Reminders

- Email me and the TAs before 5pm tonight if you have any questions/concerns about grading of the quiz or homework

Warm Up

- A telephone survey of 852 randomly selected households and cell phones throughout the U.S., conducted by the University of Michigan in April 2013, reports that 63% of adults in the sample believe that there is solid evidence of global warming.
  - ⋆ **What is the parameter?**
  - ⋆ **What is the statistic?**
  - ⋆ **What is the value of the statistic?**

Chapter 5: Experiments, Good and Bad

Aaron Zimmerman
STAT 220 - Summer 2014

Department of Statistics
University of Washington - Seattle

Motivating Example

- Headline from two years ago in Science News: "Drinking Coffee Gives Jolt to Life Span"
  - ⋆ About 18 years ago, researchers collected surveys from 402,260 individuals
  - ⋆ Respondents were asked how many cups of coffee they drink every day
  - ⋆ After 15 years, researchers checked death records to see how many respondents had died
  - ⋆ **Finding**: Respondents who drank two or more cups of coffee per day were 10 to 16 percent *less likely* to have died after 15 years than respondents who didn't drink coffee
- **Question**: Does this mean that drinking coffee causes people to live longer?

Terminology

- A **response variable** is a variable that measures an outcome or result of a study
  - ⋆ In the coffee study, the response variable is whether each respondent was still alive after 15 years
- An **explanatory variable** is a variable that we think explains or causes changes in the response variable
  - ⋆ In the coffee study, the explanatory variable is the number of cups of coffee each respondent drinks per day

Rule of thumb: If your research question is "Does A cause B?", the explanatory variable is what you're using to measure A and the response variable is what you're using to measure B

Association vs. Causation

- The response variable and explanatory variable are **associated** if there is a relationship between the variables*
  - ⋆ There is an association between drinking coffee and life span: people who drink coffee tend to live longer than people who dont drink coffee
  - ⋆ *This definition is not in Chapter 5

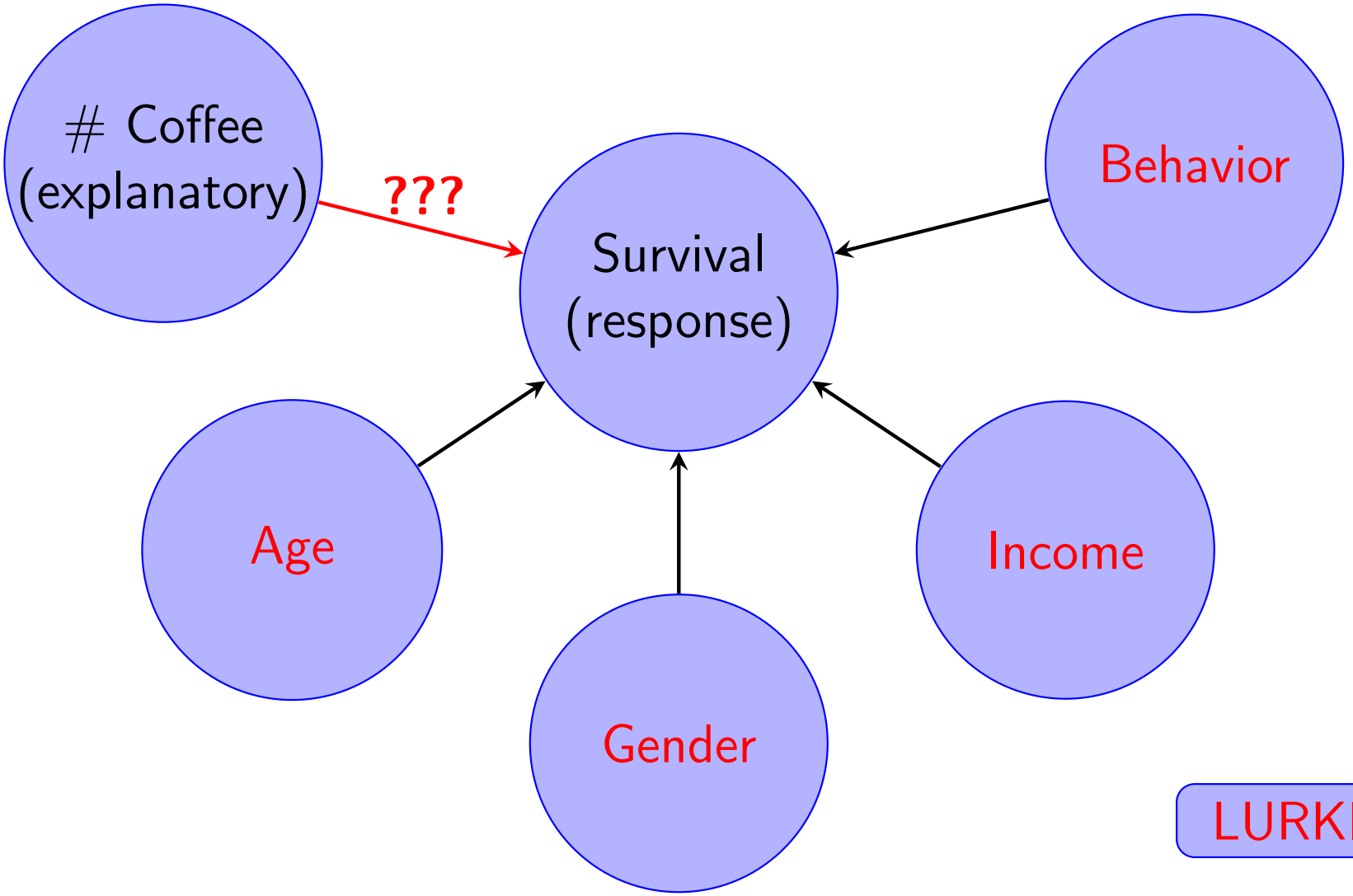
Key point #1: If two variables are associated, it does **NOT** mean that one causes the other!

