

DIRECTED ACYCLIC GRAPHS

Phần 2: Overadjustment,
Unnecessary Adjustment &
Collider-Stratification Bias

Khương Quỳnh Long
Đại Học Y Tế Công Cộng, 22/1/2019

1. Overadjustment



Định nghĩa Overadjustment

- “Statistical adjustment by an **excessive number of variables**, uninformed by substantive knowledge (e.g., lacking coherence with biological, clinical, epidemiological, or social knowledge) can obscure a true effect or create an apparent effect when none exists”.

-- A Dictionary of Epidemiology, page 207 --

Định nghĩa Overadjustment

- “Intermediate variables, if controlled for in an analysis, would usually bias results **towards the null**...such **control of an intermediate** may be viewed as a form of overadjustment.”

-- Modern epidemiology - Rothman & Greenland 2008 --

Overadjustment

- Định nghĩa chưa thống nhất
- Mang ý nghĩa miêu tả tình huống hiệu chỉnh cho biến số mà
 - Gia tăng net **bias**
 - Giảm **precision** mà không ảnh hưởng tới **bias**
- Áp dụng cho trường hợp regression adjustment, stratification, or restriction

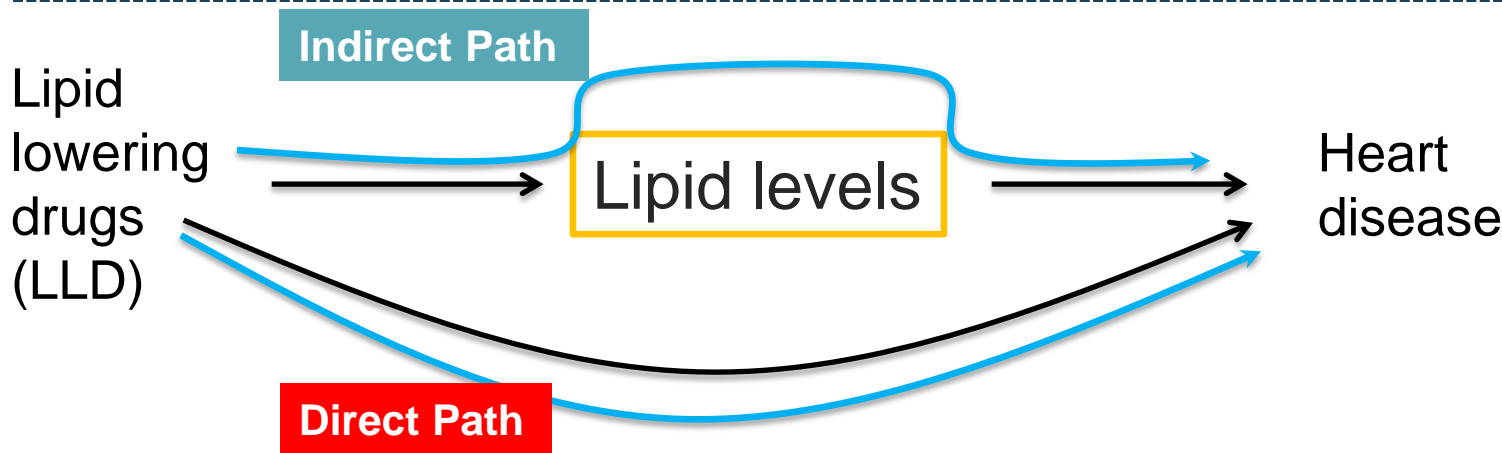
Mục tiêu

- Phân biệt **overadjustment** và **unnecessary adjustment** sử dụng DAGs
- Trình bày một số tình huống **overadjustment** thường gặp

Định nghĩa

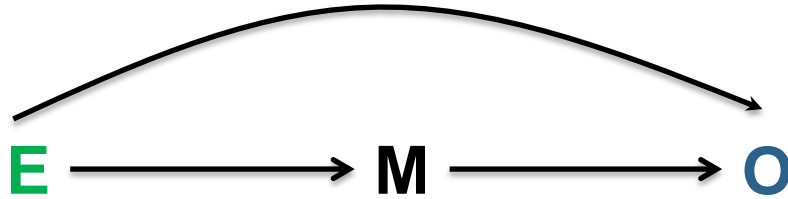
- **Overadjustment**
 - Hiệu chỉnh cho biến trung gian (intermediate)
 - Hoặc **descending proxy** của intermediate
- **Unnecessary adjustment**
 - Hiệu chỉnh cho biến số mà không ảnh hưởng tới **bias**, nhưng có thể ảnh hưởng tới **precision**

Direct and Indirect Effects: DAGs



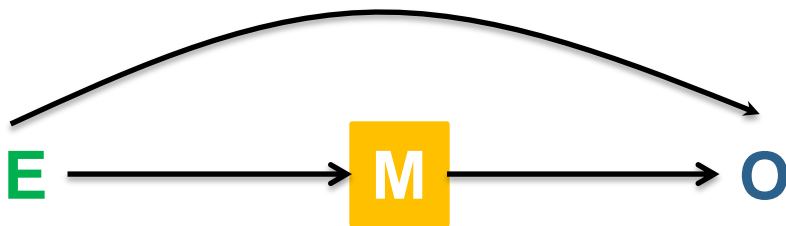
- **Total effect:** $HD = \alpha + \beta_1 LLD$
- **Direct** và **indirect effect:** $HD = \alpha + \beta_2 LLD + \beta_3 \text{lipids}$

DAG 1: Control cho intermediate



- Trường hợp đơn giản nhất của Overadjustment
- M chính là Intermediate giữa tác động của E lên O

DAG 1: Hiệu chỉnh cho intermediate

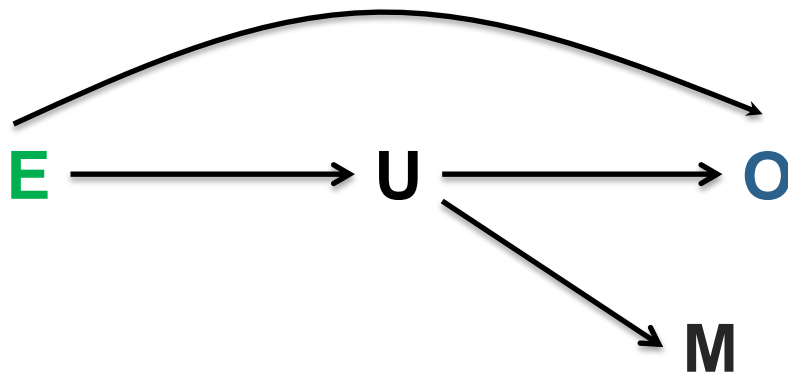


- Ước lượng **TOTAL** effect của E lên O nếu bỏ qua M
- Hiệu chỉnh cho M → chỉ ước lượng **DIRECT** effect cho E lên O

DAG 2: Control cho Descending proxy

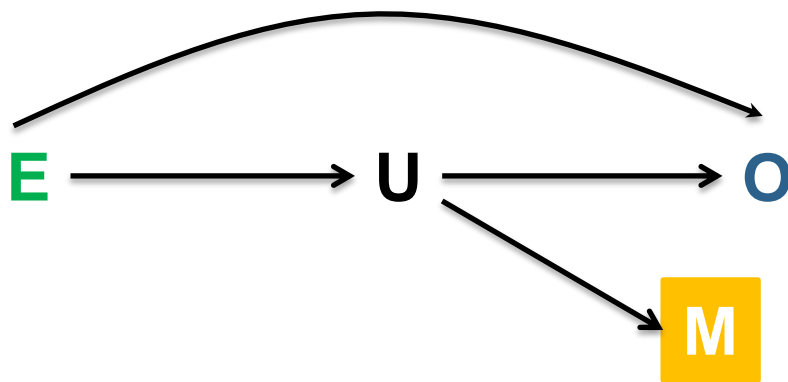
- Quay lại ví dụ 1 (phần confounding)
- Ảnh hưởng của **hút thuốc lá** tới **sảy thai**
- Cần hiệu chỉnh cho **tiền sử sảy thai** hay không?

