International Journal of Epidemiology, 2020, 1509–1516
doi: 10.1093/ije/dyaa127
Advance Access Publication Date: 30 September 2020
Original article

E-values

Bias factor, maximum bias and the E-value: insight and extended applications

Alexandre Cusson[1] and Claire Infante-Rivard [2]*

[1]Research Centre, Centre Hospitalier Universitaire Sainte-Justine, Université de Montréal, Montréal, Québec, Canada and [2]Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine, McGill University, Montréal, Québec, Canada

*Corresponding author. Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine, McGill University, 1020 Avenue Des Pins Ouest, Montréal, Québec H3A1A2, Canada. E-mail: claire.infante-rivard@mcgill.ca

Editorial decision 22 June 2020; Accepted 23 June 2020

Abstract

Background: Unmeasured confounding can bias the relationship between exposure and outcome. Sensitivity analyses generate bias-adjusted measures but these are not much used; this may change with the availability of the E-value (for evidence for causality in observational studies), appealing for its ease of calculation. However, as currently proposed, the E-value has some practical limitations that may reduce its use.

Methods: We first provide some insight into the relationship between two established measures for unmeasured confounding: 'the bias factor' and the maximum value this bias factor can take ('the B bias'). These measures are the statistical foundation for the E-value. We use them to develop new E-value formulas for situations when it is not currently applicable such as e.g. when, not unusually, a negative relation between unmeasured confounder and outcome and a positive one with exposure are postulated. We also provide E-values on the odds ratio scale because, currently, even when using the odds ratio as the study measure in the calculation of E-value, the result is to be interpreted as a relative risk, which is somewhat inconvenient.

Results: The additional formulas for the E-value measure make it applicable in all possible scenarios defined by the combined directions between unmeasured confounder and both the exposure and outcome. In addition, E-value measures can now be interpreted as odds ratios if the observed results are reported on the same scale.

Conclusions: The E-value is part of newer sensitivity analyses methods for unmeasured confounding. We provide insight into its structure, underscoring its advantages and limitations, and expand its applications.

Key words: Unmeasured confounding, E-value, bias factor, sensitivity analysis

Key Messages

- Sensitivity analyses for unmeasured confounding are not used frequently, but the availability of the E-value, particularly easy to obtain and calculate, may change that.
- As currently available, the E-value cannot be used with all possible direction combinations between the unmeasured confounder and both the exposure and the outcome; new formulas we propose make this possible.
- The E-value is currently interpretable only on the relative risk scale; with new formulas, it can be interpreted on the odds ratio scale to match the scale of the observed measure of effect.

Introduction

Although unmeasured confounding is likely to affect most observational studies, sensitivity analyses providing bias-adjusted measures[1] are not often applied. Results from such analyses are speculative because the strength of the unmeasured variable with exposure and outcome is not well known or possibly unknown; nevertheless, such results, especially when many plausible options are used, provide a 'what if' perspective with the potential of significantly enriching interpretation of the results.

We first provide some insight into the relationship between two recently proposed measures of bias due to unmeasured confounding: the bias factor (BF),[2] and the maximum value this bias factor can take, namely the B measure.[3] To estimate these measures, statistical assumptions are required, and the investigator must specify values for the relationship between the unmeasured confounder with both the outcome and the exposure. The BF, and its maximum the B value, are related to the E-value, a relative risk measure that provides estimates of the amount of unmeasured confounding needed to explain away the observed exposure–outcome association.[4] Because it was first presented in the mainstream medical literature and offered as a simple tool that can be estimated using only the observed risk measure, the E-value is likely to become a widely applied tool in epidemiology. However, as proposed, it is not directly applicable to the situation when, not unusually, the putative direction of the relationship between the unmeasured confounder and the outcome is not positive. Therefore, we extend the application of the E-value to this situation which occurs together with a negative or positive relationship between the unmeasured confounder and the exposure. We also provide E-values as odds ratios (ORs); currently, even when using observed ORs in the E-value formula, the result is to be interpreted as a relative risk, which is inconvenient.

For simplicity we consider an unmeasured confounder (U) as a binary variable. The risk ratio (RR) between U and the outcome Y is denoted RR_{UY}, while RR_{XU} denotes the risk ratio between the exposure X and U. RR (without subscripts) is for the risk ratio between exposure and outcome; in the equations, RRobs and RRtrue are the observed and the true risk ratios between X and Y, respectively. The true RR between X and Y is assumed adjusted for U.

