REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

E-value Formulae for the Odds Ratio

Authors: Kalle Leppälä ✉
– The Organismal and Evolutionary Biology Research Programme,
University of Helsinki,
Finland
kalle.m.leppala@gmail.com

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- The E-value is the minimum strength of association that a potential confounder needs to have with either the exposure or the outcome, if it were to alone explain an observed association between the two. Originally defined in terms of the risk ratio (relative risk), this short note explores the E-value formulae where one or several of the three relationships is measured using the odds ratio instead. A rigorous but elegant alternative is presented to the square root approximation that has been proposed earlier as a convenient but wrong way to combine odds ratios and E-values.

Keywords:

- *E-value; odds ratio; sensitivity analysis.*

AMS Subject Classification:

- 62D20, 92D30.

✉ Corresponding author

1. INTRODUCTION

This work concerns two measures of association between binary random variables X and Y, the risk ratio (relative risk) and the odds ratio, respectively defined as

$$\mathrm{RR}_{XY} = \frac{\mathbb{P}(Y|X)}{\mathbb{P}(Y|\overline{X})}\,, \qquad \mathrm{OR}_{XY} = \frac{\mathbb{P}(Y|X)\mathbb{P}(\overline{Y}|\overline{X})}{\mathbb{P}(Y|\overline{X})\mathbb{P}(\overline{Y}|X)}\,.$$

Let X and Y be independent conditional on a third binary random variable Z, which is not negatively associated with either X or Y:

$$(X \perp\!\!\!\perp Y)\,|\,Z\,, \qquad \begin{cases} \mathrm{RR}_{XZ} \geq 1\,, \\ \mathrm{RR}_{ZY} \geq 1\,. \end{cases} \tag{1.1}$$

The second assumption is not restrictive as we can replace a variable with its opposite if necessary, and that consequently $\mathrm{RR}_{XY} \geq 1$ as well. In terms of graphical causal models (1.1) means that Z d-separates X and Y, which is the case for example when X and Y have no direct causal relationship, Z is their common cause or a mediator (but not a common consequence), and no other variables play a role.

The classical Cornfield inequalities by Cornfield [1] and Schlesselman [6] now state that

$$\min\{\mathrm{RR}_{XZ}, \mathrm{RR}_{ZY}\} \geq \mathrm{RR}_{XY}\,.$$

The causal interpretation is as follows: if and observed association between X and Y was entirely due to a confounder or a mediator Z, then *both* the risk ratio between X and Z *and* between Z and Y must be at least as big as the risk ratio between X and Y. The first development beyond the classical Cornfield inequalities was by Lee and Wang [4], who instead of the minimum bounded a quantity called *confouding rate ratio* to assess sensitivity to stratification. A different but natural perspective is to swap the minimum into a maximum. Now *either* the risk ratio between X and Z *or* between Z and Y must be at least the so-called E-value $\mathrm{RR}_{XY} + \sqrt{\mathrm{RR}_{XY}(\mathrm{RR}_{XY}-1)}$, as formalized in the following theorem of VanderWeele and Ding.

Theorem 1.1 (VanderWeele & Ding [8]). *Assuming* (1.1),

$$\max\{\mathrm{RR}_{XZ}, \mathrm{RR}_{ZY}\} \geq \mathrm{RR}_{XY} + \sqrt{\mathrm{RR}_{XY}(\mathrm{RR}_{XY}-1)}\,.$$

VanderWeele and Ding also suggest [8] that it should become a convention in all observational science to report the E-value or some comparable form of sensitivity analysis. Another E-value formula, although not dubbed as such, has been published in the form of the following theorem by Lee.

Theorem 1.2 (Lee [3]). *Assuming* (1.1),

$$\max\{\mathrm{OR}_{XZ}, \mathrm{RR}_{ZY}\} \geq \Omega + \sqrt{(\Omega+1)(\Omega-1)},$$

where $\Omega = 1 + 2(\mathrm{RR}_{XY} - 1)$.

This time one of the three relationships is measured by the odds ratio instead of the risk ratio. On that same note, if we want to bound $\max\{\mathrm{RR}_{XZ}, \mathrm{OR}_{ZY}\}$ or $\max\{\mathrm{OR}_{XZ}, \mathrm{OR}_{ZY}\}$ with a function of RR_{XY} we can still use Theorems 1.1 and 1.2, respectively, because $\mathrm{RR}_{ZY} \leq \mathrm{OR}_{ZY}$. The bounds are optimal.

