Lab 3: Sensitivity Analysis and Measurement Error

PHS2000B, Spring 2023

Susan Rattigan

February 6, 2023

Learning Objectives

1. **Sensitivity Analysis for Unmeasured Confounding**
   a. Why Sensitivity Analyses?
   b. Cornfield Conditions
   c. Bias Factors
   d. E-Values
   e. Bias Factors with Prevalence Specification
2. **Measurement Error**
   a. Overview
   b. Types of Measurement Error
   c. Sensitivity, Specificity, NPV, PPV
   d. Effects of Measurement Error

1. Sensitivity Analysis for Unmeasured Confounding

Strong Assumptions are Required for Causal Inference

In order for an observed association to represent a true causal effect, we (generally) require a specific **set of assumptions**:

1. **Exchangeability**
2. Consistency
3. Positivity

Exchangeability implies independence between the potential outcomes Y_a and the exposure/treatment received:

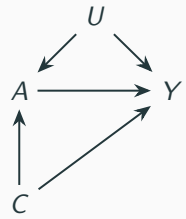
$$Y_a \perp\!\!\!\perp A$$

In observational research, we often assume that treatment/exposure groups are **conditionally exchangeable** within levels of a sufficient set of covariates C:

$$Y_a \perp\!\!\!\perp A|C$$

Beware the Open Backdoor Path

However, in practice, we can't be certain that conditional exchangeability holds. There may be **open backdoor path(s)** through unmeasured covariates U, even after conditioning on C.



In the presence of open backdoor path(s), the (conditional) estimate of the association between A and Y from the observed data provides a **biased estimate** of the true causal effect of A on Y.

$$Pr(Y_{a=1} = 1|c) - Pr(Y_{a=0} = 1|c) \neq Pr(Y = 1|A = 1, c) - Pr(Y = 1|A = 0, c)$$

Sensitivity Analyses for Unmeasured Confounding

So . . . **what do we do?!?** $\Longrightarrow$ **sensitivity analyses for unmeasured confounding!**

There are many different sensitivity analysis approaches.[1] In PHS 2000B, we have focused on **quantitative techniques** which allow us to develop **reasonable bounds on the amount of bias** that an unmeasured confounder U must produce to explain away an observed result. Specifically, we have covered the following methods:

1. **Cornfield Conditions**
2. **Bias Factors**
3. **E-values**
4. **Bias Factors with Prevalence Specification**

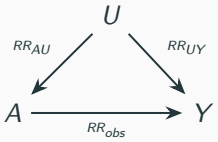
[1] Other sensitivity analysis approaches are discussed in the Supplemental Slides.

Cornfield Conditions

Cornfield Conditions[2]

In order for a single confounder to explain away an observed association, the association between i) the confounder and the exposure (RR_{AU}) and ii) the confounder and the outcome (RR_{UY}) must be at least as large as the observed association between the exposure and outcome (RR_{obs}):

$$RR_{AU} \geq RR_{obs} \quad \text{and} \quad RR_{UY} \geq RR_{obs}$$



[2]Cornfield J et al. Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions. JNCI. 1959;22:173- 203.

Cornfield Conditions: Assumptions

Cornfield's Conditions provide an easy and intuitive approach to assess whether an observed effect may be completely explained away by an unmeasured confounder. However, these conditions require several **restrictive assumptions**:

1. **Only one unmeasured confounder** U
2. **No effect heterogeneity between exposure** A **and confounder** U
3. **Exposure** A**, outcome** Y**, and confounder** U **are all binary**

Alternatively, we can generate *more nuanced bounds that make no assumptions* about the structure of the confounding bias and the nature of the variables under consideration.

Bias Factor

The **observed risk ratio** (RR_{obs}) can be expressed using: i) the **true causal risk ratio**(RR_{true}) and ii) a **bias factor**(BF) which captures unmeasured confounding.

Bias Factor: Equation 1[3]

$$RR_{obs} = BF \times RR_{true} \quad \Longleftrightarrow \quad BF = \frac{RR_{obs}}{RR_{true}}$$

[3] Miettinen OS. Components of the crude risk ratio. Am J Epidemiol. 1972;96:168-172.

Bias Factor

Bounds for the bias factor can be represented by just two parameters: RR_{UY} and RR_{AU}

Bias Factor: Equation 2[4]

$$BF \leq \frac{RR_{UY} \times RR_{AU}}{RR_{UY} + RR_{AU} - 1}$$

[4] Ding P, VanderWeele TJ. Sensitivity Analysis Without Assumptions. Epidemiology. 2016;27(3):368-77.

