Study Design

A statisticians point of view with focus on power and sample size

DATE: March 3, 2022

PRESENTED BY: Emile Latour, MS, Associate Biostatistician, OHSU Knight Cancer Institute

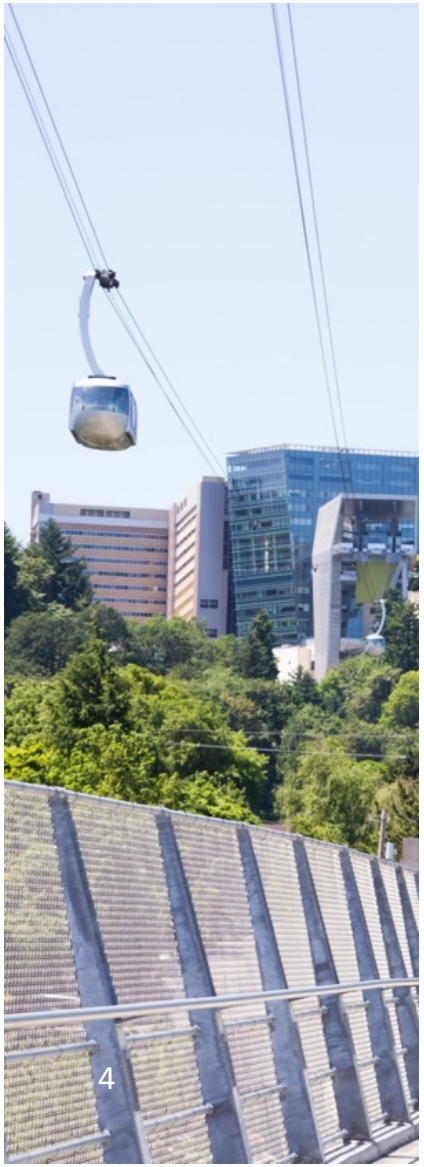
Acknowledgement

- Meike Niederhausen, PhD, Biostatistics & Design Program (BDP)
- Biostatistics, Epidemiology, & Research Design (BERD) Seminar, OCTRI Research Forum
  - Slides http://bit.ly/BERD-PSS-101
  - Recording https://echo360.org/media/c14ae529-27fe-4208-ae27-95c7c1adc928/public
- Upcoming Thursday March 31, 2022 over Zoom.
  - Register through Compass: "Power and Sample Size 101"

OHSU

About me

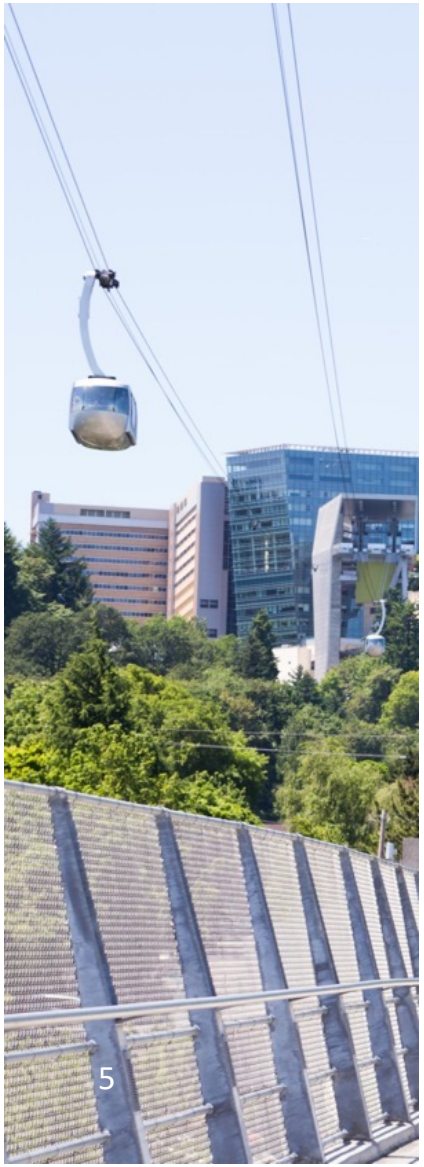
- Pension actuary for 16 years
- MS in Biostatistics from OHSU in 2017
- Associate Biostatistician with OHSU Knight Cancer Institute since 2016
- Long term support for OHSU Dermatology Department since 2018