Measuring a BF for unmeasured confounding

Bross[5] proposed a method to quantify the bias due to unmeasured confounding when the true RR is equal to 1. Schlesselman[6] proposed a BF measure applicable when the true RR is different from 1, which was used later by Vanderweele and Arah.[2] The BF measure assumes no triple interaction: the RR is the same for U = 0 as for U = 1, i.e. $RR_{XY|U=0} = RR_{XY|U=1}$. This assumption also implies that RR_{UY} is common to X = 0 and X = 1 ($RR_{UY|X=0} = RR_{UY|X=1}$); it is referred to as $RR_{UY|X}$. In the BF equation (1) shown below (and relevant others), we use RR_{UY} that is to be interpreted as the value taken by either $RR_{UY|X=0}$ or $RR_{UY|X=1}$, as done elsewhere.[7,8]

Estimating the BF requires the investigator to provide plausible values for RR_{UY} as well as for P(U|X = 1) and P(U|X = 0). The BF equation is:

$$\mathbf{BF} = (1 + ((\mathbf{RR}_{UY} - 1) * \mathbf{P}(\mathbf{U}|\mathbf{X} = 1))) \; / \; (1 + ((\mathbf{RR}_{UY} - 1) * \mathbf{P}(\mathbf{U}|\mathbf{X} = 0))) \quad (1)$$

As an example, setting RR_{UY}, P(U|X = 1), and P(U|X = 0) at 1.5, 0.5 and 0.25, respectively, with the observed RR = 1.80, the BF is calculated as 1.11.

The BF is then used to get a bias-adjusted RR.

$$\text{true } \mathbf{RR} = \text{observed } \mathbf{RR} \; / \; \mathbf{BF} \quad (2)$$

The bias-adjusted RR is 1.80/1.11 = 1.62.

With the observed RR and assuming a plausible value for the true RR, from equation (2), the BF is known:

$$\mathbf{BF} = \text{observed } \mathbf{RR} \ / \ \text{true } \mathbf{RR} \quad (3)$$

Substituting the known BF value in equation (1), the values of RR_{UY} and RR_{XU} [defined as $P(U|X=1)/P(U|X=0)$], which together generated the level of bias in the observed exposure–outcome RR, can be calculated. This is done by setting a value for RR_{UY} and, e.g. a value for $P(U|X=1)$, and then estimating $P(U|X=0)$ as the only unknown parameter in equation (1). The resulting RR_{UY} and RR_{XU} values correspond to a given BF. The same BF can be estimated from a multitude of different combinations of RR_{UY} and RR_{XU}.

When the true RR is assumed equal to 1 (a frequent assumption using the methods that we discuss later), $P(Y|X=1, U=u) = P(Y|X=0, U=u)$ for $u=0,1$; these probabilities are also equal to $P(Y|U=u)$ and as a result $RR_{UY} = RR_{UY|X}$.

When, in equation (3), the $BF=1$ (the null value for the BF), no unmeasured confounding is assumed and the observed RR is equal to the true RR. When $BF \neq 1$, the true RR is different from the observed RR. The BF can be smaller or larger than 1. The difference between the true and observed RR gets larger as the BF moves further away from 1 in either direction.

A related BF measure requiring similar but more convenient parameters

Walker[9] proposed an approach later used by Schneeweiss[8] that is algebraically equivalent to the BF (equation 1); it depends on RR_{UY}, P(U), P(X) and the odds ratio (OR) OR_{XU}. The Walker–Schneeweiss BF_{WS} equation[8] is:

$$\begin{aligned}\mathbf{BF_{WS}} = [&\left(1 - \mathbf{P(X)}\right) \ / \ \mathbf{P(X)}] \\ &* \ [\left(\left(\mathbf{P(X,U)} * (\mathbf{RR_{UY}} - 1)\right)\right. \\ &\left. + \mathbf{P(X)}\right) \ / \ \left(\left(\left(\mathbf{P(U)} - \mathbf{P(X,U)}\right)\right.\right. \\ &\left.\left. * \ (\mathbf{RR_{UY}} - 1)\right) - \mathbf{P(X)} + 1\right)] \end{aligned} \quad (4)$$

where P(X, U) can be replaced by a function that depends on OR_{XU} as well as on other parameters already present in the BF_{WS} equation. Equation (4) uses a RR for the relationship between U and Y, and then an OR for U as related to X, but an equation with RR_{XU} is algebraically possible. Equation (4) uses one more parameter than equation (1); however, a value for P(U) is perhaps easier to find than one for P(U|X) and, as for P(X), it is likely available, e.g. in health databases.

