

Research and Evidence Webinar Series:

Power Analysis for Program Evaluation I: An Introduction

March 30, 2021

Welcome!

Power Analysis for Program Evaluation I: An Introduction



OFFICE OF
RESEARCH AND
EVALUATION



AmeriCorps

We'll get started in just a couple of minutes.

Please be advised that there is no dial-in for this webinar; all audio is provided directly through the Adobe Connect platform.

Welcome!

Power Analysis for Program Evaluation I: An Introduction



Dr. Lily Zandniapour

Research and Evaluation Manager

Office of Research and Evaluation, AmeriCorps

Introductory Remarks

Power Analysis for Program Evaluation I: An Introduction



Dr. Carrie Markovitz

Principal Research Scientist

NORC

Speaker

Power Analysis for Program Evaluation I: An Introduction



Dr. Eric Hedberg

Senior Data Scientist

NORC



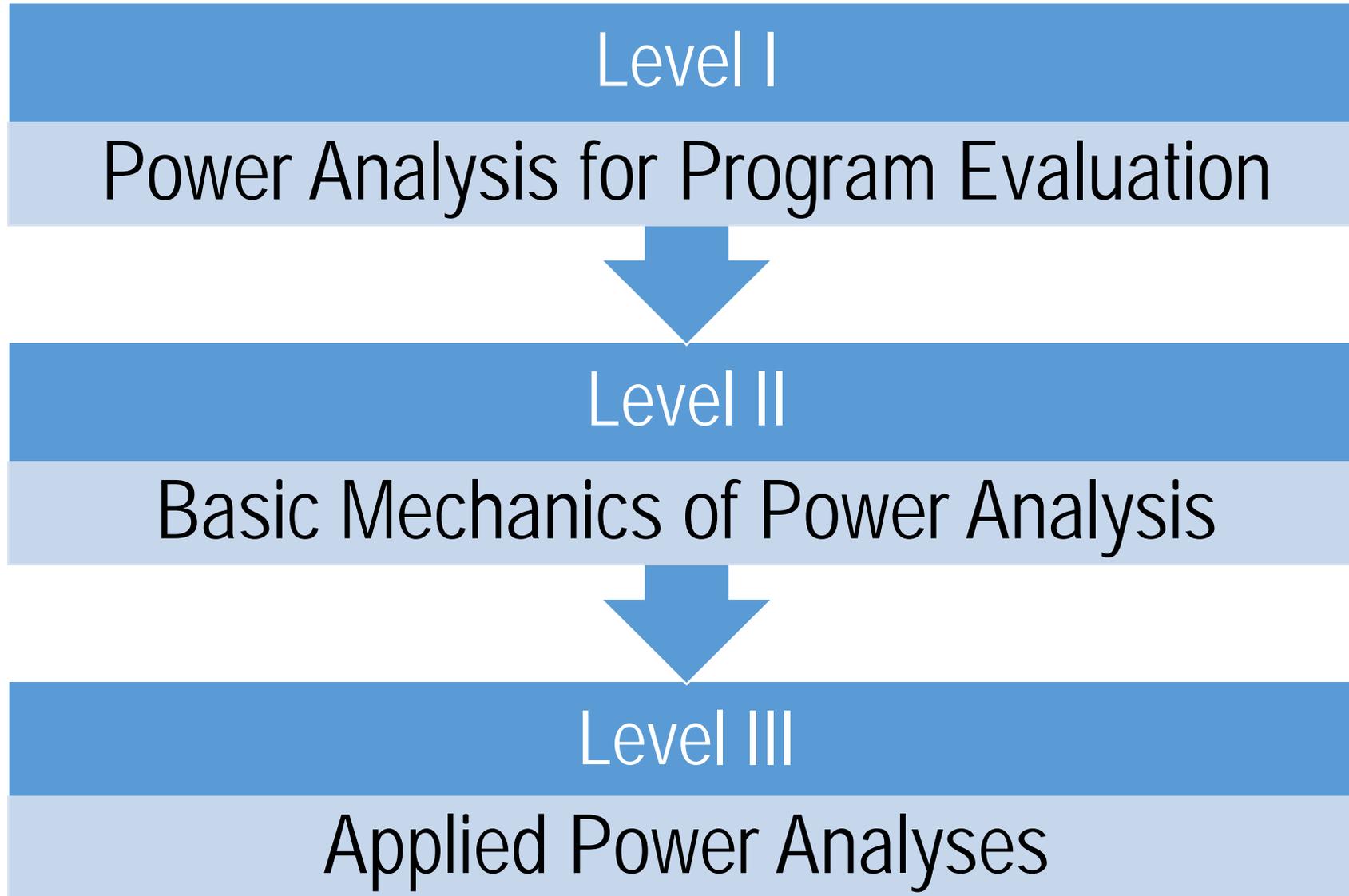
AmeriCorps



