Biology/Statistics 2244

Statistics for Science

STUDY DESIGNS & CONSIDERATIONS

Objectives

By the end of this topic, you should be able to:

- apply vocabulary associated with study designs (e.g. observational, case-cohort, treatment, etc.);
- use methods to prevent confounding;
- design common study designs to address a research question.

Plan: Thinking about measurement

Question: *Can outdoor cycling improve well-being in seniors?*

Describe what you will use to characterize/ measure *well-being*.

Plan

Create a plan—including data collection and analysis—to address the research question(s)

- What are the sampling frame and sampling strategies?
- What will be measured for the response variable(s)?
- How will you deal with (potential) explanatory variables?
- What statistical procedures do you plan to use?

Confounding

presence of additional ("confounding") variables whose effects on the response cannot be separated from that of the factor(s) of interest

Example: *Can outdoor cycling improve cognitive function in seniors?*

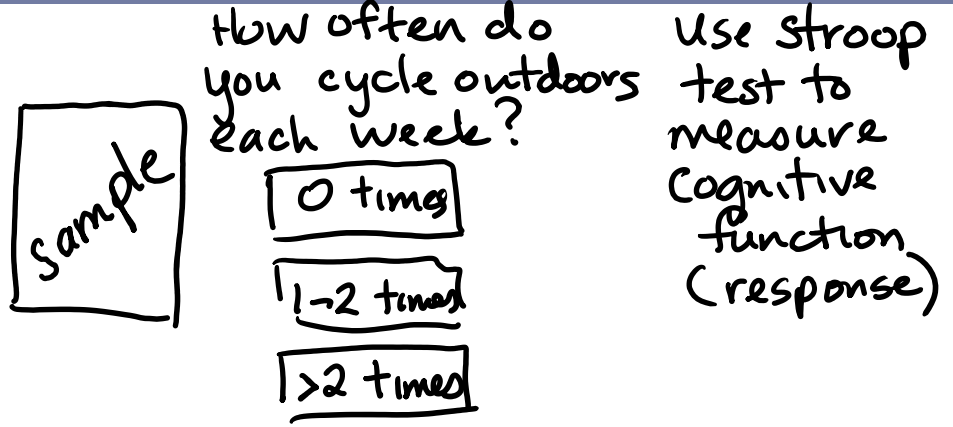
Group 1	Group 2	Group 3
		
Mean: +4	Mean: +1	Mean: +4

Images from Microsoft 365

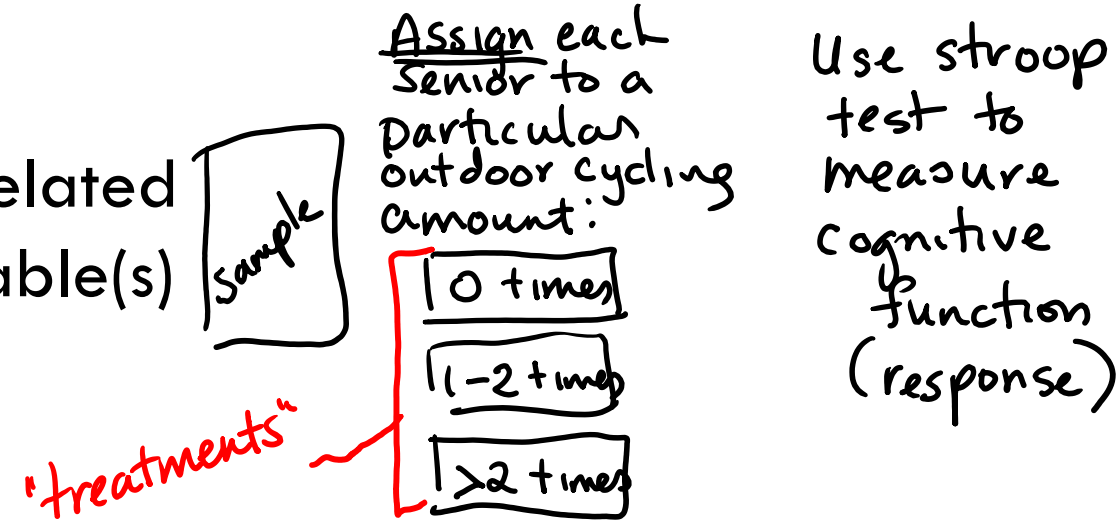
What is responsible for the observed group differences?