Goals

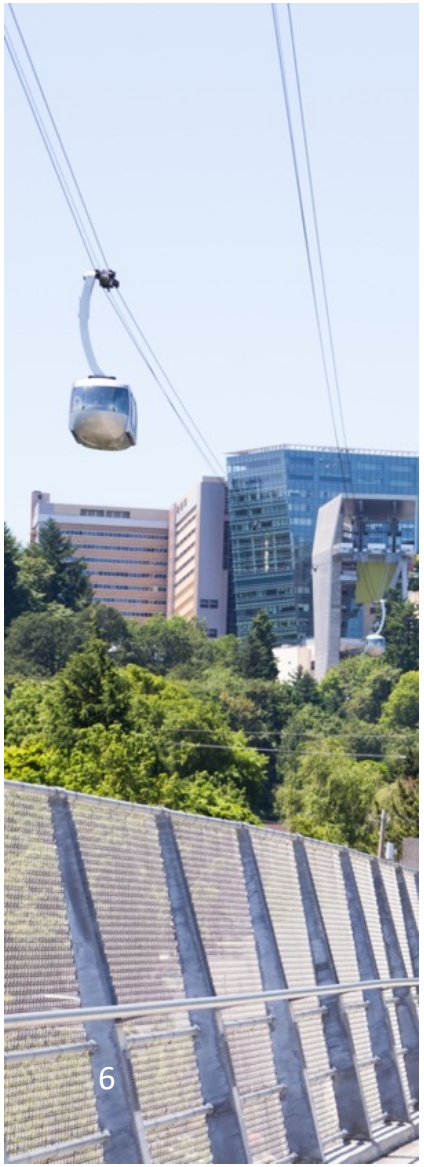
- Learn best practices in study design to aid in statistical analysis
- Share tips and suggestions to improve communications with statisticians

OHSU



Goals

- Understand statistical terms involved in study design:
  - Hypothesis tests (null and alternative),
  - Type I error rate,
  - Type II error rate and Power,
  - Effect size,
  - Sample size



Goals

- Perform simple power and sample size calculations
  - Paired t-test (one sample)
  - 2 sample t-test
  - Two proportions test

OHSU


It all starts with a
research question

The research question drives…

- Hypothesis
- The study design
  - Pilot/preliminary, Retrospective, Prospective
  - Case-Control, Cross-sectional, Cohort, RCT
- Data collection
- Statistical analysis

OHSU

Primary objectives / aims

- The stated principal purpose(s) of the study
- Expressed as a statement of purpose
  - General: efficacy, effectiveness, safety
  - Specific: dose-response, superiority, disease severity, effect on disease incidence, etc.
- “This study seeks to …”
  - To answer, To investigate, To determine, To compare, To assess….
- Coupled with the research hypothesis

Link: Tip of the week, Aims/Objectives versus Outcome Measures

OHSU

Example from literature

- **Study:** Strober B, Mallya UG, Yang M, et al. Treatment Outcomes Associated With Dupilumab Use in Patients With Atopic Dermatitis: 1-Year Results From the RELIEVE-AD Study. *JAMA Dermatol.* 2022;158(2):142–150. doi:10.1001/jamadermatol.2021.4778

- **Research question:** What are the benefits of treatment with dupilumab in the clinical practice setting from the perspective of patients with moderate-to-severe atopic dermatitis?

- **Research hypothesis:** Disease control and quality of life in patients with atopic dermatitis (AD) will improve after the initiation of dupilumab treatment.

- **Objective:** To evaluate self-reported disease control and quality of life after initiating dupilumab treatment in patients with atopic dermatitis (AD) in the the clinical setting.

Link to source paper

OHSU

Secondary objectives/aims

- These are goals that provide further information about principal purpose of the study

OHSU

Primary endpoints/outcomes

- Most important **measurement** gathered by the study
- Corresponds to the Primary objective/aim
- Used to assess the effect of the study or whether the aim/objective is met
- Used to determine design and sample size

OHSU

Secondary endpoints/outcomes

- Less important **measurements** that are part of the pre-specified analysis plan
- Each corresponds to a Secondary objective/aim
- Intended to be collected and analyzed (i.e. different from “Other/Exploratory”)

OHSU

As statisticians

- Most interested in Primary Objective and Endpoint
  - Used for Power and Sample Size
- Also interested in Secondary Objectives and Endpoints
  - Statistical analysis plan
  - ClinicalTrials.gov reporting

OHSU

Recommendations

- Clear and simple statements
- Bulleted or numbered lists are helpful
  - Don’t “clump”
- Consider what is secondary vs. exploratory

OHSU


Hypothesis testing

Hypothesis (to a statistician)

- In a statistical hypothesis test, there is a single null hypothesis (H_0) that specifies the value of the parameter being tested **if there is no effect**.

- There is also an alternative hypothesis (H_1) that contradicts the statement made under H_0 . H_1 Is a statement about what you hope to show or prove.

Lasarev, M., BSTA 511 Lecture Notes, Fall 2013

Hypothesis (to a statistician)

- Hypothesis tests can be one-sided or two-sided, depending on how H_1 is specified
  - $H_0: \mu = \mu_0$ vs. $H_1: \mu > \mu_0$ -- one-sided, upper tail
  - $H_0: \mu = \mu_0$ vs. $H_1: \mu < \mu_0$ -- one-sided, lower tail
  - $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$ -- two-sided

OHSU

Hypothesize first, then collect the data

- Specify parameters of interest, null and alternative hypothesis
- Specify significance level (α)
- Collect data; compute test statistic and *p*-value