⋆ "Association does not imply causation"

Key point #2: Observational studies only measure the association between two variables

⋆ To measure causation, you must perform an experiment

Does drinking coffee cause people to live longer?



Terminology

- A **lurking variable** is a variable that has an important effect on the relationship among the variables in a study but is not one of the explanatory variables studied

  ⋆ In the coffee study, age, income, gender, and other behavioral choices are all lurking variables because they could all affect a person's chance of survival after 15 years

- Two variables are **confounded** when their effects on a response variable cannot be distinguished from each other.

  ⋆ Confounding between lurking variables and explanatory variables can lead to bias in observational studies

Does drinking coffee cause people to live longer?

- Each of the following is a plausible explanation for how confounding could lead to bias in the coffee study:
  - ⋆ Young people may be more likely to drink coffee than old people, so coffee drinkers may be less likely to die because they because they are younger, not because they drink coffee
  - ⋆ Wealthy people may be more likely to drink coffee than poor people, so coffee drinkers may be less likely to die because they they are more wealthy, not because they drink coffee
  - ⋆ People who exercise frequently may be more likely to drink coffee than people who exercise less frequently, so coffee drinkers may be less likely to die because they exercise more often, not because they drink coffee
- **Conclusion**: we cannot conclude from this study that drinking coffee causes people to live longer

Example #1: Medicine

- A doctor wants to know whether wearing a nicotine patch helps patients who smoke to quit smoking. He asks each of his patients who smokes whether he or she is using the patch, and then asks three months later if he or she has quit smoking. He finds that patients who smoke and use the patch are five times more likely to quit smoking than patients who smoke and do not use the patch.

