

## Practice of Epidemiology

### Sensitivity Analysis on Odds Ratios

Kalle Leppälä\*

\* Correspondence to Dr. Kalle Leppälä, Organismal and Evolutionary Biology Research Programme, Viikinkaari 1 P.O. Box 65, University of Helsinki, FI-00014 Helsinki, Finland (e-mail: kalle.m.leppala@gmail.com).

Initially submitted January 23, 2023; accepted for publication June 8, 2023.

The classical Cornfield inequalities state that if a third confounding variable is fully responsible for an observed association between the exposure and the outcome variables, then the association between both the exposure and the confounder, and the confounder and the outcome, must be at least as strong as the association between the exposure and the outcome, as measured by the risk ratio. The work of Ding and VanderWeele on assumption-free sensitivity analysis sharpens this bound to a bivariate function of the 2 risk ratios involving the confounder. Analogous results are nonexistent for the odds ratio, even though the conversion from odds ratios to risk ratios can sometimes be problematic. We present a version of the classical Cornfield inequalities for the odds ratio. The proof is based on the mediant inequality, dating back to ancient Alexandria. We also develop several sharp bivariate bounds of the observed association, where the 2 variables are either risk ratios or odds ratios involving the confounder.

Cornfield conditions; Cornfield inequalities; mediant inequality; odds ratio; sensitivity analysis

In the classic 1959 paper (1), Cornfield et al. dealt a death-blow to Fisher’s constitutional hypothesis that the observed association between cigarette smoking and lung cancer is noncausal, and gave birth to the field of sensitivity analysis (2). The key idea derived from Cornfield’s work is that if a strong observed association was merely due to confounding, the confounding variable needs to be strongly associated with both the exposure and the outcome.

Precisely, let the risk ratio (RR) between binary random variables  $X$  and  $Y$  be defined as:

$$RR_{XY} = \frac{\mathbb{P}(Y|X)}{\mathbb{P}(Y|\bar{X})}.$$

Make the assumption that  $X$  and  $Y$  are independent conditional on a third binary random variable  $Z$ , which is not negatively associated with either  $X$  or  $Y$ :

$$(X \perp\!\!\!\perp Y) | Z, \quad \begin{cases} RR_{XZ} \geq 1, \\ RR_{ZY} \geq 1. \end{cases} \quad (1)$$

Note that the second assumption can always be satisfied by recoding the variables  $X$  or  $Y$ , and that as a consequence  $RR_{XY} \geq 1$  as well. The following result is often referred to

as the classical Cornfield inequalities (3), despite only the first half being proven by Cornfield (1, 4).

Theorem 1. Under Assumption 1,

$$RR_{XY} \leq RR_{XZ}, \quad RR_{XY} \leq RR_{ZY}.$$

Even though Theorem 1 is stated in terms of  $RR_{XZ}$  and  $RR_{ZY}$ ,  $X$  does not need to be a cause of  $Z$  or  $Z$  of  $Y$ , because the risk ratio is a measure of association, not causation. Assumption 1 is valid in many causal graphs; for example, in a fork where  $Z$  is a common cause of  $X$  and  $Y$ , or in a chain where  $Z$  is a mediator between  $X$  and  $Y$  (or between  $Y$  and  $X$ ). It might feel more “natural” to use  $RR_{ZX}$  instead of  $RR_{XZ}$  in a fork, for example, but a result like Theorem 1 where  $RR_{XY}$  is bounded by a function of  $RR_{ZX}$  is not possible. The reason is that no matter how large  $RR_{XY}$  is,  $RR_{ZX}$  can be arbitrarily close to 1.