OHSU

Test the hypothesis

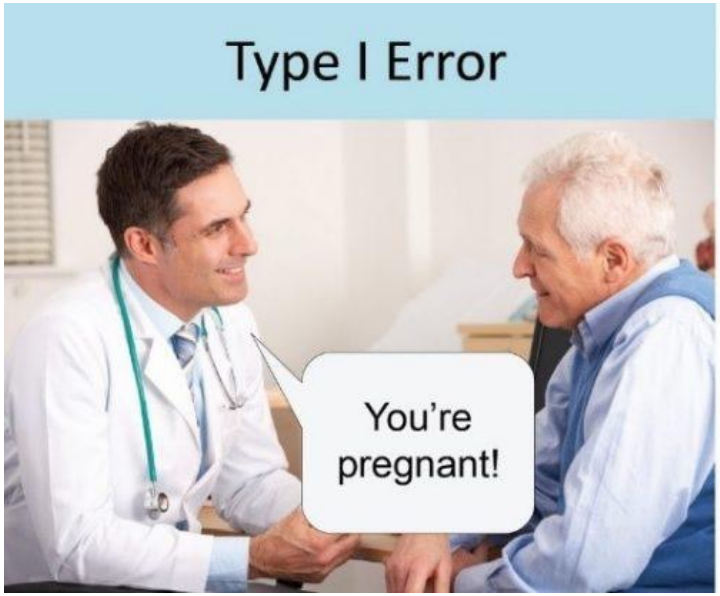
- *p*-value measures the probability of obtaining sample data as extreme or more extreme than that observed, given that H_0 (Null) is true
- Compare *p*-value to significance level (α), and if the *p*-value is tiny compared to α, then we reject H_0

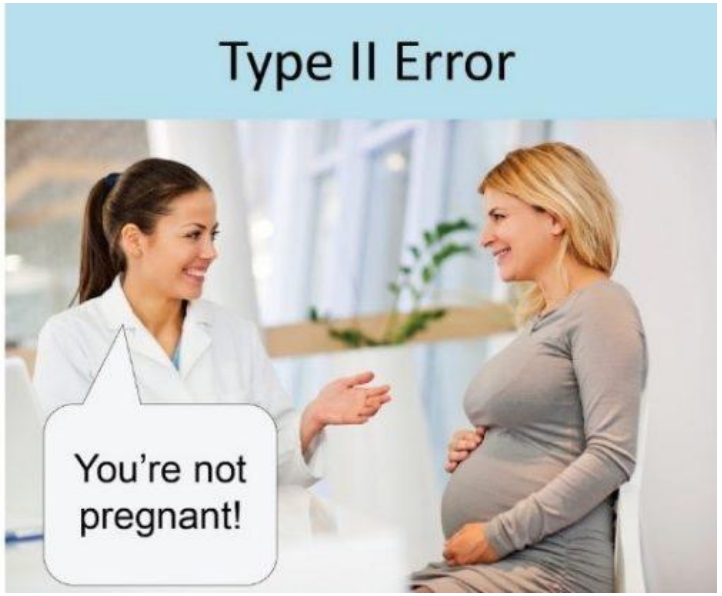
Errors can happen

- There exists an underlying, correct truth.
- There are decisions that are made based on evidence.
- Incorrect decision can be made depending on the what is genuinely true

OHSU

Type I and Type II Errors

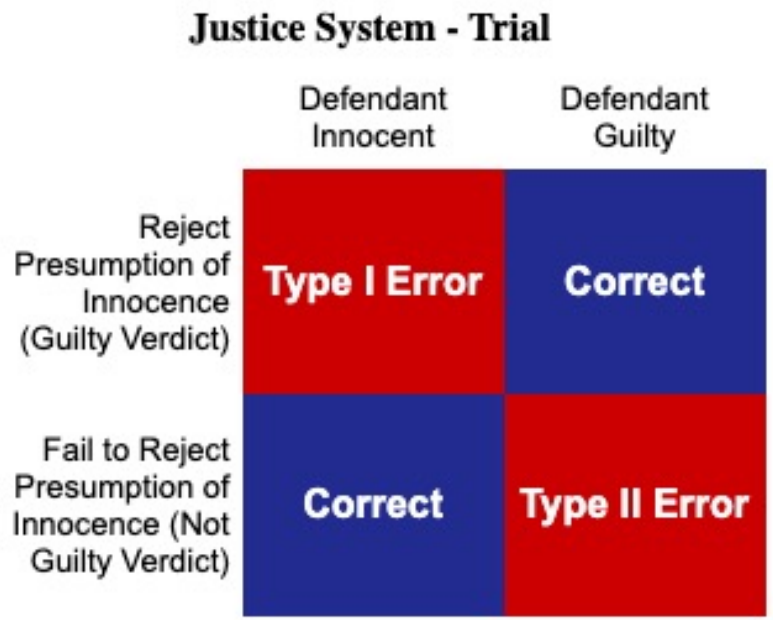




https://www.statisticssolutions.com/to-err-is-human-what-are-type-i-and-ii-errors/

OHSU

Justice system analogy



http://www.intuitor.com/statistics/T1T2Errors.html

OHSU

Statistical hypothesis test

	"Truth"	
Test decision	Null is true	Alternative is true
Reject Null	Type I (α)	Correct
Don't reject Null	Correct	Type II (β)

Type I Error Rate

- Formal statement:
  - $\alpha = P(Reject\ H_0 | H_0\ is\ true)$

- Informal statement:
  - Significance level (alpha) is the probability that a statistical test can demonstrate that there is a difference (e.g. difference in means between two groups) given that there is a difference.

