Public Health Sciences 310
Epidemiologic Methods

Lecture 7
Confounding

January 25, 2024

Brian C.-H. Chiu

Confounding

- When a noncausal association between a given exposure and an outcome is observed or the strength of the association is distorted (either too high or too low) as a result of the influence of a third factor (the confounder).

- Can occur in: Often in observational studies
  - Descriptive studies (e.g., comparing mortality rates across different geographic regions). ex) Mortality in IL needs to be adjusted for FL.
  - Analytic studies (e.g., assessing associations).
  - Randomized studies (by random chance).

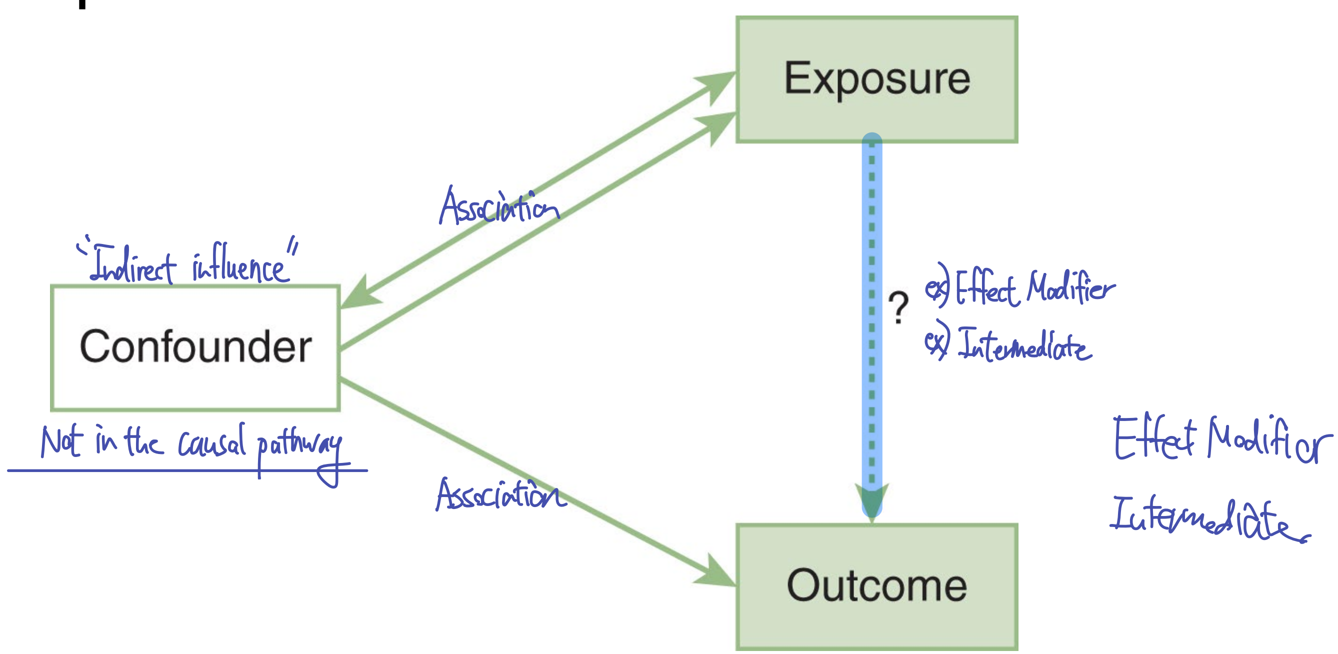
    Process to remove certain degree of confounding

General Rule for Confounding

- A variable can be a confounder if <u>all</u> the following conditions are met:
  - Causally associated with the outcome, AND
  - Causally or non-causally associated with the exposure, BUT

Intermediate is in the causal pathway; Confounding is not in the causal pathway.

  - NOT an intermediate variable in the causal pathway between exposure and outcome



Visually to Determine Confounding?

