausibly controls for much of the confounding. Said another way, the design of the study, and the collection of data on measured and known confounders, is essential in whether an estimate is plausible or not.

Lastly, it is to be remembered that the E-value is conservative insofar as, if the parameters RR_{UD} and RR_{EU} are in fact as large as the E-value, then it is possible to construct scenarios in which an unmeasured confounder U with those parameters would suffice to bring the observed association down to the null [3]. However, there are also many other scenarios in which an unmeasured confounder has confounding parameters RR_{UD} and RR_{EU} that are equal to the E-value and yet the unmeasured confounder would not suffice to reduce the observed association to the null. The inequality for the maximum bias $\frac{RR_{obs}}{RR_{true}} \leq \frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1}$ is an inequality, not an equality. The inequality is sharp in that it is always possible to construct a variable U with those confounding associations that attains the bound, but, with an actual unmeasured confounder, the bias will often be less. This is especially the case when, for example, the unmeasured confounder is rare in both exposure groups [1, 3]. The E-value essentially assumes that the distribution of U is as unfavorable as possible. Indeed when it is known in advance that the unmeasured confounder is rare, this is one scenario in which the E-value calculation is perhaps of less use, and is perhaps to be avoided, as it will, in that setting, be exceedingly conservative.

3 The E-value as a transformation of the estimate and confidence interval

In our paper, we recommend reporting the E-value for the estimate and for the limit of the confidence interval closest to the null. The former E-value reports how much unmeasured confounding would be needed to shift the estimate itself (one's best guess given the data) to the null. The latter E-value is perhaps a more adequate measure related to the actual strength of the evidence for an effect, since a large E-value for the limit of the confidence closest to the null suggests that even allowing for uncertainty in the estimation of the observed association, the entire range of plausible values for the estimate are all relatively robust to potential unmeasured confounding. We will return more explicitly to issues of inference for and with the E-value in the following section. However, with regard to our recommended practices of reporting the E-value for the estimate and for the limit of the confidence interval closest to the null, another question that has sometimes arisen concerns the E-value simply being a transformation of the estimate and confidence interval itself and thus not really providing any additional information beyond that estimate and confidence interval.

While it certainly is the case that the E-value for the estimate is just a transformation of the observed risk ratio, and the E-value for the limit of the confidence interval closest to the null is just a transformation of that limit, we still believe the reporting of these metrics is useful for interpretative purposes. The E-value gives the interpretation of the estimate and confidence interval with respect to the minimum strength of confounding associations that would be needed to explain away the estimate. It is a more intuitive assessment after the transformation to the confounding association scale, and one which we believe makes it easier to evaluate the robustness of results to potential unmeasured confounding. Most people cannot simply compute E-values in their head, nor necessarily have a clear sense as to how much confounding would be needed to explain away an estimate of a given magnitude. While the E-value, simply taken as a number, conveys nothing that is not already there in the estimate itself, we think the reporting of the E-value may assist substantially in the actual practice of science, in interpretation, and in the assessment of the robustness of conclusions.

As an analogy, in many settings, the p-value in fact conveys no additional information beyond the estimate and the confidence interval and can be derived from it [12]. While the use of the p-value has been at times controversial, it arguably is still a valuable measure of evidence for an association when properly interpreted as a continuous metric (rather than say as being dichotomized at the 0.05 level). While the p-value, as a number, likewise often does not convey any information that is not already there in the confidence interval, it can still be helpful for the practical purposes of trying to understand the strength of the evidence [13–15].

Most people cannot simply automatically compute a p-value in their head when given the estimate and confidence interval. The scale on which something is reported does make a difference in trying to understand and interpret, and this is the case with the p-value [14, 15]. As another example, instead of reporting risk ratios, we could report the hundredth root of the risk ratios that were obtained so that a risk ratio of 4 was reported as 1.014 and a risk ratio of 1.6 as 1.0047. As numbers, the information conveyed in these two forms of reporting is exactly the same, but the interpretation of the latter is arguably not very intuitive, nor as useful as the former; and again, most people cannot simply do the conversion in their head.