OHSU

Type II Error Rate

- Formal statement:
  - $\beta = P(Fail\ to\ reject\ H_0 | H_1\ is\ true)$

- Tend to talk about it in terms of Power:
  - $Power = 1 - \beta = P(Reject\ H_0 | H_1\ is\ true)$

OHSU

Power

- Formal statement:
  - $Power = 1 - \beta = P(Reject\ H_0 | H_1\ is\ true)$

- Informal statement:
  - The probability *before* conducting a statistical test that the procedure used (e.g. t-test, z-test) together with the sample data will lead to a correct decision, i.e. the test can show a difference given that there is a diffference

OHSU

What we control

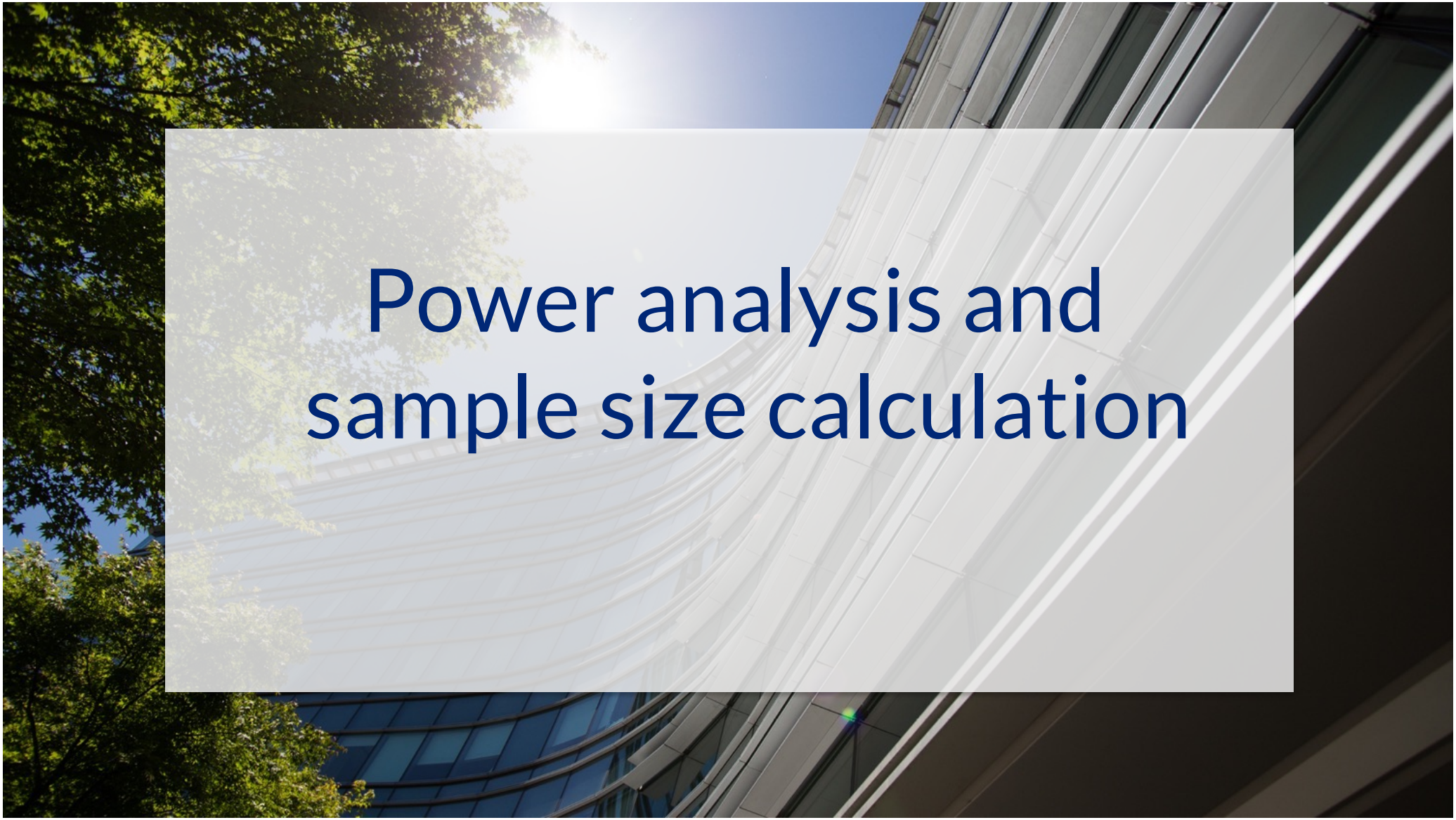
- We have direct control over α in a hypothesis test
- We have indirect control over β.
- For a fixed sample size:
  - As α increases, β decreases
  - As α decreases, β increases

What we control

- At the design stage of an experiment before data collection, we attempt to control power ($1 - \beta$) and therefore the Type II Error Rate (β).