General classes of study design

Observational: measure both explanatory and response variables from units, as they naturally occur



Experimental: impose a condition ("*treatment*") related to the explanatory variable(s) to effect change on the response variable(s)



1 Control

accounting for variation in potential explanatory variables, to isolate impact of factor(s) of interest

A. Limit variation in the variables

Hold other variables constant

variable of concern: Social structure of activity

hold constant to all seniors performing activity alone

downfall: limits "ecological validity" (ie realism)

Select a narrower sampling frame

variable of concern: age

reduce/restrict Sampling frame to only seniors aged 70-75 y

downfall: reduces generalizability to entire pop^n. (ie undercoverage)

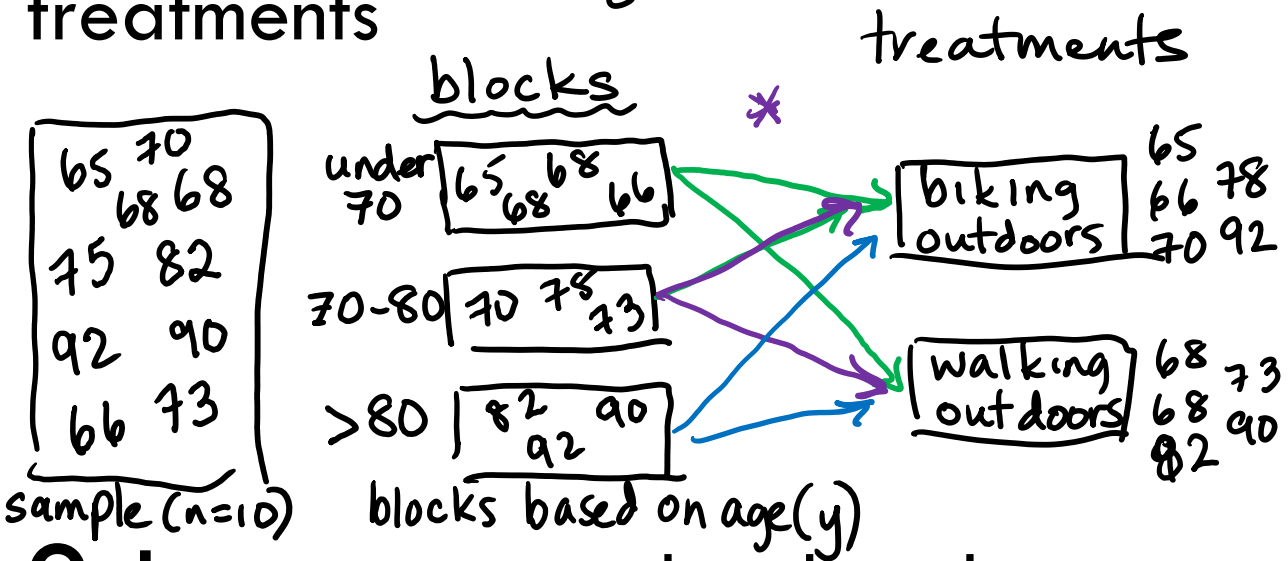
Control

* don't confuse with stratified sampling

implies experiment.