It is also possible to assume a plausible value for the true RR, in which case with the observed RR, the BF_{WS} is then known as observed RR/true RR. With BF_{WS} known, and selected values for RR_{UY}, P(U) and P(X), OR_{XU} is the only unknown parameter in the equation and it can be estimated. Equations to directly estimate OR_{XU} when the other parameters are fixed are given in Schneeweiss.[8] A multitude of different combinations of RR_{UY} and OR_{XU} correspond to a unique BF_{WS} value.

The maximum bias B: fewer supplied parameters and fewer assumptions

Ding and VanderWeele[3] proposed an equation based only on two investigator-determined values: RR_{UY} and RR_{XU}. In addition, the no triple-interaction assumption is not required. The measure, called the B bias, represents the maximum amount of bias that could affect an observed risk ratio.

$$\mathbf{B} = \mathbf{RR_{UY}} * \ \mathbf{RR_{XU}} / \ (\mathbf{RR_{UY}} + \mathbf{RR_{XU}} - 1) \quad (5)$$

An alternative (and simpler) way to introduce the B equation is to derive it from the BF equation. Since the latter requires the no triple interaction assumption, the assumption is also required for B. More information on how the B equation is derived from the BF equation is given in the Supplementary data, available at *IJE* online.

Although both BF and B are bias measures resulting from U, their difference can be understood as follows: BF [equation (1)] represents the actual amount of bias in a study, specific to RR_{UY} and values of the U conditional probabilities [$P(U|X=1)$ and $P(U|X=0)$]. Now assume the information is not sufficient to select study-specific U conditional probabilities, but is sufficient to select reasonable RR_{UY} and RR_{XU} values. The BF then does not represent bias based on a specific unmeasured confounder prevalence such as [$P(U|X=1)$ and $P(U|X=0)$]; indeed, the same RR_{XU} can result from many different prevalence values of P(U|X). However, among the multiple possible BF values, one is the maximum value (upper bound for all the BF values): it corresponds to B, and is estimated using equation (5). For details on why the B equation estimates the maximum BF value, see the Supplementary data, available at *IJE* online.

To illustrate the previous point, suppose an investigator selected: $RR_{UY} = 2$, $P(U|X=1) = 0.3$ and $P(U|X=0) = 0.1$. The BF is 1.18. In this scenario, $RR_{XU} = 3$ [where $RR_{XU} = P(U|X=1)/P(U|X=0)$]. Suppose now the selected values are: $RR_{UY} = 2$ as before, $P(U|X=1) = 0.6$ and $P(U|X=0) = 0.2$. The BF is 1.33. In both scenarios $RR_{UY} = 2$ and $RR_{XU} = 3$. In equation (5) RR_{XU} is directly used and results in only one possible B value. With $RR_{UY} = 2$ and $RR_{XU} = 3$, $B = (2x3)/(2+3-1) = 1.50$. This is the BF value furthest away from the no bias situation, or the

maximum possible bias from unmeasured confounding. Equation (5) is valid when $RR_{XU} > 1$.

B can be defined as: observed RR/true RR.

$$\mathbf{B} = \text{observed } \mathbf{RR} \ / \ \text{true } \mathbf{RR} \quad (6)$$

In equation (6) the ratio is the largest possible for assumed values of RR_{UY} and RR_{XU}.

The E-value: no investigator-selected parameters required

An even simpler method was introduced by Vanderweele and Ding[4] with the benefit of not requiring the specification of assumed numerical values for parameters based on U. The measure is called an E-value and is based on the assumption that $RR_{UY} = RR_{XU}$. Although this may not be explicit in Vanderweele and Ding,[4] to estimate the E-value, there must be an investigator-assigned value to the true RR (which in the original paper is essentially assumed = 1). The E-value equation[4] is directly derived from the B bias equation (5) (see section 'B formulas to use in the E-value' in the Supplementary data, available at *IJE* online) and is given by:

$$\text{E-value} = \mathbf{B} + \text{sqrt}\Big(\mathbf{B} \ * \ (\mathbf{B} - 1)\Big) \quad (7)$$

