

# STA130H1S TUT0109

## W11: Confounding and Ethics

**Mar 29, 2019**

(Materials used in this presentation are provided by the UofT Statistical Sciences Department)

Email: [sonia.chhay@mail.utoronto.ca](mailto:sonia.chhay@mail.utoronto.ca)

Website: [soniachhay.github.io](https://soniachhay.github.io)

# Overview

- Announcement:
  - Everyone should have received an email with their group number
  - Evaluation schedule (for evaluating other groups) have been posted to Quercus
- Vocabulary
- This week's material/homework
- Presentations

# Vocabulary for this Week

Observational study

Confounding

Confounded

Association

Causation

Experiment

Randomized (controlled) trial

Control group

Ethics/ ethical issue/ ethical consideration

Consent

Information

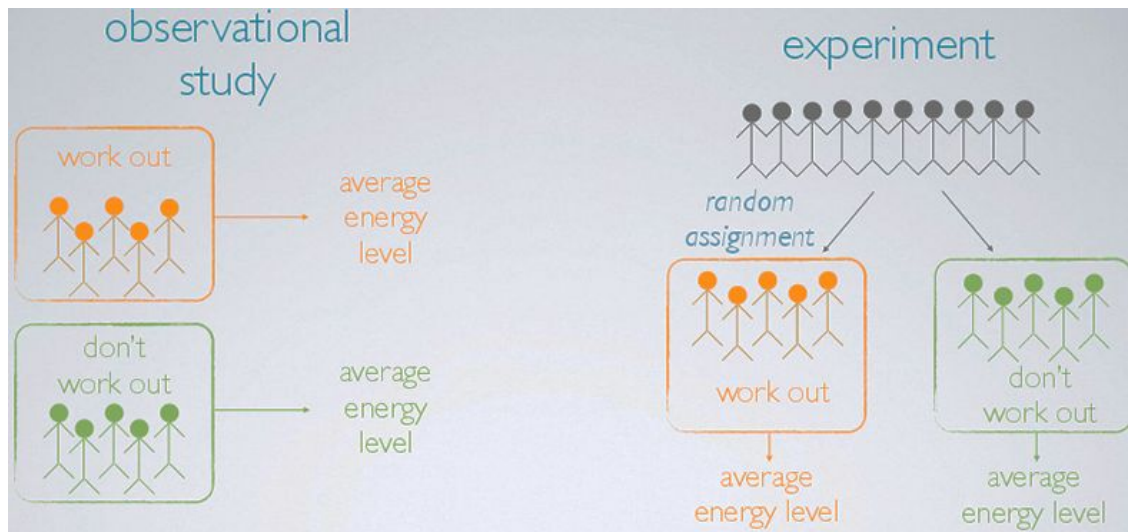
Comprehension

Voluntariness

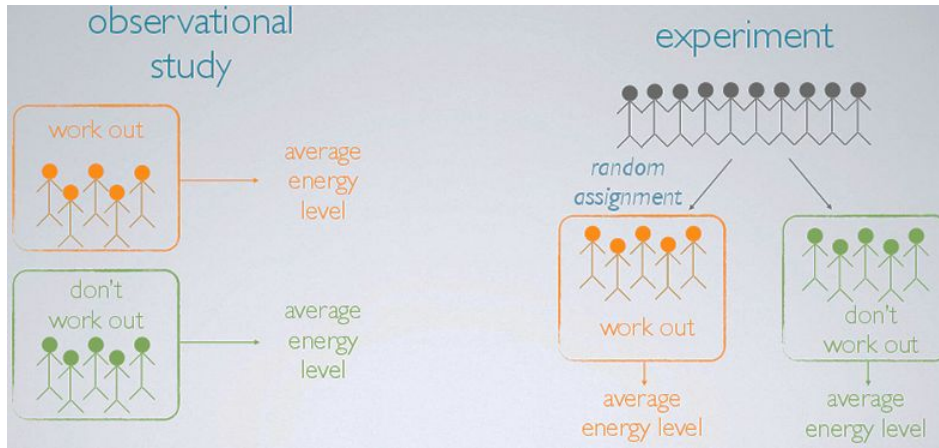
Transparent

# **Observational Study vs. Experiment**

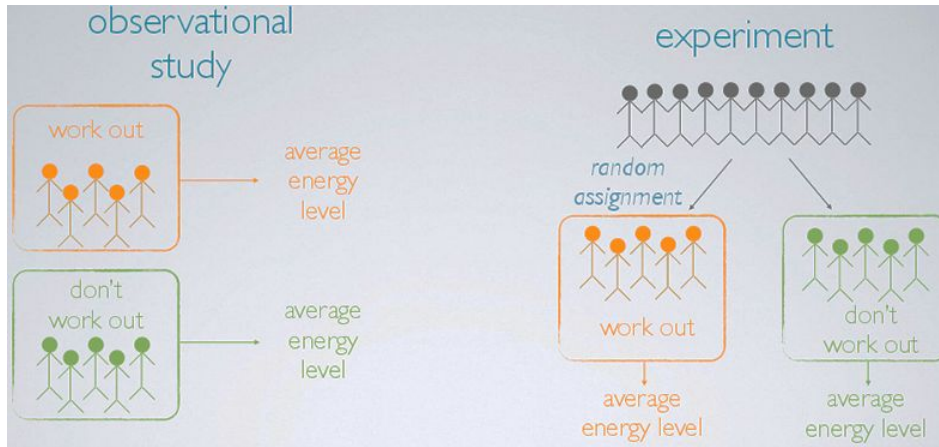
- Observational study:
  - Research simply observes natural processes
  - Observe effect of risk factor, diagnostic test, treatment or other intervention without trying to change who is or isn't exposed to it
- Experiment:
  - Researcher imposes certain conditions
  - Introduce an intervention and study the effects



- Suppose we are interested in investigating the relationship between physical activity and weight among adults