B. Distribute variation across treatments

Blocking: subdivide units into homogeneous groups based on variable of concern (age.) prior to assigning treatments



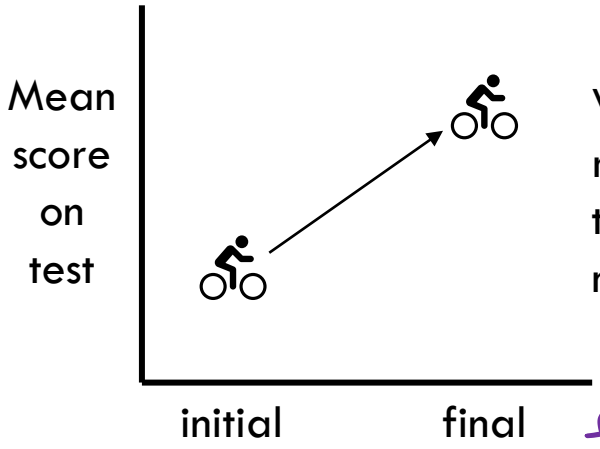
* ideally we assign to treatments randomly from each block.

Outcome: ensures treatment groups are similarly composed with respect to variable of concern

Control

C. Use comparison groups

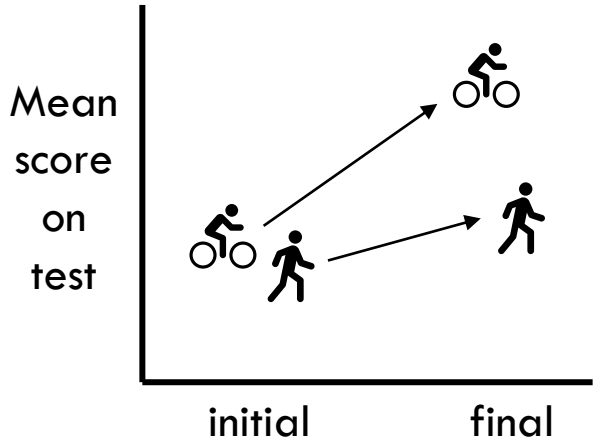
Non-comparative design



What could be responsible for the increase in mean score?

could be cycling OR familiarity with test OR change in seasons

Comparative design

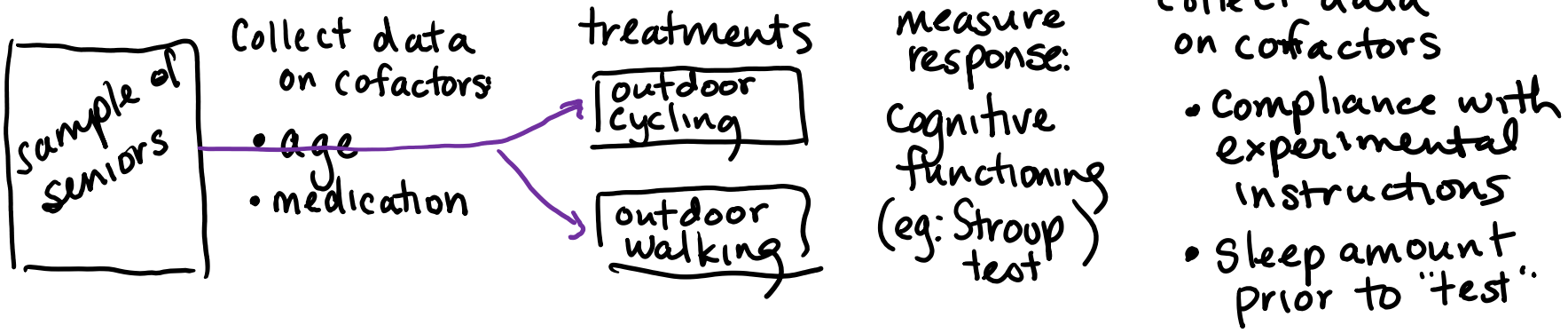


Why? Provides opportunity to attribute change to factor of interest, rather than 'natural' change

Control

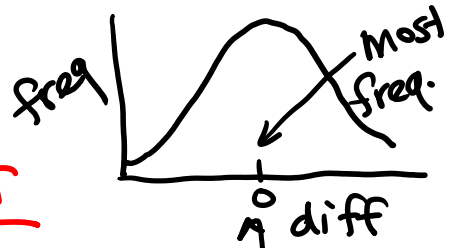
D. Collect data on variables of concern

Cofactors: Additional variables for which data is collected for comparison/explanatory purposes, but are not the factors of interest



Note: doesn't really control (prevent) for confounding variables, but could help us account for them

2 Randomization



implies experiment

using chance to assign individuals/units from the sample to treatments, or, the treatment order

- variation still exists in the sample for potential confounding variables
- similarly composed treatment groups are more frequent (ie probable) than dissimilar groups when randomizing
  - true for variables we didn't anticipate in advance
  - for our single randomization, we expect (trust) to get a similarly composed outcome.