In equation (7) B is equal to observed RR/true RR. This E-value is equivalent to RR_{UY} and RR_{XU} (since they are postulated as equal) in a scenario where the true RR is investigator-determined. When the true RR of interest is set equal to 1, B is equal to the observed RR, so the E-value can be written using the observed RR [as in equation (7′)]; equations (7) and (7′) are then obviously similar.

$$\text{E-value} = \text{RRobs} + \text{sqrt}\Big(\text{RRobs} \ * \ (\text{RRobs} - 1)\Big) \quad (7')$$

For example if the observed RR in a study is 2.60, and the true RR is assumed = 1, the E-value is = 2.6 + sqrt [2.6 x (2.6 − 1)] = 4.64. It captures the minimum strength of association, on the RR scale, that an unmeasured confounder needs to have with both Y and X for the observed RR of 2.60 to be, in this case, nullified (true RR = 1). Said differently, to nullify the observed RR of 2.6, the minimum value RR_{UY} and RR_{XU} would need to take is 4.64. On the other hand, from equation (6) B = observed RR/true RR = 2.6 [or from equation (5) = (4.64 x 4.64)/(4.64 + 4.64 – 1) = 2.60]. This is the maximum value that the BF can take given $RR_{UY} = RR_{XU} = 4.64$.

While Vanderweele and Ding[4] mostly assume a true RR value of 1, they also discuss in their Supplementary data, available at *IJE* online, a scenario where the true RR is not set to 1 [a non-null E-value corresponding to equation (7)]. We have already shown how the BF can be used when the true RR is different from 1 and discussed how the E-value is directly related to B (maximum value that the BF can take). It can therefore be assumed that all the new formulas we will propose apply to the scenario when the true RR is different from 1.

The interest in setting the true RR to 1 is to determine whether an observed association can be completely explained away by an unmeasured confounder. If the E-value corresponding to a true RR of 1 is too large for plausibility, the conclusion is that the association is still non-null even after correcting for U, conferring some robustness of the observed results.

The E-value can also be used with observed confidence interval (CI) bounds.[4]

Some practical limitations to the E-value

The E-value has practical limitations that are not explicit in the original paper.[4] We present them now, and address them with additional material in subsequent sections.

First we consider the relationship between BF and the RR parameters. Since BF = observed RR/true RR (equation 3), we have that BF > 1 when true RR < observed RR, and BF < 1 when true RR > observed RR. Consider also equation (1) for the BF based on RR_{UY} and RR_{XU} [P(U|X = 1)/P(U|X = 0)]; it changes direction depending on the direction of RR_{UY} with respect to the null value, and on the relationship between P(U|X = 1) and P(U|X = 0). Table 1 shows the BF direction as > or < 1 depending on all these parameters.

Equation (7) for the E-value uses B which is valid only when $RR_{XU} > 1$. Recall that the E-value assumes $RR_{UY} = RR_{XU}$. Together, these two conditions ($RR_{UY} = RR_{XU} > 1$) imply that the E-value equation is strictly valid only when both RR_{UY} and RR_{XU} are >1. In other words, the E-value is directly applicable when the true RR is smaller than the observed RR. This scenario is described on line 1 in Table 1. However, given that RR_{XU} can naturally be <1, additional B formulas to estimate the E-value for this scenario will be necessary. These are presented in the section 'B formulas to use in the E-value' with additional comments in the Supplementary data, available at *IJE* online.

When RRtrue > RRobs, the BF is <1, therefore RR_{UY} and RR_{XU} cannot both be >1. Line 2 in Table 1 corresponds to this scenario where $RR_{UY} > 1$ and $RR_{XU} < 1$. In their presentation, as mentioned, Vanderweele and Ding[4] essentially assume that the true RR is equal to 1. Given this assumption, in the case of line 2 in Table 1, they recommend replacing the observed RR by RR* = 1/RRobs and

Table 1 Direction of the BF and B formulas for all combinations of RR_{UY} and RR_{XU}