- We can design the study as an observation study or an experiment
  - Identify a group of people, hopefully representative of the population you're interested in



- For observational study:
  - Investigate whether less physically active people were more likely to gain weight than active people (e.g. over 5, 10, ?? years)
- For experiment:
  - Assign (at random) some people to be physically active and others to not be physically active (called the control group)
  - Investigate whether the physically active group were less/more likely to gain weight (over X years) than the less physically active



- Difference:
  - In the observational study, decision of whether to work out or not is the subjects' own choice
  - In an experiment, this decision is imposed by the researcher
- Could look at the weight gain difference in each study
- When we have data from an observational study, we can only conclude association between variables; *not* causation

# Association vs. Causation

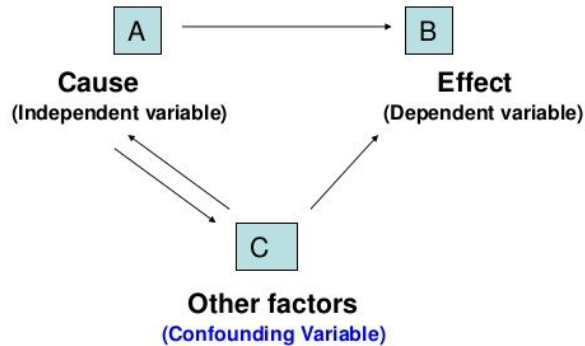
- Association:
  - Only means that two things are related
- Causation:
  - Means that one thing *causes* the other thing to occur
- Association does not mean causation!
- *Must* be very careful when making causal statements
  - Several factors must be considered (refer to Bradford-Hill Criteria discussed in class)



