Introduction to Epidemiological and Biostatistical Thinking

UW Neurology Fellowship

Marlena Bannick

6/18/2020

PhD Student, University of Washington Dept. of Biostatistics
Researcher, Institute for Health Metrics and Evaluation

Goal

Introduce you to epidemiological thinking and key (bio)statistical concepts that you can use to critically interpret scientific studies in health and medicine.

Learning Objectives I.

1. **Basics**. Identify key elements of an epidemiological study and how they relate to the scientific question
2. **Study Design**. Recognize the basic types of epidemiological study design and identify when each design is appropriate for the scientific question
3. **Bias**. Recognize sources of bias in study designs or measurements and understand how they might affect your ability to answer the scientific question

Learning Objectives II.

4. **Modeling**. Understand how you can formulate your understanding about a data generating process, assumptions, and a hypothesis to test in a statistical model
5. **Inference**. Recognize the distinction between an effect size, a confidence interval, and a p-value as they relate to parameters that are estimated in a statistical model

Basics

A epidemiological study should be generated by a *scientific question of interest*. Broadly, you can think of these scientific questions falling into two main categories:

- **Descriptive**: What is the incidence rate of ischemic stroke (IS) in women aged 45 - 60 years old?
- **Inferential**: What is the effect of an experimental treatment on mortality following ischemic stroke in women aged 45 - 60?

From a statistical point of view it is not a clean distinction because you still use statistical tools to *infer* the incidence rate for a descriptive study.

Basics

The questions *who, what, where, when* have never been more important than in the context of epidemiology!

Having a well-defined scientific question means having clear answers for the following components:

- **Exposure**: What is the group in study exposed to that you want to measure the effect of, and over what period of time?
- **Population**: Who is the group being studied?
- **Outcome**: What outcome is being studied (either in relation to the exposure or on its own) and over what period of time?

The *why* is also important! Epidemiological studies should serve some purpose.

Basics: Measures

Once you've defined your target exposure, outcome, and population that makes up your scientific questions, understanding **measurement** of the outcomes is of utmost importance.

Some common outcome measurements in the context of health sciences are

- **prevalence**: proportion of a population with an outcome
- **incidence**: rate of getting the outcome among individuals in a population that did not already have the outcome ("risk")
- **remission**: rate of returning to be outcome-free among those that had the outcome

Think about denominators!

Basics

What are the exposure, outcome, and population for each of these scientific questions?

- **Descriptive**: What is the incidence rate of ischemic stroke (IS) in women age 45 - 60 years old?
- **Inferential**: What is the effect of an experimental treatment on mortality following ischemic stroke in women age 45 - 60?

Table 1: Basic Elements of Study Design

	Descriptive	Inferential
Exposure		
Outcome		
Population		

Basics

What are the exposure, outcome, and population for each of these scientific questions?

- **Descriptive**: What is the incidence rate of ischemic stroke (IS) in women age 45 - 60 years old?
- **Inferential**: What is the effect of an experimental treatment on mortality following ischemic stroke in women age 45 - 60?

	Descriptive	Inferential
Exposure		experimental treatment
Outcome	ischemic stroke (IS)	death from IS
Population	women age 45-60 without IS	women age 45-60 with IS

How would you make these questions more precise?

Study Design

Biases

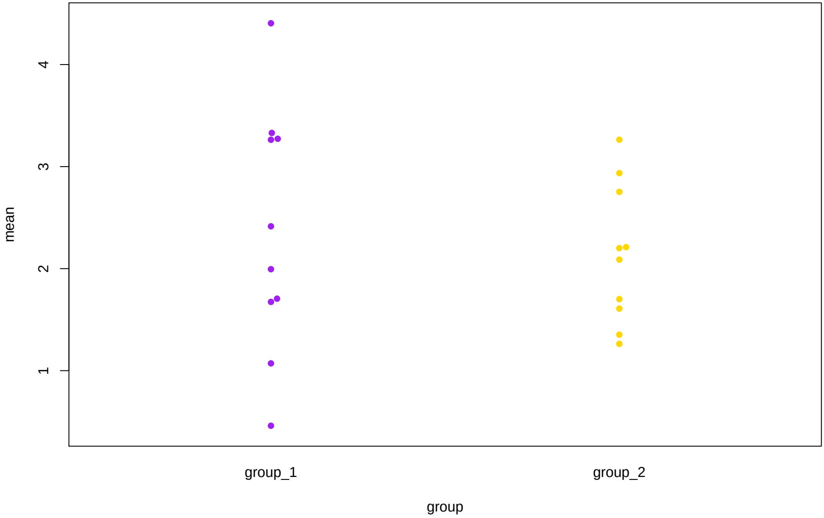
Biases in the epidemiological context are any factors in your study that *prevent* you from being able to answer your precise scientific question.

Biases may result from systematically incorrect measurements of the outcome, the exposure, or the population.

