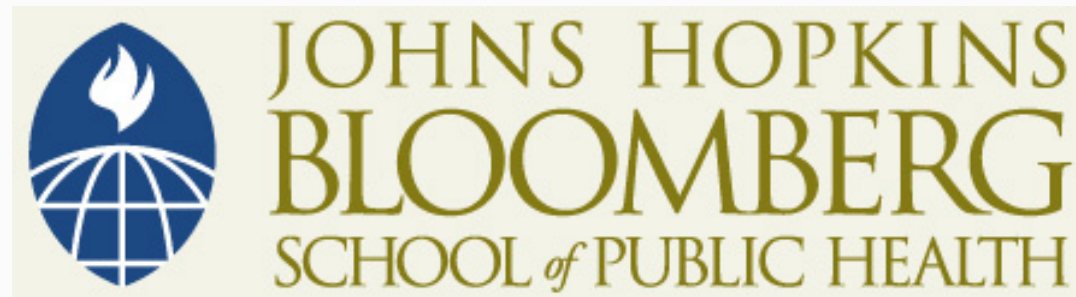


This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2009, The Johns Hopkins University and John McGready. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Describing Data: Part I

John McGready
Johns Hopkins University

Lecture Topics

- What role does statistics have in public health?
- Types of data: continuous, binary, categorical, time-to event
- Continuous data: numerical summary measures
- Continuous data: visual summary measure
- Sample data versus population (process) level data



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section A

What Role Does Biostatistics Play in Public Health (Why Do I Need this Stuff?)

Data Is Everywhere!

- Data is utilized and summarized frequently in research literature
- From *Archives of Surgery* article, August 2000:
 - “**Hypothesis:** Surgeon-directed institutional peer review, associated with positive physician feedback, can decrease the morbidity and mortality rates associated with carotid endarterectomy.”
 - “**Results:** Stroke rate decreased from **3.8%** (1993-1994) to **0%** (1997-1998). The mortality rate decreased from **2.8%** (1993-1994) to **0%** (1997-1998). (Average) length of stay decreased from **4.7 days** (1993-1994) to **2.6 days** (1997-1998). The (average) total cost decreased from **\$13,344** (1993-1994) to **\$9,548** (1997-1998).”

Data Is Everywhere!

- Data is utilized and summarized with statistics frequently in popular media
- From cnn.com, Monday July 8th, 2008:
 - “For the first time, an influential doctors group is recommending that some children as young as eight be given cholesterol-fighting drugs to ward off future heart problems . . . With **one-third** of U.S. children overweight and about **17 percent obese**, the new recommendations are important,’ said Dr. Jennifer Li, a Duke University children's heart specialist.”

Data Is Everywhere!

- Data is utilized and summarized with statistics frequently in popular media
- From *Washington Post*, June 27th, 2008:
 - “The number of young homosexual men being newly diagnosed with HIV infection is rising by **12 percent a year**, with the steepest upward trend in young black men, according to a new report.”

Data Provides Information

- Good data can be analyzed and summarized to provide useful information
- Bad data can be analyzed and summarized to provide incorrect/harmful/non-informative information

Steps in a Research Project

- Planning/design of study
- Data collection
- Data analysis
- Presentation
- Interpretation
- Biostatistics CAN play a role in each of these steps! (but sometimes is only called upon for the data analysis part)

Biostatistics Issues

- Planning/design of studies
 - Primary question(s) of interest:
 - ▶ Quantifying information about a single group?
 - ▶ Comparing multiple groups?
 - Sample size
 - ▶ How many subjects needed total?
 - ▶ How many in each of the groups to be compared?
 - Selecting study participants
 - ▶ Randomly chosen from “master list?”
 - ▶ Selected from a pool of interested persons?
 - ▶ Take whoever shows up?
 - If group comparison of interest, how to assign to groups?