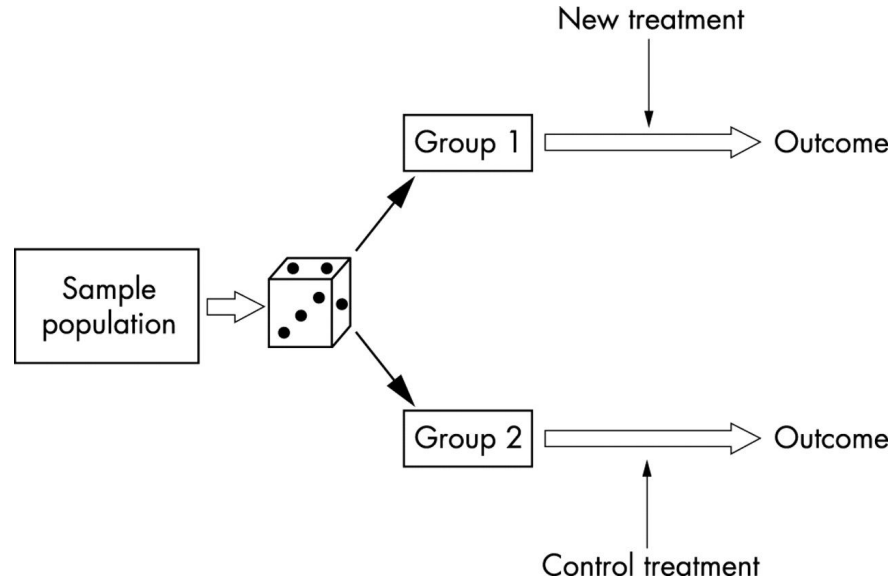
# Confounding

- There is a variable that influences both the dependent variable and independent variable

For a factor to be a **potential confounding variable** there has to be a **triangular relationship** between the **first risk factor**, the **potential confounding factor** and the **problem under investigation**, as shown in Figure

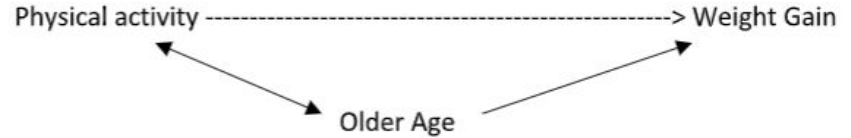
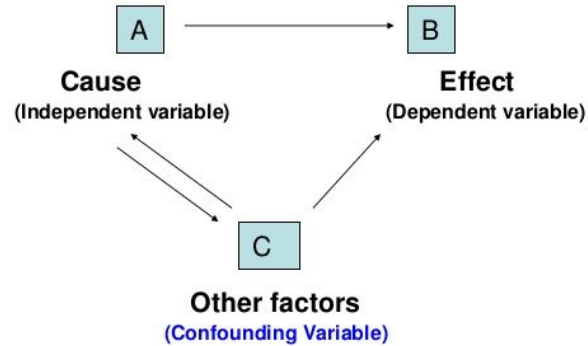


(The apparent association between A and B may be due to a third variable, C which associates with both A and B)



- In randomized controlled trials, confounding should theoretically not exist because groups are assigned at random
- This is not the case in observational studies

For a factor to be a **potential confounding variable** there has to be a **triangular relationship** between the first risk factor, the **potential confounding factor** and the **problem under investigation**, as shown in Figure

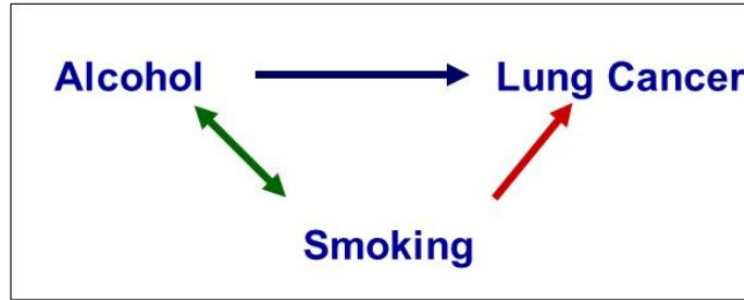


(The apparent association between A and B may be due to a third variable, C which associates with both A and B)

Dr. RS Mehta, MSND

- There's MANY reasons why people engage in different behaviors and these may be related to both your independent and dependent variable
- Consider *age*
  - Older people less likely to be physically active and more likely to gain weight (due to biological/other reasons)