DAG 2: Control cho Descending proxy



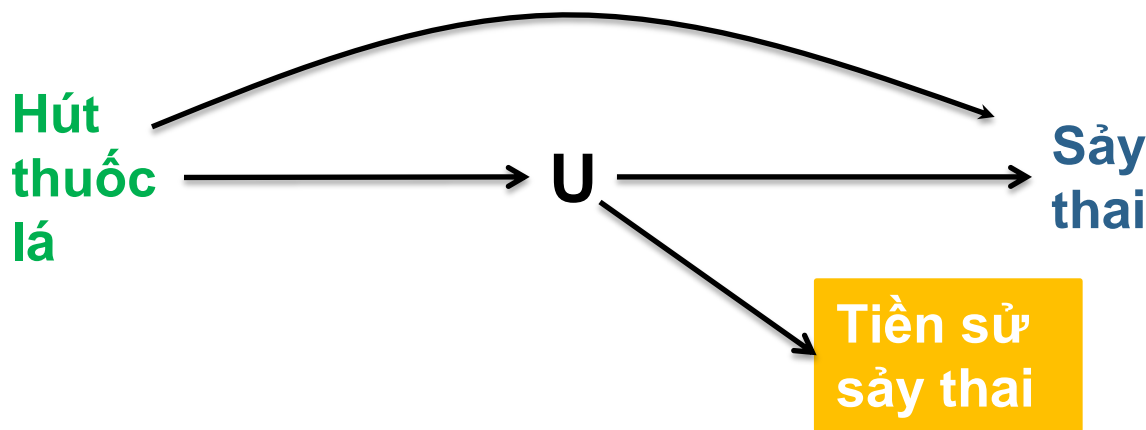
- Tương tự, ước lượng **TOTAL** effect của E lên O nếu bỏ qua M

DAG 2: Control cho Descending proxy



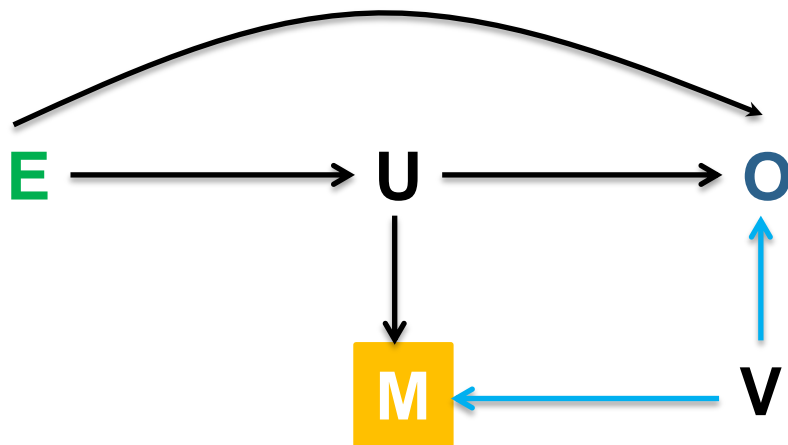
- Tương tự, ước lượng **TOTAL** effect của E lên O nếu bỏ qua M
- Hiệu chỉnh cho M ảnh hưởng tới **TOTAL** effect của E lên O. Nhưng ít hiệu ứng Null-bias như DAG1 (vì chỉ ảnh hưởng 1 phần)

DAG 2: Control cho Descending proxy



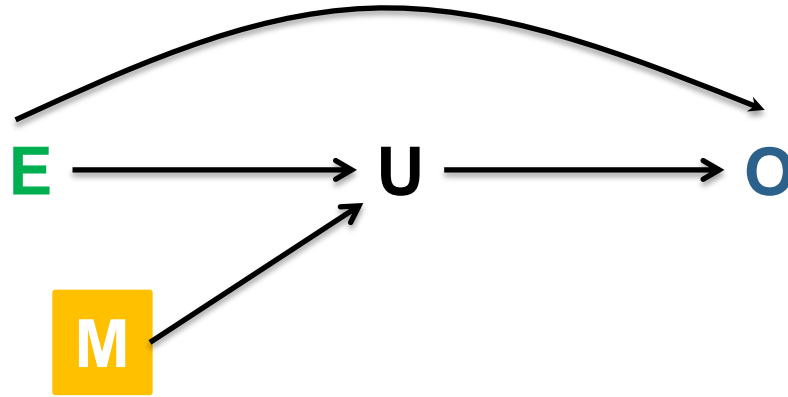
- U đại diện cho biến ẩn (unmeasured), ví dụ bất thường nội mạc tử cung...
- Cần hiệu chỉnh cho **tiền sử sảy thai** hay không?

DAG 3: Control cho Descending proxy



- Hiệu chỉnh cho descending proxy của intermediate dễ dẫn tới tình huống **collider-stratification bias** (giới thiệu phần sau)

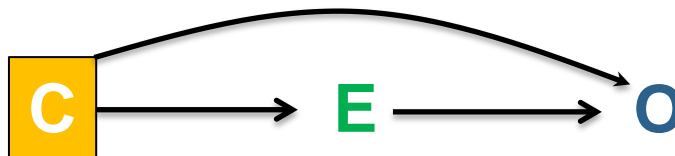
DAG 4: Control cho **ASCENDING** proxy



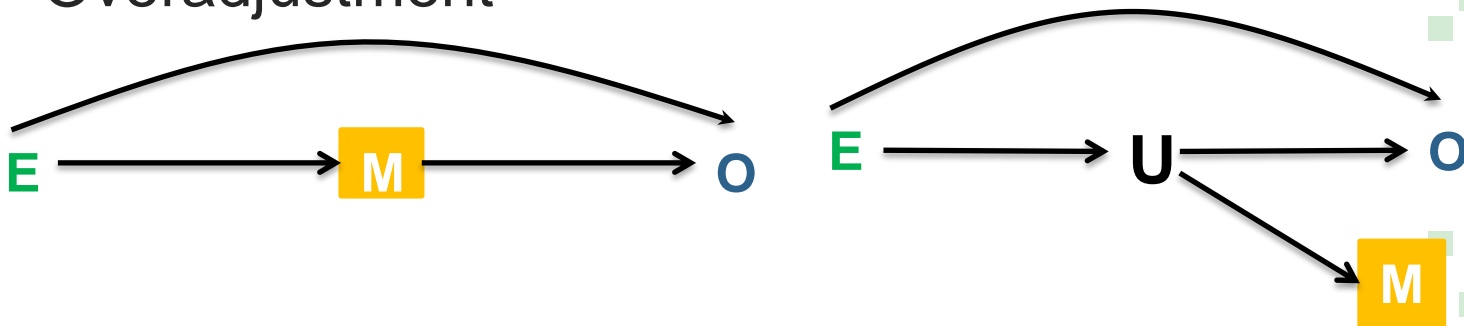
- Overadjustment ??
- Hiệu chỉnh cho Ascending proxy không làm ảnh hưởng tới ảnh hưởng của **E** lên **O**

Tóm tắt

- Confounding



- Overadjustment



- unnecessary adjustment
- Collider-stratification bias

2. Unnecessary adjustment



Unnecessary adjustment¹

- Hiệu chỉnh cho biến số mà không ảnh hưởng tới **bias**, nhưng có thể ảnh hưởng tới **precision**
- Còn gọi là “Bias-neutral adjustment”
- Tăng hay giảm **precision** phụ thuộc vào
 - Mối liên quan giữa C_i , E và O
 - Sample size

1. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009;20(4):488-95.