Above two out of the three relationships were allowed to be measured by the odds ratio instead of the risk ratio, while RR_{XY} remained fixed. The purpose of this short note is to explore the cases where the association between X and Y is quantified with OR_{XY}. The odds ratio arises naturally in the context of logistic regression or case-control design, and sensitivity analysis results tailored for odds ratio should be preferred over unnecessary conversions to risk ratios. In a recent paper [5], the author published an odds ratio analogue

$$\min\{\mathrm{OR}_{XZ}, \mathrm{OR}_{ZY}\} \geq \mathrm{OR}_{XY}$$

of the classical Cornfield inequalities and completed the collection of joint bounding formulae of the type (4.1) and (4.2) by Ding and VanderWeele [2] and Lee [3], respectively. Such joint bounding formulae are easily inverted into E-value formulae. Suppose $t \leq f(r, s)$, where t, r and s are the relevant risk- or odds ratios in question. Provided that f is increasing with respect to both r and s in the domain $r > 1$, $s > 1$, we can bound $t \leq f(\max(r, s), \max(r, s)) = F(\max(r, s))$ and then solve $\max(r, s) \geq F^{-1}(t)$. The author however failed to seize the opportunity to supplement the joint bounding formulae in [5] with the corresponding E-value formulae—this omission is fixed in Section 2. However, some E-value formulae remain impossible, such as bounding of $\max\{\mathrm{RR}_{XZ}, \mathrm{RR}_{ZY}\}$ with a function of OR_{XY}, because not all combinations of risk ratios and odds ratios allow a joint bounding formula [5]. In this instance VanderWeele recommends [8] approximating OR_{XY} with either RR_{XY} or RR_{XY}^2 [7] depending on whether the outcome Y is rare or common, respectively. Section 3 proposes a rigorous but elegant alternative to the approximation, in terms of a non-zero parameter α controlling the least and largest of the four elements in the probability mass function of Y given X (incidentally also quantifying how much error the square approximation would introduce). Finally, section 4 takes a new look at the joint bounding formulae missing from the author's previous work due to them not existing in general form [5], and presents restricted versions of them that use the parameter α. The results of section 3 could also be derived from these new joint bounding formulae of section 4 using the inversion procedure described above.

2. *E*-VALUE FORMULAE FOR THE ODDS RATIO

Theorem 2.1. *Assuming* (1.1),

$$\max\{\mathrm{OR}_{XZ}, \mathrm{OR}_{ZY}\} \geq \Omega + \sqrt{(\Omega+1)(\Omega-1)}\,,$$

where $\Omega = 1 + 2(\mathrm{OR}_{XY} - 1)$.

Proof: The statement is similar to that of Theorem 1.2 but we present a proof nevertheless. Under assumption (1.1) the joint bounding formula

$$\mathrm{OR}_{XY} \leq \left(\frac{\sqrt{\mathrm{OR}_{XZ}\mathrm{OR}_{ZY}} + 1}{\sqrt{\mathrm{OR}_{XZ}} + \sqrt{\mathrm{OR}_{ZY}}} \right)^2$$

holds by another theorem of the author [5]. Denote $M = \max\{\mathrm{OR}_{XZ}, \mathrm{OR}_{ZY}\}$. Define and differentiate

$$f(r) = \left(\frac{rs+1}{r+s} \right)^2, \qquad f_r(r) = \frac{2(rs+1)(s^2-1)}{(r+s)^3} > 0\,,$$

where $s > 1$. We therefore get

$$\text{(2.1)} \qquad \mathrm{OR}_{XY} \leq \left(\frac{M+1}{2\sqrt{M}} \right)^2 \qquad \Leftrightarrow \qquad \mathrm{OR}_{XY} - 1 \leq (\sqrt{M} - \sqrt{\mathrm{OR}_{XY}})^2\,.$$