  - ⋆ **What is the explanatory variable?**
  - ⋆ **What is the response variable?**
  - ⋆ **What is a potential lurking variable?**
  - ⋆ **Why could it lead to bias?**

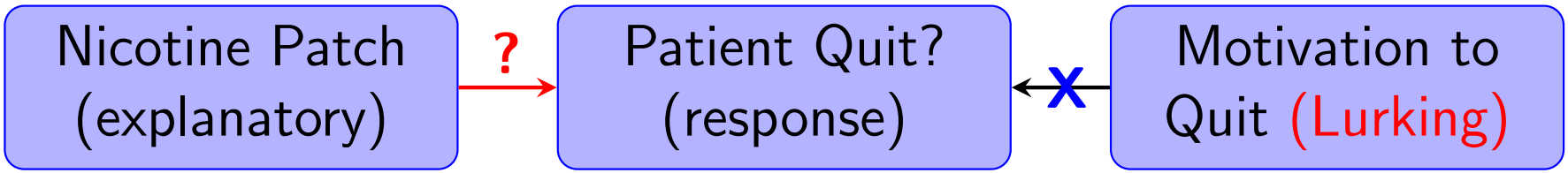
Example #2: Plant Ecology

- An ecologist wants to know whether higher pollinator visitation rates cause an increase in the proportion of flowers on Mount Rainier that set fruit. She takes a random sample of flowers and records the number of pollinator visits per day for each flower and whether each flower sets fruit by the end of the summer. She finds that flowers with that are visited by pollinators more than once per day are three times more likely to set fruit by the end of the summer.
  - ⋆ **What is the explanatory variable?**
  - ⋆ **What is the response variable?**
  - ⋆ **What is a potential lurking variable?**
  - ⋆ **Why could it lead to bias?**

Example #3: Education

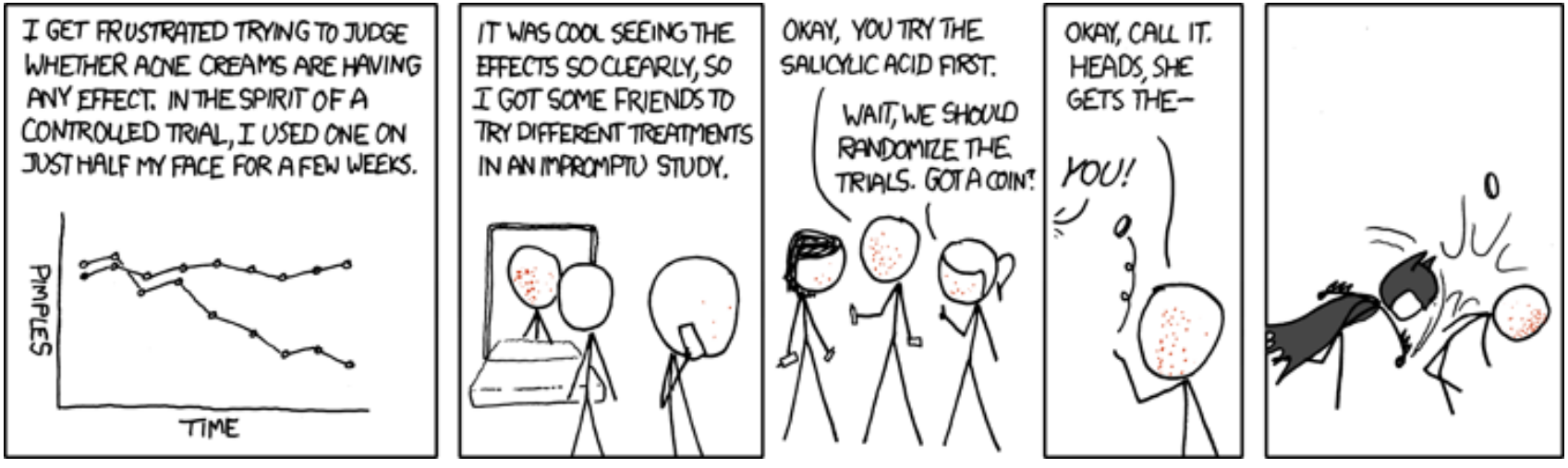
- A school district superintendent wants to know whether a professional development program improves the math knowledge of elementary school teachers. He records whether each teacher in his district attended a voluntary professional development meeting, and then gives a math content quiz to each teacher in the district. He finds that teachers who attended the professional development scored 20 percentage points higher on the content quiz than teachers who did not attend.
  - ⋆ **What is the explanatory variable?**
  - ⋆ **What is the response variable?**
  - ⋆ **What is a potential lurking variable?**
  - ⋆ **Why could it lead to bias?**