Unnecessary adjustment

- C1: biến số nằm hoàn toàn phía ngoài mối liên hệ **Exposure** → **Outcome**

C_1



Unnecessary adjustment

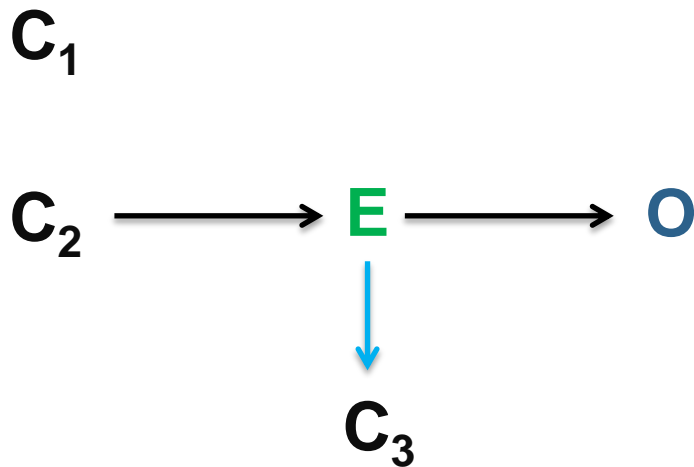
- C2: biến gây nên chỉ riêng **Exposure**

C_1



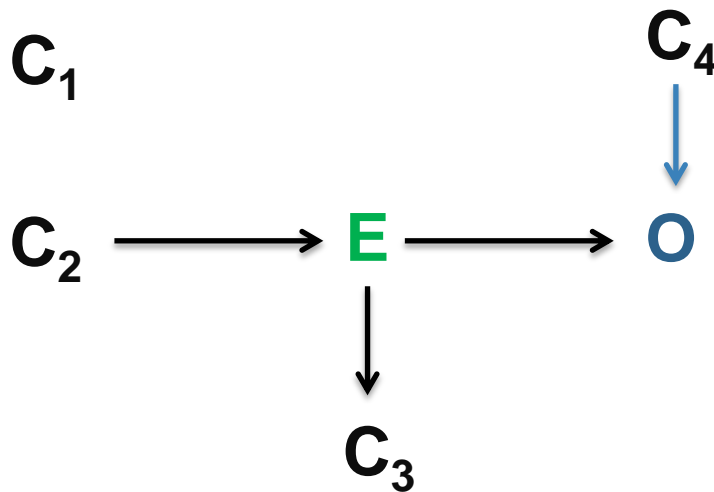
Unnecessary adjustment

- C3: là descendent **Exposure** nhưng không nằm trên đường **E** → **O**



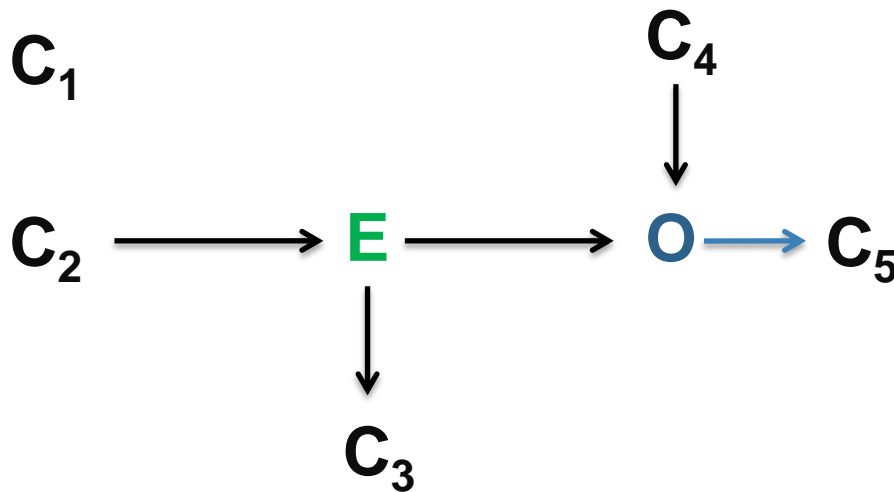
Unnecessary adjustment

- C4: gây ra chỉ riêng **Outcome**



Unnecessary adjustment

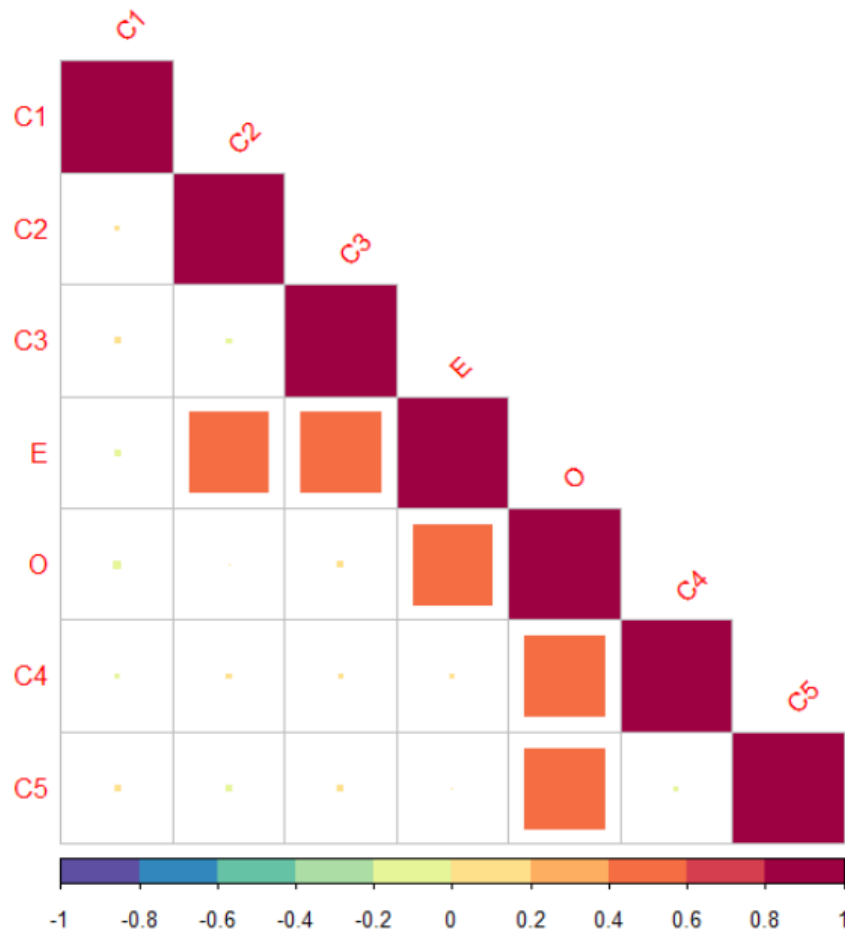
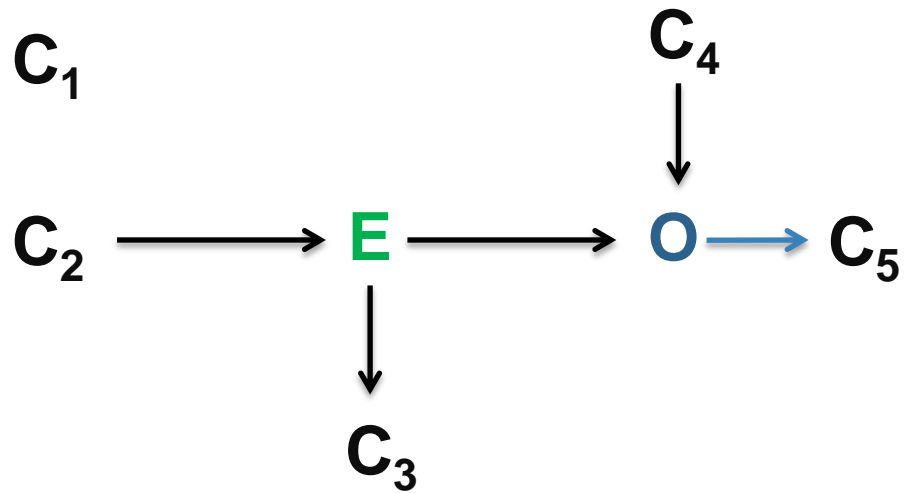
- C5: là descendent của Outcome