By the odds ratio versions [5] of the Cornfield inequalities we have $M \geq \mathrm{OR}_{XY}$, and thereby also $|\sqrt{M} - \sqrt{\mathrm{OR}_{XY}}| = \sqrt{M} - \sqrt{\mathrm{OR}_{XY}}$. Taking a square root of both sides of (2.1) and solving for $\sqrt{M}$ yields

$$\sqrt{M} \geq \sqrt{\mathrm{OR}_{XY}} + \sqrt{\mathrm{OR}_{XY} - 1}\,,$$

from which the claim follows by squaring. □

Theorem 2.2. *Assuming* (1.1),

$$\max\{\mathrm{RR}_{XZ}, \mathrm{OR}_{ZY}\} \geq \Omega + \sqrt{(\Omega+1)(\Omega-1)}\,,$$

where $\Omega = 1 + (\mathrm{OR}_{XY} - 1)/2$

Proof: Under assumption (1.1) the joint bounding formula

$$\mathrm{OR}_{XY} \leq \frac{1 + (\mathrm{RR}_{XZ} - 1)\mathrm{OR}_{ZY}}{\mathrm{RR}_{XZ}} = 1 + \frac{(\mathrm{RR}_{XZ} - 1)(\mathrm{OR}_{ZY} - 1)}{\mathrm{RR}_{XZ}}$$

holds by another theorem of the author [5]. Denote $M = \max\{\mathrm{RR}_{XZ}, \mathrm{OR}_{ZY}\}$. Define and differentiate

$$f(r,s) = \frac{1 + (r-1)s}{r}\,, \qquad f_r(r,s) = \frac{s(r-1)}{r^2} > 0\,, \qquad f_s(r,s) = \frac{r-1}{r} > 0\,,$$

where $r > 1$. Now

$$\text{OR}_{XY} \le \frac{M^2 - M + 1}{M} \quad \Leftrightarrow \quad (\Omega + 1)(\Omega - 1) \le (M - \Omega)^2 \,. \tag{2.2}$$

If it was true that $M < \Omega$, then

$$2M - 1 < 2\Omega - 1 = \text{OR}_{XY} \le \frac{M^2 - M + 1}{M} \quad \Leftrightarrow \quad M^2 < 1 \,,$$

which is a contradiction. Therefore $|M - \Omega| = M - \Omega$, and the theorem is proved by taking a square root and adding Ω to both sides of (2.2). □

Due to the symmetry of the odds ratio, Theorem 2.2 can also be used to bound $\max\{\text{OR}_{XZ}, \text{RR}_{YZ}\}$.

3. *E*-VALUE FORMULAE UNDER AN EXTRA ASSUMPTION

As pointed out earlier [5], OR_{XY} cannot be bounded from above by a joint function of RR_{XZ} and RR_{ZY}, or of OR_{XZ} and RR_{ZY}. Consequently, neither $\max\{\text{RR}_{XZ}, \text{RR}_{ZY}\}$ nor $\max\{\text{OR}_{XZ}, \text{RR}_{ZY}\}$ can be bounded from below by a function of OR_{XY}—joint bounding will be returned to in Section 4. Both the risk ratio and the odds ratio are nevertheless common measures of association, and the desire to do sensitivity analysis doesn't vanish when relationship between X and Y is quantified by the odds ratio. At this point we're aware of the option to use Theorem 2.1, but if one still prefers to measure the associations between X and Z and between Z and Y using the risk ratio, some auxiliary parameter is needed. Assume that the conditional probabilities of Y or $\overline{Y}$ given X or $\overline{X}$ are not close to zero or one, or precisely for some $\alpha > 0$

$$\begin{cases} \min\{\mathbb{P}(Y|X), \mathbb{P}(Y|\overline{X}), \mathbb{P}(\overline{Y}|X), \mathbb{P}(\overline{Y}|\overline{X})\} \ge \alpha \,, \\ \max\{\mathbb{P}(Y|X), \mathbb{P}(Y|\overline{X}), \mathbb{P}(\overline{Y}|X), \mathbb{P}(\overline{Y}|\overline{X})\} \le 1 - \alpha \,. \end{cases} \tag{3.1}$$

Now we can bound the odds ratio with the risk ratio by scaling the portion exceeding one using the parameter α, and as corollaries derive the required *E*-value formulae.