It is similar with the E-value. The proposed E-value calculations, as numbers, do not provide additional information beyond what is already present in the estimate and limit of the confidence interval closest to the null. However, the transformation of these estimates, carried out by the E-value computation, provides the appropriate scale on which to interpret robustness to confounding. Most people again cannot carry out such computations in their head and will thus have more difficulty in interpreting robustness to potential unmeasured confounding when using the untransformed numbers. What is the E-value for a lower limit of the confidence interval which is 1.12? How much confounding would at the minimum be needed to bring such a risk ratio to the null? Again, without going through the computation it is not entirely easy to see or guess. In this case we obtain an E-value of nearly 1.5.

We believe the E-value computations, if routinely carried out are likely to affect interpretative practices with regard to robustness to unmeasured confounding. Consider two hypothetical estimates of a causal effect from two different studies that have adjusted for similar, and all known, confounders: one study obtains an estimate as RR = 1.18 (95 % CI: 1.04, 1.33) and the other as RR = 1.18 (95 % CI: 1.12, 1.24). In our current set of practices, we believe, all other things being equal, the evidence for a causal effect in these two studies would be interpreted in a relatively similar manner. Both obtained similar effect sizes; both had confidence intervals somewhat bounded away from the null so that it seemed unlikely that it was simply a matter of "p-hacking" to get the confidence interval just above 1; the p-value in the latter study is smaller, but both are relatively extreme. Current practices for both studies would probably suggest evidence for association, with the caveat that association is not causation and that there may be unmeasured confounding. However, the types of confounders that would alter inference in these two studies are quite different in strength. The E-value for the confidence interval of the former study is 1.24 and for the latter it is 1.49. While we routinely see risk ratios of 1.24 in the research literature, those of a magnitude of 1.5 are somewhat rarer, and to have a risk ratio of magnitude 1.5 with both the outcome and the exposure, conditional on the measured covariates, rarer still. We believe if the E-values for the lower limit of the confidence intervals for these two studies were reported, along with the estimates and confidence intervals themselves, the robustness to potential unmeasured confounding would be more appropriately evaluated, discussed, and assessed. And this is not simply a matter of also reporting the p-value. We have given elsewhere an example of two studies, one with a more extreme p-value, but the other having the more extreme E-value for the confidence interval [1]. So while our proposed reporting practices for the E-value are indeed just a transformation of the estimate and the limit of the confidence interval closest to the null, we believe this will prove helpful in interpretation and will improve assessments of robustness.

4 Inference for and using E-values

As noted above, we recommend reporting the E-value for the estimate and for the limit of the confidence interval closest to the null (provided the confidence interval excludes the null; otherwise the E-value for the confidence interval is defined as simply 1) [1]. Questions have arisen as to whether it might be good to provide a confidence interval for the E-value itself. Note that our recommendation is to provide an E-value for the limit of the confidence interval closest to the null; it is not to provide a confidence interval for the E-value itself. The distinction is subtle, but important, and concerns the goal of inference. Our perspective is that, in settings in which the E-value may be of use, the goal of inference is the causal effect itself of the exposure on the outcome. The E-value is a tool, not the goal, of inference. The E-value is a tool, a tool to assess the

robustness of one's conclusions to potential unmeasured confounding when trying to draw inferences about causal effects. The goal and object of inference does not concern the E-value itself, but rather the causal effect.