RR_{UY}	RR_{XU}	P(U|X)	BF	RR true vs obs	BF function	B that maximizes bias
>1	>1	P(U|X = 1) > P(U|X = 0)	>1	True < obs	Ascending	$(RR_{UY} \times RR_{XU})/(RR_{UY} + RR_{XU} - 1)$
>1	<1	P(U|X = 0) > P(U|X = 1)	<1	True > obs	Descending	$[1 + (RR_{XU} \times RR_{UY}) - RR_{XU}]/RR_{UY}$
<1	>1	P(U|X = 1) > P(U|X = 0)	<1	True > obs	Descending	$(RR_{UY} \times RR_{XU})/(RR_{UY} + RR_{XU} - 1)$
<1	<1	P(U|X = 0) > P(U|X = 1)	>1	True < obs	Ascending	$[1 + (RR_{XU} \times RR_{UY}) - RR_{XU}]/RR_{UY}$

applying the E-value equation to this new ratio [E-value = RR* + sqrt(RR* × (RR* − 1)]. However, the calculated E-value for this scenario must be interpreted as E-value = RR_{UY} and E-value = $1/RR_{XU}$. The reason is that inverting the observed RR is equivalent to flipping the code for the exposure X such that exposed = 0 and unexposed = 1 (which reverses the direction for both RR_{XY} and RR_{XU}). This is now equivalent to the first scenario as RRobs is >1 (larger than the true RR = 1) and RR_{XU} is >1. Take an observed RR of 0.3846, while assuming the true RR = 1, such that B = RRobs/RRtrue = 0.3846. As suggested by Vanderweele and Ding[4] the E-value can be obtained using 1/RRobs, which is 1/0.3846 = 2.60. The E-value was estimated earlier with this observed RR with a result of 4.64 and is interpreted as: RR_{UY} = 4.64 and RR_{XU} = 1/4.64 = 0.2155. Said differently, it takes a RR_{UY} of 4.64 and a RR_{XU} of 0.2155 to nullify the observed RR of 0.3846 (transformed into 2.6 to use in the formula).

However when RR_{UY} is <1 and RR_{XU} >1, as on line 3 in Table 1, the E-value cannot be used since its results are applicable to a scenario when RR_{UY} is equal to RR_{XU}, which is not the case here. We cannot reverse the exposure variable since the equation for B is not valid when RR_{XU} < 1. Similarly, when RR_{UY} is <1 and RR_{XU} < 1, as on line 4 in Table 1, the E-value cannot be used since the equation for B is not valid. We cannot reverse the coding for U since this would not affect RR_{UY} and RR_{XU} similarly.

We now address the estimation of the E-value according to the combined directions taken by RR_{UY} and RR_{XU}, filling the gap for when such an estimation is not currently available.

B Formula to use in the E-value

The B bias as given in formula (5) is estimated as the maximum BF only when $RR_{XU} > 1$. An alternative formula is necessary to calculate the E-value in scenarios where $RR_{XU} < 1$. In the Supplementary data, available at *IJE* online [where equations (8)–(11) are provided] we first show the known B formula for $RR_{XU} > 1$ (equation 9), and then develop one for $RR_{XU} < 1$ (equation 11). These are also summarized in the last column in Table 1. All our additional B equations are derived from the BF equation and assume, as the latter, no triple interaction. For discussion of the B equation with fewer assumptions, see Ding and Vanderweele.[3]

The E-value, as currently proposed, uses known B from equation (5) but is directly applicable only when RR_{UY} and RR_{XU} are both >1 (line 1 in Table 1). RR_{UY} is selected equal to RR_{XU} (denoted E), and algebra solves for E. When $RR_{UY} > 1$ and $RR_{XU} < 1$ (line 2 in Table 1), VanderWeele and Ding[4] offer a practical solution as previously mentioned. We first show this and then additional equations (summarized in Table 2) corresponding to scenarios described on lines 3 and 4 in Table 1. In each scenario, B represents observed RR/true RR, where the true RR can be fixed to a value different from 1.

E-Value when RR_{UY} and RR_{XU} are not both >1

a. RR_{UY} >1 and RR_{XU} <1

Since $RR_{XU} < 1$, equation (11) shown in the Supplementary data, available at *IJE* online, is used, setting $RR_{UY} = 1/RR_{XU}$, such that RR_{UY} = E and RR_{XU} = 1/E, the equation for B becomes:

$$\mathbf{B} = \left(1 + \left(\mathbf{E} * (1/\mathbf{E})\right) - (1/\mathbf{E})\right) / \mathbf{E}$$
$$\rightarrow (\mathbf{E}^2 * \mathbf{B}) - (2 * \mathbf{E}) + 1 = 0$$

