Multiple-Bias Sensitivity Analysis Using Bounds

Louisa H. Smith

louisa_h_smith@g.harvard.edu

with Maya Mathur (Stanford) and Tyler VanderWeele (Harvard)

August 9, 2021

JSM

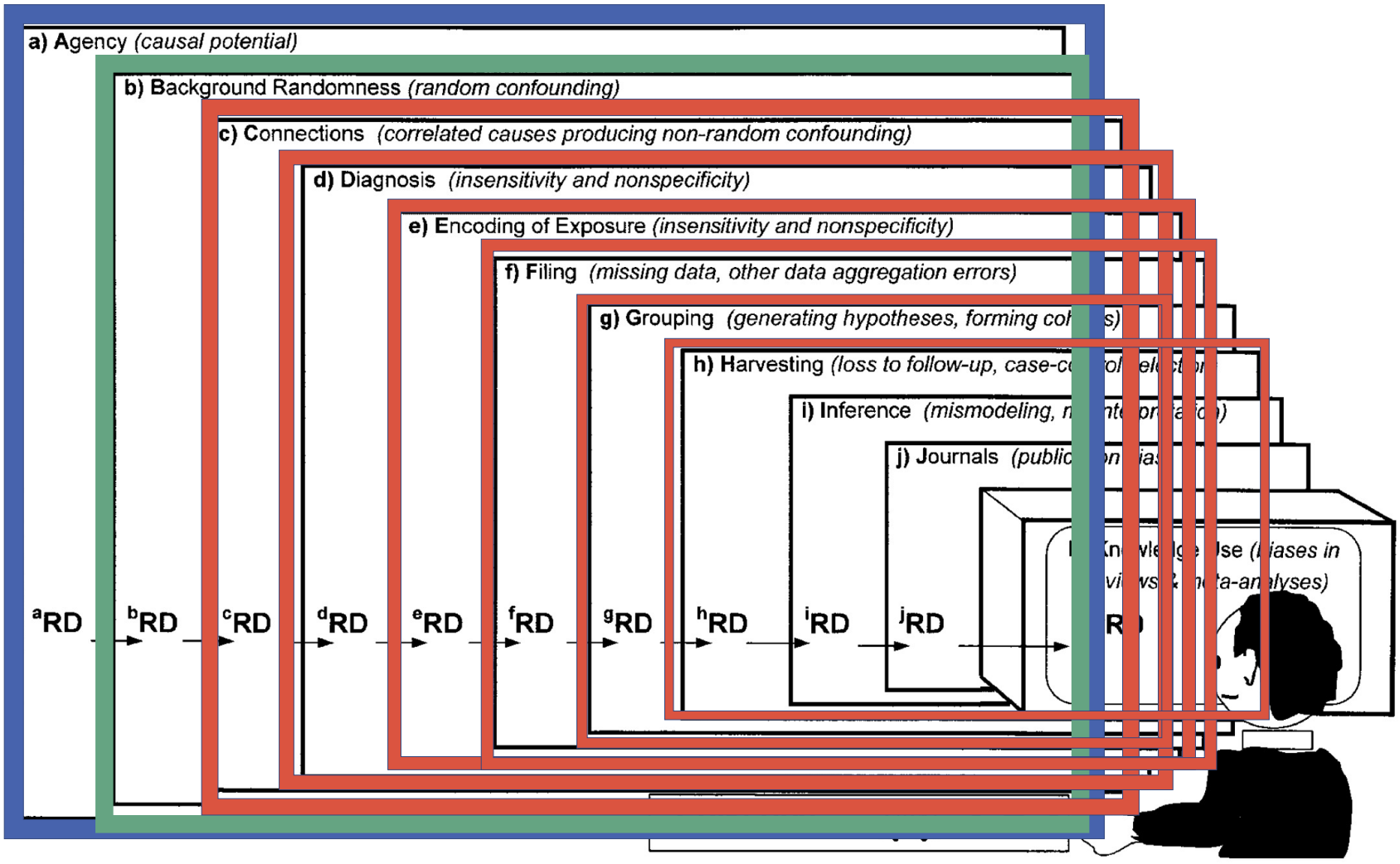
Goal of epidemiology

*or other fields that attempt to observe and quantify natural phenomena

Some truth about a population

Published/ shared conclusions about that population

Layers of distortion



Maclure & Schneeweiss, 2001

Layers of distortion

"The main lesson from our causal model of bias is that we should question the adequacy of qualitative assessments of bias in the Discussion sections of study reports. Even with the help of user-friendly notation and illustrations, it is hard to think about all these causes of bias simultaneously."

Can't rely on heuristics…

Maclure & Schneeweiss, 2001

Multiple-bias modelling

J. R. Statist. Soc. A (2005)
168, *Part* 2, *pp.* 267–306

Multiple-bias modelling for analysis of observational data

Sander Greenland

University of California, Los Angeles, USA

[*Read before* The Royal Statistical Society *on Wednesday, September 29th, 2004, the President*, Professor A. P. Grieve, *in the Chair*]

Summary. Conventional analytic results do not reflect any source of uncertainty other than random error, and as a result readers must rely on informal judgments regarding the effect of possible biases. When standard errors are small these judgments often fail to capture sources of uncertainty and their interactions adequately. Multiple-bias models provide alternatives that allow one systematically to integrate major sources of uncertainty, and thus to provide better input to research planning and policy analysis. Typically, the bias parameters in the model are not identified by the analysis data and so the results depend completely on priors for those parameters. A Bayesian analysis is then natural, but several alternatives based on sensitivity analysis have appeared in the risk assessment and epidemiologic literature. Under some circumstances these methods approximate a Bayesian analysis and can be modified to do so even better. These points are illustrated with a pooled analysis of case–control studies of residential magnetic field exposure and childhood leukaemia, which highlights the diminishing value of conventional studies conducted after the early 1990s. It is argued that multiple-bias modelling should become part of the core training of anyone who will be entrusted with the analysis of observational data, and should become standard procedure when random error is not the only important source of uncertainty (as in meta-analysis and pooled analysis).