Another way to formulate Theorem 1 is  $RR_{XY} \leq \min(RR_{XZ}, RR_{ZY})$ . A natural question is to ask how the maximum function behaves, and the answer is that it is the upper bound of the so-called E-value  $RR_{XY} + \sqrt{RR_{XY}(RR_{XY} - 1)} \leq \max(RR_{XZ}, RR_{ZY})$  (5). That is, under Assumption 1 not only do both  $RR_{XZ}$  and  $RR_{ZY}$  have to be at least  $RR_{XY}$ ; in addition, at least one of them has to be

even bigger. A second natural question is to ask whether the minimum function is optimal, and the answer is negative. Ding and VanderWeele (6) have shown that the sharp bound of  $RR_{XY}$  using both variables  $RR_{XZ}$  and  $RR_{ZY}$  is

$$RR_{XY} \leq \frac{RR_{XZ}RR_{ZY}}{RR_{XZ} + RR_{ZY} - 1}. \quad (2)$$

Another popular measure of the strength of association between binary variables is the odds ratio (OR):

$$OR_{XY} = \frac{\mathbb{P}(Y|X)\mathbb{P}(\bar{Y}|\bar{X})}{\mathbb{P}(Y|\bar{X})\mathbb{P}(\bar{Y}|X)}.$$

The odds ratio arises naturally in the context of logistic regression or case-control design. In some situations where the outcome is rare, the risk ratio and the odds ratio are very close, and a square root conversion has been suggested for other situations (7). To do a precise conversion, however, one needs some additional parameter, such as the prevalence among the nonexposed. The additional parameter can introduce additional uncertainty, in underdiagnosed disease in particular. It would therefore be preferable to use results explicitly designed for the odds ratio rather than converting to risk ratios and using Cornfield's inequalities or inequality (2).

In this work, we derive an odds ratio analogue of the classical Cornfield inequalities using the mediant inequality dating back to ancient Alexandria. We also compile a complete collection of 2-variable bounds analogous to inequality (2), where some of the 3 risk ratios are odds ratios instead. One such bound by Lee (8) already exists in the literature. Many combinations of risk and odds ratios do not admit a bound, and there is also symmetry in play, so we present 4 original results (2 of which are almost trivial corollaries).

For bounds involving risk differences, see Ding and VanderWeele (3).

## RESULTS

### Bounding with a single ratio

The proof of Theorem 1 will serve as a template for original results.

By the law of total probability and Assumption 1,

$$RR_{XY} = \frac{\mathbb{P}(Y|X)}{\mathbb{P}(Y|\bar{X})} = \frac{\mathbb{P}(Y|Z)\mathbb{P}(Z|X) + \mathbb{P}(Y|\bar{Z})\mathbb{P}(\bar{Z}|X)}{\mathbb{P}(Y|Z)\mathbb{P}(Z|\bar{X}) + \mathbb{P}(Y|\bar{Z})\mathbb{P}(\bar{Z}|\bar{X})}. \quad (3)$$

Cornfield calls the first inequality of Theorem 1 "rather obvious" and derives it algebraically (1, Appendix A) from Ding and VanderWeele (3). We wish to give this classical inequality a classical proof.

Lemma 2. Inequality. If  $bd > 0$ , then

$$\min\left(\frac{a}{b}, \frac{c}{d}\right) \leq \frac{a+c}{b+d} \leq \max\left(\frac{a}{b}, \frac{c}{d}\right).$$

Special cases of the mediant inequality were already known by Plato and Euclid, but the first known formal proof is given by Pappus of Alexandria (9, 10). The validity of the mediant inequality is easy to see geometrically: The slope of the sum vector is between the slopes of the summands. Applying the mediant inequality to equation 3 immediately gives

$$RR_{XY} \leq \max(RR_{XZ}, RR_{X\bar{Z}}) = RR_{XZ}.$$

Schlesselman (4) proves the second classical Cornfield inequality by writing equation 3 as

$$RR_{XY} = \frac{RR_{ZY}\mathbb{P}(Z|X) + 1 - \mathbb{P}(Z|X)}{RR_{ZY}\mathbb{P}(Z|\bar{X}) + 1 - \mathbb{P}(Z|\bar{X})},$$

and maximizing the numerator and minimizing the denominator at the point  $\mathbb{P}(Z|X) = 1, \mathbb{P}(Z|\bar{X}) = 0$ .