Example 1 Revisited



- The observational study in Example 1 might be biased because patients who choose to wear a nicotine patch may be more motivated to quit smoking than patients who dont wear a nicotine patch
- However, what if we randomly decide which patients wear the nicotine patch?
  - ⋆ Then, patients who wear the patch shouldn't be any more motivated, on average, than patients who dont wear the patch
  - ⋆ **This is the key idea behind randomized experiments**

XKCD comic #552



Terminology for Experiments

- The individuals studied in an experiment are often called **subjects**
- A **treatment** is any specific experimental condition applied to the subjects.
  - ⋆ Subjects who receive the same treatment are in the same **treatment group**

Terminology for Experiments

- In a **randomized comparative experiment**, the *average* response for subjects in each treatment group of interest is compared to the average response for subjects in the **control group**

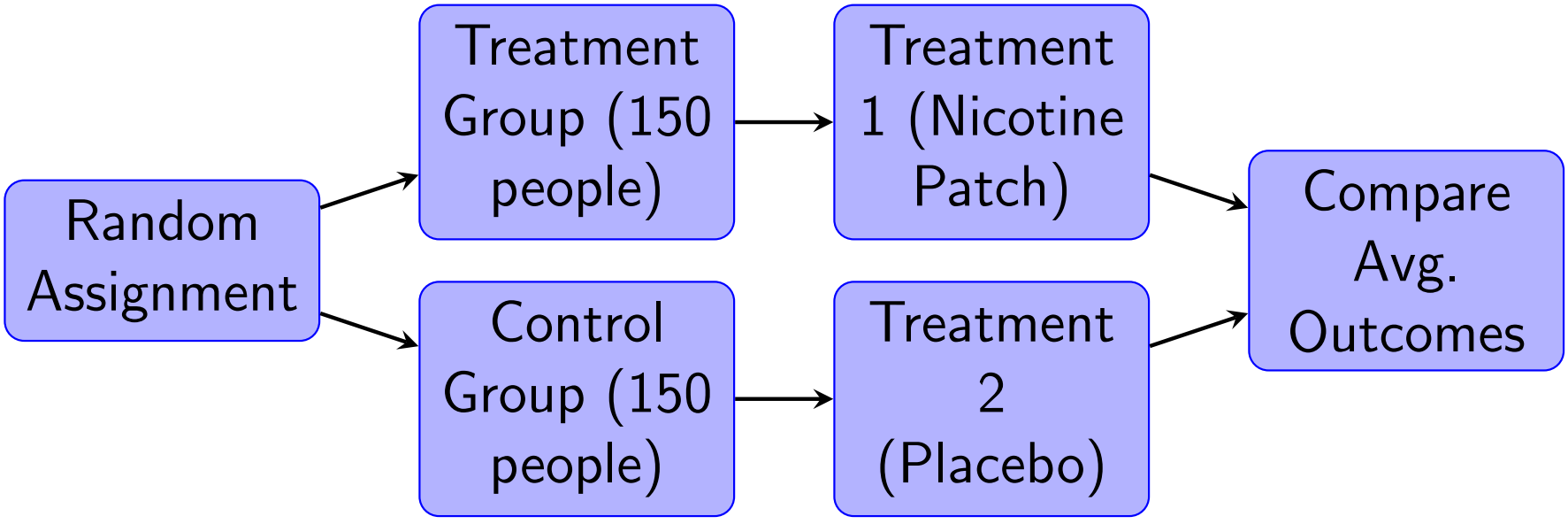
  ⋆ Depending on the question, subjects in the control group may receive no treatment, an alternative treatment, or a **placebo**, which is a dummy treatment with no active ingredients

  ⋆ Many patients respond favorably to any treatment, even a placebo. This response to a dummy treatment is the **placebo effect**