The distinction between the E-value for the confidence interval versus the confidence interval for the E-value becomes clearer when we think about the type of inferential statements one is able to make in repeated sampling. Suppose one calculated a 95 % confidence interval for the E-value for the confounded association. In that case, one could make statements along the lines of "Across repeated samples, at least 95 % of the time, the minimum strength of association on the risk ratio scale that an unmeasured confounder would have to have with both the exposure and the outcome, conditional on the measured covariates, to explain away the actual confounded exposure-outcome association will lie in the confidence interval provided." Such statements may be of some interest, but they are statements concerning, over repeated samples, minimum unmeasured confounding associations, rather than statements directly about the causal effect itself. Suppose instead of calculating a confidence interval for the E-value, one alternatively, as we advocate, calculated the E-value for the limit of the confidence interval closest to the null and did this across samples and settings. One could then make statements along the following lines: "Across repeated samples, at least 95 % of the time it is the case that: if the actual confounding parameters RR_{UY} and RR_{EU} are both less than the E-value for the confidence interval that was calculated, then the association adjusted by the unmeasured confounder(s) will be in the same direction as the observed association."[1] This is a statement more directly about the presence of a true causal effect and for this reason we believe that in most settings it is the type of statement that is of interest. It makes the causal effect, not the E-value, the target of inference. See the Appendix B for greater formality.

Again, as above, one could in principle obtain a confidence interval for the E-value for the estimate, perhaps by bootstrapping or by the delta method. It is not difficult to derive an asymptotic standard error using the delta method for the E-value of the estimate, when that E-value is computed by $RR_{obs} + \sqrt{RR_{obs}(RR_{obs} - 1)}$. Typically, estimation and inference for risk ratios are carried out using symmetric confidence intervals around $\beta = log(RR_{obs})$. Suppose we have an estimate $\hat{\beta}$ of β with estimated standard error $\hat{\sigma}$, then the E-value for the estimate is $e^{\hat{\beta}} + \sqrt{e^{\hat{\beta}}(e^{\hat{\beta}} - 1)}$ and its standard error is, by the delta method, $\hat{\sigma}\left\{ e^{\hat{\beta}} + \frac{2e^{2\hat{\beta}} - e^{\hat{\beta}}}{2\sqrt{e^{\hat{\beta}}(e^{\hat{\beta}} - 1)}} \right\}$ and from there one could obtain an asymptotic 95 % confidence interval for the E-value as $e^{\hat{\beta}} + \sqrt{e^{\hat{\beta}}(e^{\hat{\beta}} - 1)} \pm 1.96\hat{\sigma}\left\{ e^{\hat{\beta}} + \frac{2e^{2\hat{\beta}} - e^{\hat{\beta}}}{2\sqrt{e^{\hat{\beta}}(e^{\hat{\beta}} - 1)}} \right\}$. However, as above, in most contexts it will not be E-value itself, but rather the causal effect, that is the target of inference.

5 Relation to Rosenbaum's design sensitivity

Questions have also arisen with respect to the relation of the E-value to what Paul Rosenbaum calls design sensitivity [16]. The two concepts are related but also have a number of important differences. The sensitivity analysis parameter, Γ, in Rosenbaum's design sensitivity is the maximum ratio by which two units with identical covariates C may differ in their odds of receiving the exposure. Under randomization conditional on C, two units with the same covariates would not differ at all in their odds of exposure and thus we would have Γ = 1. If however there were an unmeasured covariate U that affected the odds of exposure, then we may have Γ > 1. Rosenbaum discusses evaluating how large the parameter would need to be for the results to be sensitive e. g. for the P-value to rise above 0.05. This is in some ways analogous to the proposed E-value for the confidence interval, but with a different parameterization.

With regard to design sensitivity specifically, for a given population, and a given design, and a proposed method of analysis, the design sensitivity is how large the sensitivity analysis parameter, Γ, would have to be in large samples to change the conclusion. What is similar with design sensitivity and the E-value is that both

---

1 Or more generally one could make statements of the form "Across repeated samples, at least 95 % of the time it is the case that: if the actual confounding parameters RR_{UY} and RR_{EU} are such that the bias factor generated by them is less than that given by having RR_{UY} and RR_{EU} equal to the E-value for the confidence interval that was calculated, then the association adjusted by the unmeasured confounder(s) will be in the same direction as the observed association."

concern the amount of unmeasured confounding that would be required to alter conclusions or to explain away an observed association as to not being due to a true causal effect of the exposure on the outcome.