Lemma 3.1. *Assuming* (3.1),

$$\text{OR}_{XY} \le 1 + \frac{\text{RR}_{XY} - 1}{\alpha} \,.$$

Proof: Denote $\mathbb{P}(Y|\overline{X}) = r$ so that

$$\mathbb{P}(Y|X) = \text{RR}_{XY} r \,, \quad 0 < r < \frac{1}{\text{RR}_{XY}} \,.$$

By definition

$$\mathrm{OR}_{XY} = \frac{\mathbb{P}(Y|X)\mathbb{P}(\overline{Y}|\overline{X})}{\mathbb{P}(Y|\overline{X})\mathbb{P}(\overline{Y}|X)} = \frac{\mathrm{RR}_{XY}(1-r)}{1-\mathrm{RR}_{XY}r} = f(r)\,,$$

and differentiating gives

$$f_r(r) = \frac{\mathrm{RR}_{XY}(\mathrm{RR}_{XY}-1)}{(1-\mathrm{RR}_{XY}r)^2} > 0\,.$$

As r approaches $1/\mathrm{RR}_{XY}$, the value $f(r)$ tends to infinity, but with the boundaries (3.1) $\mathbb{P}(\overline{Y}|X) = 1-\mathrm{RR}_{XY}r$ actually hits α first at the point $r = (1-\alpha)/\mathrm{RR}_{XY}$. This happens before $\mathbb{P}(\overline{Y}|\overline{X}) = 1-r$ hits α at $r = 1-\alpha$. Under (3.1), the odds ratio is therefore bounded from above by $f((1-\alpha)/\mathrm{RR}_{XY})$. □

Theorem 3.1. *Assuming* (1.1) *and* (3.1),

$$\max\{\mathrm{RR}_{XZ}, \mathrm{RR}_{ZY}\} \geq \Omega + \sqrt{\Omega(\Omega-1)}\,,$$

where $\Omega = 1+\alpha(\mathrm{OR}_{XY}-1)$.

Proof: Assuming (3.1), $\mathrm{RR}_{XY} \geq \Omega$ by Lemma 3.1. Assuming (1.1), we can just plug this to the E-value formula of VanderWeele & Ding in Theorem 1.1. □

Theorem 3.2. *Assuming* (1.1) *and* (3.1),

$$\max\{\mathrm{OR}_{XZ}, \mathrm{RR}_{ZY}\} \geq \Omega + \sqrt{(\Omega+1)(\Omega-1)}\,,$$

where $\Omega = 1+2\alpha(\mathrm{OR}_{XY}-1)$.

Proof: Assuming (3.1), $\mathrm{RR}_{XY} \geq \Omega$ by Lemma 3.1. Assuming (1.1), we can just plug this to the E-value formula of Lee in Theorem 1.2. □

VanderWeele and Ding have earlier suggested [8] that for rare outcomes the original E-value formula of Theorem 1.1 is applicable for the odds ratio as well, and that for common outcomes it becomes applicable after approximating $\sqrt{\mathrm{OR}_{XY}} \approx \mathrm{RR}_{XY}$ [7]. The rare–common distinction concerning the outcome Y is somewhat misleading as the error factor introduced by the approximation can be arbitrarily large even if the prevalence of Y is one half, and in fact approximation was mathematically motivated using the assumption (3.1) instead. VanderWeele implicitly derives [7] the upper bound for the introduced error as

$$\mathrm{OR}_{XY} \leq \frac{\mathrm{RR}_{XY}^2}{4\alpha(1-\alpha)}\,, \qquad (3.2)$$

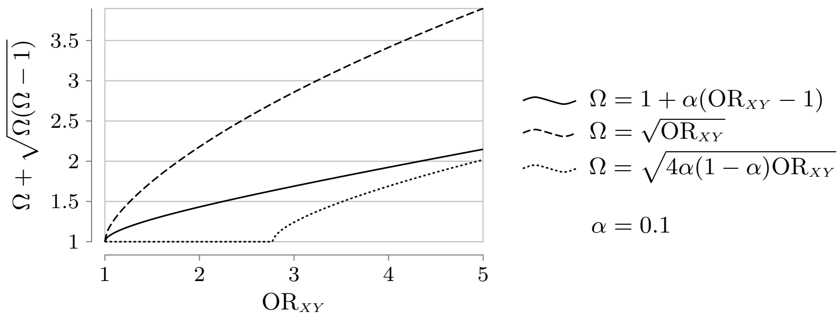

Figure 1: Visualization of different lower bounds of $\max\{RR_{XZ}, RR_{ZY}\}$. The solid line is Theorem 3.1. The dashed line is the currently recommended [8] raw square root approximation; plugging $\sqrt{OR_{XY}}$ in place of RR_{XY} in Theorem 1.1. It has the drawback that it's not actually a bound, and as can be observed the error introduced by approximation can be big. The dotted line corrects the square root approximation using inequality (3.2); it's rigorous but weaker than the solid line bound given by Theorem 3.1.

and so it would be more appropriate to plug in $\sqrt{4\alpha(1-\alpha)OR_{XY}}$ to Theorem 1.1 than the raw square root approximation. Compared to this strictly correct version of the square root approximation Theorem 3.1 gives a stronger statement, and the new *E*-value formula can be seen as quite naturally scaling the portion exceeding one. The comparison is visualized in the figure above.