"Formulae can be applied in sequence to correct multiple biases… One can imagine each correction moving a step from the biased data back to the unbiased structure, as if hypothetically 'unwrapping the truth from the data package'."

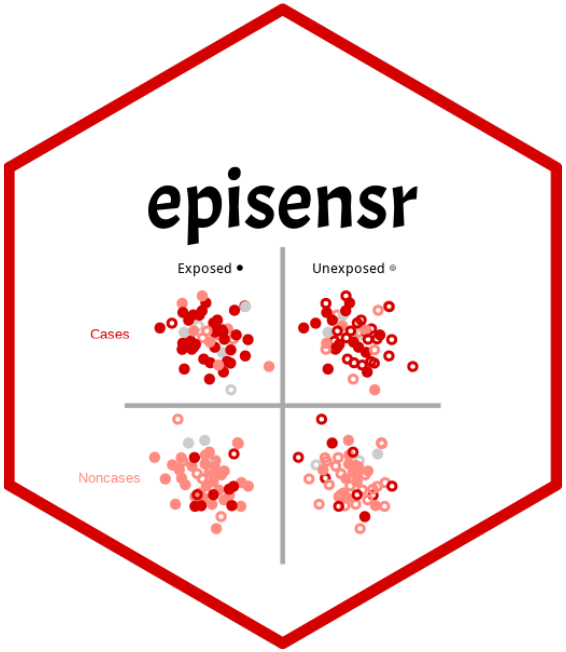
Multiple-bias modelling

"multiple-bias modelling should become part of the core training of anyone who will be entrusted with the analysis of observational data"

"For a sensitivity analysis to be useful, it is surely necessary that the assumptions which drive the different conclusions are sufficiently transparent that they can be communicated. Even to a statistical audience, Professor Greenland's bias models have taken several pages to explain…."

Greenland, 2005, discussion by Copas

Multiple-bias modelling



https://dhaine.github.io/episensr/articles/c_multiple_bias.html

Current state of bias analysis?

A systematic review of quantitative bias analysis applied to epidemiological research

Julie M Petersen [1]* Lynsie R Ranker [1] Ruby Barnard-Mayers [1] Richard F MacLehose [2] and Matthew P Fox [1,3]

[1]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA, [2]Division of Epidemiology and Community Health, University of Minnesota, School of Public Health, Minneapolis, MN, USA and [3]Department of Global Health, Boston University School of Public Health, Boston, MA, USA

*Corresponding author. 715 Albany St, Talbot Building Rm 324E, Boston, MA 02118, USA. E-mail: juliemo@bu.edu

Editorial decision 9 February 2021; Accepted 5 March 2021

- 238 quantitative bias analyses in epidemiologic literature 2006-2019
- 23% were an applied example in a methods paper or an entire paper devoted to the bias analysis
- 87% only modelled one bias

Toward a ”simpler” sensitivity analysis

ORIGINAL ARTICLE

OPEN

Sensitivity Analysis Without Assumptions

Peng Ding[a] *and Tyler J. VanderWeele*[b]

Abstract: Unmeasured confounding may undermine the validity of causal inference with observational studies. Sensitivity analysis provides an attractive way to partially circumvent this issue by assessing the potential influence of unmeasured confounding on causal conclusions. However, previ
and untestable
that is binary, c
sure and the co
confounder. W
confounder or
inequality such
inequality if an
effect estimate
implement and
our bounding f
more conservat
niques that do r
only the traditio
exposure on th
must satisfy bu
risks must satis
as a measure o
the outcome in

(*Epidemiology*

causal inferences even without full control of the confounders of the relationship between the exposure and outcome.

Sensitivity analysis plays a central role in assessing the influence of the unmeasured confounding on the causal con-

RESEARCH AND REPORTING METHODS **Annals of Internal Medicine**

Sensitivity Analysis in Observational Research: Introducing the E-Value

Tyler J. VanderWeele, PhD, and Peng Ding, PhD

Sensitivity analysis is useful in assessing how robust an association is to potential unmeasured or uncontrolled confounding. This article introduces a new measure called the "E-value," which is related to the evidence for causality in observational studies that are potentially subject to confounding. The E-value is defined as the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and the outcome to fully explain away a specific treatment–outcome association, conditional on the measured covariates. A large E-value implies that considerable unmeasured confounding would be needed to explain away an effect estimate. A small E-value implies little unmeasured confounding would be needed to explain away an effect estimate. The authors propose that in all observational studies intended to produce evidence for causality, the E-value be reported or some other sensitivity analysis be used. They suggest calculating the E-value for both the observed association estimate (after adjustments for measured confounders) and the limit of the confidence interval closest to the null. If this were to become standard practice, the ability of the scientific community to assess evidence from observational studies would improve considerably, and ultimately, science would be strengthened.