For the odds ratio, the law of total probability and Assumption 1 give

$$\begin{aligned} OR_{XY} &= \frac{\mathbb{P}(Y|X)\mathbb{P}(\bar{Y}|\bar{X})}{\mathbb{P}(Y|\bar{X})\mathbb{P}(\bar{Y}|X)} \\ &= \frac{\mathbb{P}(Y|Z)\mathbb{P}(Z|X) + \mathbb{P}(Y|\bar{Z})\mathbb{P}(\bar{Z}|X)}{\mathbb{P}(Y|Z)\mathbb{P}(Z|\bar{X}) + \mathbb{P}(Y|\bar{Z})\mathbb{P}(\bar{Z}|\bar{X})} \\ &\quad \times \frac{\mathbb{P}(\bar{Y}|\bar{Z})\mathbb{P}(\bar{Z}|\bar{X}) + \mathbb{P}(\bar{Y}|Z)\mathbb{P}(Z|\bar{X})}{\mathbb{P}(\bar{Y}|Z)\mathbb{P}(Z|X) + \mathbb{P}(\bar{Y}|\bar{Z})\mathbb{P}(\bar{Z}|X)}. \end{aligned} \quad (4)$$

Using the mediant inequality (Lemma 2) to both fractions yields

$$OR_{XY} \leq \max(RR_{XZ}, RR_{X\bar{Z}}) \times \max(RR_{\bar{X}Z}, RR_{\bar{X}\bar{Z}}) = OR_{XZ}.$$

On the other hand, we can rearrange equation 4 into

$$\begin{aligned} &\frac{\mathbb{P}(Y|Z)\mathbb{P}(Z|X) + \mathbb{P}(Y|\bar{Z})\mathbb{P}(\bar{Z}|X)}{\mathbb{P}(\bar{Y}|Z)\mathbb{P}(Z|X) + \mathbb{P}(\bar{Y}|\bar{Z})\mathbb{P}(\bar{Z}|X)} \\ &\quad \times \frac{\mathbb{P}(\bar{Y}|\bar{Z})\mathbb{P}(\bar{Z}|\bar{X}) + \mathbb{P}(\bar{Y}|Z)\mathbb{P}(Z|\bar{X})}{\mathbb{P}(Y|Z)\mathbb{P}(Z|\bar{X}) + \mathbb{P}(Y|\bar{Z})\mathbb{P}(\bar{Z}|\bar{X})}. \end{aligned}$$

This time the 2 applications of the mediant inequality (Lemma 2) give

$$\begin{aligned} OR_{XY} &\leq \max\left(\frac{\mathbb{P}(Y|Z)}{\mathbb{P}(\bar{Y}|Z)}, \frac{\mathbb{P}(Y|\bar{Z})}{\mathbb{P}(\bar{Y}|\bar{Z})}\right) \times \max\left(\frac{\mathbb{P}(\bar{Y}|\bar{Z})}{\mathbb{P}(Y|\bar{Z})}, \frac{\mathbb{P}(\bar{Y}|Z)}{\mathbb{P}(Y|Z)}\right) \\ &= OR_{ZY}. \end{aligned}$$

We have shown analogues of the classical Cornfield inequalities for the odds ratio.

Theorem 3. Under Assumption 1,

$$OR_{XY} \leq OR_{XZ}, \quad OR_{XY} \leq OR_{ZY}.$$

**Bounding with 2 ratios**

We set out to complete the works of Ding and VanderWeele (6) and Lee (8) with an exhaustive collection of formulae bounding a risk or odds ratio between  $X$  and  $Y$  with any pair of risk or odds ratios between  $X$  and  $Z$  and between  $Z$  and  $Y$ . Not all combinations admit a bound, however.

**Bounding a risk ratio**

It is not possible to bound  $RR_{XY}$  with  $RR_{ZX}$  or  $RR_{YZ}$ , so we only have 4 combinations, 2 of which are existing results (Theorems 4 and 6) and 2 their immediate corollaries (Theorems 5 and 7). The bounds are sharp.