However, there are several differences between the design sensitivity and the E-value. First, different associations are used to characterize unmeasured confounding in the two approaches. In Rosenbaum's design sensitivity the strength of the unmeasured confounding relates to how much an unmeasured covariate might increase the odds of exposure. With the E-value, the sensitivity analysis parameters are the associations relating the exposure to the unmeasured confounder, *and* also relating the unmeasured confounder to the outcome. Rosenbaum's design sensitivity does not make explicit reference to the effect of the unmeasured confounder on the outcome. In further work Rosenbaum and Silber [17] propose what they call an amplification of the sensitivity analysis that re-expresses the sensitivity analysis parameter Γ in terms of effects of an unmeasured confounder on the exposure and outcome. However, it does so under a particular model for the effect of the confounder on the outcome. In contrast, the sensitivity analysis parameters that are used in the E-value, RR_{UD} and RR_{EU}, do not presuppose a model for the effect of the unmeasured confounder on the outcome, nor for the relation between the exposure and the unmeasured confounder. The sensitivity analysis parameters RR_{UD} and RR_{EU} are defined non-parametrically, as above, using maximums.

A second difference between the approaches is that Rosenbaum's design sensitivity was developed to evaluate the *sharp* null hypothesis of no causal effect for any individual. The E-value can be used to assess the strength of unmeasured confounding that is needed to move the estimate to the null of no *average* causal effect; but the E-value can also be used to assess the strength of unmeasured confounding that is needed to move the estimate of the average causal effect to any other value of the causal effect as well, for example to a scientifically meaningful threshold for which a causal effect of lesser magnitude would simply not be of substantive interest [1, 3].

A third difference between the approaches is that for the design sensitivity Rosenbaum proposes that the sensitivity parameter that would explain away the observed association be assessed as the sample size tends to infinity, whereas our proposal is that the E-value be calculated for the actual sample. Rosenbaum's design sensitivity is intended to be a property of the design, not the sample size. Using the design sensitivity, one can compare different designs for large sample sizes to determine which designs may be more robust to potential unmeasured confounding. Our proposed approach using E-values is calculated with the actual data and estimates. As above, we propose calculating E-values for both the estimate and for the limit of the confidence interval closest to the null [1]. The E-value for the limit of the confidence interval closest to the null will of course vary across samples and will vary by sample size. There is also an E-value for the actual confounded association between exposure and the outcome conditional on C i. e. the risk ratio one would obtain in an infinite sample size relating the exposure and the outcome, conditional on the measured covariates, but not adjusting for unmeasured covariates U. That E-value for the actual confounded risk in an infinite sample is more closely analogous to Rosenbaum's design sensitivity. It is also what would be the target of inference if one were to calculate a confidence interval for the E-value of the estimated risk ratio. However, as argued in the previous section, this seems of less use in evaluating the actual evidence for a causal effect from a given study than the E-value for the confidence interval itself. Again, as argued above, the target of inference is generally the causal effect, not the E-value.

A fourth difference between design sensitivity and the E-value is the scale used. The design sensitivity is defined on an odds ratio scale. The E-value is defined on the risk ratio scale. In principle, this difference is only a matter of mathematical definition of scale. However, in practice, we think it is often an important difference. In practice, odds ratios are not infrequently interpreted, often inadvertently, as risk ratios. When the variable under consideration is rare, odds ratios in fact approximate risk ratios and this is then unproblematic [12]. However, when the variable for which the odds is being considered is common, then odds ratios can vastly overestimate risk ratios. In many scenarios, odds ratios are roughly the square of risk ratios [18]. When the probability of the variable being considered lies in the range 0.2 to 0.8, the odds ratio can exaggerate the risk ratio by a factor as large as 400 % [18]! In these cases, interpreting odds ratios as risk ratios is highly problematic. The sensitivity analysis parameters in Rosenbaum's design sensitivity are defined in terms of odds ratios for the exposure. The exposures being examined in many studies are of course often relatively common. Sensitivity analysis using odds ratio scales in these settings can be problematic [19], and one must take due