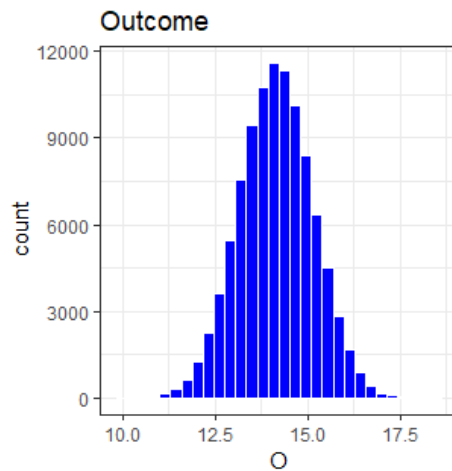
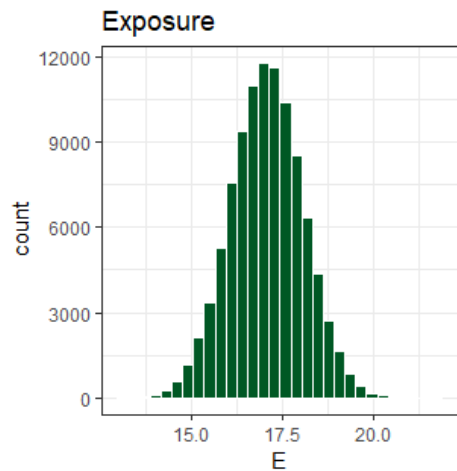
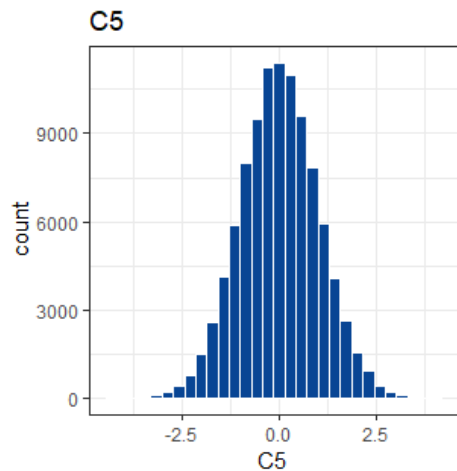
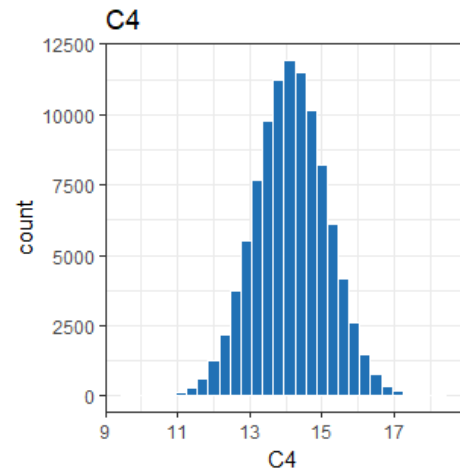
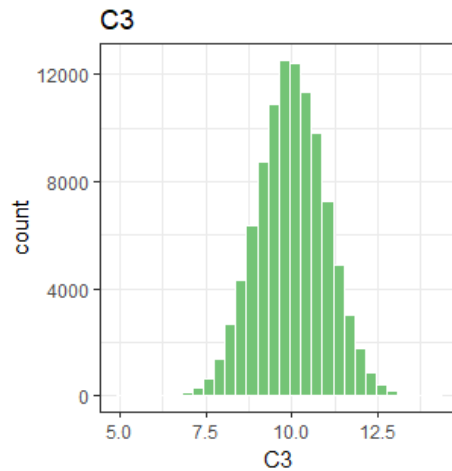
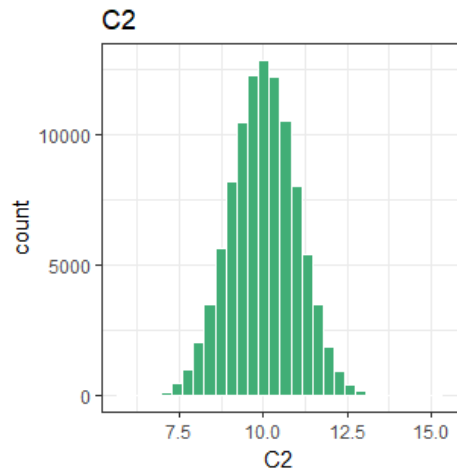
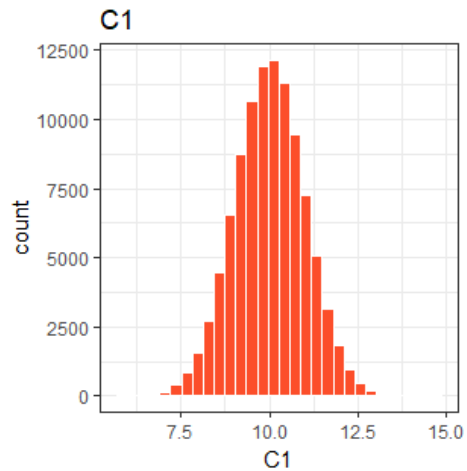


Mô phỏng tác động của C_i

❖ Các bước:

- Tạo biến $C_1, C_2, C_4, E \sim N(10,1)$
- Giả định tất cả các mối liên hệ đều là linear
- Hệ số phương trình được đặt $= 0.5$
- Sample size được đặt từ 100 – 100,000
- Mỗi sample size, fit hồi quy tuyến tính $O \sim E + C_i$ (mỗi C_i được fit one at a time)
- Tính relative bias và slope variance





```
> m = lm(O ~ E + C4, data = raw)
> summary(m)
```

```
Call:
lm(formula = O ~ E + C4, data = raw)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-2.8633	-0.4770	-0.0035	0.4807	3.2686