Graphical Aids to Frame Confounding

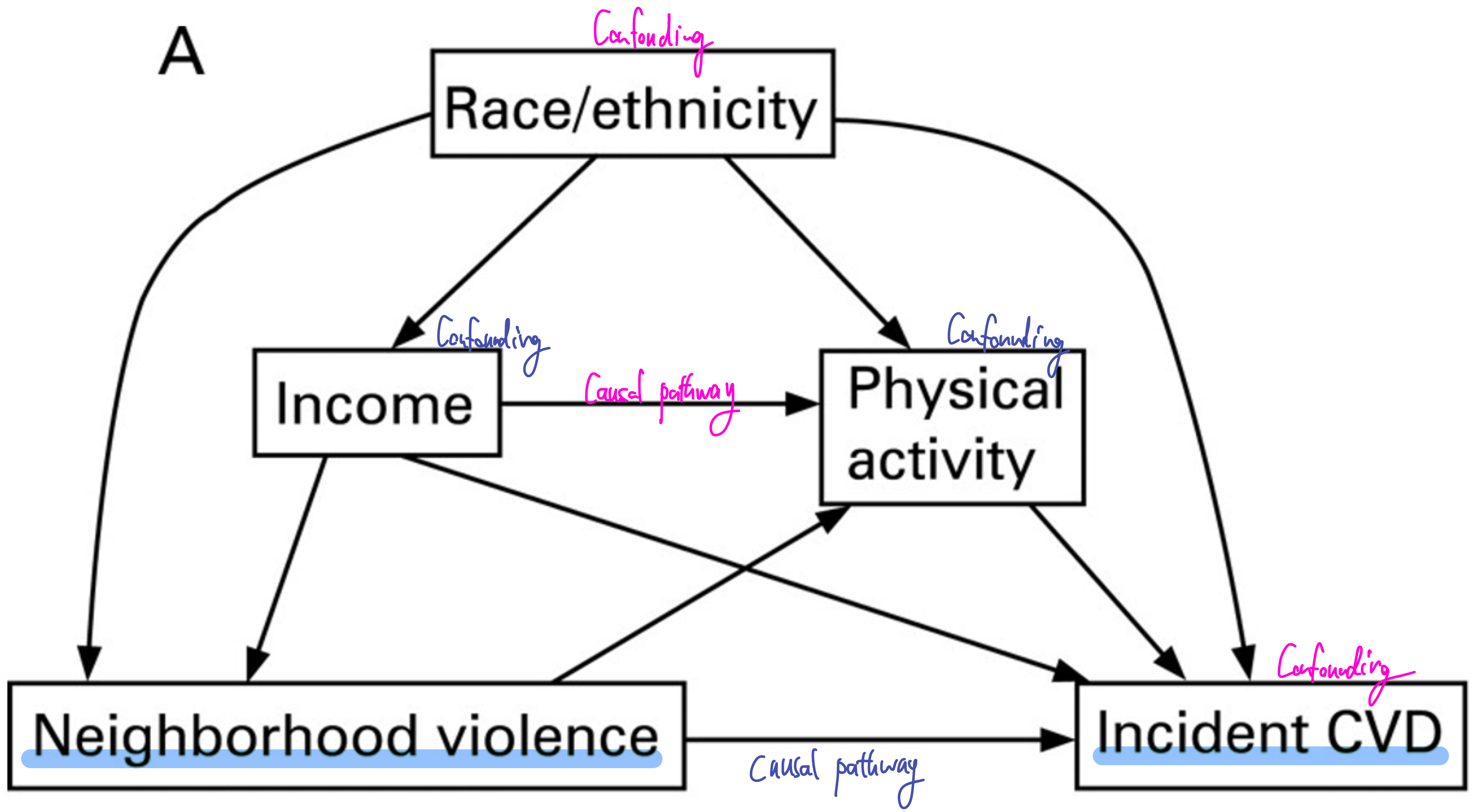
- Directed Acyclic Graphs (DAG) ("causal diagram")
  - a formal and more elaborate extension of traditional graphs to represent confounding.
  - Facilitate and guide the causal inference process
  - the direction of the association between the variables of interests and other unknown confounders is explicitly displayed to facilitate and guide the casual inference process.

Good tool for find paper + presentation

Examples of DAGs

Using DAGs to identify variables that need to be controlled for in estimating neighborhood health effects

Factors defined : Depends on the research question of interest.



Fleischer and Diez Roux. J Epidemiol Community Health 2008;62:842-6

Assessing Confounding

1. Is the confounding variable related to both the exposure and the outcome in the study?
2. Does the exposure–outcome association seen in the crude analysis have the same direction and similar magnitude as the associations seen within strata of the confounding variable?
3. Does the exposure–outcome association seen in the crude analysis have the same direction and similar magnitude as that seen after controlling (adjusting) for the confounding variable?

Same after adjusting: If yes, then likely effect modifier / intermediate.

Step #1: Does the variable meet the criteria to be a confounder?