caution. For example, if 50 % of the population is exposed and 50 % is unexposed and there are no measured covariates but one unmeasured binary confounder U with 50 % prevalence in the population such that when U = 1, the exposure occurs with 70 % probability and when U = 0 it occurs with 30 % probability, then the sensitivity analysis parameter relevant in a design sensitivity calculation would be: $(0.7/0.3)/(0.3/0.7) = 5.4$. One could correctly say that two units with identical measured covariates could differ in odds of treatment by at most 5.4-fold. If, however, this is inadvertently interpreted as a risk ratio, then this is problematic, since in fact it is the case that two units with identical measured covariates could differ in probability treatment by at most $0.7/0.3 = 2.33$-fold. In this example, the parameter RR_{EU} is also 2.33 (but this will not be the case if there are not equal numbers of the exposed versus unexposed). The point here is only the obvious one that odds ratios should not be interpreted as risk ratios; if they are, then robustness will be exaggerated. If investigators are careful not to interpret odds ratios as risk ratios, then this need not necessarily be problematic. However, we believe such misinterpretation of odds ratios as risk ratios is common in practice and for that reason we would in general advocate for using sensitivity analysis parameters on risk ratio scales. It should, however, be noted that in the formulation of E-values, the parameter RR_{EU} is, as noted above, the risk ratio for U conditional on E, rather than for E conditional on U, which likewise must be taken into account in interpretation. The parameter is thus on the risk ratio scale but in the reverse direction that is sometimes expected. We also have endeavored to provide a variety of approximations so that E-values, with parameters reported on risk ratio scales, can be obtained regardless of the initial method of analysis or effect measure employed in estimation [1, 18].

In summary, while design sensitivity is somewhat analogous to the E-value for the actual confounded association between the exposure and outcome, the reporting practices for the E-value that we advocate for [1] differ from design sensitivity in their considerations of the relations between the unmeasured confounder and the outcome; differ in considering the null of no average causal effect rather than the sharp null; differ in considering the actual sample versus an infinite sample; and differ in using risk ratio rather than odds ratio scales.

6 Concluding remarks

It is our hope, by addressing these questions concerning the interpretation of the confounding association parameters, the nature of the E-value transformation, questions of statistical inference using the E-value, and distinctions from design sensitivity, that the interpretation of the E-value metric is clearer and that its use will thereby be further facilitated.

Appendix A. E-values under coarsening

We will consider an example with an unbounded U such that one of sensitivity parameters RR_{EU} is infinite and the other RR_{UD} is very large, but such that a coarsening U' of U into five categories suffices to reduce the bias to less than 1 % and for which two sensitivity parameters $RR_{U'D}$ and $RR_{EU'}$ are finite and relatively moderate. For simplicity will assume no measured covariates. Suppose E is binary with 50 % exposed and 50 % unexposed, and that U takes values among the non-negative integers with distribution conditional on E that is Poisson with mean $(1.5 + 0.5E)$. Suppose further that Y follows a logistic model with

$$\Pr(Y = 1|E, U) = \frac{\exp(-4 + 0.5E + 0.1U)}{1 + \exp(-4 + 0.5E + 0.1U)}$$

In this case,

$$RR_{EU} = max_u \frac{P(U = u|E = 1)}{P(U = u|E = 0)} = max_u \frac{2^u e^{-2}}{u!} \frac{u!}{1.5^u e^{-1.5}} = max_u \frac{(\frac{4}{3})^u}{e^{0.5}} = \infty$$