4. JOINT BOUNDING FORMULAE UNDER AN EXTRA ASSUMPTION

The last two theorems are joint bounding formulae that only became possible under the newly introduced assumption (3.1), included here for completeness sake.

Theorem 4.1. *Assuming* (1.1) *and* (3.1),

$$OR_{XY} \leq 1 + \frac{(RR_{XZ} - 1)(RR_{ZY} - 1)}{\alpha(RR_{XZ} + RR_{ZY} - 1)} .$$

Proof: Assuming (1.1),

$$RR_{XY} \leq \frac{RR_{XZ}RR_{ZY}}{RR_{XZ} + RR_{ZY} - 1} = 1 + \frac{(RR_{XZ} - 1)(RR_{ZY} - 1)}{RR_{XZ} + RR_{ZY} - 1} \tag{4.1}$$

by a theorem of Ding and VanderWeele [2]; the second form is obtained with simple algebraic manipulation. Assuming (3.1), we can plug (4.1) into the bound of Lemma 3.1 and simplify. □

Theorem 4.2. *Assuming* (1.1) *and* (3.1),

$$\mathrm{OR}_{XY} \leq 1 + \frac{(\mathrm{OR}_{XZ} - 1)(\mathrm{RR}_{ZY} - 1)}{\alpha(\sqrt{\mathrm{OR}_{XZ}} + \sqrt{\mathrm{RR}_{ZY}})^2} .$$

Proof: Assuming (1.1),

$$\mathrm{RR}_{XY} \leq \left(\frac{\sqrt{\mathrm{OR}_{XZ}\mathrm{RR}_{ZY}} + 1}{\sqrt{\mathrm{OR}_{XZ}} + \sqrt{\mathrm{RR}_{ZY}}} \right)^2 = 1 + \frac{(\mathrm{RR}_{XZ} - 1)(\mathrm{RR}_{ZY} - 1)}{(\sqrt{\mathrm{OR}_{XZ}} + \sqrt{\mathrm{RR}_{ZY}})^2} \tag{4.2}$$

by a theorem of Lee [3]; again the second form is but algebraic manipulation. Assuming (3.1), we can plug (4.2) into the bound of Lemma 3.1 and simplify. □

It's easy if somewhat laborious to use the law of total probability and differential calculus in the same manner as employed in [5] to show that the bounds in Theorems 4.1 and 4.2 are sharp given the bounds (3.1). Likewise, using Theorems 4.1 and 4.2 as a staring point instead of Lemma 3.1 wouldn't improve the bounds in Theorems 3.1 and 3.2.

REFERENCES

[1] Cornfield, J. et al. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *J. Natl. Cancer Inst.* **22**,1,173–203.

[2] Ding, P. and VanderWeele, T. (2016). Sensitivity analysis without assumptions. *Epidemiology (Cambridge, Mass.)* **27**,3,368–377.

[3] Lee, W.C. (2011). Bounding the bias of unmeasured factors with confounding and effect-modifying potentials. *Stat. Med.* **30**,9,1007–1017.

[4] Lee, W.C. and Wang, L.-Y. (2008). Simple formulas for gauging the potential impacts of population stratification bias. *Am. J. Epidemiol.* **167**,1,86–89.

[5] Leppälä, K. (2023). Sensitivity analysis on odds ratios. *Am. J. Epidemiol.* **192**,9,1882–1886.

[6] Schlesselman, J.J. (1978). Assessing effects of confounding variables. *Am. J. Epidemiol.* **108**,1,3–8.

[7] VanderWeele T.J. (2017) On a square-root transformation of the odds ratio for a common outcome. *Epidemiology.* **28**,6,e58–e60.

[8] VanderWeele T.J., Ding P. (2017) Sensitivity analysis in observational research: introducing the *E*-value. *Ann. Intern. Med.* **167**,4,268–274.