Bias Factor

RR_{UY} is the **largest possible effect of the unmeasured confounder on the outcome** within either stratum of the exposure:

$$RR_{UY} = \max\left(\frac{\max_u P(Y=1|A=0,c,u)}{\min_u P(Y=1|A=0,c,u)}, \frac{\max_u P(Y=1|A=1,c,u)}{\min_u P(Y=1|A=1,c,u)}\right)$$

RR_{AU} is the **maximum risk ratio for a single value of the unmeasured confounder** comparing the two exposure levels:

$$RR_{AU} = \max_u \frac{P(u|A=1,c)}{P(u|A=0,c)}$$

Bias Factor: Summary

A **general road map** for developing bounds for confounding bias using the bias factor approach can be summarized in 4 steps:

- **Step 1:** Calculate the adjusted risk ratio $RR_{obs,c}$ relating the exposure to the outcome after conditioning on all measured confounders C.
- **Step 2:** Using subject matter expertise, come up with suitable estimates of RR_{UY} and RR_{AU} for the most likely unmeasured confounders U. Alternatively, if not much is known about the structure of U, explore a range of possible values.
- **Step 3:** Calculate the largest bias factor that could result from these values of RR_{UY} and RR_{AU} using bias factor equation 2.
- **Step 4:** Estimate potential values for the true risk ratio, $RR_{true,c}$, using bias factor equation 1.

E-value

It may be helpful to consider the special case in which RR_{UY} and RR_{AU} are equivalent (i.e., $RR_{UY} = RR_{AU} = RR_{eq}$).

The bias factor then simplifies to:

$$BF \leq \frac{RR_{eq}^2}{2RR_{eq} - 1}$$

Using bias factor equation 1 (and doing some algebra), we can now define the E-value for unmeasured confounding:[5]

E-value for Unmeasured Confounding

$$\boxed{\text{E-value} = RR_{eq} = RR_{obs} + \sqrt{RR_{obs}(RR_{obs} - 1)}}$$

[5] VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017;167(4):268-274. A full proof of this result is provided in the Supplemental Slides.

E-value: Protective Associations

To estimate an E-value for a protective association:

- Take the inverse of the protective RR
- Plug this number into the E-value equation
- The interpretation remains the same as previously discussed
- **Example:** $RR_{obs} = 0.75 \implies RR_{obs}^{-1} = 1/0.75 = 1.33$

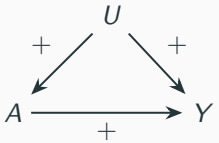
$$\begin{aligned} \text{E-value} &= RR_{obs}^{-1} + \sqrt{RR_{obs}^{-1}(RR_{obs}^{-1} - 1)} \\ &= 1.33 + \sqrt{1.33(1.33 - 1)} \\ &= \mathbf{2} \end{aligned}$$

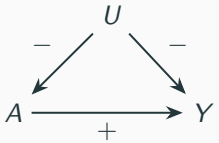
Bias Factor: Protective associations

To estimate a bias factor that would fully explain an association, test the rules of confounding:

$RR_{obs} > 1$

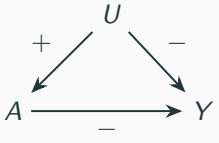
RR_{AU} and RR_{UY} that could explain away RR_{obs} must indicate the A-Y relationship is too high

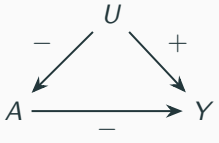




$RR_{obs} < 1$

RR_{AU} and RR_{UY} that could explain away RR_{obs} must indicate the A-Y relationship is too low





Bias Factor with Prevalence Specification

We can calculate the **exact amount of bias** (rather than bounds) if we are willing to make a few additional assumptions and specify the prevalence of the unmeasured confounder.