The solution is:

$$\mathbf{E} = [1 + sqrt(1 - \mathbf{B})] / \mathbf{B} \quad (12)$$

b. RR_{UY} <1 and RR_{XU} >1

In this scenario (corresponding to line 3 in Tables 1 and 2) the original equation for B presented by Ding and Vanderweele[3] can be used because B requires $RR_{XU} > 1$, which is applicable here. However, the E-value cannot be used, as it requires $RR_{UY} = RR_{XU}$, and coding for exposure cannot be reversed as the equation for B is not valid for $RR_{XU} < 1$. Instead, we set $RR_{XU} = 1/RR_{UY}$, such that RR_{XU} = E and RR_{UY} = 1/E. B is the ratio RRobs/RRtrue. The equation for B becomes:

Table 2 E-Value formulas on the RR scale for all combinations of RR_{UY} and RR_{XU}

RR_{UY}	RR_{XU}	B that maximizes bias	E-value	RR_{UY}	RR_{XU}
>1	>1	$(RR_{UY} \times RR_{XU})/(RR_{UY} + RR_{XU} - 1)$	B + sqrt [B x (B − 1)]	E	E
>1	<1	$[1 + (RR_{XU} \times RR_{UY}) - RR_{XU}]/RR_{UY}$	[1 + sqrt(1 − B)]/B	E	1/E
<1	>1	$(RR_{UY} \times RR_{XU})/(RR_{UY} + RR_{XU} - 1)$	$\{(B + 1) + sqrt\ [(B+1)^2 - (4 \times B^2)]\}/(2B)$	1/E	E
<1	<1	$[1 + (RR_{XU} \times RR_{UY}) - RR_{XU}]/RR_{UY}$	$\{(B+1)—sqrt\ [((B+1)^2 - 4]\}/2$	E	E

Table 3 E-Value formulas on the OR scale for all combinations of OR_{UY} and OR_{XU}

OR_{UY}	OR_{XU}	B that maximizes bias	E-value	OR_{UY}	OR_{XU}
>1	>1	$(E + 1)^2/(4E)$	$(2B - 1) + sqrt\ [(2B - 1)^2 - 1]$	E	E
>1	<1	$(4E)/(E + 1)^2$	[(2 − B) + 2 x sqrt (1 − B)]/B	E	1/E
<1	>1	$(4E)/(E + 1)^2$	[(2 − B) + 2 x sqrt (1 − B)]/B	1/E	E
<1	<1	$(E + 1)^2/(4E)$	$(2B - 1) - sqrt\ [(2B - 1)^2 - 1]$	E	E

$$\mathbf{B} = \left(\mathbf{E} * (1/\mathbf{E})\right) / \left(\mathbf{E} + (1/\mathbf{E}) - 1\right)$$
$$\rightarrow (\mathbf{E}^2 * \mathbf{B}) - \left((\mathbf{B} + 1) * \mathbf{E}\right) + \mathbf{B} = 0$$

The solution is:

$$\mathbf{E} = [(\mathbf{B} + 1) + sqrt\left((\mathbf{B} + 1)^2 - 4 * \mathbf{B}^2\right)] / (2B) \quad (13)$$

c. RR_{UY} <1 and RR_{XU} <1

This scenario corresponds to line 4 in Tables 1 and 2. Since $RR_{XU} < 1$, we use equation (11) from the Supplementary data, available at *IJE* online. Here, E will be equal to both RR_{UY} and RR_{XU} ($E = RR_{UY} = RR_{XU}$). The equation for B becomes:

$$\mathbf{B} = (1 + \mathbf{E}^2 - \mathbf{E}) / \mathbf{E}$$
$$\rightarrow \mathbf{E}^2 - \left((\mathbf{B} + 1) * \mathbf{E}\right) + 1 = 0$$

The solution is:

$$\mathbf{E} = [(\mathbf{B} + 1) - sqrt\left((\mathbf{B} + 1)^2 - 4\right)] / 2 \quad (14)$$

Application to ORs

When the observed measure of effect is the OR instead of the RR, the E-value approach can still be used.[4] When the outcome is rare, the observed OR is directly used in the E-value equation as it is assumed approximately equal to the RR. When the outcome is common, the observed RR can be estimated using the sqrt(OR) in the E-value equation.[4] However the E-value still corresponds to RR_{UY} and RR_{XU}. It would therefore be of interest to have, mainly for ease of interpretation, an E-value returning the values on the OR scale.

