MODULE 8: CONFOUNDING AND INTERACTION

MODULE 8.3: *CONFOUNDING*

PHC 6051 Biostatistics II

MODULE 8.3 OBJECTIVES

PHC 6051

- **Students will be able to**:
  - Explain the three conditions or properties of a ***confounder***.
  - Define and identify ***confounding***.
  - Explain and calculate crude and adjusted-effect measures of association to identify whether ***confounding*** is present.
  - Distinguish that ***confounding*** is not evaluated with statistical testing, while interaction is.

CONFOUNDING

IS THE ESTIMATE OF THE ASSOCIATION BETWEEN X AND Y MEANINGFULLY DIFFERENT DEPENDING ON WHETHER OR NOT WE IGNORE X_2?

WHAT IS CONFOUNDING?

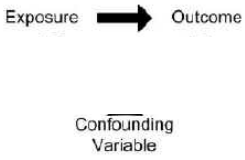
PHC 6051

- It is a distortion in the estimated measure of association between an exposure (independent variable) and an outcome (dependent variable) that occurs when the study groups differ with respect to other factors (variables) that influence the outcome.

WHAT IS CONFOUNDING?

PHC 6051

- Conditions for a confounding variable:
  1. Confounder must be independently associated with *Y*
  2. Confounder must be associated with primary *X* variable
  3. Confounder must not be an intermediary step in the causal pathway between *X and Y*



WHAT IS CONFOUNDING?

PHC 6051

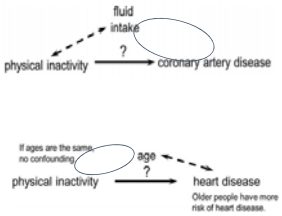
- In the diagram below, the primary goal is to ascertain the strength of association between physical inactivity and heart disease.
- Age is a confounding factor because:
  1. it is associated with the outcome (because older people are at greater risk of developing heart disease),
  2. it is associated with the exposure (meaning that older people are more likely to be inactive), and
  3. it is not an intermediate step between physical inactivity and heart disease.



WHAT IS CONFOUNDING?

PHC 6051

- In this scenario, what would NOT be confounders based on three criteria?
  - subjects who are physically active may drink more fluids (water) than inactive people, but drinking more fluid has no effect on the risk of heart disease, so fluid intake is not a confounding factor.
  - if the age distribution is similar in the exposure groups being compared, then age will not cause confounding.



WHAT IS CONFOUNDING?

PHC 6051

- Other scenarios for confounding variables:
  - A confounder can be another *risk factor* for the disease: Testing the association between *exercise* and *heart disease*, *age* is a confounder because it is a risk factor for heart disease.
  - A confounder can also be *surrogate or a marker* for some other cause of disease: Socioeconomic status (SES) may be a confounder because lower SES is a marker for a complex set of poorly understood factors that seem to carry a higher risk of heart disease.
  - There may be *many* possible confounding factors that could influence an association. For example, in looking at the association between exercise and heart disease, other possible confounders might include age, diet, smoking status and a variety of other risk factors.

CONFOUNDING IN REGRESSION

PHC 6051

- Identifying Confounding
  - No statistical test to determine if confounding is present.
  - A simple, direct way to determine whether a risk factor is a confounder is to compare the estimated measure of association *before* and *after* adjusting for a potential confounding variable.
  - If meaningful difference between the two measures of association exist, then confounding is present.

CONFOUNDING IN REGRESSION

PHC 6051

- Effects of Confounding
  - May account for all or part of an apparent association
  - May cause an overestimate of the true association (*positive confounding*)
  - May cause an underestimate of the true association (*negative confounding*)

CONTROLLING FOR ONE EXTRANEOUS VARIABLE

PHC 6051

- To assess effect of confounding, we compare a crude estimate and an adjusted estimate in terms of the following two regression models:

$$Y = \beta_0 + \beta_1 X + \beta_2 C + E \qquad \text{(Model 1)}$$

$$Y = \beta_0 + \beta_1 X + E \qquad \text{(Model 2)}$$

- Model 1 expresses the relationship between *X* and *Y*, *adjusted* for the variable *C*.
  - The estimate $\hat{\beta}_{1|C}$ of β_1 is an adjusted-effect measure in the sense that it gives the estimated change in *Y* per unit change in *T* after accounting for *C*.
- Model 2 expresses the relationship between X and Y - the crude estimate is the estimated coefficient of *X* (namely, $\hat{\beta}_1$) without involving *C*.