```
Coefficients:
```

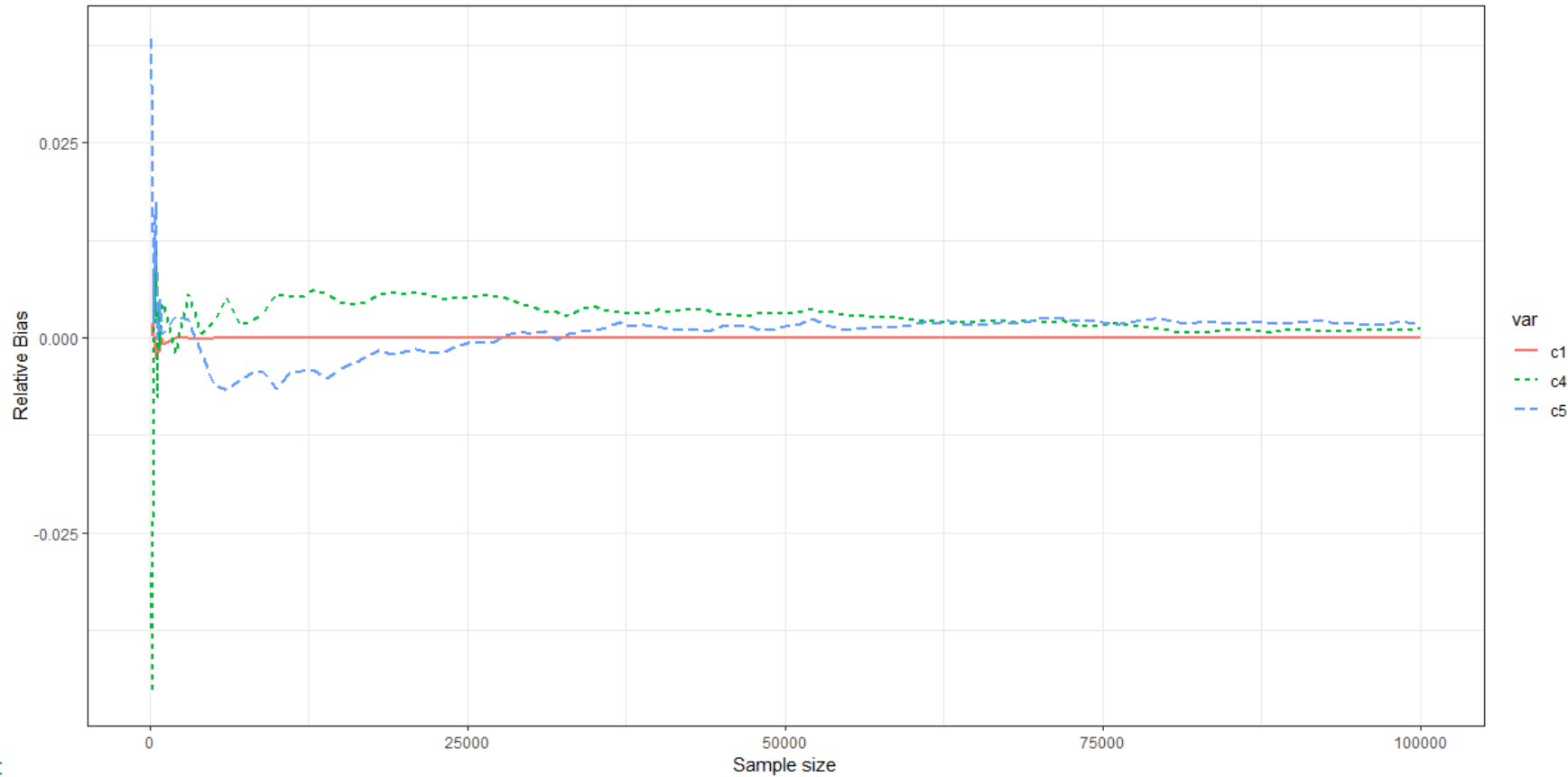
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1.476179	0.049664	-29.72	<2e-16	***
E	0.501457	0.002240	223.89	<2e-16	***
C4	0.499216	0.002237	223.15	<2e-16	***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.7091 on 99997 degrees of freedom
Multiple R-squared:  0.5,    Adjusted R-squared:  0.4999
F-statistic: 4.999e+04 on 2 and 99997 DF,  p-value: < 2.2e-16
```