- Most often done by selecting an appropriate sample size and an appropriate α level.

OHSU


Power analysis and
sample size calculation

4 components

- Significance level (α)
- Statistical Power ($1 - \beta$)
- Sample size (n)
- Effect size (Δ)

- Must specify 3 in order to solve for the 4^{th}.

OHSU

Effect sizes

- An effect size is a value that measures the strength of a relationship
- Experimental effect
- Magnitude of a difference that study aims to detect (difference in means between two groups for example)

OHSU

Estimate for effect size

- Pilot or preliminary data
- Published literature
- Clinically or biologically meaningful difference

OHSU

Estimate for effect size

- When considering effect sizes, it helps to ask:
  - What is an effect size that must be detected?
  - What is a biologically or clinically meaningful effect?

OHSU

Sample size

- Sample size seeks to ensure that the study design will have sufficient power to detect an specified effect size.

OHSU

Study settings: Prospective

- Significance level (α) – Typically set at 0.05
- Statistical Power ($1 - \beta$) – Typically set at 80%
- **Sample size (n) – ?????**
- Effect size (Δ) – determined a priori

OHSU

Study settings: Pilot/Preliminary

- Significance level (α) – Typically set at 0.05
- Statistical Power ($1 - \beta$) – Typically set at 80%
- Sample size (n) – Usually fixed due to budget, ethical considerations
- **Effect size (Δ) – ????**

OHSU

Study settings: Retrospective

- Significance level (α) – Typically set at 0.05
- **Statistical Power ($1 - \beta$) – ????**
- Sample size (n) – Usually fixed
- Effect size (Δ) – determined a priori

OHSU

4 components

- Significance level (α)
- Statistical Power ($1 - \beta$)
- Sample size (n)
- Effect size (Δ)

- Must specify 3 in order to solve for the 4^{th}.

OHSU


Study design with a
statistician

Preparation

- Background
- A stated hypothesis, if there is one
- Aims/objectives, especially the primary
- Endpoints (how the aim/objective is measured)
- Effect size, if there is one

OHSU

Power analysis for a new study Example #1 Paired t-test

Hypothesis and outcome

- Primary objective
  - To assess the change in quality of life in patients with atopic dermatitis (AD) one year after the initiation of dupilumab treatment
- Hypothesis:
  - Quality of life in patients with AD will improve one year after the initiation of dupilumab treatment
- Primary endpoint:
  - Change in Dermatology Life Quality Index (DLQI) from baseline to 12 months

OHSU

Study design

- Single group
- DLQI at baseline and 12-months (continuous, paired data)
- Statistical analysis plan:
  - Paired t-test

OHSU

What sample size is needed?

- To solve for sample size (n), we will need:
  - Alpha = 0.05
  - Power = 80.0%
  - Effect size = ?????

OHSU

Published literature

** Smaller DLQI = Better Quality of Life

Table 3. Patient-Reported Health-Related Quality of Life and Work and Activity Impairment

Outcome	Mean (SD) [PMM range][a]						
	Baseline (n = 699)	Month 1 (n = 632)[b]	Month 2 (n = 626)[b]	Month 3 (n = 596)[b]	Month 6 (n = 543)[b]	Month 9 (n = 477)[b]	Month 12 (n = 483)[b]
DLQI[c]							
Total score	14.4 (7.3)	5.9 (5.8) [5.6-6.7]	5.1 (5.5) [4.9-6.1]	4.8 (5.5) [4.4-6.5]	4.1 (5.0) [3.4-6.7]	3.8 (5.2) [2.9-7.8]	3.5 (4.9) [2.7-7.8]

Link to paper Strober, et al. Treatment Outcomes Associated With Dupilumab Use in Patients With Atopic Dermatitis: 1-Year Results From the RELIEVE-AD Study. *JAMA Dermatol.* 2022;158(2):142–150

OHSU

Problem

- For an effect size, we need both estimates for
  - Chane in the means, and
  - Standard deviation (SD) of the difference

- From the paper, we have the means and can calculate the change. We have standard deviations, but…

- Issue #1: we don't know the SD of the difference
- Issue #2: need to know correlation between DLQI values

OHSU

Solution

- Vary the correlation and estimate possible scenarios for SD of the differences
- Involves statistical formula; not shown here

OHSU

Sample size: Paired t-test

	Baseline		12-months		Assumed	For the estimated difference		Estimated
Scenario	Mean	SD	Mean	SD	Correlation	Mean	SD	Sample Size
1	14.4	7.3	3.5	4.9	0.20	10.9	7.9	7
2	14.4	7.3	3.5	4.9	0.50	10.9	6.4	5
3	14.4	7.3	3.5	4.9	0.80	10.9	4.5	4

- Assuming: Paired t-test, Power = 0.80, $\alpha = 0.05$

https://sample-size.net/calculator-finder/

OHSU

Are the estimates realistic?

- Sample size estimates are very small, even for a paired test
- The effects size from the Published literature is very big! 10 points on a 0—30 scale.