Conclusion: randomization creates similarly composed treatment groups, reducing/eliminating confounding (when we can't block or otherwise explicitly control).

applet: https://www.rossmanchance.com/applets/2021/randomassignment/Subjects.html?subjects=2

3

Replication

* be careful not to confuse this with reproducing or repeating a study on ~~a~~ a new sample.

*inclusion of more than one **individual/unit** from the sample in a treatment or comparison group*

Treatment 1

3 replicates

Treatment 2

2 replicates

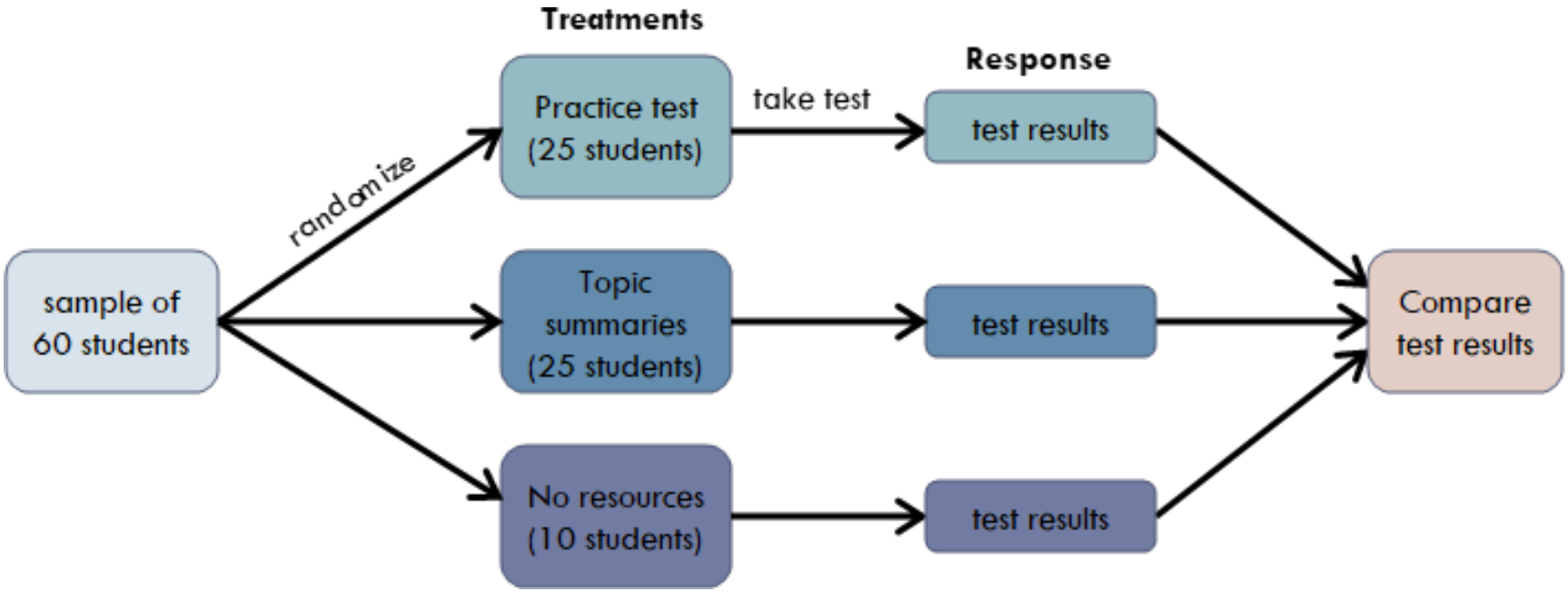
- accounts for variation in response among units
- can focus on typical or most frequent response