and

$$RR_{UD} = max \left\{ \frac{max_u P(D = 1|E = 1, c, u)}{min_u P(D = 1|E = 1, c, u)}, \frac{max_u P(D = 1|E = 0, c, u)}{min_u P(D = 1|E = 0, c, u)} \right\} =$$

$$\frac{max_u P(D = 1|E = 0, c, u)}{min_u P(D = 1|E = 0, c, u)} = \frac{max_u \frac{\exp(-4+0.1u)}{1+\exp(-4+0.1u)}}{min_u P \frac{\exp(-4+0.1u)}{1+\exp(-4+0.1u)}} = \frac{1}{\frac{\exp(-4)}{1+\exp(-4)}} = 55.6.$$

However, with a coarsening U' of U such that U' = U if U ∈ {0, 1, 2, 3} and U' = 4 if U ≥ 4 then one can show by numeric integration that the bias from standardization by U' rather than U is such that

$$\frac{\frac{\sum_u P(D=1|E=1,u)P(u)}{\sum_u P(D=1|E=0,u)P(u)}}{\frac{\sum_{u'} P(D=1|E=1,u')P(u')}{\sum_{u'} P(D=1|E=0,u')P(u')}} < 1.01$$

and moreover that

$$RR_{EU'} = max_{u'} \frac{P(U' = u'|E = 1)}{P(U' = u'|E = 0)} = \frac{P(U' = 4|E = 1)}{P(U' = 4|E = 0)} \approx 2.17$$

and by numerical integration that

$$RR_{U'D} = max \left\{ \frac{max_{u'} P(D = 1|E = 1, u')}{min_{u'} P(D = 1|E = 1, u')}, \frac{max_{u'} P(D = 1|E = 0, u')}{min_{u'} P(D = 1|E = 0, u')} \right\}$$

$$= \frac{P(D = 1|E = 1, U' = 4)}{P(D = 1|E = 1, U' = 0)} \approx 1.55.$$

Thus a coarsening of an unbounded U into 5 categories that would suffice to move the estimate to less than 1 % of the true causal risk ratio, has relatively bounded sensitivity parameters. The conservative bound generated by these sensitivity parameters for U' would be

$$\frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1} \approx 1.23.$$

The actual bias is

$$\frac{RR_{obs}}{RR_{true}} = \frac{\frac{P(D=1|E=1)}{P(D=1|E=0)}}{\frac{\sum_u P(D=1|E=1,u)P(u)}{\sum_u P(D=1|E=0,u)P(u)}} \approx 1.05.$$

Again the bound is conservative.

Appendix B. Interpretation of E-values for the confidence interval

Suppose we have obtained an estimate of

$$RR_{obs} = \frac{P(D = 1|E = 1, c)}{P(D = 1|E = 0, c)}$$

and a 95 % confidence interval, $(\hat{V}, \hat{W})$, as a function of the data. We will consider the case wherein the entire confidence interval is greater than 1. The case where the entire confidence interval is less than 1 is analogous. Suppose we calculate $\hat{Q}$ as the E-value for the limit of the confidence interval closest to the null so that $\hat{Q} = \hat{V} + \sqrt{\hat{V}(\hat{V} - 1)}$. Let $B_c = \frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1}$. We then have that

$$P(\hat{V} < RR_{obs} < \hat{W}) \geq 0.95$$

Thus

$$P(RR_{obs} < \hat{V}) < 0.05$$

$$P(1\{B_c < \hat{V}\}1(RR_{obs} < \hat{V})) < 0.05$$

$$P(1\{B_c < \hat{V}\}1(B_c * RR_{true} < \hat{V})) < 0.05$$

$$P(1\{B_c < \hat{V}\}1(RR_{true} < \hat{V}/B_c)) < 0.05$$

$$P(1\{B_c < \hat{V}\}1(RR_{true} < 1)) < 0.05$$

$$P(1\{\max(RR_{UD}, RR_{EU}) < \hat{Q}\}1(RR_{true} < 1)) < 0.05$$

where the fourth to last line follows because $RR_{obs}/B_c \leq RR_{true}$ and the second to last line follows because if $B_c < \hat{V}$ then $\hat{V}/B_c > 1$, and the last line follows because if $\max(RR_{UD}, RR_{EU}) < \hat{Q}$ then $B_c = \frac{RR_{UD} \times RR_{EU}}{RR_{UD} + RR_{EU} - 1} <$