Ann Intern Med. 2017;167:268-274. doi:10.7326/M16-2607 **Annals.org**
For author affiliations, see end of text.
This article was published at Annals.org on 11 July 2017.

Framework using bounds

bias parameters X $\Rightarrow$

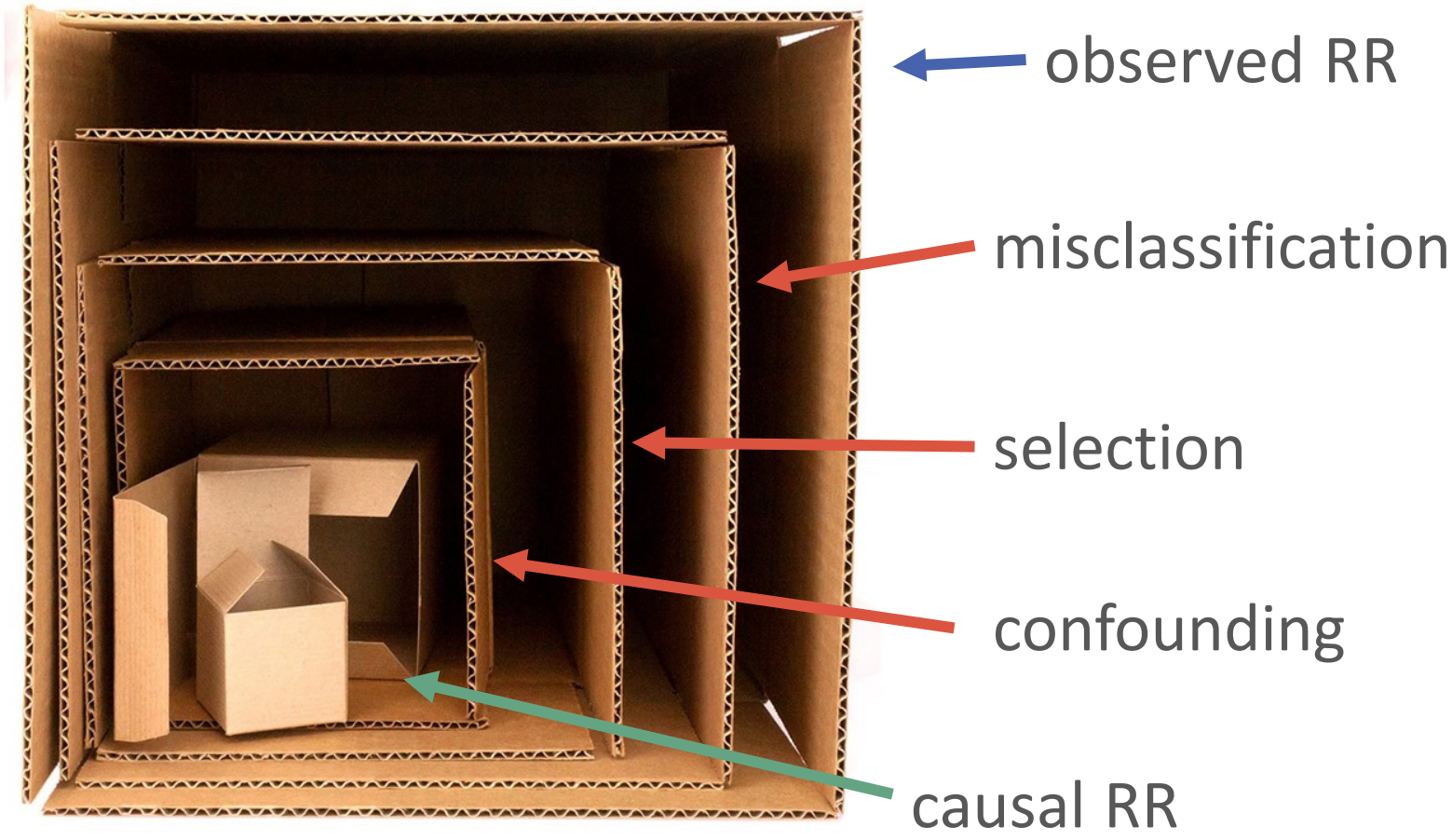
$$\text{observed risk ratio (RR)} \leq f(\text{X}) \times \text{causal RR} \Rightarrow$$

$$\text{causal RR} \geq {}^{\text{observed RR}}/_{f(\text{X})}$$

We can make statements of the form: "If the bias is of magnitude X, the true causal RR must be *at least* as large as ${}^{\text{observed RR}}/_{f(\text{X})}$."

The E-value inverts that statement and tells us what the minimum X would have to be for the observed RR to be compatible with a certain causal RR, usually the null.