- Assume $Y_a \perp\!\!\!\perp A \mid (C, U)$ and U is a single binary confounder with the same risk ratio for Y among both the exposed and unexposed:[6]

Bias Factor with Prevalence Specification

$$B_{mult}(c) = \frac{1 + (\gamma - 1)P(U = 1 \mid a_1, c)}{1 + (\gamma - 1)P(U = 1 \mid a_0, c)}$$

- $\gamma = \frac{P(Y=1|U=1,a,c)}{P(Y=1|U=0,a,c)}$ = risk ratio between U and Y, homogeneous with respect to exposure
- $P(U = 1 \mid a_1, c)$ = prevalence of confounder U among the exposed, conditional on C
- $P(U = 1 \mid a_0, c)$ = prevalence of confounder U among the unexposed, conditional on C

[6] Schlesselman JJ. Assessing effects of confounding variables. Am J Epidemiol. 1978;108:3-8.

Other Effect Measures

VanderWeele and Ding (2017) generated approximations which can be used to convert other effect measures to risk ratios. These approximations can then be used in the previous equations.

Measure	Conditions	Point Estimate Approximation	Confidence Interval Approximation
OR	Prevalence of $Y < 15\%$	$RR = OR$	same as point estimate
OR	Prevalence of $Y > 15\%$	$RR = \sqrt{OR}$	same as point estimate
HR	Prevalence of $Y < 15\%$	$RR = HR$	same as point estimate
HR	Prevalence of $Y > 15\%$	$RR = \frac{1-0.5^{\sqrt{HR}}}{1-0.5^{1/\sqrt{HR}}}$	same as point estimate
IRR	In all cases	$RR = IRR$	same as point estimate
Cohen's d	In all cases	$RR = e^{0.91d}$	$(e^{0.91d-1.78s_d}, e^{0.91d+1.78s_d})$
RD	p_1 and p_0 between 0.2 - 0.8	$RR = e^{0.91RD}$	
RD	p_1 and p_0 not between 0.2 - 0.8	$RR = \frac{p_1}{p_0}$	... complicated

Strengths and Limitations

	Cornfield Conditions	Bias Factor	E-value	Bias Factor w/ Prevalence
Strengths	Intuitive, easy. Can be helpful in assessing if an effect could be completely explained away.	No structural assumptions. Works for more than one U. Provides bounds for bias.	No structural assumptions. Works for more than one U. Provides bounds for bias. Single number summary. Standardized metric.	Results are exact, not conservative. Better if confounder is rare. Helpful if calculating corrected estimates.
Limitations	Assumes one binary U, binary A and Y, no effect heterogeneity for effect of A on Y by U. Very conservative.	Calculated under worst-case confounder prevalence.	Calculated under worst-case confounder prevalence. The single number summary assumes $RR_{AU} = RR_{UY}$.	Must specify prevalences of U. Assumes same RR for U and Y among exposed and unexposed.

2. Measurement Error

Overview

Measurement error is a ubiquitous problem which can occur for a myriad of reasons, including intentional or unintentional misreporting, accidental miscodings, faulty equipment, etc.

Possible consequences of measurement error:

- Biased estimates
- Incorrect confidence intervals
- Lower than expected power
- Effects that appear heterogeneous when they truly are not

Possible solutions:

- Argue that bias is towards the null and may not affect your conclusion. But remember that bias towards the null is not guaranteed, and is only expected in certain cases!
- Perform a sensitivity analysis
- Correct effect estimates using validation study data

Types of Measurement Error

Let A and Y be correctly measured variables and A^* and Y^* be their mismeasured versions. We have the following **types of measurement error:**[7]

- **Non-differential**: given the true value A, the distribution of the mismeasured variable A^* is the same no matter the value of Y: $\Pr(A^*|A, Y) = \Pr(A^*|A)$.
- **Differential**: non-differential measurement error does not hold, i.e., Y provides additional information about the distribution of A^*, even if we know the value of A.
- **Independent**: the error in one variable doesn't tell you anything about the error in another variable: $\Pr(A^*|A, Y, Y^*) = \Pr(A^*|A, Y)$.
- **Dependent**: independent measurement error does not hold, i.e., knowing the value of mismeasured Y^* tells you something about A^*, conditional on both true values.

[7] Differential vs. non-differential applies whenever one or more variables are mismeasured. By contrast, independent vs. dependent only applies when there are at least two mismeasured variables.

Types of Measurement Error

Question: What type of measurement error is represented in this DAG?