CONTROLLING FOR ONE EXTRANEOUS VARIABLE

PHC 6051

- General rule for assessing the presence of confounding when only one independent variable is to be controlled:
  - Confounding is present if the estimate of the coefficient (β_1) of the study variable *X* meaningfully changes when the variable *C* is removed from model; that is, $\hat{\beta}_{1|C} \neq \hat{\beta}_1$.
  - The ≠ sign indicates that a *subjective decision* is required to determine whether the two estimates each describe a different interpretation of the *X*–*Y* association in question.
  - A statistical test is neither required nor appropriate.

CONTROLLING FOR ONE EXTRANEOUS VARIABLE

PHC 6051

- Example: Let Y =SBP, X = PAL, and C =AGE.

CRUDE $\hat{\beta}_1$	ADJUSTED $\hat{\beta}_{1\|AGE}$	CONFOUNDING?
15.9	4.1	YES
6.1	6.2	NO
5.5	4.1	DEPENDS

- One approach to this problem is to control for a confounder that changes the crude effect estimate by some pre-specified amount determined by clinical judgment (e.g., 10%).

CONTROLLING FOR ONE EXTRANEOUS VARIABLE

PHC 6051

- One approach sometimes used (**incorrectly**) to assess confounding is, for example, to conduct a statistical test of H_0: $\beta_2 = 0$ in model
  - This test does not address confounding but rather *precision*.
  - This test evaluates whether significant additional variation in Y is explained by adding C to a model already containing X.
  - An equivalent approach is to determine whether a confidence interval for b1, the coefficient of X, is considerably narrower when C is in the model than when it is not.
  - Precision is often an important issue to assess when considering extraneous factors, but it differs fundamentally from confounding.

CONTROLLING FOR SEVERAL EXTRANEOUS VARIABLES

PHC 6051

- We can extend controlling for one extraneous variable to controlling for more than one extraneous variable.
- An example: Consider model
  $$SBP = \beta_0 + \beta_1(PAL) + \beta_2(AGE) + \beta_3(SEX) + E$$
  - A hypothetical study to assess the relationship between physical activity level (PAL) and systolic blood pressure (SBP) while controlling for both AGE and SEX.
- Assessment of confounding: Does the coefficient of PAL change when AGE and/or SEX is dropped from the model?

CONTROLLING FOR SEVERAL EXTRANEOUS VARIABLES

PHC 6051

- To answer this, we can examine the estimate of the coefficient of PAL in four models
  1. One including both AGE and SEX.
  2. One involving AGE.
  3. One involving SEX.
  4. One involving neither.
- The gold standard model for comparison contains both control variables and PAL.

CONTROLLING FOR SEVERAL EXTRANEOUS VARIABLES

PHC 6051

- If the estimate of β_1 changes considerably when at least one control variable is dropped from this gold standard model, we need to control for both AGE and SEX.
- If we obtain essentially the same estimate of β_1 (as obtained using the gold standard model) when only AGE is in the model, we do not need to retain SEX in the model to control for confounding.
- Including SEX variable in addition to AGE may increase or decrease precision. Thus, the decision as to whether to control for just AGE or for both depends, for example, on a comparison of confidence intervals for β_1 .
- If the confidence interval is considerably narrower when only AGE is controlled, we should not retain SEX in the model.

CONTROLLING FOR SEVERAL EXTRANEOUS VARIABLES

PHC 6051

- Once we make a decision about which variables to control (best model selected), we make statistical inferences about the true PAL–SBP relationship.
- Given the no-interaction model, this involves testing H_0: β_1 =0 in the best model and then obtaining an interval estimate of β_1.

SUMMARY

PHC 6051

- Confounding and interaction are two methodological concepts that pertain to assessing a relationship between independent and dependent variables.
- Confounding, which is not evaluated with statistical testing, is present when the effect of interest differs depending on whether an extraneous variable is ignored or retained in the analysis.
- In regression terms, confounding is assessed by comparing crude versus adjusted regression coefficients from different models.
- Interaction, which takes precedence over confounding, exists when the relationship of interest differs at different levels of extraneous (control) variables.

SUMMARY

PHC 6051

- When several potential confounders are being considered, it may be worthwhile to identify non-confounders that can be dropped from the model to gain precision; this may not be possible (i.e., precision may be lost by dropping variables) in some situations.
- Moreover, in such a situation, the assessment of confounding involving other extraneous variables, though possible, is quite complex and extremely subjective.
- Consequently, the assessment of confounding is not usually recommended when important interaction effects have been identified.