BIOSTAT 702: Module 1

Selecting and Describing Study Participants; Part 1: Flow of Study Participants

Dr. Marissa Ashner

Department of Biostatistics and Bioinformatics

Fall 2025

Module Goals

- Understand systematic approaches to describing how research participants were selected and analyzed
- Be able to use information about participant selection to assess the potential for selection bias, including concerns related to internal and external validity

Resources for this Module

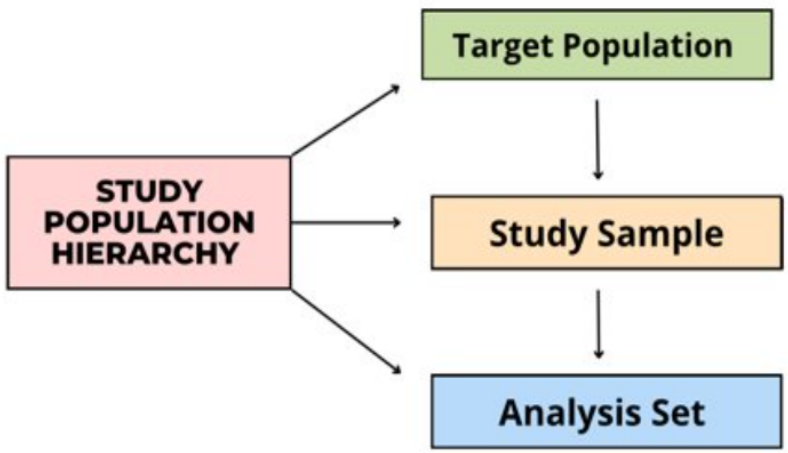
Canvas

- The Short Report that explains the STROBE statement
- The 'explanation and elaboration' paper on the STROBE statement
- The CONSORT 'explanation and elaboration' paper

Websites

- Article: The CONSORT 2010 statement
- Article: Where to look for the most frequent biases?
- Article: Value of Flow Diagrams in Reports of RCTs

Three Nested Populations in Biostatistical Studies



Target Population
– The full group of interest that you want to make conclusions about

Study Sample
– The group that you actually enroll / observe in the study

Analysis Set
– The group that you actually perform data analysis on

Ideal to Feasible: Target Population vs. Study Sample

study sample → target population

<

- You want the study sample to be representative of the target population
- This is the first narrowing step
  - practicality may limit who you are able to sample and therefore affect representativeness
- Usually described by a diagram or figure that starts with potentially eligible subjects and displays when, why, and how many subjects were lost at each stage

Example of a Target Population

- Investigators are studying a new drug for SAH (Sub-arachnoid hemorrhage: sudden bleed between brain & meninges)
  - Drug intended to shrink the volume of the "lesion" (i.e., the area of bleeding)
- They plan to include patients with very small lesions in a randomized trial of their drug versus usual care

Let's Talk about this Target Population

Who Actually Gets Analyzed: Study Sample vs. Analysis Set

- You also want the analysis set to be representative of the study sample
- Some causes of removing participants from the study sample when analyzing the data:
  - Missing Data
    - e.g., lost to follow up, lab samples being lost, device fails, skipping survey questions / “prefer not to answer”
    - *Note:* Missing Data is often handled other ways aside from removing it – this is highly dependent on the situation and beyond the scope of this course
  - Exclusion Criteria
    - e.g., post-hoc protocol deviation, safety stop, ineligibility based on pre-determined characteristics

Why does this matter?

- *Internal Validity:*
  - The degree to which the study is free from error
    - Random Error: occurs by chance and may lead to poor precision
    - Systematic Errors: Bias!
- *External Validity:*
  - The degree to which results of a study may apply to populations/groups that did not participate (i.e., the target population)
    - Can also be affected by bias!

Types of Bias

Selection Bias

- The relationship between the variables of interest differ in those who participate in the study and those who do not
  - e.g., sampling bias, confounding by indication, incidence-prevalence/survivor bias, attrition bias

Information Bias

- Errors in the measurement, collection, or interpretations of variables being analyzed
  - e.g., recall bias, interviewer bias, observer bias

Your Role

- When collaborating with investigators, it is your responsibility to investigate potential sources of bias
  - If entering the study at the design stage, perhaps you have a role in determining how to minimize bias from the target population to study sample
  - If entering at the analysis stage, perhaps you have a role in determining how to minimize bias from the study sample to analysis set
- Regardless, it is important to ensure interpretations are made correctly such that you are transparent about the populations that you anticipate your results are actually generalizable to

Adequate Reporting

- There are some guidelines for different study types regarding what you should report in a manuscript in order to increase transparency and rigor

Observational Studies

- Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)
  - 22 item checklist for reporting on cohort, case-control, and cross-sectional studies

Randomized Trials

- CONsolidated Standards Of Reporting Trials (CONSORT) Checklist
  - 25 item checklist for reporting on randomized trials

Example of the Flow of Participants

- Aim: identify unmet medical needs among U.S. veterans with spinal-cord injury (SCI)
- Data: VA admin file (1955–2024)
  - 2 billion records from 10 million veterans

How might we begin the process of choosing a sample here?

determine the age range, region (target population)

What are potential biases?

filter...

Example of the Flow of Participants

- Aim: identify unmet medical needs among U.S. veterans with spinal-cord injury (SCI)
- Data: VA admin file (1955–2024)
  - 2 billion records from 10 million veterans

How might we begin the process of choosing a sample here?

What are potential biases?

Example of the Flow of Participants

- Aim: identify unmet medical needs among U.S. veterans with spinal-cord injury (SCI)
- Data: VA admin file (1955–2024)
  - 2 billion records from 10 million veterans

How might we begin the process of choosing a sample here?

What are potential biases?