Theorem 4. By Ding and VanderWeele (6). Under Assumption 1,

$$RR_{XY} \leq \frac{RR_{XZ}RR_{ZY}}{RR_{XZ} + RR_{ZY} - 1}.$$

Theorem 5. Under Assumption 1,

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

Proof. Define and differentiate

$$f(r) = \frac{RR_{XZ}r}{RR_{XZ} + r - 1}, \quad f'(r) = \frac{RR_{XZ}^2 - 1}{(RR_{XZ} + r - 1)^2}.$$

The claim follows from  $RR_{XZ} > 1$ , Theorem 4, and  $RR_{ZY} < OR_{ZY}$ .

Theorem 6. By Lee (8). Under Assumption 1,

$$RR_{XY} \leq \left( \frac{\sqrt{OR_{XZ}RR_{ZY}} + 1}{\sqrt{OR_{XZ}} + \sqrt{RR_{ZY}}} \right)^2.$$

Theorem 7. Under Assumption 1,

$$RR_{XY} \leq \left( \frac{\sqrt{OR_{XZ}OR_{ZY}} + 1}{\sqrt{OR_{XZ}} + \sqrt{OR_{ZY}}} \right)^2.$$

Proof. Define and differentiate

$$f(r) = \left( \frac{\sqrt{OR_{XZ}r} + 1}{\sqrt{OR_{XZ}} + r} \right)^2, \\ f'(r) = \frac{2(\sqrt{OR_{XZ}r} + 1)(OR_{XZ} - 1)}{(\sqrt{OR_{XZ}} + r)^3}.$$

The claim follows from  $\sqrt{OR_{XZ}} > 1$ , Theorem 6, and  $\sqrt{RR_{ZY}} < \sqrt{OR_{ZY}}$ .

**Bounding an odds ratio**

It is not possible to bound  $OR_{XY}$  with risk ratios between  $X$  and  $Z$  and between  $Z$  and  $Y$ . Theorem 8 uses a familiar function of  $OR_{XZ}$  and  $OR_{ZY}$ . Theorem 9 provides a bound as a function of  $RR_{XZ}$  and  $OR_{ZY}$ , and by the symmetry of the odds ratio can also be used for  $OR_{XZ}$  and  $RR_{YZ}$ . For the combination of  $OR_{XZ}$  and  $RR_{ZY}$ , and by symmetry  $RR_{ZX}$  and  $OR_{ZY}$ , the bound of Theorem 3 is already sharp.

Theorem 8. Under Assumption 1,

$$OR_{XY} \leq \left( \frac{\sqrt{OR_{XZ}OR_{ZY}} + 1}{\sqrt{OR_{XZ}} + \sqrt{OR_{ZY}}} \right)^2.$$

Proof. Denote  $\mathbb{P}(Z|\bar{X}) = r$  and  $\mathbb{P}(Y|\bar{Z}) = s$ , which implies

$$\begin{cases} \mathbb{P}(Z|X) = \frac{rOR_{XZ}}{r(OR_{XZ}-1)+1}, & 0 < r < 1, \\ \mathbb{P}(Y|Z) = \frac{sOR_{ZY}}{s(OR_{ZY}-1)+1}, & 0 < s < 1. \end{cases}$$

Plugging this into equation 4 gives

$$OR_{XY} = f(r, s) = \frac{(OR_{XZ}OR_{ZY}r + g(r, s))(r + g(r, s))}{(OR_{XZ}r + g(r, s))(OR_{ZY}r + g(r, s))},$$

where  $g(r, s) = (1 - r)(s(OR_{ZY} - 1) + 1)$ . By the chain rule,

$$f_s(r, s) = \frac{(OR_{XZ} - 1)(OR_{ZY} - 1)rg_s(r, s)}{(OR_{XZ}r + g(r, s))^2(OR_{ZY}r + g(r, s))^2} \times (OR_{XZ}OR_{ZY}r^2 - g(r, s)^2).$$