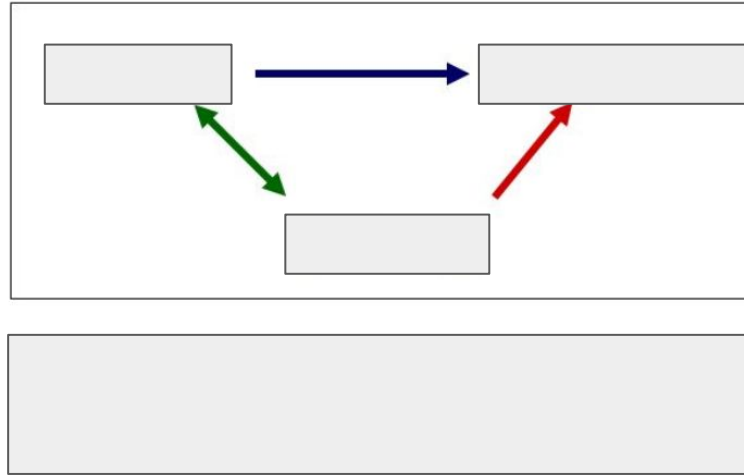
# Example



**Smoking is correlated with alcohol consumption and a risk factor even for those who do not drink alcohol**

- Apparent association between *Alcohol* and *Lung Cancer* may be due to our confounding variable, *Smoking*

# Example



- Suppose you were interested in investigating whether a Bachelor degree from certain universities, e.g. UofT, Oxford, McMaster, UWaterloo, Harvard, Ryerson, etc., were associated with higher job earnings
- Are there any confounders you can think of, why may this may be a confounder?

# **Ethics**

- Research involving humans and animals requires ethics board approval
- Such work must reach ethical and scientific standards
- For example, the research must not pose unnecessary risks to the participants
  - E.g. We could not conduct an experiment where we assign to people to smoke or not smoke.
    - Why do you think this is?

# Stanford Prison Experiment (1971)

- Social psychology experiment that attempted to investigate the psychological effects of perceived power
- 24 participants recruited; 12 “prisoners” and 12 “prison guards”
  - Guards became abusive
  - Prisoners became depressed and showed signs of extreme stress

## Ethics:

- Lack of debriefing
- Lack of protection to prisoners/guards
- Poor debriefing of prisoners and training of guards

## Example (Homework Q4)

- A data scientist compiled data from several public sources (voter registration, political contributions, tax records) that were used to predict sexual orientation of individuals in a community
- What ethical considerations arise that should guide use of such data sets?

Homework Q3 and Q5 have other ethics examples.



# **Presentations**

30 min

# Presentation = Clear, Concise, Cohesive

1. The **purpose**. What are you studying? Why should we care about the analysis you've done?
2. A **summary of the methods** used. What did you do? Why did you do it this way?
3. A **summary of the results**. We don't need to know everything you found – only the most critical things relating to your purpose! Remember, sometimes less is more! If they include a figure, it should be clear and able to stand on its own (e.g. contain proper titles).
4. A **conclusion**. What is your take away message? Remember, a conclusion is not the place to present new findings.

# Presentation Topics

## **Group 1. Consider question 1a, specifically 1a iv).**

Interpret the p-value of this test to compare the mean improvement for Lumosity versus crossword puzzles. How does it compare to the p-value estimated using the randomization test earlier in this question? Is this surprising? Why or why not? Make sure to explain the methods you used.

## **Group 2. Consider question 1b.**

What type of study did Hardy et al. conduct? What were the conclusions? Are there any limitations?

## **Group 3. Consider question 1c.**

Is age a confounder of this association? Why or why not?

## **Group 4. Consider Question 1d and 2.**

What ethical considerations did Hardy et al. make in their study? Why were these steps necessary? Consider the Statistical Society of Canada (SSC) Code of Ethical Statistical Practice, what practices should you consider while completing your poster project?

# During Presentations

- While listening to presentations, one member from each group should complete the oral presentation rubric
- Write down at least one question for the group