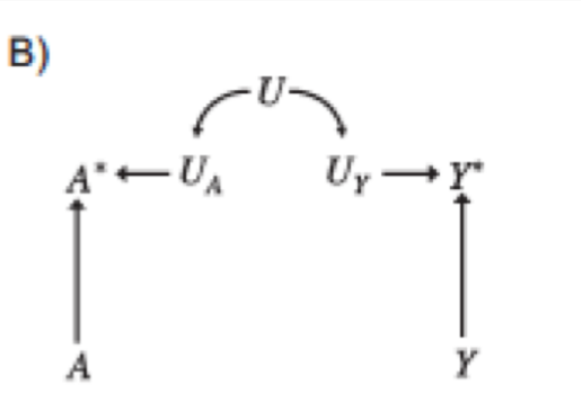


Figure 1: Types of Measurement Error, DAG B[8]

[8] Hernán MA, Cole SR. 2009;170(8):959-962. VanderWeele TJ, Hernán MA. 2012;175(12):1303-10.

Sensitivity and Specificity

When measuring a binary variable, we are concerned about two types of error: **false positives** and **false negatives**. Sensitivity and specificity quantify the probability of each of these.

- **Sensitivity (Se):** $\mathbf{1 - \text{false negative probability} = \Pr(X^* = 1 | X = 1)}$
  A highly sensitive measure is one that picks up on a lot (like a sensitive person) – it will pick up most of the true positives, but some false positives too.

- **Specificity (Sp):** $\mathbf{1 - \text{false positive probability} = \Pr(X^* = 0 | X = 0)}$
  A highly specific measure is very particular (like a person with very specific tastes) – it successfully excludes most of the true negatives but will also fail to capture some true positives.

Positive and Negative Predictive Values

Positive and negative predictive values answer the question: "If I know the *measured* value of a given variable, what is the probability that that measurement is correct?"

- **Positive Predictive Value (PPV):** $\mathbf{Pr}(\boldsymbol{X} = 1|\boldsymbol{X}^* = 1) = \frac{\boldsymbol{Se}\cdot\boldsymbol{p_x}}{\boldsymbol{Se}\cdot\boldsymbol{p_x}+(1-\boldsymbol{Sp})\cdot(1-\boldsymbol{p_x})}$
- **Negative Predictive Value (NPV):** $\mathbf{Pr}(\boldsymbol{X} = 0|\boldsymbol{X}^* = 0) = \frac{\boldsymbol{Sp}\cdot(1-\boldsymbol{p_x})}{(1-\boldsymbol{Se})\cdot\boldsymbol{p_x}+\boldsymbol{Sp}\cdot(1-\boldsymbol{p_x}))}$

These values depend on both the sensitivity and specificity of the measurement and on the prevalence of X, denoted p_x.

- If you're testing for a rare disease, even with a highly specific test, there's a good probability that any given positive test result is a false positive. To understand why, let's look at a brief example.

Effects of Measurement Error

In general, **differential measurement error may cause bias in any direction** (either towards or away from the null).

The **effects of non-differential measurement error vary depending on:**[9]

- Whether the variable is binary, continuous, or categorical
- Whether the variable is an exposure, outcome, or confounder

[9] The following slides summarize the effects of non-differential measurement errror and are adapted from Schwartz, G. (2017). Magnum Opus. Cambridge, MA: PHS.

Effect of Non-differential Measurement Error in an Exposure A

Type	Direction	Sign	Correctable?	Notes
Binary	Towards null	Correct as long as $\text{Sp} + \text{Se} > 1$	Yes (with Se and Sp)	If $\text{Se} + \text{Sp} < 1$ then effect may go beyond the null
Categorical	Depends	Correct if we can assume monotonicity	Yes (with all inter-category Se and Sp values)	Monotonicity: $E[Y \mid A]$ increasing in A and $E[A^* \mid A]$ increasing in A
Continuous	Towards null	Correct if we can assume monotonicity	Yes (through regression calibration, Rosner or Carroll method)	Beware of overcorrecting if gold standard is an approximation

Effect of Non-differential Measurement Error in an Outcome Y

Type	Direction	Sign	Correctable?	Notes
Binary	Towards null	Correct as long as Sp + Se > 1	Yes (with Se and Sp, or just Sp for risk ratios)	RD = RD* / (Se + Sp - 1) RR = (p_1^* - 1 + Sp) / (p_0^* - 1 + Sp) Can use PPV/NPV weighting in logit models
Continuous	No bias!	Correct sign!	N/A	Assume $Y^* = Y + \epsilon$ and then the error is incorporated into the error term on the right-hand side of the regression