Are the estimates realistic?

- Further reading in the article, they state that a clinically meaningful change is 4 or more points. And they provide reference (Basra et al, 2015)
- What if we used that effect size with the standard deviation we already calculated?

OHSU

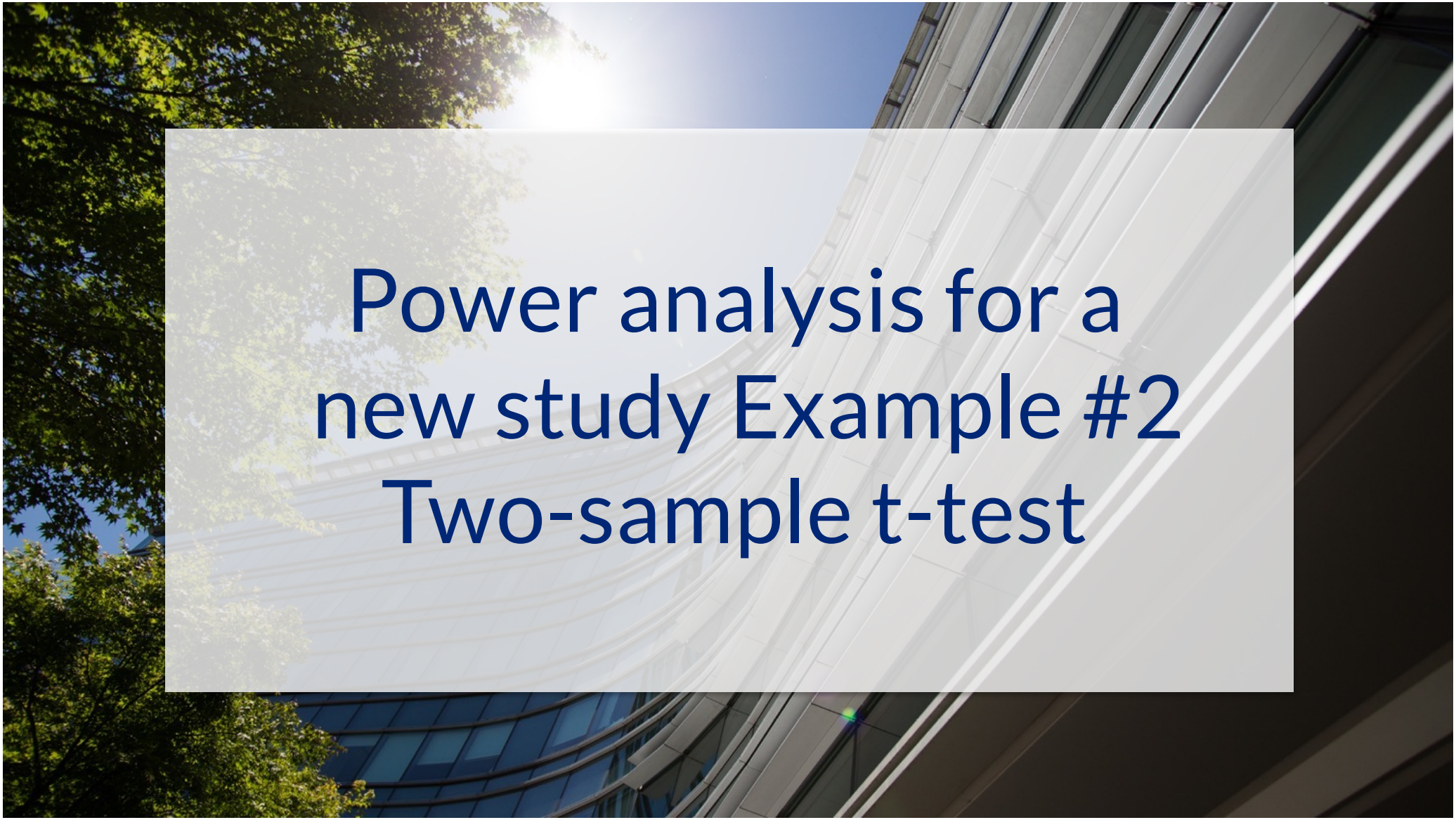
Revised sample size: Paired t-test

Scenario	Baseline SD	12-months SD	Assumed Correlation	For the estimated difference Mean	For the estimated difference SD	Estimated Sample Size
1	7.3	4.9	0.20	4.0	7.9	33
2	7.3	4.9	0.50	4.0	6.4	23
3	7.3	4.9	0.80	4.0	4.5	12

- Assuming: Paired t-test, Power = 0.80, $\alpha = 0.05$

https://sample-size.net/calculator-finder/

OHSU



Power analysis for a new study Example #2 Two-sample t-test

Hypothesis and outcome

- Primary objective
  - To compare the quality of life in patients with atopic dermatitis (AD) treated with Novel Treatment and Dupilumab after one year
- Hypothesis:
  - After one year, the quality of life in patients with AD treated with Novel Treatment will be better than those treated with Dupilumab
- Primary endpoint:
  - Difference in Dermatology Life Quality Index (DLQI) after 12 months between treated with Novel Treatment and Dupilumab