Images from Microsoft 365

Completely randomized design

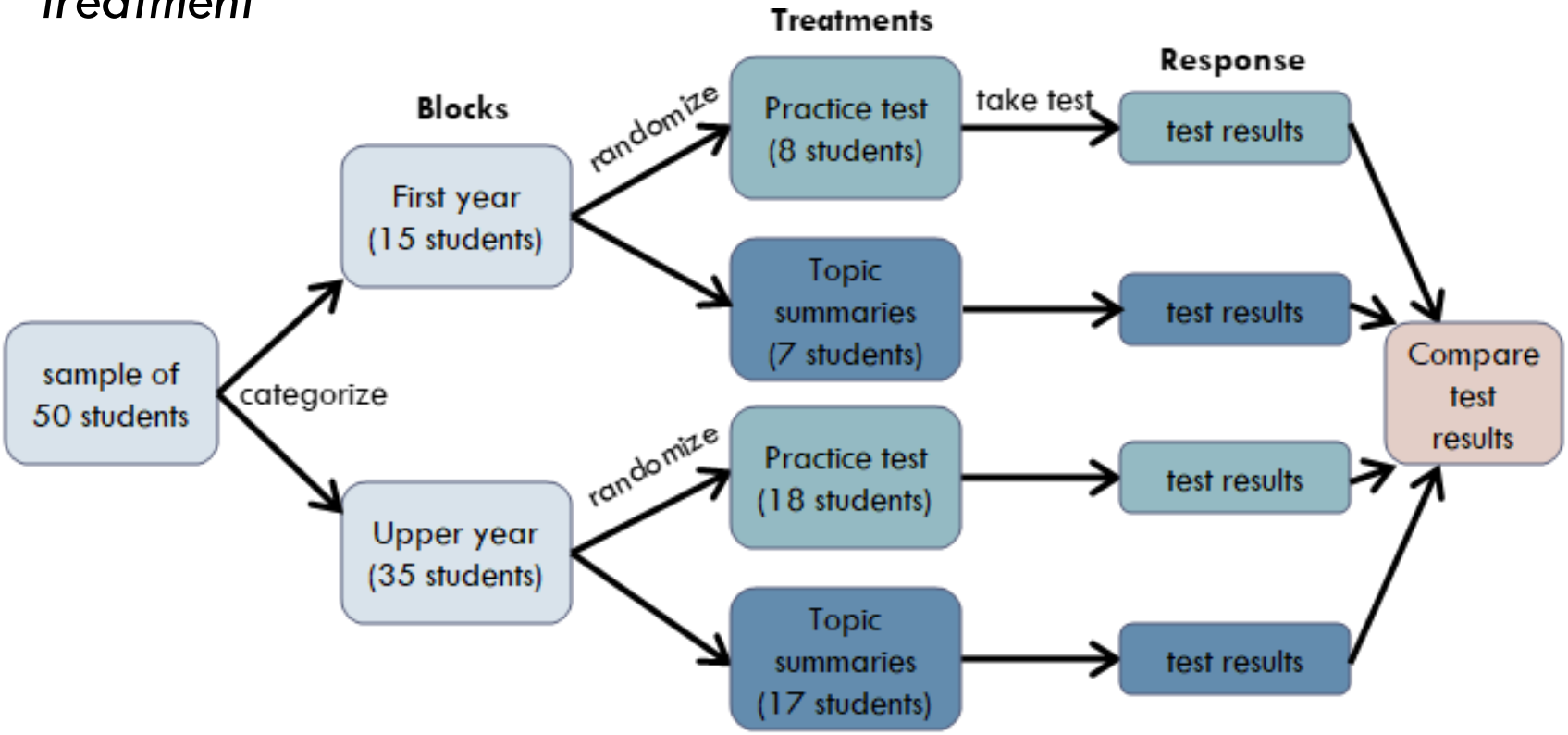
units from the sample are each randomly assigned ('randomized') to a treatment

Research Question: What impact do different studying resources have on academic performance in undergrads?