Effect of Non-differential Measurement Error in a Confounder C

Type	Conditions	Notes
Binary	If the effect of C on Y is monotonic, partial control	Monotonicity conditions sufficient but not necessary for partial control
Categorical	If the effect of C on Y is monotonic and the effect of C on A is monotonic, partial control	
Continuous	Can correct coefficient on A with proportion of variance in C^* explained by true C, $\lambda = \text{Var}(C \mid A, X)/\text{Var}(C^* \mid A, X)$	Coefficients from linear or logistic regression with rare outcome can be corrected

Supplemental Slides

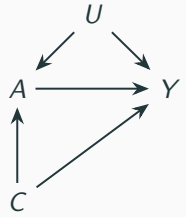
Bias Factor

We can calculate RR_{obs} as the **observed risk ratio conditional on C**:

$$RR_{obs,c} = \frac{P(Y = 1 \mid A = 1, c)}{P(Y = 1 \mid A = 0, c)}$$

But RR_{true}, **the true effect of A on Y**, is identified if and only if we condition on C and U:

$$RR_{true,c,u} = \frac{P(Y_1 \mid c, u)}{P(Y_0 \mid c, u)} = \frac{P(Y = 1 \mid A = 1, c, u)}{P(Y = 1 \mid A = 0, c, u)}$$



Bias Factor

We want to use $RR_{obs,c}$ and $RR_{true,c,u}$ to develop an **expression for the bias factor.** However, we can't yet, because the former is conditional on C while the latter is conditional on C and U.

Let's use the basic rules of probability to obtain $RR_{true,c}$, i.e., **the true risk ratio conditional on C alone,** by standardizing (taking the weighted sum over all values of U):

$$RR_{true,c} = \frac{\sum_u P(Y = 1 \mid A = 1, c, u)P(u \mid c)}{\sum_u P(Y = 1 \mid A = 0, c, u)P(u \mid c)}$$

The **bias factor** BF is then:

$$BF = \frac{RR_{obs,c}}{RR_{true,c}} = \frac{\frac{P(Y=1|A=1,c)}{P(Y=1|A=0,c)}}{\frac{\sum_u P(Y=1|A=1,c,u)P(u|c)}{\sum_u P(Y=1|A=0,c,u)P(u|c)}}$$

E-value Proof

Recall that the E-value refers to the special case in which $RR_{UY} = RR_{AU} = RR_{eq}$. The bias factor then simplifies to:

$$BF \leq \frac{RR_{eq}^2}{2RR_{eq} - 1}$$

Using bias factor equation 1, we can substitute in this value for BF and 1 for the true risk ratio, RR_{true}. We obtain:

$$\begin{aligned}
RR_{obs} &= BF \times RR_{true} \\
RR_{obs} &= \frac{RR_{eq}^2}{2RR_{eq} - 1} \times (1) \\
RR_{obs}(2RR_{eq} - 1) &= RR_{eq}^2 \\
2RR_{obs}RR_{eq} - RR_{obs} &= RR_{eq}^2 \\
0 &= RR_{eq}^2 - 2RR_{obs}RR_{eq} + RR_{obs}
\end{aligned}$$

E-value Proof

Now we will need to use the quadratic equation to further simplify this expression:

$$\begin{aligned} RR_{eq} &= \frac{2RR_{obs} \pm \sqrt{4RR_{obs}^2 - 4RR_{obs}}}{2} \\ &= \frac{2RR_{obs} \pm 2\sqrt{RR_{obs}^2 - RR_{obs}}}{2} \\ &= \boldsymbol{RR_{obs} \pm \sqrt{RR_{obs}(RR_{obs} - 1)}} \end{aligned}$$

The positive root $RR_{obs} + \sqrt{RR_{obs}(RR_{obs} - 1)}$ is the only one that makes sense in this scenario, so we define it as the E-value.

Other E-values

Although we reviewed an E-value for unmeasured confounding, **many other sources of bias** may affect the validity of our inferences. Fortunately, similar approaches can be used to develop E-values for other forms of bias and even for multiple forms of bias:

- **E-value for selection bias**: Smith and VanderWeele (2019)
- **E-value for differential measurement error**: VanderWeele and Li (2019)
- **E-value for unmeasured confounding, selection bias, and measurement error**: Smith and VanderWeele (2021)
- **E-value for unmeasured confounding in meta-analyses**: Mathur and VanderWeele (2020a) and Mathur and VanderWeele (2020b)

Other Sensitivity Analyses for Unmeasured Confounding

There are also **other sensitivity analysis approaches** that can be used to assess potential bias resulting from unmeasured confounding. For your reference, here are some additional methods that can be considered:

- **Negative outcome and exposure controls**[10]
- **Monte Carlo simulation-based sensitivity analyses**[11]
- **Leverage populations with different confounding structures or methods which employ alternative assumptions** (e.g., instrumental variable estimation, which we will discuss later this spring!)