Apply bounding framework to sequence of biases



https://amazon.com/Give-Gift-Frustration-Practical-Christmas/dp/B0779KYSLQ/

Notation

A: exposure (consider two values, 0 and 1)

Y: binary outcome

Y_a: counterfactual outcome under exposure $A = a$

C: measured covariates – best attempt to avoid bias

What we want: causal risk ratio

$$\text{RR}_{AY}^{\text{true}} = \frac{\Pr(Y_1 = 1 \mid c)}{\Pr(Y_0 = 1 \mid c)}$$

Notation

What we can estimate: observed risk ratio

$$\mathrm{RR}_{AY}^{\mathrm{obs}} = \frac{\Pr(Y^* = 1 \mid A = 1, S = 1, c)}{\Pr(Y^* = 1 \mid A = 0, S = 1, c)}$$

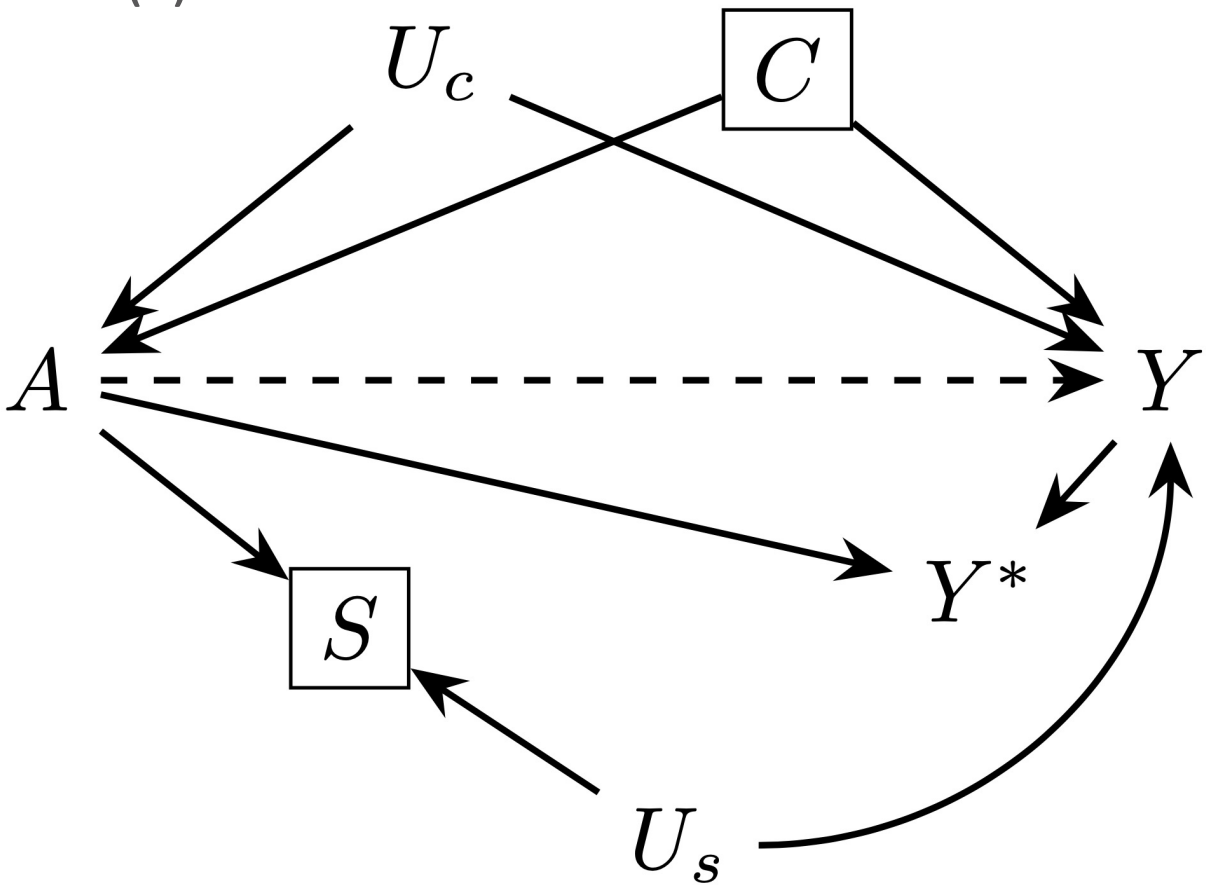
Y^*: observed Y

S: indicator of selection into the study, such that we only have data on the subset of the population for which $S = 1$

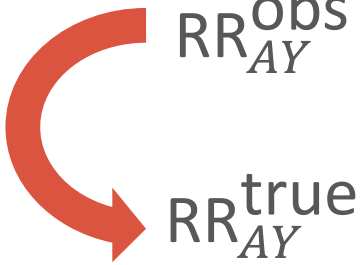
What's the problem?

U_c: unmeasured confounder(s)

U_s: unmeasured cause(s) of selection



Peeling off the layers



$$\mathrm{RR}_{AY}^{\mathrm{obs}} = \frac{\Pr(Y^* = 1 \mid A = 1, S = 1, c)}{\Pr(Y^* = 1 \mid A = 0, S = 1, c)}$$

$$\mathrm{RR}_{AY}^{\mathrm{true}} = \frac{\Pr(Y_1 = 1 \mid c)}{\Pr(Y_0 = 1 \mid c)}$$