You are the emergency room physician of a local hospital. After Sunday's NFL games, you see a number of patients who had a motor vehicle injury. At intake, patients often mention that they had been out watching the games with friends. You suspect that watching football game might be associated with motor vehicle injury. So, you design a case-control study to investigate this association.

Watch football game	Case Motor Injury	Control
Yes	91 a	19 b
No	19 c	91 d
Total	110	110

$$\text{OR (crude)} = \frac{\frac{91}{19}\ (91 \times 91)}{\frac{19}{91}\ (19 \times 19)} = 23$$

Is watching football game related to the risk of motor vehicle injury?

Since OR is very high (=23), very unlikely to have been influenced by confounding. But still need to double check.

First Criterion: Is the Confounding Variable Associated with the Outcome?

Confounding need to influence outcome: "founding"

Being the long time NFL fan that you are, you begin to suspect that perhaps football viewers are drinking beer when watching the game, and perhaps the association of watching football with motor vehicle injury is confounded by beer drinking. So, the first thing you do is determine whether beer drinking is associated with motor vehicle injury.

Possible Confounder Outcome

Beer drinking	Case	Control
Yes	$\frac{90}{110}$ 90 (82%)	$\frac{17}{110}$ 17 (15%)
No	$\frac{20}{110}$ 20 (18%)	$\frac{93}{110}$ 93 (85%)

Chi-square test

$X^2 = 96.96$ (p = < 0.0001)

Is beer drinking related to motor vehicle injury? **YES**

Reject Ho. Beer drinking is significantly different among cases and control of motor injury groups.

Second criterion: Is the Confounding Variable (i.e., Drinking Beer) ~~Non-Causally or Causally Associated with the Exposure~~ (i.e., Watching Football game)?

Confounding is associated with the "cause" : "confound" the results together.

Then you want to determine whether there is an association between watching football game and drinking beer in your study sample.

Possible Confounder	Watch Football Game (Exposure)	
Beer drinking	Yes	No
Yes	96 (87%)	11 (10%)
No	14 (13%)	99 (90%)

$X^2 = 131.46$ ($p = < 0.0001$)

Is drinking beer associated with watching football game? **YES**

Reject Ho. Beer drinking is significantly different among groups that watch / doesn't watch football.

Third criterion: Is the confounding an intermediate variable in the causal pathway between exposure and outcome?

- Note: Judgment and knowledge about the pathophysiology of the disease process are critical to answer this question.

- Drinking beer is probably not a mediator of the relationship between watching the game and motor vehicle injury.

Expected Magnitude of Association

Provided that:

- Crude association between watching football game and motor vehicle injury: OR = 23
- Drinking beer is more frequent among game viewers, and
  ↳ ex) Proved by chi-square
- Drinking beer is associated with greater risk of motor vehicle injury…
  ↳ ex) Proved by chi-square

What would be the expected magnitude of the association between watching football game and motor vehicle injury after controlling for beer drinking?

The (adjusted) association estimate will be:

"Removes the effect of beer drinking"
OR should ↓.

- $\boxed{< 23?}$ ↓
- = 23?
- > 23?

Step 2: Does the association seen in the crude analysis have the same direction and similar magnitude as the associations seen within strata of the confounding variable?

- Stratification according to the confounder represents one of the strategies to control for its effect

Association between Watching Football Game and Motor Vehicle Injury Stratified by Beer Drinking

Confounding

NO BEER DRINKER		
Watch Game	Case	Control
Yes	10	10
No	10	82
Total	20	92

$$OR = \frac{(10 \times 82)}{(10 \times 10)} = 8.2$$

Stratum-specific ORs

Similar ORs, but different from crude.

(Likely homo: $p > 0.05$)

BEER DRINKER		
Watch Game	Case	Control
Yes	81	9
No	9	9
Total	90	18

$$OR = \frac{(81 \times 9)}{(9 \times 9)} = 9$$

If confounding is present, the strata specific associations are similar but they are different from the crude.

Step 3: Does controlling for the putative confounder change the magnitude of the exposure-outcome association?

RQ: Is there an association between watching football game and motor vehicle injury?