Experiment for Example 1

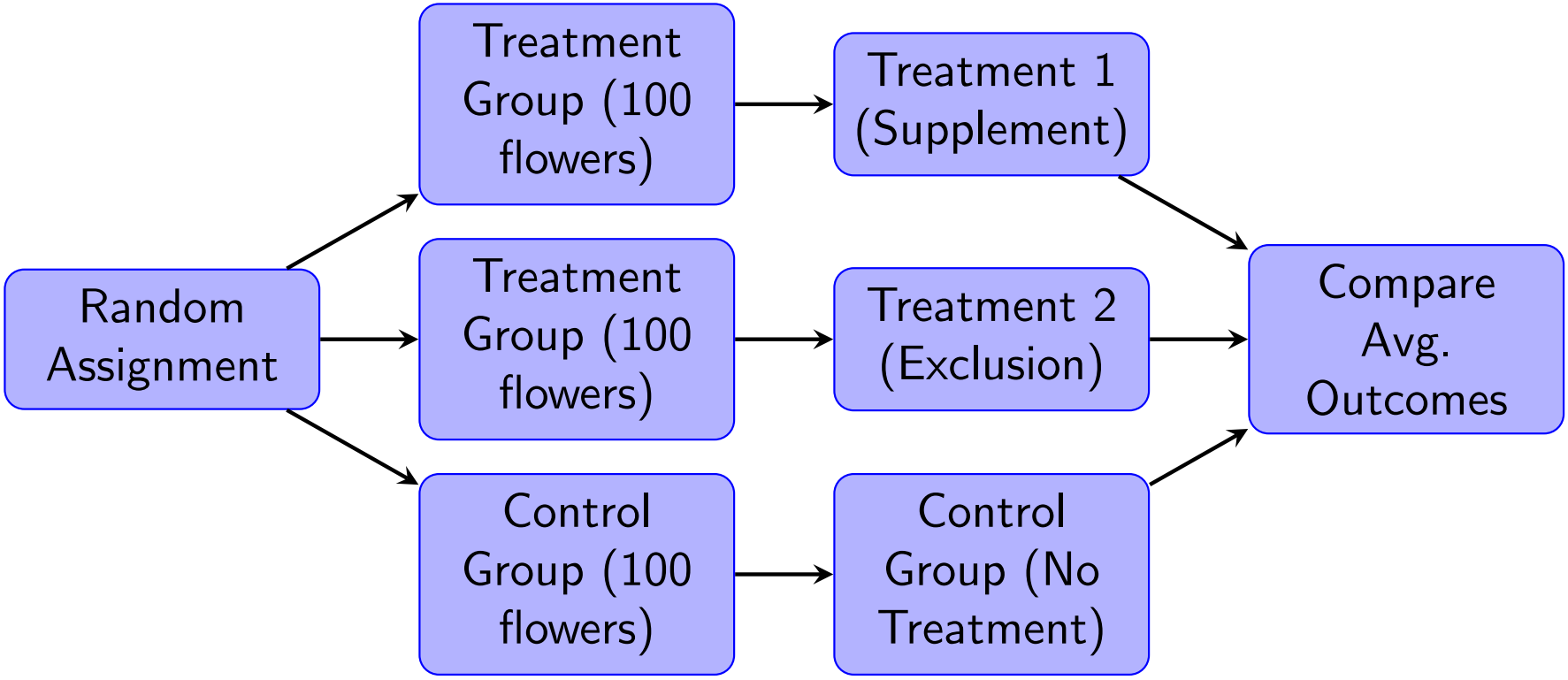
- The doctor randomly selects 300 patients who smoke and asks each to wear a patch for three months.
  - ⋆ The 300 patients are the **subjects** in the study
- He randomly assigns 150 of them to wear the nicotine patch, and the other 150 to wear a patch that injects nothing into the patient
  - ⋆ The 150 subjects selected to wear the nicotine patch are in the **treatment group**. The nicotine patch is the **treatment** in this group
  - ⋆ The other 150 subjects are the **control group**. The treatment in this group is a **placebo**
- At the end of three months, the doctor compares the percent of patients who quit smoking in the treatment and control groups