Assume that there exist U_s and U_c such that

$Y \coprod S \mid A, C, U_s$ and $Y_a \coprod A \mid C, U_c$,

but that it is not necessarily true that

$Y \coprod S \mid A, C$ or $Y_a \coprod A \mid C$

Layer 1: outcome misclassification

$$\mathrm{RR}_{AY}^{\mathrm{obs}} \leq \max_{y} \frac{\Pr(Y^* = 1 \mid Y = y, A = 1, S = 1, c)}{\Pr(Y^* = 1 \mid Y = y, A = 0, S = 1, c)} \times \frac{\Pr(Y = 1 \mid A = 1, S = 1, c)}{\Pr(Y = 1 \mid A = 0, S = 1, c)}$$

misclassification bounding factor

$$\mathrm{BF}_m = \mathrm{RR}_{AY^* \mid y, S=1} = \max_{y} \frac{\Pr(Y^* = 1 \mid Y = y, A = 1, S = 1, c)}{\Pr(Y^* = 1 \mid Y = y, A = 0, S = 1, c)}$$

$$\mathrm{RR}_{AY}^{\mathrm{obs}} \leq \mathrm{BF}_m \times \frac{\Pr(Y = 1 \mid A = 1, S = 1, c)}{\Pr(Y = 1 \mid A = 0, S = 1, c)}$$

the part of the package with no misclassification

Layers 2 and 3: selection and confounding

$$\frac{\Pr(Y=1 \mid A=1, S=1, c)}{\Pr(Y=1 \mid A=0, S=1, c)} \leq \mathrm{BF}_S \times \frac{\Pr(Y=1 \mid A=1, c)}{\Pr(Y=1 \mid A=0, c)}$$

selection bias bounding factor

the part of the package with no selection bias

$$\frac{\Pr(Y=1 \mid A=1, c)}{\Pr(Y=1 \mid A=0, c)} \leq \mathrm{BF}_c \times \frac{\Pr(Y_1=1 \mid c)}{\Pr(Y_0=1 \mid c)}$$

confounding bounding factor

$\mathrm{RR}_{AY}^{\mathrm{true}}$

In summary

$$
\begin{aligned}
\mathrm{RR}_{AY}^{\mathrm{obs}} &\leq \mathrm{BF}_m \times \frac{\Pr(Y=1 \mid A=1, S=1, c)}{\Pr(Y=1 \mid A=0, S=1, c)} \\
&\leq \mathrm{BF}_m \times \mathrm{BF}_s \times \frac{\Pr(Y=1 \mid A=1, c)}{\Pr(Y=1 \mid A=0, c)} \\
&\leq \mathrm{BF}_m \times \mathrm{BF}_s \times \mathrm{BF}_c \times \frac{\Pr(Y_1=1 \mid c)}{\Pr(Y_0=1 \mid c)} \\
&= \mathrm{BF}_m \times \mathrm{BF}_s \times \mathrm{BF}_c \times \mathrm{RR}_{AY}^{\mathrm{true}}
\end{aligned}
$$

Proof?

ORIGINAL ARTICLE

OPEN

OPEN

Practice of Epidemiology

Simple Sensitivity Analysis for Differential Measurement Error

Tyler J. VanderWeele* and Yige Li

* Correspondence to Dr. Tyler J. VanderWeele, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115 (e-mail: tvanderw@hsph.harvard.edu).

Initially submitted November 1, 2018; accepted for publication May 17, 2019.

Sensitivity analysis results are given for differential measurement error of either the exposure or outcome. In the case of differential measurement error of the outcome, it is shown that the true effect of the exposure on the outcome on the risk ratio scale must be at least as large as the observed association between the exposure and the mismeasured outcome divided by the maximum strength of differential measurement error. This maximum strength of differential measurement error is itself assessed as the risk ratio of the controlled direct effect of the exposure on the mismeasured outcome not through the true outcome. In the case of differential measurement error of the exposure, under certain assumptions concerning classification probabilities, the true effect on the odds ratio scale of the exposure on the outcome must be at least as large as the observed odds ratio between the mismeasured exposure and the outcome divided by the maximum odds ratio of the effect of the outcome on mismeasured exposure conditional on the true exposure. The results can be immediately used to indicate the minimum strength of differential measurement error that would be needed to explain away an observed association between an exposure measurement and an outcome measurement for this to be solely due to measurement error.

bias analysis; differential; measurement error; misclassification; sensitivity analysis

esis of no causal effect. Although some sensitivity analysis

Layer 1: outcome misclassification

$$\text{BF}_m = \text{RR}_{AY^*|y,S=1} = \max_y \frac{\Pr(Y^* = 1 \mid Y = y, A = 1, S = 1, c)}{\Pr(Y^* = 1 \mid Y = y, A = 0, S = 1, c)}$$

How much greater is the outcome sensitivity among the exposed vs. the unexposed?

or

How much more likely are false-positives among the exposed vs. the unexposed?

This is specific to the selected group.

Layer 2: selection

$$\mathrm{BF}_S = \frac{\mathrm{RR}_{U_S Y|A=1} \times \mathrm{RR}_{SU_S|A=1}}{\mathrm{RR}_{U_S Y|A=1} + \mathrm{RR}_{SU_S|A=1} - 1} \times \frac{\mathrm{RR}_{U_S Y|A=0} \times \mathrm{RR}_{SU_S|A=0}}{\mathrm{RR}_{U_S Y|A=0} + \mathrm{RR}_{SU_S|A=0} - 1}$$

where

$$\mathrm{RR}_{U_S Y|A=a} = \frac{\max_u \Pr(Y = 1 \mid A = a, c, U_S = u)}{\min_u \Pr(Y = 1 \mid A = a, c, U_S = u)} \quad \text{for } a = 0,1$$

$$\mathrm{RR}_{SU_S|A=1} = \max_u \frac{\Pr(U_S = u \mid A = 1, S = 1, c)}{\Pr(U_S = u \mid A = 1, S = 0, c)}$$

$$\mathrm{RR}_{SU_S|A=0} = \max_u \frac{\Pr(U_S = u \mid A = 0, S = 0, c)}{\Pr(U_S = u \mid A = 0, S = 1, c)} .$$