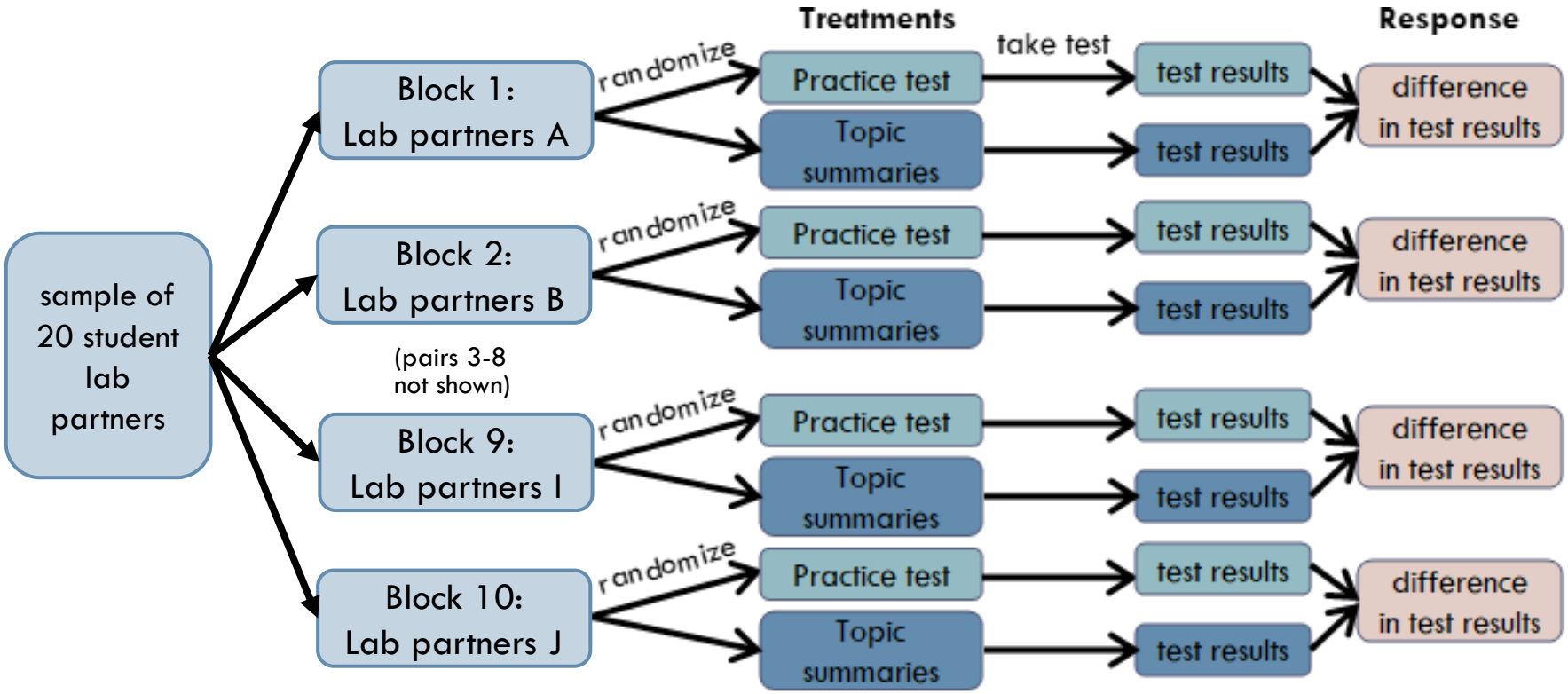
Randomized block design

units from the sample are subdivided into 'blocks' based on pre-existing characteristic(s), then randomly assigned from blocks to a treatment