We assume a rare disease such that RR_{XY} and RR_{UY} are approximated by OR_{XY} and OR_{UY}, respectively. The order of the subscripts has implications for the RR such that RR_{UY} is $P(Y|U=1)/P(Y|U=0)$ while RR_{YU} is $P(U|Y=1)/P(U|Y=0)$. However, it has no implication for the OR because the OR is symmetric.

E-value estimates on the odds ratio scale are shown in Table 3.

The BF equation (1) can be used as before to estimate the bias contrasting the true and an observed OR:

$$\mathbf{BF} = (1 + ((\mathbf{OR}_{UY} - 1) * \mathbf{P}(\mathbf{U}|\mathbf{X} = 1))) / (1 + ((\mathbf{OR}_{UY} - 1) * \mathbf{P}(\mathbf{U}|\mathbf{X} = 0))) \quad (15)$$

where true OR = observed OR/BF.

The OR_{XU} is measured as $[P(U|X = 1) * \left(1 - P(U|X = 0)\right)] / [P(U|X = 0) * \left(1 - P(U|X = 1)\right)]$ and so $P(U|X = 0)$ is measured as $[P(U|X = 1)] / [OR_{XU} - \left(OR_{XU} * P(U|X = 1)\right) + P(U|X = 1)]$.

As with the RR, the goal is to define a B function that captures the maximum value of the BF for a fixed value of OR_{UY} and OR_{XU}. Equation (15) is rewritten as:

$$\mathbf{BF} = \left(1 + \left((\mathbf{OR}_{UY} - 1) * \mathbf{P}(\mathbf{U}|\mathbf{X} = 1)\right)\right) / (1 + ((\mathbf{OR}_{UY} - 1) * (\mathbf{P}(\mathbf{U}|\mathbf{X} = 1) / [\mathbf{OR}_{XU} - \left(\mathbf{OR}_{XU} * \mathbf{P}(\mathbf{U}|\mathbf{X} = 1)\right) + \mathbf{P}(\mathbf{U}|\mathbf{X} = 1)])))) \quad (16)$$

When $OR_{XU} = OR_{UY}$, E-value equations are obtained by replacing in equation (16) OR_{XU} and OR_{UY} by E. When OR_{XU} is in the opposite direction of OR_{UY}, one of the terms is replaced by E and the other by 1/E. Next, the first and

second derivatives of the function on $P(U|X=1)$ are estimated to determine the $P(U|X=1)$ value for which the BF function (and consequently the bias) is maximized. By replacing $P(U|X=1)$ with that value, we get the equation for B. An equation for E can then be estimated algebraically using the same approach as Vanderweele and Ding[4] for RRs.

E-value formulas on the OR scale are shown in Table 3. Methodological details are found in the Supplementary data, available at *IJE* online.

Despite its limitations (e.g. not collapsible) the wide use of OR in itself justifies adapting new sensitivity methods for it. Moreover, the need for E-value formulas on the OR scale is supported by the fact that while $OR_{UY} \approx RR_{UY}$ when the outcome is rare, OR_{XU} is not approximately equal to RR_{XU}. It may also be useful for comparison with the Walker-Schneeweiss method[8] and similar ones that report the X–U association on the OR scale.

Application to hazard ratio and risk difference

Ding and Vanderweele[3] showed that, under the assumption of a rare time-to-event outcome, the B equation applies to an exposure–outcome association measured with a hazard ratio (HR) by replacing RR_{UY} with the hazard ratio HR_{UY}. However, to use the E-value, both the confounder-outcome and confounder–exposure associations need to be on the same scale. Vanderweele and Ding[4] recommend estimating the RR_{XY} and then applying the RR-derived E-value. When the outcome is rare, the RR approximates the HR; if not, the RR approximates the HR as $RR \approx (1–0.5^{sqrt(HR)})/(1–0.5^{sqrt(1/HR)})$. As for an exposure–outcome association measured as a risk difference (RD), Ding and Vanderweele[3] provide indirect methods to obtain the E-value. Software has been developed to estimate the results.[10]

Sensitivity analyses and multiple bias analysis

A probabilistic sensitivity analysis can be used.[11] Parameters such as RR_{UY}, $P(U|X=1)$ and $P(U|X=0)$ from the BF equation would each be assigned a prior probabilistic distribution (e.g. beta). One advantage of using a probabilistic distribution as opposed to a deterministic approach (one parameter value at a time) is that the estimation accounts for parameter uncertainty. However, selecting the distribution can be arbitrary. On the other hand, such an approach is not now applicable to the E-value, as it does not account for assumptions about the prevalence of U, or the strength of the association between U and both X and Y.