Layer 2: selection

$$\mathrm{RR}_{U_S Y|A=a} = \frac{\max_u \Pr(Y = 1 \mid A = a, c, U_s = u)}{\min_u \Pr(Y = 1 \mid A = a, c, U_s = u)} \quad \text{for } a = 0,1$$

$$\mathrm{RR}_{SU_s|A=1} = \max_u \frac{\Pr(U_s = u \mid A = 1, S = 1, c)}{\Pr(U_s = u \mid A = 1, S = 0, c)}$$

$$\mathrm{RR}_{SU_s|A=0} = \max_u \frac{\Pr(U_s = u \mid A = 0, S = 0, c)}{\Pr(U_s = u \mid A = 0, S = 1, c)}$$

To what extent is the outcome risk increased by the unmeasured factor, within a single level of the exposure?

To what extent is some value of the unmeasured factor more prevalent among the selected compared to the non-selected group?

Layer 3: confounding

$$\mathrm{BF}_C = \frac{\mathrm{RR}_{AU_c} \times \mathrm{RR}_{U_cY}}{\mathrm{RR}_{AU_c} + \mathrm{RR}_{U_cY} - 1}$$

where

$$\mathrm{RR}_{AU_c} = \max_u \frac{\Pr(U_c = u \mid A = 1, c)}{\Pr(U_c = u \mid A = 0, c)}$$

$$\mathrm{RR}_{U_cY} = \max_a \frac{\max_u \Pr(Y = 1 \mid A = a, c, U_c = u)}{\min_u \Pr(Y = 1 \mid A = a, c, U_c = u)}$$

How much more prevalent is the confounder among the exposure than the unexposed?

How much extra risk of the outcome does the confounder confer among either the exposed or unexposed?

Peeling off the layers in a different order?

Assume that there exist U_s and U_c such that

$Y^* \coprod S \mid A, C, U_s$ and $Y_a \coprod A \mid C, U_c$.

This may be a more interpretable assumption if, for example, selection into the study is based on a factor related to the (mis)measured outcome, not the true outcome.

Recall the earlier assumptions were:

$Y \coprod S \mid A, C, U_s$ and $Y_a \coprod A \mid C, U_c$

Example: confounding and selection

Does infection to HIV in utero increase the risk of wasting?

Sample: Children of participants of a vitamin A supplementation trial in Zimbabwe

Exposure: Exposed vs. unexposed to HIV in utero

Outcome: Weight-for-length Z-score of < -2 as toddlers

$$RR_{AY}^{\text{obs}} = 6.75 \text{ (95\% CI, 2.79, 16.31)}$$

Omoni et al. 2017

Concern about confounding and selection bias

Observational study with respect to HIV infection, but no adjustment for confounders

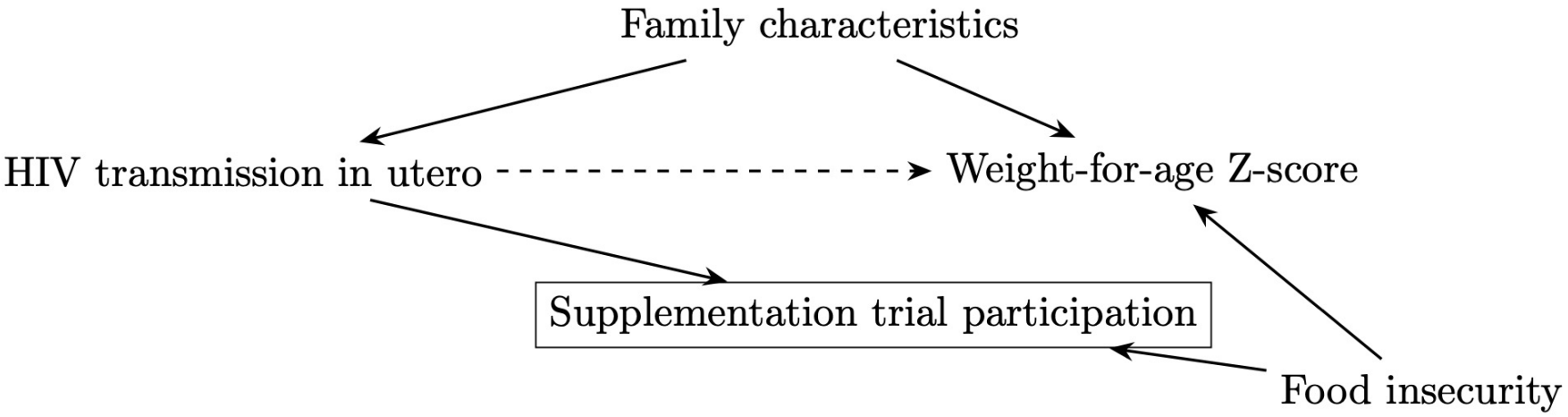
- The authors did not adjust for parity or marital status, though they report that primiparous women were less likely to have HIV, as were married women

- We may be concerned that children in single-parent households and those with more siblings are at higher risk of wasting

Enrollment occurred at delivery, after possible HIV exposure and transmission

- The choice of whether to participate in the trial could have been affected by HIV status as well as other factors (e.g., food insecurity), leading to selection bias if those factors affect future child growth

DAG


Family characteristics
HIV transmission in utero
Weight-for-age Z-score
Supplementation trial participation
Food insecurity

Plausible parameters: confounding

- The authors reported proportions of these characteristics stratified by exposure, which can aid in coming up with a reasonable value for RR_{AU_c}.