$\frac{\hat{Q} \times \hat{Q}}{\hat{Q} + \hat{Q} - 1} = \frac{\{\hat{V} + \sqrt{\hat{V}(\hat{V}-1)}\}\{\hat{V} + \sqrt{\hat{V}(\hat{V}-1)}\}}{\hat{V} + \sqrt{\hat{V}(\hat{V}-1)} + \hat{V} + \sqrt{\hat{V}(\hat{V}-1)} - 1} = \hat{V}$. Thus across repeated samples, less than 5 % of the time will it be the case that both RR_{EU} and RR_{UD} are less than the E-value for the limit of the confidence interval closest to the null, and that RR_{true} is in the opposite direction of RR_{obs}. From this it follows that across repeated samples, at least 95 % of the time it is the case that: if the actual confounding parameters RR_{UY} and RR_{EU} are both less than the E-value for the confidence interval that was calculated, then the association adjusted by the unmeasured confounder(s) will be in the same direction as the observed association.

References

1. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 2017;167:268–74.
2. Mathur MB, Ding P, Riddell CA, VanderWeele TJ. Website and R package for computing E-values. Epidemiology. 2017, in press.
3. Ding P, VanderWeele TJ. Sensitivity analysis without assumptions. Epidemiology. 2016;27(3):368–77.
4. Pearl J. Causality: models, reasoning, and inference. 2nd ed. Cambridge: Cambridge University Press; 2009.
5. Sjølander A. Letter to the editor. Stat Med. 2009;28:1416–20.
6. Ding P, Miratrix LW. To adjust or not to adjust? Sensitivity analysis of M-Bias and Butterfly-Bias (with comments). J Causal Infer. 2015;3:41–57.
7. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004 Sep;15(5):615–25.
8. Wooldridge J. Should instrumental variables be used as matching variables? Res Econ. 2016;70:232–7.
9. Ding P, VanderWeele TJ, Robins JM. Instrumental variables as bias amplifiers with general outcome and confounding. Biometrika. 2017;104:291–302.
10. Pearl J. On a class of bias-amplifying variables that endanger effect estimates. In: Grunwald P, Spirtes P, editors. Proc. 26th conf. uncert. artif. intel. (UAI 2010). Corvallis, Oregon: Association for Uncertainty in Artificial Intelligence; 2010. p. 425–32.
11. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics. 1968;24:295–313.
12. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Lippincott Williams & Wilkins; 2008.
13. Wasserstein RL, Lazar NL. The ASA's statement on p-values: context, process, and purpose. Am Stat. 2017;70:129–33. 2016.
14. Greenland S. Invited commentary: the need for cognitive science in methodology. Am J Epidemiol. 2017;186:639–45.
15. VanderWeele TJ. Re: The ongoing tyranny of statistical significance testing in biomedical research. Eur J Epidemiol. 2010;25:843.
16. Rosenbaum PR. Design sensitivity in observational studies. Biometrika. 2004;91:153–64.
17. Rosenbaum PR, Silber JH. Amplification of sensitivity analysis in observational studies. J Am Stat Assoc. 2009;104:1398–405.
18. VanderWeele TJ. On a square-root transformation of the odds ratio for a common outcome. Epidemiology. 2017;28:e58–60.
19. Robins JM. Comment on "Covariance adjustment in randomized experiments and observational studies." by Paul R. Rosenbaum. Stat Sci. 2002;17(3):286–327.