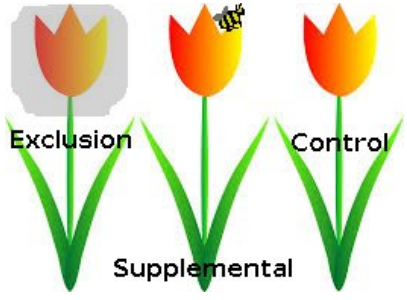
Experiment for Example 1



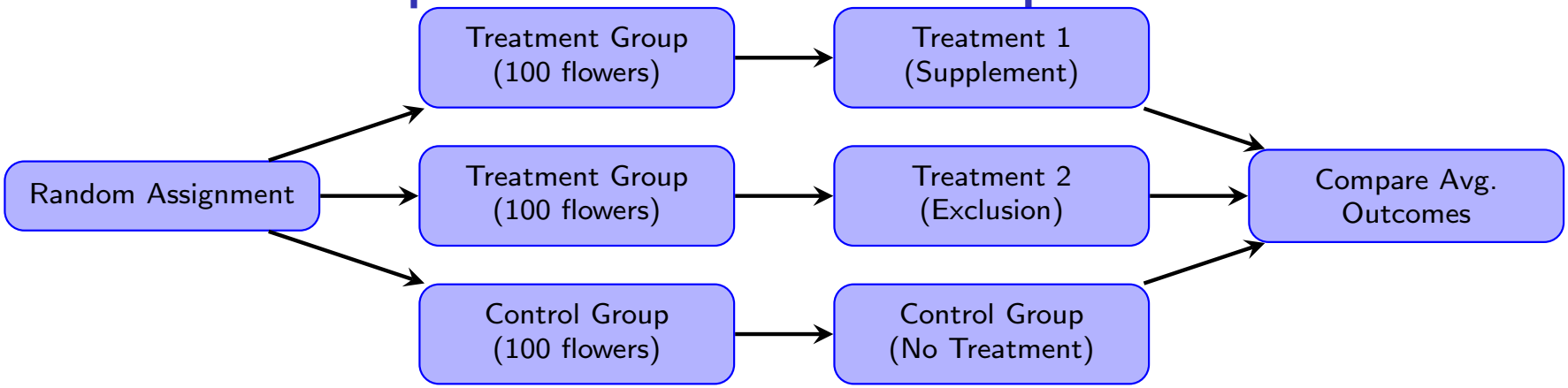
Lurking variables do not bias the results because the treatment is randomly assigned. Patients in the treatment group shouldn't be different, on average, than patients in the control group.

Experiment for Example 2





Experiment for Example 2



- The **subjects** are the 300 flowers
- There are two **treatment groups**: the **treatments** are supplemental pollen and pollen exclusion
- The **control group** receives no treatment
- There is no **placebo** in this study

Lurking variables do not bias the results because the treatment is randomly assigned. Flowers in the treatment groups shouldn't be different, on average, than flowers in the control group.

Experiment for Example 3

- The school superintendent selects 100 elementary teachers and randomly selects 50 of them to attend the professional development training. After the training, he gives the content quiz to all 100 teachers and compares the average score for teachers who attended the training and teachers who did not attend.

  - ⋆ **Who are the subjects?**
  - ⋆ **What is the treatment and the treatment group?**
  - ⋆ **What is the control group?**
  - ⋆ **Is there a placebo?**
  - ⋆ **Why do lurking variables not bias the results of the experiment?**