- Suppose we estimate that 3% of the women whose infants are infected with HIV are multiparous and unmarried, but that this is true of 7% of the women without HIV.
  - If this is the family situation with the largest disparity between exposure groups, then we can specify $RR_{AU} = 2.3$.

- Now suppose that children in these most precarious families have 2.5 times the risk of wasting than those in the least precarious, so that $RR_{UY} = 2.5$

Plausible parameters: selection

We assume that wasting is more likely in children of participants than of non-participants, both among those with HIV as well as those without

It turns out that if this is the case, we only need two of the parameters:

- Suppose that children of the most food-insecure mothers are 3 times as likely to have extremely low weight-for-length scores than the least likely group, so that $RR_{UY|A=1} = 3$

- Suppose mothers with HIV infection in the study compared to those not in the study are twice as likely to be food insecure, so that $RR_{SU_s|A=1} = 2$

Computing the bound

$$RR_{AY}^{obs} \leq B_s \times B_c \times RR_{\mathrm{AY}}^{true}$$

Plugging in the plausible parameters:

$$RR_{AY}^{obs} \leq \frac{3 \times 2}{3 + 2 - 1} \times \frac{2.3 \times 2.5}{2.3 + 2.5 - 1} \times RR_{\mathrm{AY}}^{true}$$

Plugging in the observed estimate:

$$6.75 \leq 2.27 \times RR_{\mathrm{AY}}^{true}$$

This amount of bias could not fully explain an observed risk ratio of 6.75.

We can also consider the lower limit of the confidence interval, 2.79.

In the worst case scenario, RR_{AY}^{obs} is still consistent with $2.79/2.27 = 1.23$

R package

```
library(EValue)

HIV_biases <- multi_bias(confounding(),
selection("general", "increased risk"))

HIV_biases


## The following arguments can be copied and
pasted into the multi_bound()
## function: RRAUc = , RRUcY = , RRUsYA1 = ,
RRSUsA1 =
```

R package

```
multi_bound(biases = HIV_biases,
            RRAUc = 2.3, RRUcY = 2.5,
            RRUsYA1 = 3, RRSUsA1 = 2)
## [1] 2.269737

summary(HIV_biases)
##           bias     output argument
## 1 confounding      RR_AUc     RRAUc
## 2 confounding      RR_UcY     RRUcY
## 3   selection RR_UsY|A=1   RRUsYA1
## 4   selection RR_SUs|A=1   RRSUsA1
```

Computing many bounds

The parameter values, though informed by background knowledge and data reported in the article, aren't known.

We should vary the parameters over a range of values.

```
# vary RRAUc from 1.25 to 3
sapply(seq(1.25, 3, by = .25), function(RRAUc) {
  multi_bound(biases = HIV_biases, RRAUc = RRAUc,
              RRUcY = 2.5, RRUsYA1 = 3, RRSUsA1 = 2)
  })
```

```
## [1] 1.704545 1.875000 2.019231 2.142857
2.250000 2.343750 2.426471 2.500000
```

```
# ?multi_evalue()
```

What about exposure misclassification?

The bound for exposure misclassification from VanderWeele & Li 2019 applies to the odds ratio, not the risk ratio, and the sensitivity parameters are also not risk ratios. That is,

$$\frac{\frac{\Pr(Y=1 \mid A^*=1,c)}{\Pr(Y=0 \mid A^*=1,c)}}{\frac{\Pr(Y=1 \mid A^*=0,c)}{\Pr(Y=0 \mid A^*=0,c)}} \leq \mathrm{BF}_m{}' \times \frac{\frac{\Pr(Y=1 \mid A=1,c)}{\Pr(Y=0 \mid A=1,c)}}{\frac{\Pr(Y=1 \mid A=0,c)}{\Pr(Y=0 \mid A=0,c)}}$$

for

$$\mathrm{BF}_m{}' = \mathrm{OR}_{YA^*|a} = \max\left(\frac{\frac{s'_1}{1-s'_1}}{\frac{s'_0}{1-s'_0}}, \frac{\frac{f'_1}{1-f'_1}}{\frac{f'_0}{1-f'_0}}, \frac{\frac{f'_1}{f'_0}}{\frac{1-s'_1}{1-s'_0}}, \frac{\frac{s'_1}{s'_0}}{\frac{1-f'_1}{1-f'_0}}\right)$$

where $s'_y = \Pr(A^*=1 \mid Y=y, A=1, c)$ and $f'_y = \Pr(A^*=1 \mid Y=y, A=0, c)$.