Biostatistics Issues

- Data collection

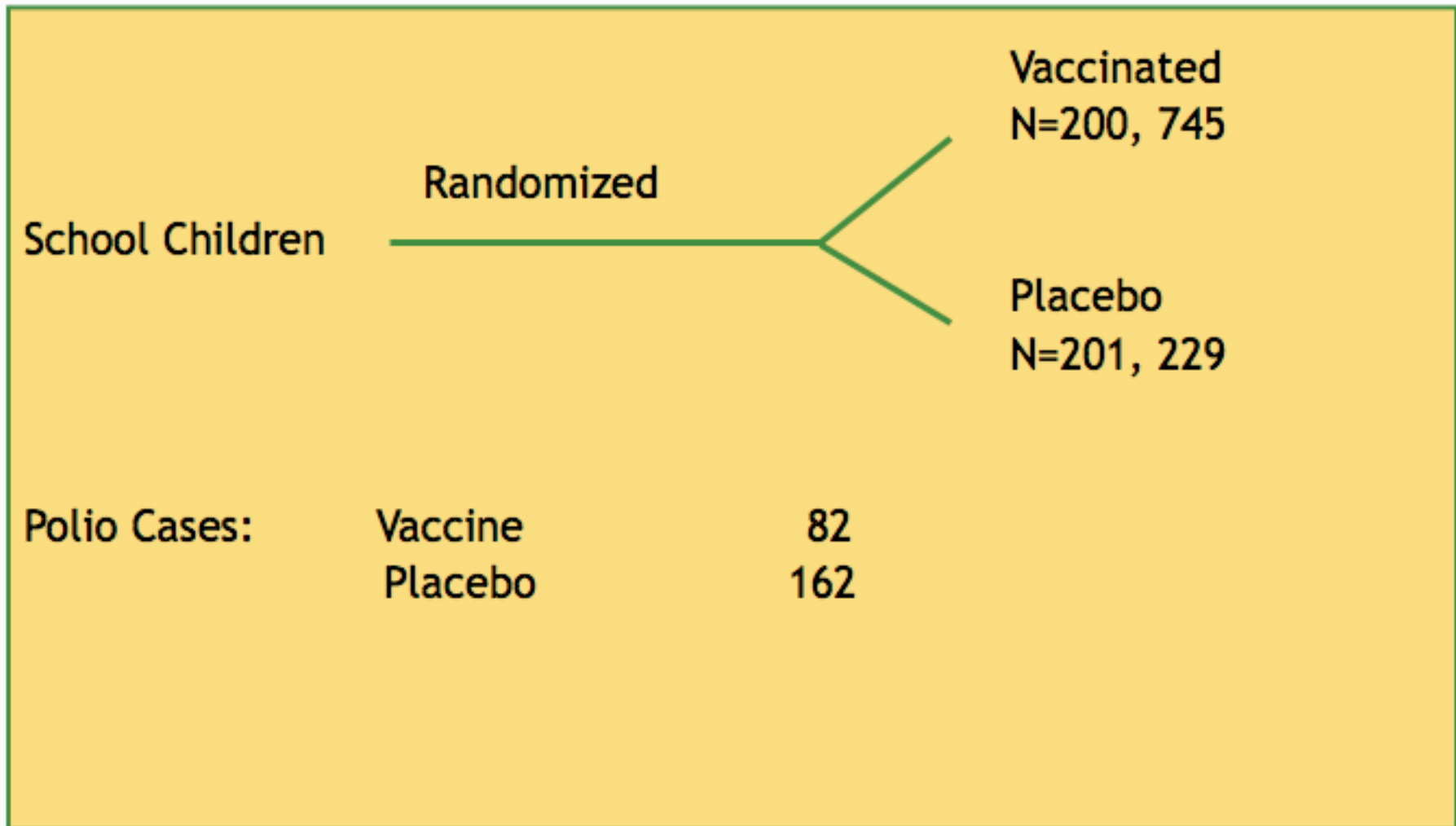
- Data analysis
 - What statistical methods are appropriate given the data collected?
 - Dealing with variability (both natural and sampling related):
 - ▶ Important patterns in data are obscured by variability
 - ▶ Distinguish real patterns from random variation
 - Inference: using information from the single study coupled with information about variability to make statement about the larger population/process of interest

Biostatistics Issues

- Presentation
 - What summary measures will best convey the “main messages” in the data about the primary (and secondary) research questions of interest
 - How to convey/ rectify uncertainty in estimates based on the data

- Interpretation
 - What do the results mean in terms of practice, the program, the population etc.?

1954 Salk Polio Vaccine Trial



Source: Meier, P. (1972), "The Biggest Public Health Experiment Ever: The 1954 Field Trial of the Salk Poliomyelitis Vaccine," In J. Tanur (Editor), *Statistics: A Guide to the Unknown*. Holden-Day.

Design: Features of the Polio Trial

- Comparison group
- Randomized
- Placebo controls
- Double blind
- Objective—the groups should be equivalent except for the factor (vaccine) being investigated

Analysis Question

- Question
 - There were almost twice as many polio cases in the placebo compared to the vaccine group
 - Could the results be due to chance?

Such Great Imbalance by Chance?

- Polio cases
 - Vaccine—82
 - Placebo—162
- Statistical methods tell us how to make these probability calculations