The most persuasive approach to determine whether there is a confounding effect is the comparison between adjusted and crude association

Watch game	Case	Control
Yes	91	19
No	19	91
Total	110	110

$$\text{OR (crude)} = \frac{(91 \times 91)}{(19 \times 19)} = 23$$

NO BEER DRINKER		
Watch game	Case	Control
Yes	10	10
No	10	82
Total	20	92

OR = 8.2 ^

BEER DRINKER		
Watch game	Case	Control
Yes	81	9
No	9	9
Total	90	18

OR = 9.0 ^

Ways to Control for Confounding

- At the design phase of the study:
  - Matching: mostly case-control studies ex) Individual matching Frequency matching
  - Randomization: clinical trials
- At the analysis phase of the study:
  - Stratification: all study designs
  - Adjustment (multivariate analysis): all study designs

ex): SLR, MLR, Log, Cox, Poisson ...etc.

Control for Confounding During Analysis

- Stratification
  - Standardization
    - -Direct adjustment
    - -Indirect adjustment
  - Mantel-Haenszel Method

- Multivariate Regression
  - Linear Regression
  - Logistic Regression
  - Cox Proportion Hazards Regression
  - Poisson Regression

Stratification Adjustment Methods to Disentangle Confounding: Mantel-Haenszel Adjustment

MHOR

- MH adjusted odds ratios are a weighted average of stratum specific odds ratios.
- Similarly, MH adjusted rate ratios are a weighted average of stratum specific rate ratios.
- Reasonable to use when number of stratification variables (i.e., confounders) is limited and can be categorical.

Mantel-Haenszel Adjustment

$$OR_{MH} = \frac{\sum_{i=1}^{k} (a_i d_i / Ni)}{\sum_{i=1}^{k} (b_i c_i / Ni)}$$

Mantel-Haenszel Adjustment

Stratum 1

a1	b1
c1	d1

N1

Stratum 2

a2	b2
c2	d2

N2

Stratum 3

a3	b3
c3	d3

N3

Stratum k

ak	bk
ck	dk

Nk

$$OR_{MH} = \frac{\dfrac{a1d1}{N1} + \dfrac{a2d2}{N2} + \dfrac{a3d3}{N3} + \dfrac{akdk}{Nk}}{\dfrac{b1c1}{N1} + \dfrac{b2c2}{N2} + \dfrac{b3c3}{N3} + \dfrac{bkck}{Nk}}$$

Mantel Haenszel Adjusted OR

NO BEER DRINKER		
Watch game	Case	Control
Yes	10	10
No	10	82
Total	20	92

BEER DRINKER		
Watch game	Case	Control
Yes	81	9
No	9	9
Total	90	18

N1 = 112

N2 = 108

A·D

$$OR_{MH} = \frac{\frac{a1d1}{N1} + \frac{a2d2}{N2}}{\frac{b1c1}{N1} + \frac{b2c2}{N2}} = \frac{\frac{10*82}{112} + \frac{81*9}{108}}{\frac{10*10}{112} + \frac{9*9}{108}} = 8.5$$

V.S. 23

SS Confounding

Stratum ORs

B·C

Interpreting Adjusted Values

MHOR

- Independent of beer drinking, the OR for motor vehicle injury among game viewers compared to non-viewers is 8.5

Test of Significance of the OR_{MH}

p<0.05, MHor is statistically different from crude OR.

$$\chi^2_{MH} = \frac{\left(\left|\sum_{i=1}^{k} a_i - \sum_{i=1}^{k} E_i\right| - 0.5\right)^2}{\sum_{i=1}^{k} V_i}$$

a_i = observed number of exposed cases in the ith stratum

E_i = expected number of exposed cases in the ith stratum

$$E_i = (m_{1i} \times n_{1i}) / N_i$$

V_i = estimate of the variance of a_i

$$E_i = (m_{1i} \times m_{2i} \times n_{1i} \times n_{2i}) / (N_i^2 \times (N_i - 1)$$

Test of significance: the absurd example

QUESTION: Can we conclude that there is an association between watching football game and motor vehicle injury after adjusting for alcohol intake?

Step 1. Null Hypothesis: H_0: $OR_{MH} = 1$ No Association.

There is no association between watching the game and motor vehicle injury after adjusting for beer drinking.

Step 2. Alternative Hypothesis: H_A: $OR_{MH} \neq 1$

There is an association between watching the game and motor vehicle injury after adjusting for beer drinking.