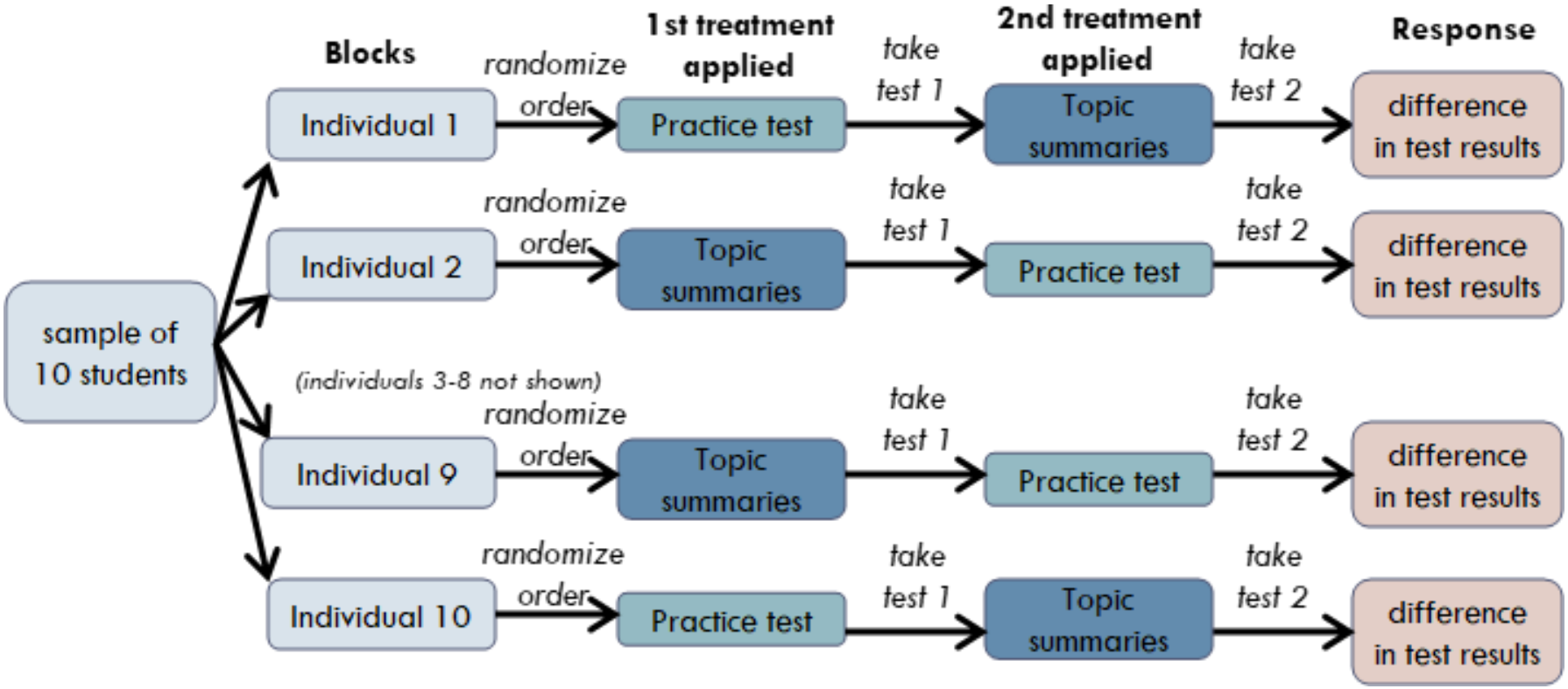
Matched pairs design

pairs of units from the sample are matched based on similarity across pre-existing characteristics, then randomly assigned from pairs to treatments.



Repeated measures design

each unit in the sample is assigned to both/all treatments, ideally with treatment order determined randomly



Survey

an observational study that collects information about variables of interest from a sample



Typically collect data on additional variables (e.g. demographics, environmental characteristics)

Cohort studies

examine the emergence of a specific condition over time in a homogeneous group of individuals

- typically prospective and longitudinal
- used to connect exposure factors to outcome
- may yield incidence rate of outcome

Case-Control study

a sample of 'cases' with the outcome of interest is selected, and compared against a sample of 'controls' known to not have the condition

- cases and controls can be 'matched'
- useful for rare conditions