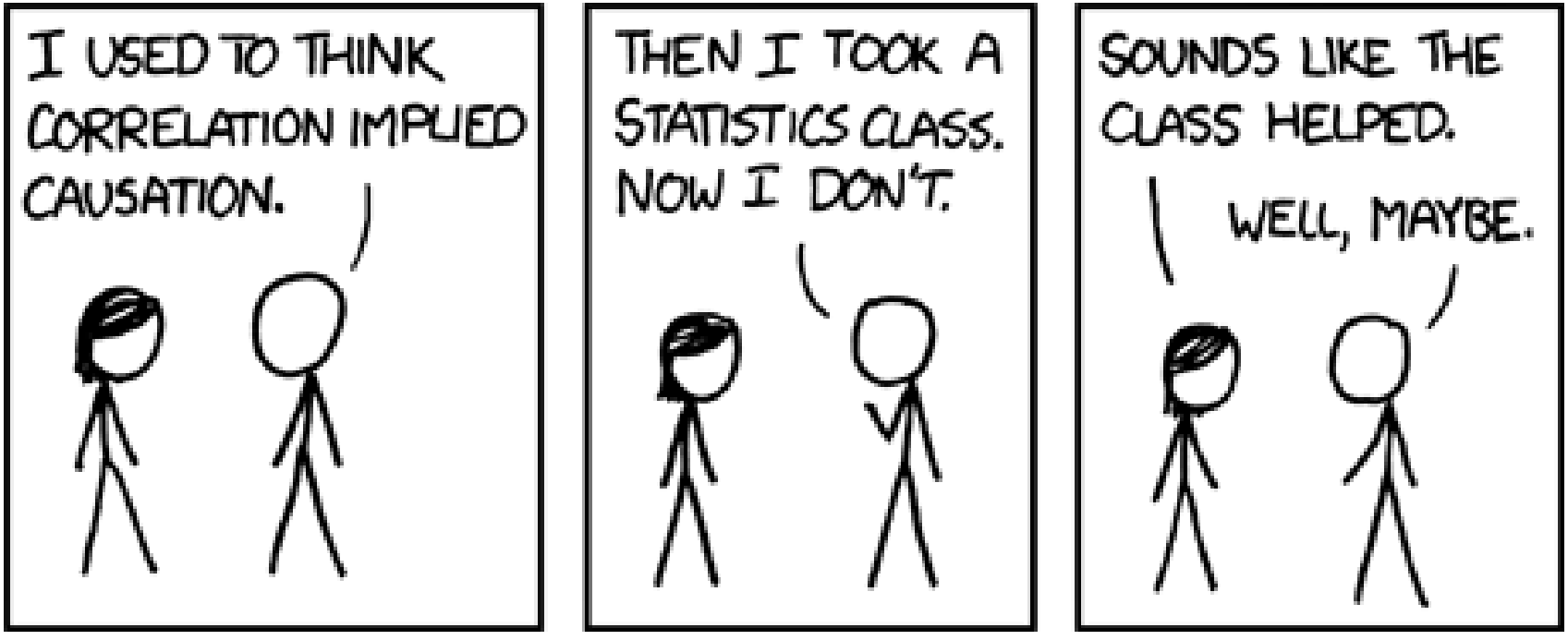
Statistical Significance

- Randomization eliminates *only* systematic differences between groups in an experiment. It eliminates bias.
- Random sampling error might still lead our results to be different than the truth
  - ⋆ *Increasing the sampling size can help*
- A observed effect from a treatment that would rarely occur by chance is called **statistically significant**
  - ⋆ We'll discuss this more at then

A Final Note

- Some questions are very difficult to address with an experiment
  - ⋆ In the coffee example, we would need to randomly select a treatment group and force those subjects to drink two cups of coffee every day for 15 years
- So, many questions still have to be answered with observational studies
  - ⋆ Good authors always stress that the results of these studies only demonstrate **association**, not **causation**
  - ⋆ Title of research article about coffee drinking: *Association of Coffee Drinking with Total and Cause-Specific Mortality*
  - ⋆ There are methods to reduce the bias in observational studies that we will discuss in week 6

XKCD comic #552



Correlation doesn't imply causation, but it does waggle its eyebrows suggestively and gesture furtively while mouthing 'look over there'.

Homework #3

- Read Chapter 5
- Complete problems 5.1, 5.2, 5.4, 5.5, 5.6ab, 5.7, 5.8, 5.11abe, 5.14, and 5.21