Step 3. Let $\alpha = 0.05$

Test of significance: the absurd example

Step 4. Test statistic is $\chi2_{MH}$

The rejection region is R: $\chi2 \geq \chi2\ (1, 1 - \alpha) = 3.84$

Step 5. Sample values and test statistic

$a_1 = 10$

$a_2 = 81$

$E_1 = (20 \times 20)/112 = 3.57$

$E_2 = (90 \times 90)/108 = 75$

$V_1 = (20 \times 92 \times 20 \times 92)/(112 \times 112) \times 111 = 2.43$

$V_2 = (90 \times 18 \times 90 \times 18)/(108 \times 108) \times 107 = 2.10$

Test of significance: the absurd example

$$\chi^2_{MH} = \frac{\left(\left|\sum_{i=1}^{k} a_i - \sum_{i=1}^{k} E_i\right| - 0.5\right)^2}{\sum_{i=1} V_i}$$

$$\chi^2_{MH} = \frac{\left(\left|(10 + 81) - (3.57 + 75)\right| - 0.5\right)^2}{2.43 + 2.10} = 31.4$$

Step 6. Conclusion

Since 31.4 > 3.84, we can reject H_0, and conclude that there is an association between watching the football game and motor vehicle injury adjusting for alcohol.

Other Issues in Confounding

- Strategies to assess confounding:

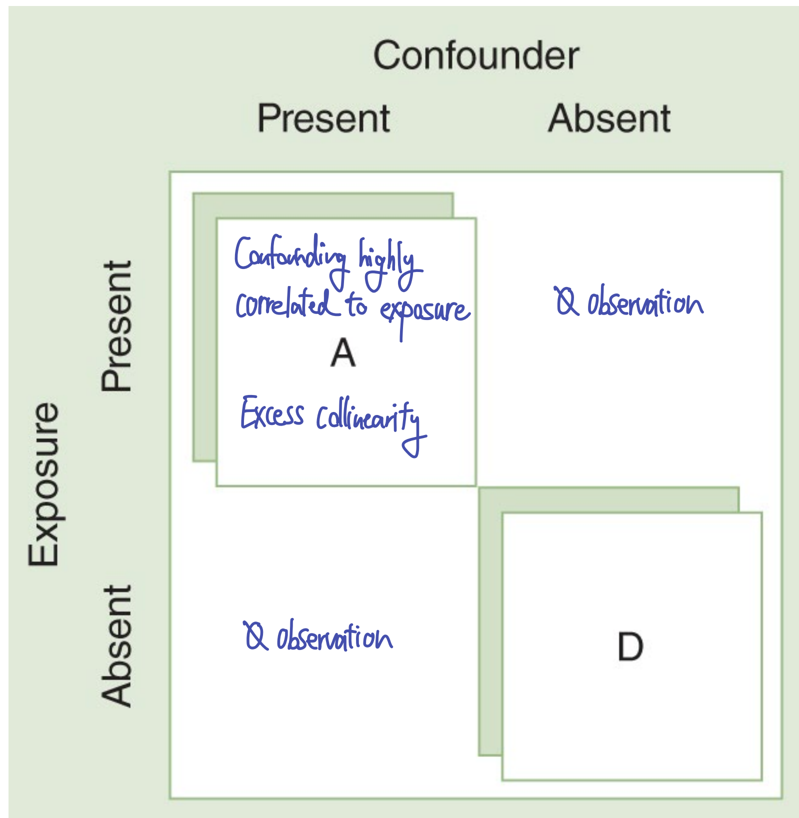
  - The crude differs from the adjusted
  - Stratification

- Excess correlation between the confounder and the exposure of interest

Issues
- Clinical research in Epi → Identify third factor without evaluation of data X
- Prognostic → Highly correlated to confounding, collinearity making data unstable X.
- Small sample → <20 X.

Excessive Correlation Between the Confounder and the Exposure of Interest

ex) Unstable Model



ex) Being vegetarian and high BP.

vegetarian don't eat meat, leading to almost ∅ observations for BP. (Collinearity)

Perfect correlation between dichotomous exposure of interest and confounding factor: There are no cells in which the exposure is present and the confounder is absent and vice versa.

- Always begin with hypothesis and evaluate based on predictions.
- Don't jump into conclusions.

Other Issues in Confounding

- Strategies to assess confounding:

  - The crude differs from the adjusted
  - Stratification

- Excess correlation between the confounder and the exposure of interest

- ~~Confounding vs. Bias~~

Is confounding a bias?