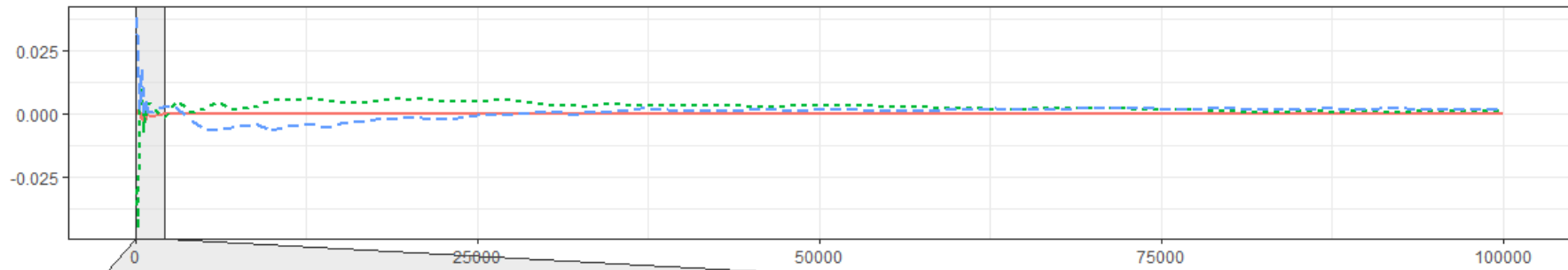
Bias

Unnecessary adjustment

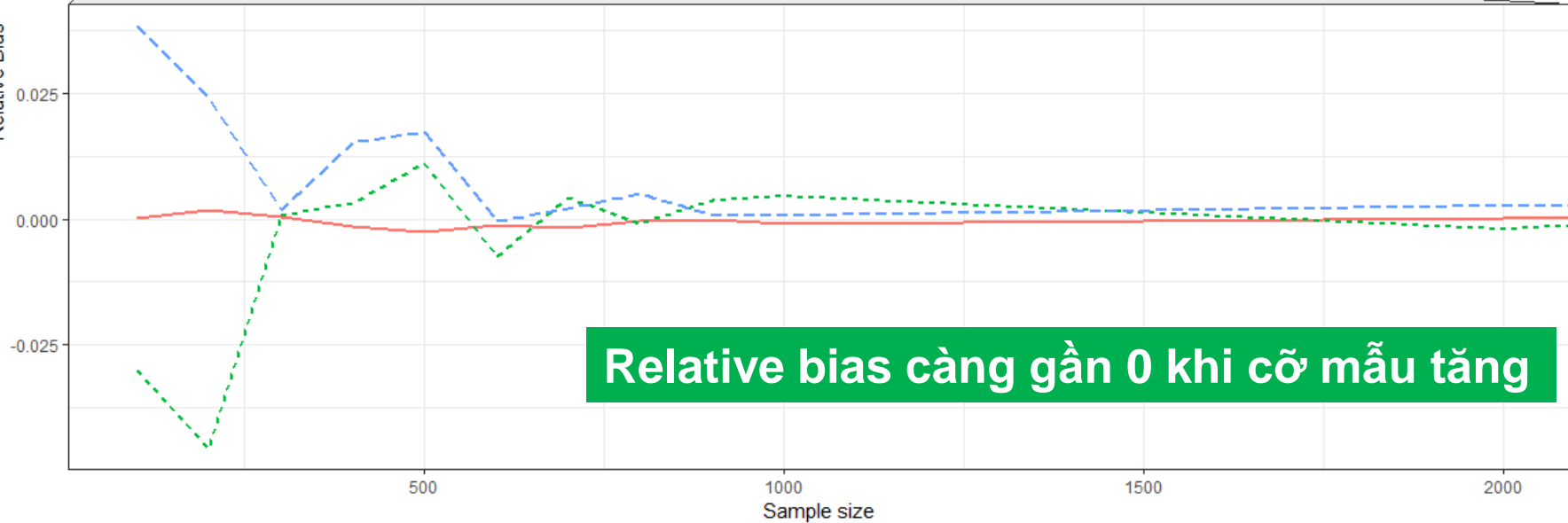




Unnecessary adjustment

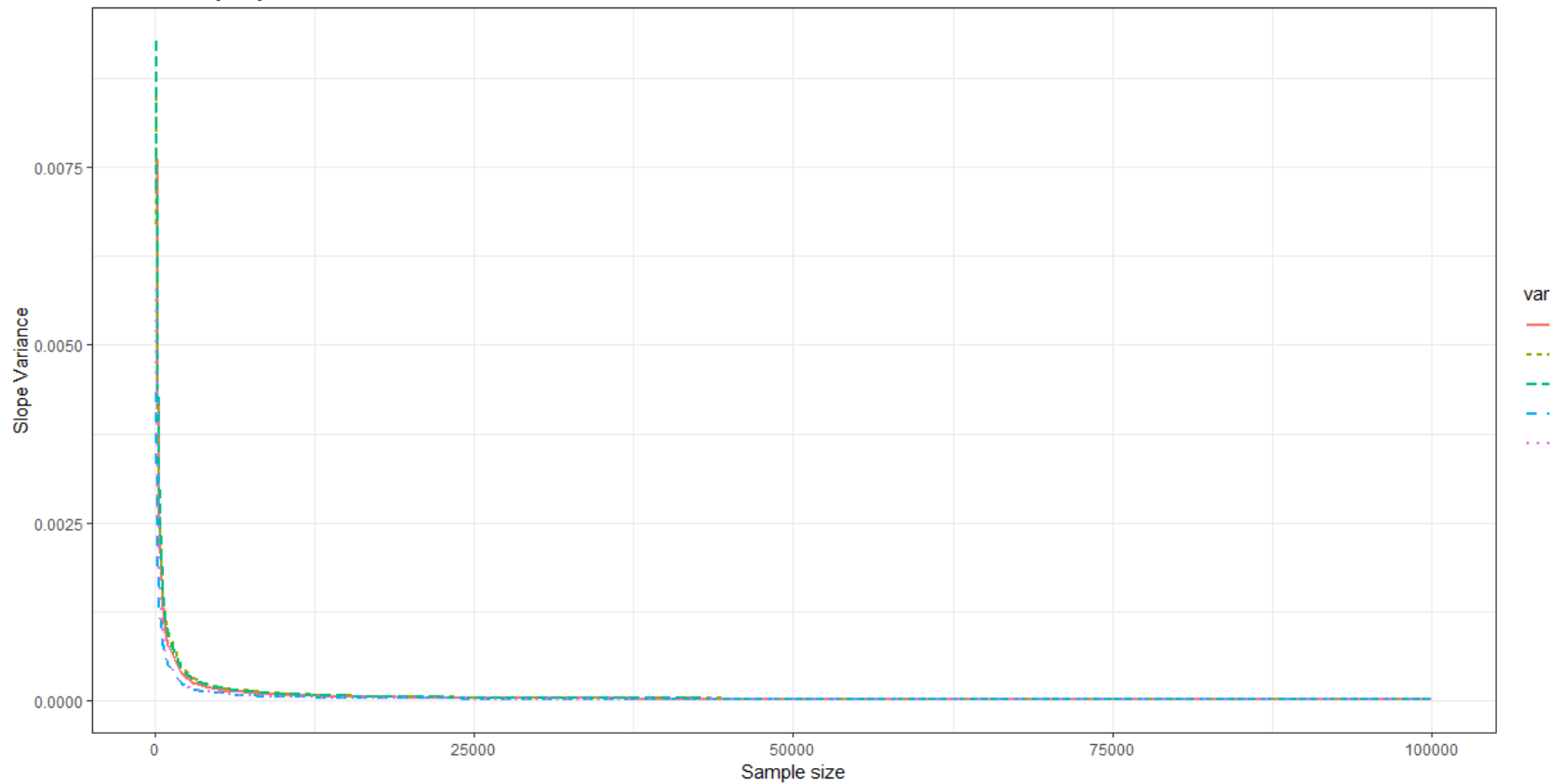


Relative Bias



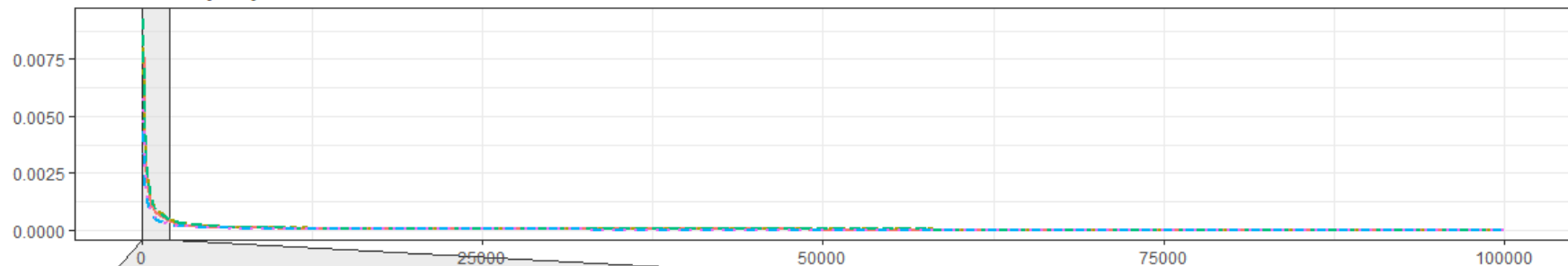
Variance

Unnecessary adjustment

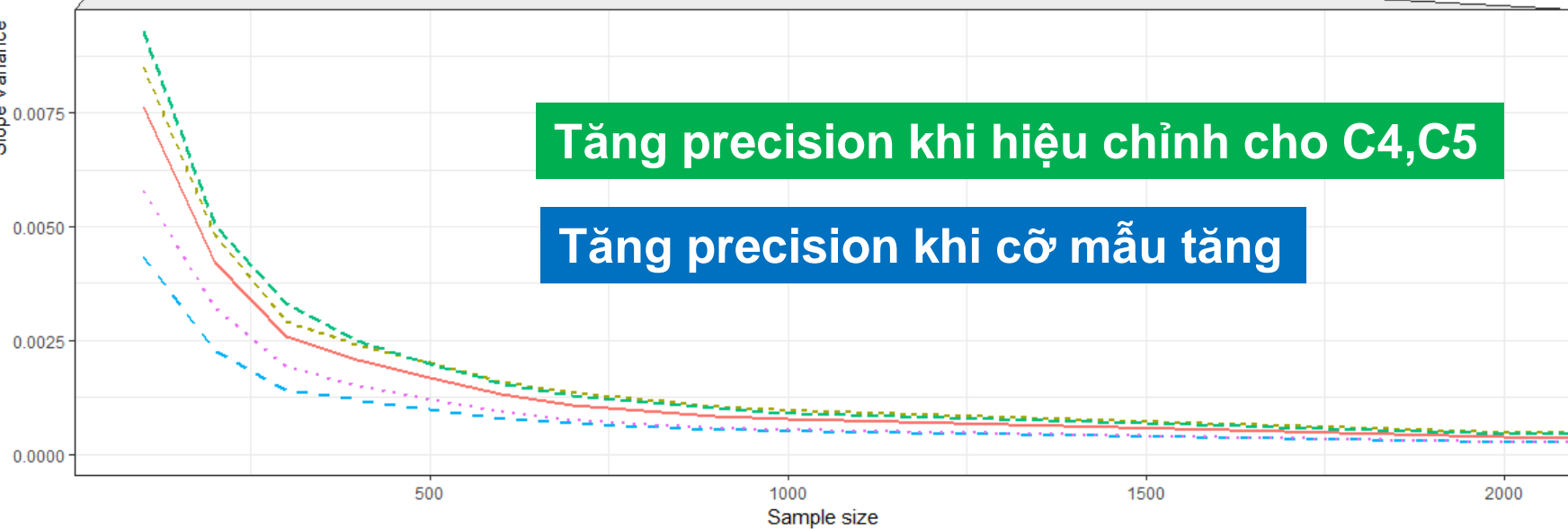




Unnecessary adjustment



Slope Variance



Tăng precision khi hiệu chỉnh cho C4,C5

Tăng precision khi cỡ mẫu tăng

var

c1

c2

c3

c4

c5



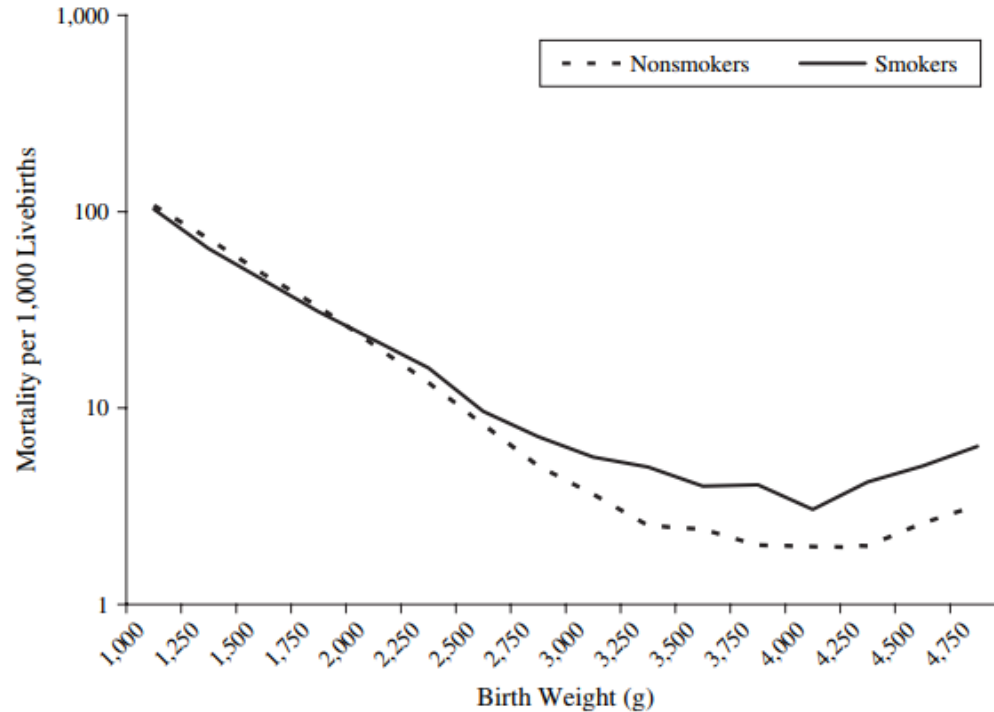
Variance

- Tăng hay giảm phụ thuộc vào loại model