OHSU

Study design

- Two treatment groups
- DLQI at 12-months (continuous, independent data)
- Statistical analysis plan:
  - Two-sample t-test (independent samples)

Published literature

** Smaller DLQI = Better Quality of Life

Table 3. Patient-Reported Health-Related Quality of Life and Work and Activity Impairment

Outcome	Mean (SD) [PMM range][a]						
	Baseline (n = 699)	Month 1 (n = 632)[b]	Month 2 (n = 626)[b]	Month 3 (n = 596)[b]	Month 6 (n = 543)[b]	Month 9 (n = 477)[b]	Month 12 (n = 483)[b]
DLQI[c]							
Total score	14.4 (7.3)	5.9 (5.8) [5.6-6.7]	5.1 (5.5) [4.9-6.1]	4.8 (5.5) [4.4-6.5]	4.1 (5.0) [3.4-6.7]	3.8 (5.2) [2.9-7.8]	3.5 (4.9) [2.7-7.8]

Link to paper Strober, et al. Treatment Outcomes Associated With Dupilumab Use in Patients With Atopic Dermatitis: 1-Year Results From the RELIEVE-AD Study. *JAMA Dermatol.* 2022;158(2):142–150

OHSU

Sample size: Two sample t-test

- Assuming: Power = 0.80, α = 0.05
- From published literature:
  - Mean (SD) DLQI at twelve months = 3.5 (4.9)
  - Use this as the mean for our control group (Dupilumab)
- Assume the same standard deviation for both groups
- Effect size = ???
  - Need to know the difference between groups
  - Ideally, there would be preliminary data or a clinically relevant difference
  - Absent this, we can estimate the sample size for varying values