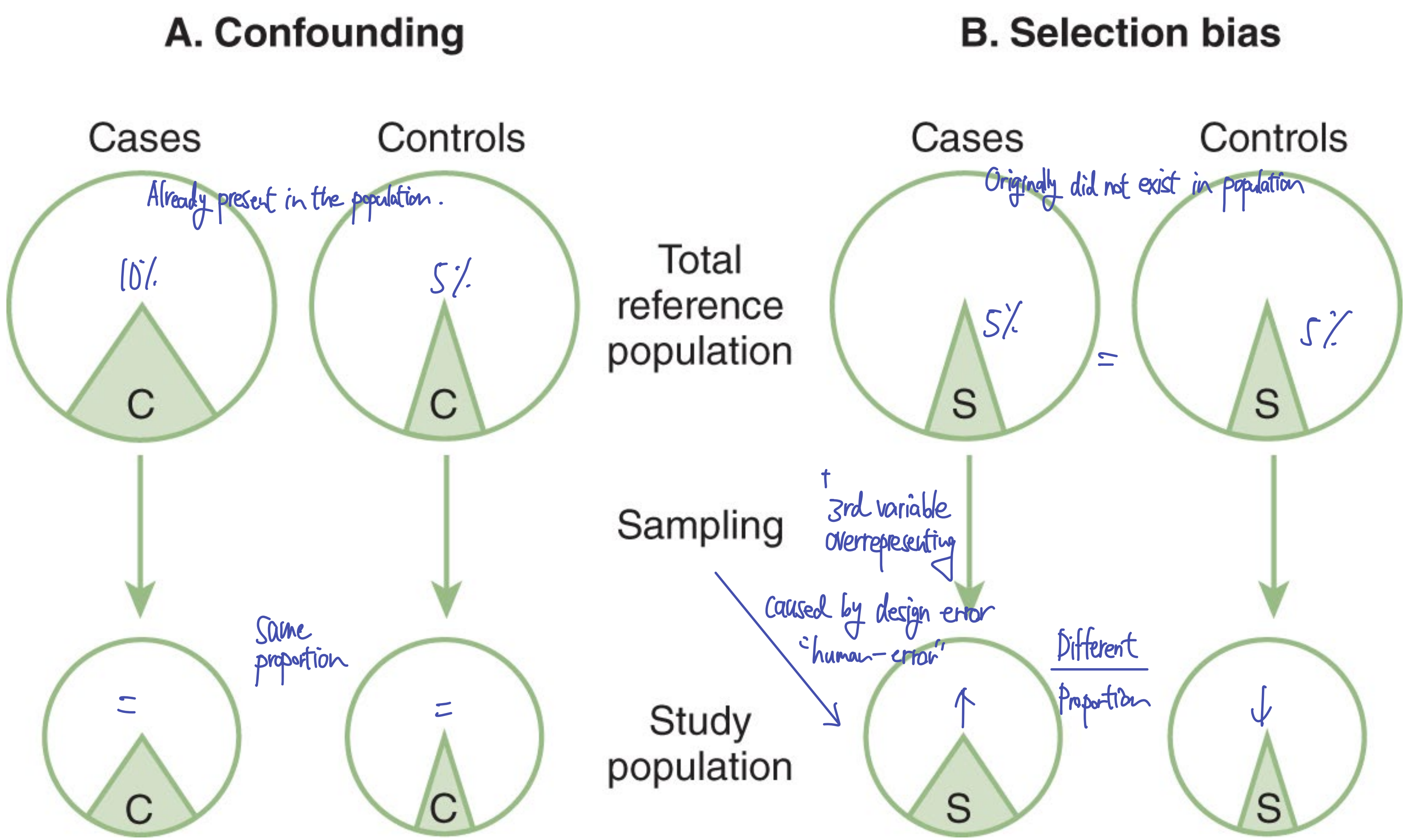
YES

- Because, if one concludes that a variable is a **causal** risk factor by missing the existence of confounding, this is an erroneous (biased) conclusion.
- A confounded estimate is a "biased estimate" of the true **causal** association between exposure and disease.

NO

- Because, 1) They are essentially different phenomena:
  - Bias: error in the design of the study that result on an estimate of an association that is not really present (or absence of one that really exists). ex): Human Error
  - Confounding: the observed association is really present (e.g., watching football game and motor vehicle injury, even if it is not causal).

Confounding and Bias



In the total reference population, confounding factor *C* is more common in cases than in controls; assuming no random variability, the study samples of cases and controls reflect the higher frequency of *C* in cases than in controls **(A)**. In selection bias **(B)**, the frequency of factor *S* is the same in cases and controls; however, through the selection process, it becomes more common in cases than in controls. Thus, confounding exists "in nature," whereas selection bias is a result of the sampling process.