In the presence of multiple sources of bias, including also measurement error and selection bias, if the errors are considered independent, corrections are applied in the reverse order of occurrence.[11] Since unmeasured confounding is considered to occur first in the population (before the selection process), its correction is carried out last. A possible empirical approach (to be tested) would be to obtain a RR, independently corrected for measurement error and selection bias, and then apply the E-value approach.

Discussion

The E-value[4] is the easiest approach to gauge the potential impact of unmeasured confounding and is likely to become much used. However, as currently available, it has limitations to its applications, some of which we addressed. To better understand the E-value, we have shown how previously proposed correction parameters such as the BF and its maximum value, the B bias, are related to it. We also report (in the Supplementary data, available at *IJE* online) estimates for the maximum bias B on the RR scale for scenarios involving certain directions in the relationship between U with Y and X, not covered in the original E-value publication.[4] In addition, we provided formulas for the E-value for scenarios when it is not currently applicable, i.e. when the association between U and Y is postulated as negative. Finally, we have shown E-values on the OR scale (instead of on a RR scale as currently) which should greatly facilitate interpretation when reporting an OR in the study. Throughout we underscored assumptions, requirements, advantages and limitations of the E-value.

The bias factor presented by Schlesselman[6] or by Vanderweele and Arah[2] was used for its simplicity with respect to algebraic manipulations. Its use however implies an underlying assumption of no triple interaction such that $RR_{XY|U=0} = RR_{XY|U=1}$, and $RR_{UY|X=0} = RR_{UY|X=1}$. This assumption requires using $RR_{UY|X}$ as a parameter in the BF equation, except for scenarios where the true $RR=1$. Other approaches, albeit more complex, do not require this assumption.[3] They could also be used to develop E-values for additional scenarios.

We did not develop B equations depending on OR_{UY} and OR_{XU} as in equation (5) leading to B as a ratio of observed OR/true OR. Instead, we replaced the ORs in these equations by E (when OR_{UY} is set equal to OR_{XU}) or by 1/E (when the ORs are in different directions). However, an equation for B similar to equation (5) as a function of OR_{UY} and OR_{XU} is possible and would be useful for scenarios where the effect of U on Y is assumed of a different strength than the effect of U on X. Lee[12] as well as Ding and Vanderweele[3] developed such equations for the observed RR using RR_{UY} and OR_{XU}. Under an assumption of

rare disease, these equations can be used for OR measures of association.

The E-value alleviates the difficulty of providing, for sensitivity analyses of unmeasured confounding, educated guesses for RR_{UY} and RR_{XU}. Whereas it does not require that provision, it does assume that RR_{UY} is equal to RR_{XU} (or its inverse); this assumption may often be quite unrealistic. Calculation simplicity should not automatically favour the E-value as opposed to other BF parameters explained in this paper, especially when good data can guide the choice of values for RR_{UY} and RR_{XU}.

Supplementary data

Supplementary data are available at *IJE* online.

Conflict of interest

None declared.

References

1. Liu W, Kuramoto SJ, Stuart E. An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. *Prev Sci* 2013;**14**:570–80.
2. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 2011;**22**:42–52.
3. Ding P, VanderWeele TJ. Sensitivity analysis without assumptions. *Epidemiology* 2016;**27**:368–77.
4. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med* 2017;**167**:268–74.
5. Bross ID. Spurious effects from an extraneous variable. *J Chronic Dis* 1966;**19**:637–47.
6. Schlesselman JJ. Assessing effects of confounding variables. *Am J Epidemiol* 1978;**108**:3–8.
7. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York, NY: Springer, 2009.
8. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf* 2006;**15**:291–303.
9. Walker AM. *Observation and Inference*. Newton: Epidemiology Resources Inc., 1991: 120–24.
10. Linden A, Mathur M, VanderWeele TJ. Conducting sensitivity analysis for unmeasured confounding in observational studies using E-values: the evalue package. *Stata J* 2020;**20**:162–75.
11. Orsini N, Bellocco R, Bottai M, Wolk A, Greenland S. A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. *Stata J* 2008;**8**:29–48.
12. Lee WC. Bounding the bias of unmeasured factors with confounding and effect-modifying potentials. *Stat Med* 2011;**30**:1007–017.