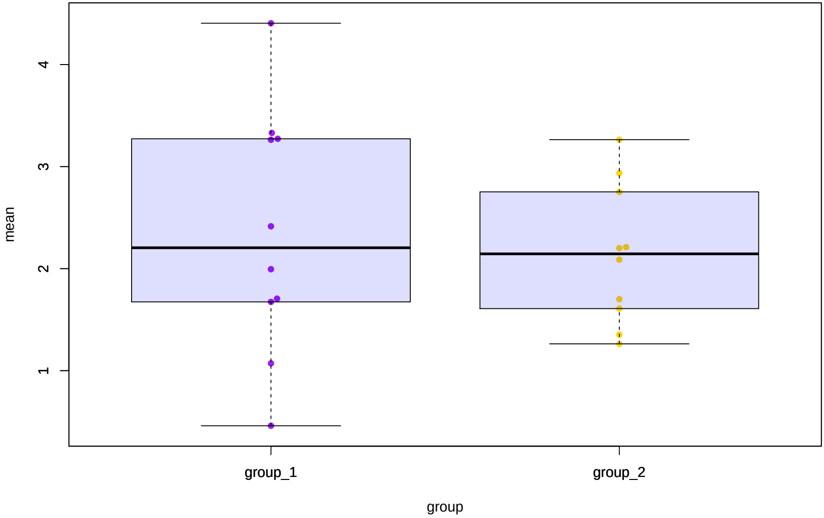
Examples of biases include:

- **Selection bias**: the population that you want to study is not the population that is actually in your study
- **Recall bias**: individuals are being asked about exposures or outcomes that they do not remember correctly
- **Social desirability bias**: individuals are not comfortable disclosing their true exposure or outcome status for fear of judgement by others

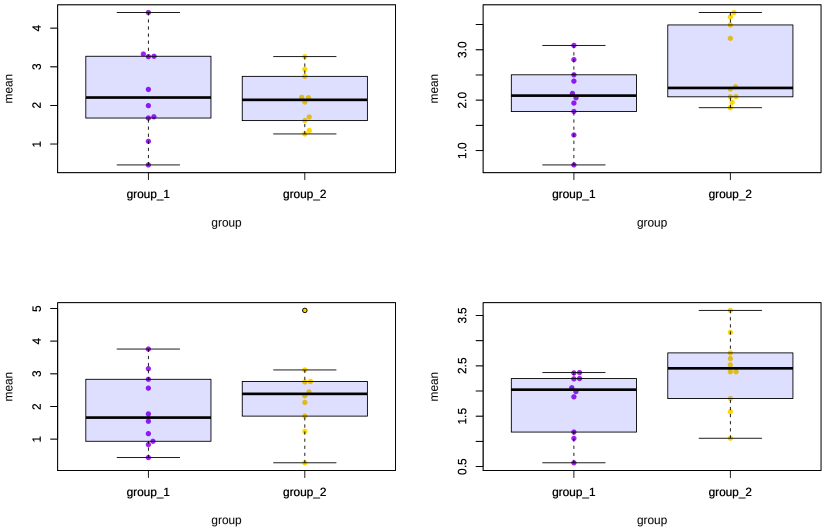
Inference: Simple Means



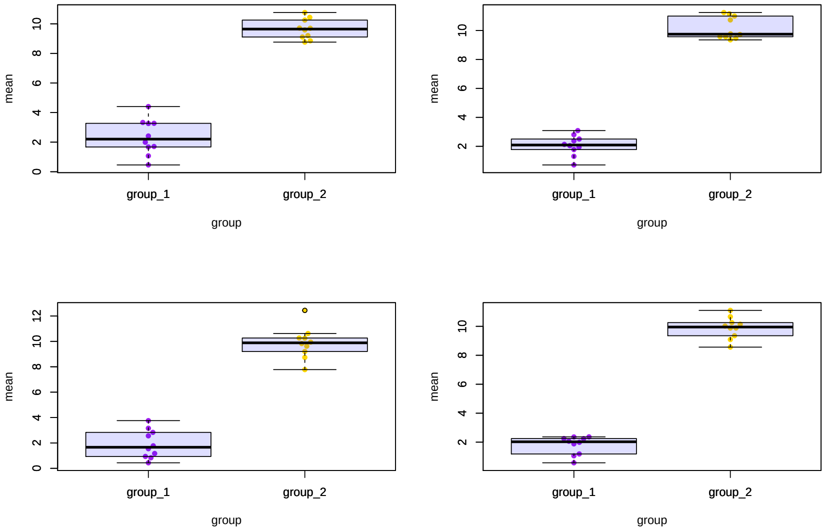
Inference: Simple Means



Inference: Small Effect Size, Small Sample Size



Inference: Large Effect Size, Small Sample Size



Inference: Small Effect Size, Large Sample Size