OHSU

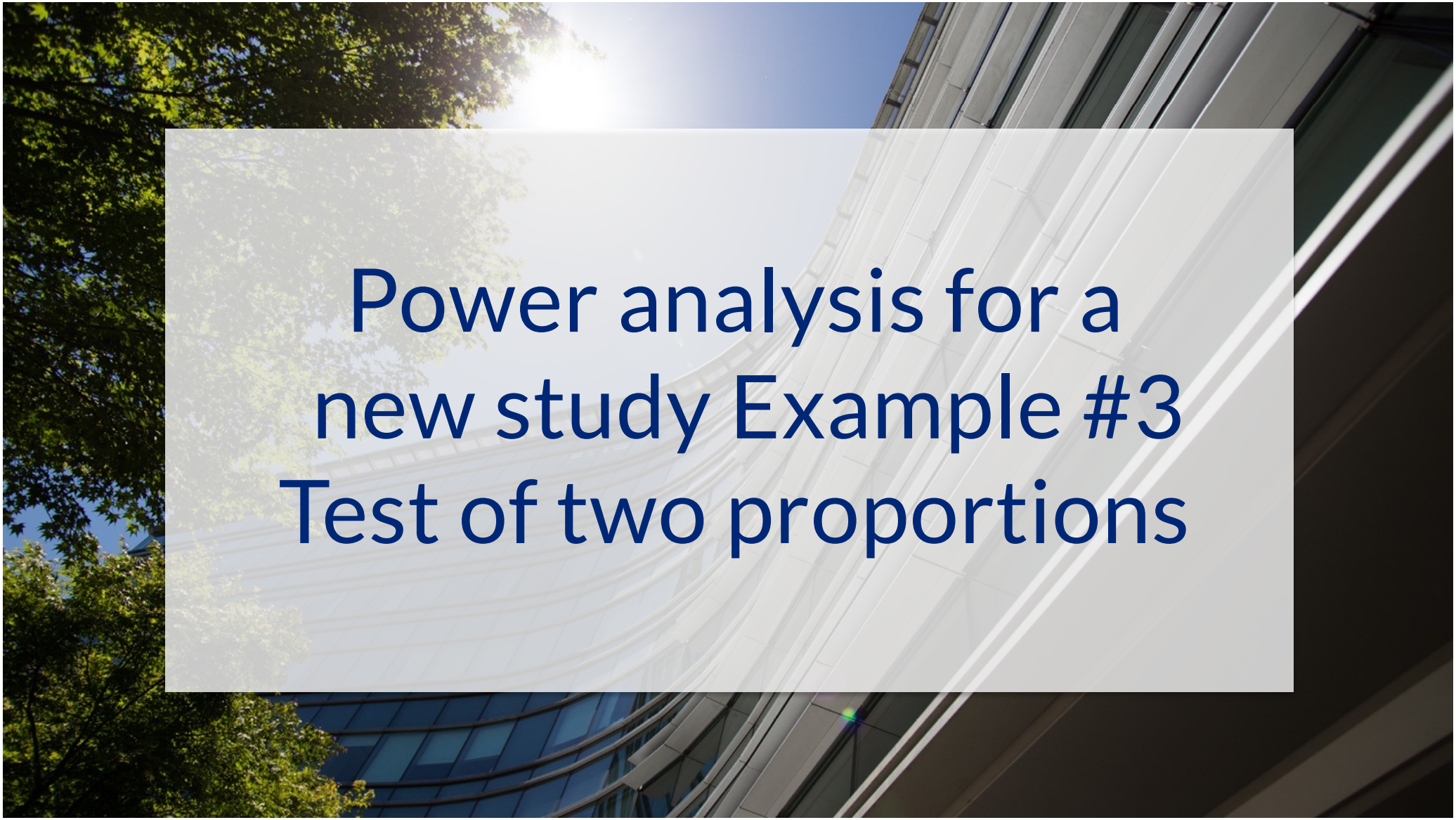
Sample size: Two sample t-test

Scenario	SD for each group (from Lit.)	Assumed Difference in means	PER GROUP Estimated Sample Size	TOTAL Estimated Sample Size
1	4.9	1	378	756
2	4.9	2	96	192
3	4.9	3	43	86

** Note: we didn't need the mean for the treatment group since the above sample size calculation is based off of the difference in means

https://sample-size.net/calculator-finder/

OHSU



Power analysis for a new study Example #3 Test of two proportions

Hypothesis and outcome

- Secondary objective
  - To compare the proportion of patients reporting “No effect” in the Novel Treatment and with those treated with Dupilumab after one year
- Hypothesis:
  - After one year, the proportion of patients reporting “No effect” will be higher in those treated with Novel Treatment than those treated with Dupilumab
- Secondary endpoint:
  - Difference in the proportions of patients in each group reporting “No effect” (DLQI = 0—1)

OHSU

Study design

- Two treatment groups
- Binary outcome (DLQI = 0 or 1, Yes/No)
- Statistical analysis plan:
  - Two Proportions Z-test

OHSU

Power analysis: two-proportion z-test

- Assume the sample size calculated in example #2 for the primary endpoint
  - Difference in means = 3
  - N = 86 (43 per group)
- Assuming 80% power, what difference of proportions can we detect?

OHSU

Power analysis: two-proportion z-test

- Assume the sample size calculated in example #2 for the primary endpoint
  - Difference in means = 3
  - N = 86 (43 per group)
- Assuming 80% power, what difference of proportions can we detect?

OHSU

Power analysis: two-proportion z-test

- Assume: 80% power, $\alpha = 0.05$

Scenario	n Ctrl. group	n Tx group	Proportion in Ctrl. Group	Estimated proportion in Tx Group	Increase in proportion Detected
1	43	43	0.300	0.596	0.296
2	43	43	0.500	0.786	0.286
3	43	43	0.700	0.931	0.231

- Ask: is this reasonable, achievable, clinically meaningful?

https://sample-size.net/calculator-finder/

OHSU


Past talks and
resources

Past talks

- Power and Sample Size 101, Mieke Niederhausen, PhD (February 2021)
  - Slides http://bit.ly/BERD-PSS-101
  - Recording https://echo360.org/media/c14ae529-27fe-4208-ae27-95c7c1adc928/public
- Planning a Study with Power and Sample Size Considerations in Mind, David Yanez, PhD (May 2019)
  - Slides https://www.ohsu.edu/sites/default/files/2019-12/PowerAndSampleSize_29MAY2019.pdf
  - Recording https://echo360.org/media/ee2a5565-1168-4941-82ea-d9eda4223281/public
- Power and Sample Size for Clinical Trials: An Introduction, Yiyi Chen, PhD (February 2021)
  - Slides https://drive.google.com/file/d/1dK4ktxQa81SNjRM4z4JC8JXH8DTgcr8Z/view
  - Recording https://echo360.org/media/c2197215-3c5f-4c9b-b3eb-3877566ad3da/public

OHSU

Resources

- PASS
  - Training videos (short and good) https://www.ncss.com/videos/pass/training/
  - Documentation https://www.ncss.com/software/pass/pass-documentation/
- Biostats4You, University of Minnesota
  - https://biostats4you.umn.edu/resources
  - "website was developed to serve medical and public health researchers and professionals who wish to learn more about biostatistics"
- Links I think are pretty good
  - Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies (Serdar et al.) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7745163/#:~:text=When%20the%20sample%20size%20is,the%20power%20of%20the%20study.
  - Understanding Statistical Power and Significance Testing, an interactive visualization https://rpsychologist.com/d3/nhst/
  - Power and Sample Size Applet https://wbakerrobinson.shinyapps.io/Paired_T_test_shiny/
  - The Relationship between Significance, Power, Sample Size & Effect Size (K. Mysiak) https://towardsdatascience.com/the-relationship-between-significance-power-sample-size-effect-size-899fcf95a76d

OHSU

Software (free)

- Sample size calculators from UCSF
  - Link https://sample-size.net/calculator-finder/
  - User friendly! Web based.
- G*Power software
  - Sample size determination and power analysis using the G*Power software (H. Kang) https://pubmed.ncbi.nlm.nih.gov/34325496/
  - Download https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html
  - More involved. Requires download.
- CRAB (Cancer Research and Biostatistics) Statistical Tools
  - Link https://stattools.crab.org/
  - Web based.

OHSU

Thank You