 - Linear model: hầu như tăng precision cho total effect
 - Non-linear (e.g., logistic): giảm precision

3. Collider-stratification



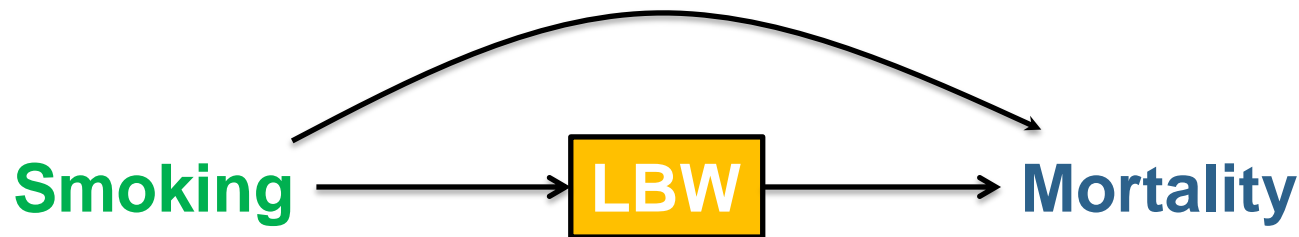
Birth weight paradox



Birth weight paradox

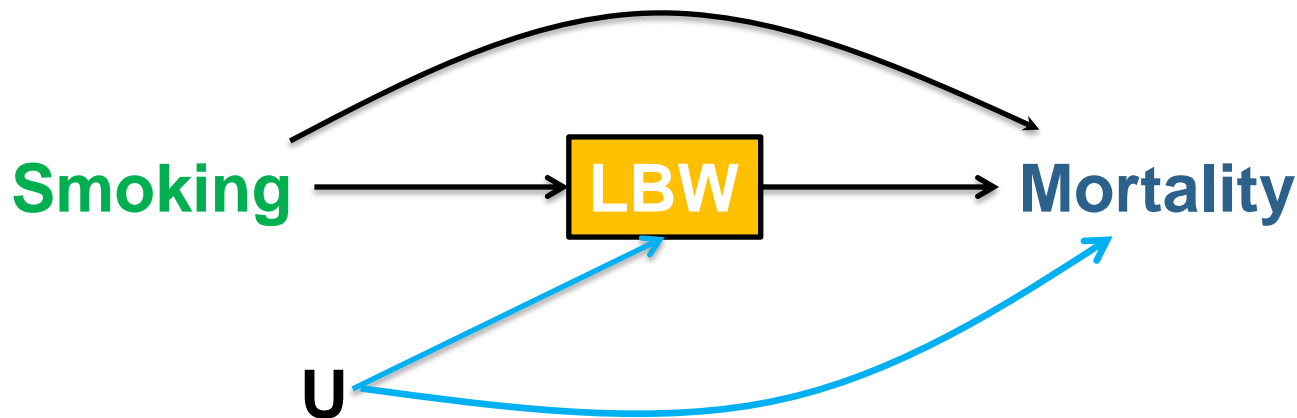
- Không hiệu chỉnh: RR 1.55 (1.50 – 1.59)
 - Hiệu chỉnh cho biến cân nặng: RR 1.09 (1.05 - 1.12)
 - Cân nặng < 2000 g: **RR 0.79 (0.76 - 0.82)**
 - Cân nặng > 2000g: RR 1.80 (1.72 - 1.88)
- ➔ Ở trẻ sơ sinh nhẹ cân (bw <2000g), mẹ hút thuốc lá làm giảm nguy cơ tử vong trẻ?