Because  $g_s(r, s) = (1 - r)(OR_{ZY} - 1) > 0$ , the partial derivative  $f_s(r, s)$  is zero on the curve  $g(r, s) = \sqrt{OR_{XZ}OR_{ZY}r}$ , where  $f(r, s)$  is actually a constant:

$$f(r, s) \leq \frac{(OR_{XZ}OR_{ZY} + \sqrt{OR_{XZ}OR_{ZY}})(1 + \sqrt{OR_{XZ}OR_{ZY}})}{(OR_{XZ} + \sqrt{OR_{XZ}OR_{ZY}})(OR_{ZY} + \sqrt{OR_{XZ}OR_{ZY}})} \\ = \left( \frac{\sqrt{OR_{XZ}OR_{ZY}} + 1}{\sqrt{OR_{XZ}} + \sqrt{OR_{ZY}}} \right)^2.$$

Theorem 9. Under Assumption 1,

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

Proof. Denote  $\mathbb{P}(Z|\bar{X}) = r$  and  $\mathbb{P}(Y|\bar{Z}) = s$ , which implies

$$\begin{cases} \mathbb{P}(Z|X) = rRR_{XZ}, & 0 < r < \frac{1}{RR_{XZ}}, \\ \mathbb{P}(Y|Z) = \frac{sOR_{ZY}}{s(OR_{ZY}-1)+1}, & 0 < s < 1. \end{cases}$$

According to equation 4 we can write

$$OR_{XY} = f(r, s) = \frac{g(RR_{XZ}, s)}{g(r, s)},$$

where

$$g(r, s) = \frac{OR_{ZY}r + (1-r)(s(OR_{ZY}-1)+1)}{r + (1-r)(s(OR_{ZY}-1)+1)}.$$

Differentiate

$$g_r(r, s) = \frac{(OR_{ZY}-1)(s(OR_{ZY}-1)+1)}{(r+(1-r)(s(OR_{ZY}-1)+1))^2} > 0,$$

$$g_{rr}(r, s) = \frac{2(OR_{ZY}-1)(s(OR_{ZY}-1)+1)s(OR_{ZY}-1)}{(r+(1-r)(s(OR_{ZY}-1)+1))^3} > 0.$$

It follows that  $f_r(r, s) > 0$  and so we set  $r = 1/RR_{XZ}$ . To maximize over  $s$  we differentiate

$$f_s\left(\frac{1}{RR_{XZ}}, s\right) = \frac{OR_{ZY}(1-RR_{XZ})(1-OR_{ZY})^2}{((RR_{XZ}-1)(OR_{ZY}-1)s+RR_{XZ}+OR_{ZY}-1)^2} > 0$$

and conclude that

$$OR_{XY} \leq f\left(\frac{1}{RR_{XZ}}, 1\right) = \frac{1+(RR_{XZ}-1)OR_{ZY}}{RR_{XZ}}.$$

## APPLICATION

As an illustrative example we use a systematic review by Owen et al. (11) on the effect of breastfeeding on type 2 diabetes; see also Ip et al. (12). Owen et al. report an aggregate odds ratio of 0.61 (95% confidence interval: 0.44, 0.85,  $P$  value = 0.003) between ever breastfed vs. exclusively formula fed and later type 2 diabetes of the child. The meta-analysis comprises 7 individual studies on Europeans and Americans, one of which has a case-control design, with others based on a cohort. Because type 2 diabetes is common and underdiagnosed, converting odds ratios into risk ratios introduces complications for both cohort and case-control designs, so basing sensitivity analysis on Cornfield's inequalities 1 or Theorems 4–7 is not advised. Instead we can use Theorems 3, 8, and 9.

To satisfy Assumption 1, we switch the order of the 2 feeding groups so that the aggregate odds ratio becomes  $1/0.61 = 1.63$ . We then ask: Can some third unmeasured binary variable  $Z$  fully explain the observed association between breastfeeding and the child's type 2 diabetes status? According to Theorem 3, the association between  $Z$  and breastfeeding and the association between  $Z$  and child's type 2 diabetes would both need to have an odds ratio of at least 1.61. Moreover, if one of the associations is weak, the other has to compensate by being strong.