[10] Lipsitch M et al. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiol. 2010;21(3):383-8.
[11] Banack HR et al. Monte Carlo Simulaiton Approaches for Quantitative Bias Analysis: A Tutorial. Epidemol Reviews. 2021;43(1):106-117.

Measurement Error Correction

There are many different formulae which can be used to perform measurement error correction. Here, we will focus on *a few of these formulae* which are commonly used in simple measurement error settings. See the lecture slides for additional formulae.

Prevalence: If we know sensitivity and specificity, we can use this information to **correct a mismeasured prevalence estimate** p_{X^*} and obtain an estimate of the true prevalence p_X:

$$p_X = (p_{X^*} - 1 + Sp)/(Se + Sp - 1)$$

Measurement Error Correction

RD, RR, or OR: If we have a binary exposure and outcome, one of which has **non-differential measurement error** with respect to the other, we can calculate corrected risk ratios and risk differences, or correct all the cells of a 2x2 table.

For example, to calculate a corrected risk difference:

$$RD = \frac{A_1^* - FpM_1}{N_1 - Fp(M_1 + M_0)} - \frac{A_0^* - FnM_1}{N_0 - Fn(M_1 + M_0)}$$

Where Fn is the false negative probability, Fp is the false positive probability, and:

	A^*	1	0	
Y	1	A_1^*	A_0^*	M_1
	0	B_1^*	B_0^*	M_0
		N_1	N_0	

Measurement Error in Continuous Variables

One perplexing feature of measurement error is that non-differential measurement error in a continuous exposure **causes bias towards the null**, but non-differential measurement error in a continuous outcome **causes no bias!**

- Note that in either case, measurement error **increases the variance** of your estimates, making them less precise.

To review why this is the case, we'll do some **mathematical derivations.** Then, we'll **simulate** continuous exposure and outcome data with measurement error to help us visualize this phenomenon.

Measurement Error in a Continuous Outcome

We'll start with **non-differential measurement error in a continuous outcome**, under the assumption that $Y^* = Y + \epsilon$ where $Y \perp\!\!\!\perp \epsilon$.

Recall that the regression coefficient from a regression of Y on A is:

$$\beta_1 = \frac{\text{Cov}(Y, A)}{\text{Var}(A)}$$

When we regress Y^* on A instead, we obtain:

$$\beta_1^* = \frac{\text{Cov}(Y + \epsilon, A)}{\text{Var}(A)} = \frac{\text{Cov}(Y, A) + \text{Cov}(\epsilon, A)}{\text{Var}(A)} = \frac{\text{Cov}(Y, A)}{\text{Var}(A)} = \beta_1$$

Because $A \perp\!\!\!\perp \epsilon \implies \text{Cov}(\epsilon, A) = 0$. We obtain an **unbiased estimate** of the effect!

Measurement Error in a Continuous Exposure

Now we'll evaluate the effects of **non-differential measurement error in a continuous exposure**, assuming $A^* = A + \epsilon$ where $A \perp\!\!\!\perp \epsilon$.

Recall that the regression coefficient from a regression of Y on A is:

$$\beta_1 = \frac{\text{Cov}(Y, A)}{\text{Var}(A)}$$

When we regress Y on A^* instead, we obtain:

$$\beta_1^* = \frac{\text{Cov}(Y, A + \epsilon)}{\text{Var}(A + \epsilon)} = \frac{\text{Cov}(Y, A) + \text{Cov}(Y, \epsilon)}{\text{Var}(A) + \text{Var}(\epsilon) + 2\text{Cov}(A, \epsilon)} = \frac{\text{Cov}(Y, A)}{\text{Var}(A) + \text{Var}(\epsilon)}$$

Because $A \perp\!\!\!\perp \epsilon$ and $Y \perp\!\!\!\perp \epsilon \implies \text{Cov}(A, \epsilon) = 0$ and $\text{Cov}(Y, \epsilon) = 0$. Therefore, β_1^* is *smaller* in absolute magnitude than the true effect $\beta_1 \implies$ **bias towards the null!**