Birth weight paradox



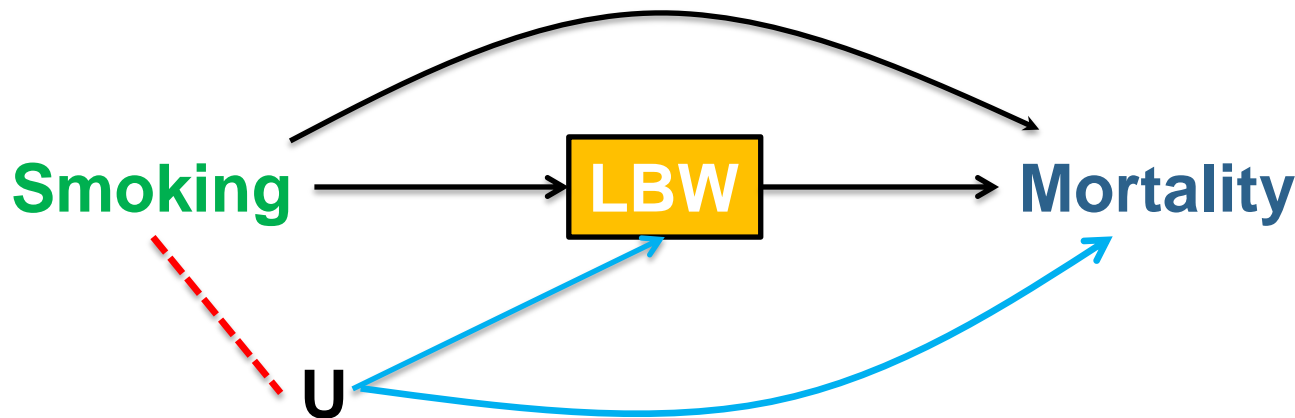
- Overadjustment do hiệu chỉnh mediator

Birth weight paradox



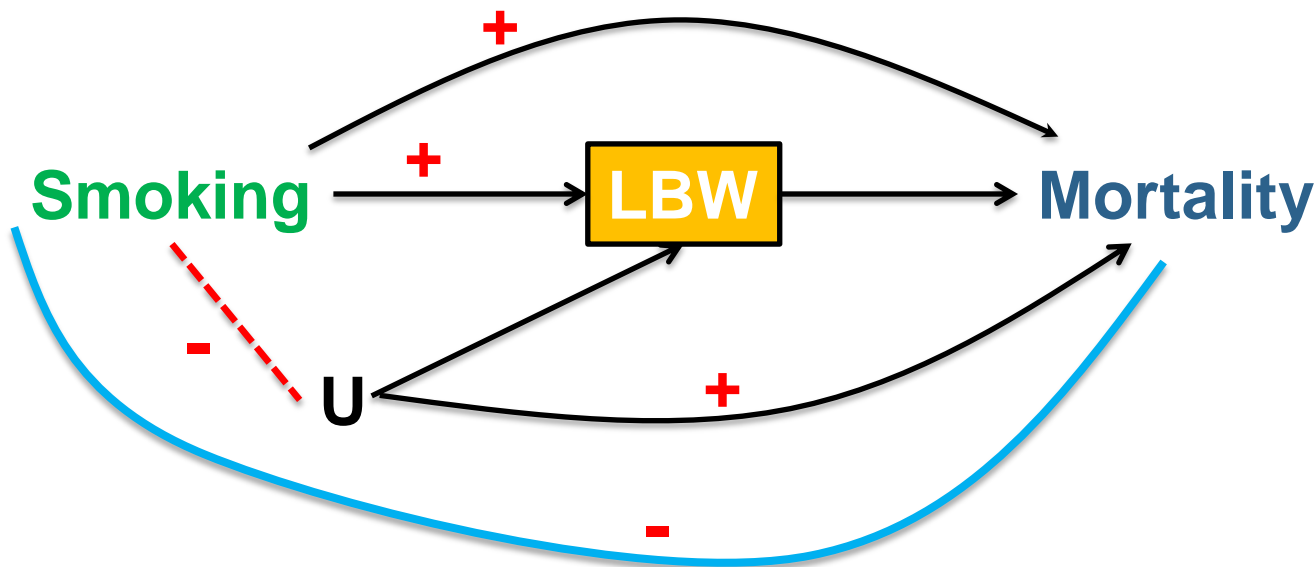
- Ví dụ U: dị tật thai nhi, suy dinh dưỡng...
- Giả sử dị tật thai nhi và smoking là 2 nguyên nhân gây ra LBW

Birth weight paradox



- Ở trẻ LBW. Nếu mẹ không hút thuốc → LBW là do nguyên nhân khác (i.e., dị tật bẩm sinh ...), dị tật bẩm sinh gây ra nguy cơ tử vong cao hơn hút thuốc

Birth weight paradox



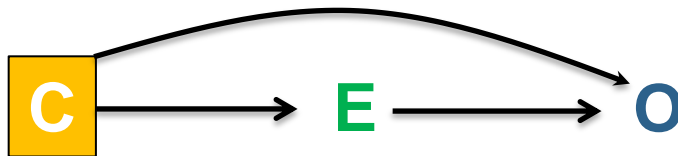
- Cho dù có mối liên hệ giữa LBW với mortality hay không, hiệu chỉnh cho LBW sẽ dẫn tới Collider-stratification bias

Collider-stratification bias

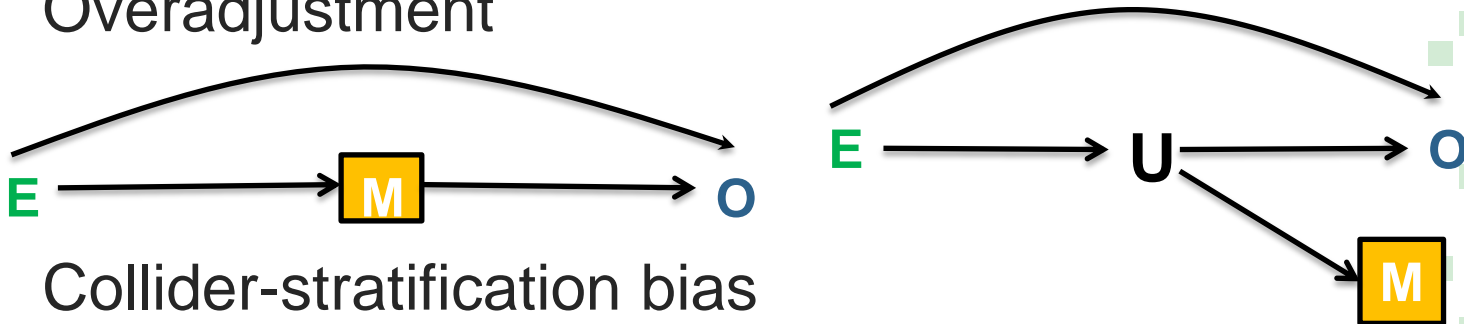
- Collider-stratification bias xảy ra khi hiệu chỉnh cho biến
 - Ảnh hưởng bởi **exposure**
 - Có cùng nguyên nhân với biến **outcome**

Tóm tắt

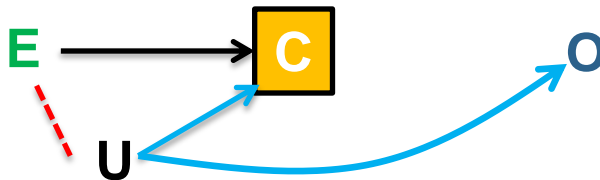
- Confounding



- Overadjustment



- Collider-stratification bias



What Do We Know?

- What is DAG?
- How do I construct a DAG?
- What is a confounder?
- What is a collider?
- What is an intermediate?
- What is a causal path?
- What is a backdoor path?
- When is the backdoor path open?
- When is the backdoor path closed?
- What is a minimal sufficient adjustment set?

THANK YOU!