Say  $Z$  is type 2 diabetes of the mother (diagnosed before or after the childbirth). According to Meigs et al. (13), the association between the mother's and the child's type 2 diabetes has an odds ratio of 3.4. By solving for  $RR_{XZ}$  using Theorem 9 with  $OR_{XY} = 1.61$ ,  $OR_{ZY} = 3.4$  we see that the risk of type 2 diabetes among mothers who do not lactate would have to be at least 1.35... times the risk among mothers who do not. Diabetics and future diabetics can have on average more trouble lactating, and lactating can have a protective effect against diabetes. A recent meta-analysis by Pinho-Gomes et al. (14) estimates that never vs. ever lactation is associated with maternal type 2 diabetes with risk ratio at  $1/0.73 = 1.36...$  (We should point out that there is a slight discrepancy because a woman can have several children and only breastfeed some of them.) A noncausal association between breastfeeding and a child's type 2 diabetes, entirely confounded by maternal type 2 diabetes, looks plausible at the moment.

However, from the proof of Theorem 9, we can see that the bound is sharpest when  $\mathbb{P}(Z|\bar{X})$ , the prevalence of type 2 diabetes among women who ever lactated, is large (close to 0.73). We can reasonably argue that it should not be more than the prevalence among men, which should be safe to bound under 30%. Thus, the association reported by Pinho-Gomes et al. (14) has an odds ratio no more than 1.63, where we have used the conversion formula

$$OR_{XZ} = \frac{(1 - \mathbb{P}(Z|\bar{X})) RR_{XZ}}{1 - \mathbb{P}(Z|\bar{X}) RR_{XZ}}$$

and the crude estimate  $\mathbb{P}(Z|\bar{X}) \leq 0.3$ . If we plug  $OR_{XY} = 1.61$ ,  $OR_{ZY} = 3.4$  into Theorem 8 and solve for  $OR_{XZ}$ , we see that the odds ratio would have to be at least 5.7 if confounding by maternal type 2 diabetes were to fully explain the observed breastfeeding-diabetes association.

Indeed, 3 of the 7 studies accounted in the meta-analysis by Owen et al. (11) used a number of relevant covariables, including parental type 2 diabetes, and when restricting to those 3 studies the aggregate odds ratio was similar ( $1/0.55 = 1.81$ ) with and without covariates.

## ACKNOWLEDGMENTS

Author affiliation: Research Unit of Mathematical Sciences, University of Oulu, Oulu, Finland (Kalle Leppälä).

Conflict of interest: none declared.

## REFERENCES

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

2. Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect*. New York, NY: Basic Books; 2018.
3. Ding P, VanderWeele TJ. Generalized Cornfield conditions for the risk difference. *Biometrika*. 2014;101(4):971–977.
4. Schlesselman JJ. Assessing effects of confounding variables. *Am J Epidemiol*. 1978;108(1):3–8.
5. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med*. 2017;167(4):268–274.
6. Ding P, VanderWeele TJ. Sensitivity analysis without assumptions. *Epidemiology*. 2016;27(3):368–377.
7. VanderWeele TJ. On a square-root transformation of the odds ratio for a common outcome. *Epidemiology*. 2017;28(6):e58.
8. Lee WC. Bounding the bias of unmeasured factors with confounding and effect-modifying potentials. *Stat Med*. 2011;30(9):1007–1017.
9. Fowler DH. An approximation technique, and its use by Wallis and Taylor. *Arch Hist Exact Sci*. 1991;41(3):189–233.
10. Pappus of Alexandria. *VIII*. Fourth century (CE).
11. Owen CG, Martin RM, Whincup PH, et al. Does breastfeeding influence risk of type 2 diabetes in later life? A quantitative analysis of published evidence. *Am J Clin Nutr*. 2006;84(5):1043–1054.
12. Ip S, Chung M, Raman G, et al. Breastfeeding and maternal and infant health outcomes in developed countries. *Evid Rep Technol Assess*. Full Report. 2007;153:1–186.
13. Meigs JB, Cupples LA, Wilson PW. Parental transmission of type 2 diabetes: the Framingham Offspring Study. *Diabetes*. 2000;49(12):2201–2207.
14. Pinho-Gomes AC, Morelli G, Jones A, et al. Association of lactation with maternal risk of type 2 diabetes: a systematic review and meta-analysis of observational studies. *Diabetes Obes Metab*. 2021;23(8):1902–1916.