Which causes a problem

Applying this bound after factoring out selection bias, we would find that we are left with

$$\mathrm{RR}_{AY}^{\mathrm{obs}} \leq \mathrm{BF}_m' \times \mathrm{BF}_s \times \mathrm{BF}_c \times \mathrm{RR}_{AY}^{\mathrm{true}} \times \frac{\Pr(Y = 0 \mid A = 0, c)}{\Pr(Y = 0 \mid A = 1, c)} \times \frac{\Pr(Y = 0 \mid A^* = 1, c)}{\Pr(Y = 0 \mid A^* = 0, c)}$$

However, if the outcome is sufficiently rare in all strata, a simpler bound holds approximately

Result for exposure misclassification

Under the previous assumptions, with a rare outcome:

$$\mathrm{RR}_{AY}^{\mathrm{obs}\prime} = \frac{\Pr(Y = 1 \mid A^* = 1, S = 1, c)}{\Pr(Y = 1 \mid A^* = 0, S = 1, c)} \lesssim \mathrm{BF}_m{}' \times \mathrm{BF}_s \times \mathrm{BF}_c \times \mathrm{RR}_{AY}^{\mathrm{true}}$$

for $\mathrm{BF}_m{}' = \mathrm{OR}_{YA^*|a,S=1}$

with $s'_y = \Pr(A^* = 1 \mid Y = y, A = 1, S = 1, c)$ and

$f'_y = \Pr(A^* = 1 \mid Y = y, A = 0, S = 1, c)$

i.e., generally odds ratios for the sensitivities/false positive probabilities in the two outcome groups

Example with exposure misclassification

ORIGINAL ARTICLE

Multiple-bias Sensitivity Analysis Using Bounds

Louisa H. Smith,[a] *Maya B. Mathur,*[b] *and Tyler J. VanderWeele*[a,c]

Abstract: Confounding, selection bias, and measurement error are well-known sources of bias in epidemiologic research. Methods for assessing these biases have their own limitations. Many quantitative sensitivity analysis approaches consider each type of bias individually, although more complex approaches are harder to implement or require numerous assumptions. By failing to consider multiple biases at once, researchers can underestimate—or overestimate—their joint impact. We show that it is possible to bound the total composite bias owing to these three sources and to use that bound to assess the sensitivity of a risk ratio to any combination of these biases. We derive bounds for the total composite bias under a variety of scenarios, providing researchers with tools to assess their total potential impact. We apply this technique to a study where unmeasured confounding and selection bias are both concerns and to another study in which possible differential exposure misclassification and confounding are concerns. The approach we describe, though conservative, is easier to implement and makes simpler assumptions than quantitative bias analysis. We provide R functions to aid implementation.

Keywords: Bias analysis; Causal inference; Differential misclassification; Selection bias; Unmeasured confounding

(*Epidemiology* 2021;32: 625–634)

better sampling schemes, blinded outcome ascertainment, more extensive covariate measurements, and so on—other times confounding, selection bias, and measurement error are unavoidable. In such situations, our next best option is to assess the extent to which a given study's conclusions might be sensitive or robust to these biases and whether they threaten its conclusions. Often, however, this is limited to a few sentences in a discussion section qualitatively assessing the possibility of bias, sometimes appealing without quantitative justification to heuristics that may or may not hold true in a particular study.[1–3]

The weak uptake of quantitative bias analysis in epidemiology belies its long history. Over a half century ago, Cornfield and then Bross argued that the extent of possible bias was quantifiable based on observed data and possibly hypothetical quantities.[4–7] Attempts to generalize these results, as well as consider other biases, sometimes simultaneously with confounding, followed.[7–12] More recently, probabilistic bias analysis methods have been developed, allowing researchers to propose distributions for various bias parameters across multiple biases and to explore how various combinations of those parameters would affect their results.[13–18]

R examples

We walk you through how this simple sensitivity analysis for multiple biases works in R code here:

https://cran.r-project.org/web/packages/EValue/vignettes/multiple-bias.html

and share examples:

https://cran.r-project.org/web/packages/EValue/vignettes/multiple-bias-examples.html

References

louisa_h_smith@g.harvard.edu
@louisahsmith

Ding P, VanderWeele TJ. Sensitivity analysis without assumptions. *Epidemiology*. 2016;27(3):368-377. doi:10.1097/EDE.0000000000000457

Greenland S. Multiple-bias modelling for analysis of observational data (with discussion). *Journal of the Royal Statistical Society Series A (Statistics in Society)*. 2005;168(2):267-306. doi:10.1111/j.1467-985X.2004.00349.x

Maclure M, Schneeweiss S. Causation of bias: The episcope. *Epidemiology*. 2001;12(1):114-122. doi:10.1097/00001648-200101000-00019

Omoni AO, Ntozini R, Evans C, et al. Child growth according to maternal and child HIV status in Zimbabwe. *Pediatric Infectious Disease Journal*. 2017;36(9):869-876. doi:10.1097/INF.0000000000001574

Petersen JM, Ranker LR, Barnard-Mayers R, MacLehose RF, Fox MP. A systematic review of quantitative bias analysis applied to epidemiological research. *International Journal of Epidemiology*. 2021;(dyab061). doi:10.1093/ije/dyab061

Smith LH, VanderWeele TJ. Bounding bias due to selection. *Epidemiology*. 2019;30(4):509-516. doi:10.1097/EDE.0000000000001032

Smith LH, Mathur MB, VanderWeele TJ. Multiple-bias sensitivity analysis using bounds. *Epidemiology*. 2021;32(5). doi:10.1097/EDE.0000000000001380

VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Annals of Internal Medicine*. 2017;167(4):268-275. doi:10.7326/M16-2607

VanderWeele TJ, Li Y. Simple sensitivity analysis for differential measurement error. *American Journal of Epidemiology*. 2019;188(10):1823-1829. doi:10.1093/aje/kwz133