Power Analysis for Program Evaluation

Level I: Defining and Understanding Statistical Power

Overview of Courses (Three Levels)



- Intended for All Audiences: program staff, funders, internal or external evaluators, and third-party evaluation/evidence reviewers
- Introduces the concept of power
- Explains why power is important for evaluation planning
 - Helps support appropriate use of resources for evaluation by providing information required to produce an efficient sample (i.e., a sample may be sufficient rather than a census)

Level II: Basic Mechanics of Power Analysis

- Intended for program and evaluation staff
- Type I and Type II errors
 - How each error threatens a study
 - Power is complement of Type II error

Level III: Applied Power Analysis

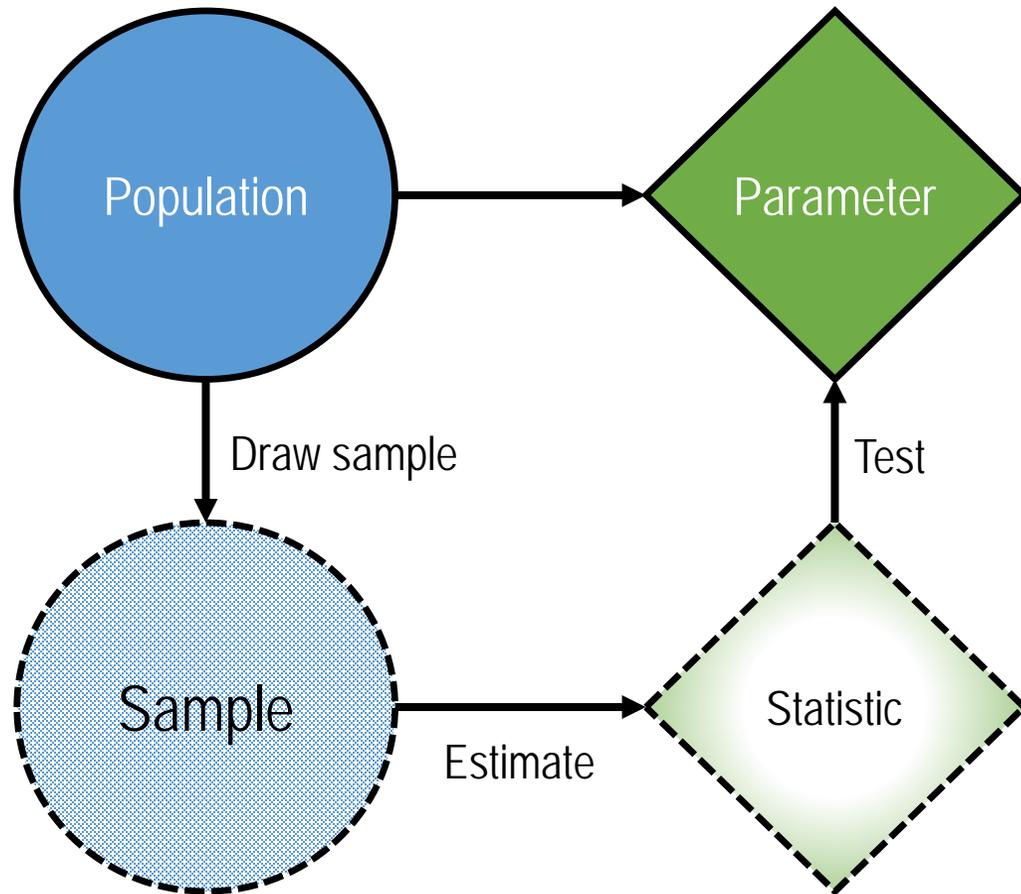
- Intended for program staff or evaluators working with statisticians
- How to conduct power analyses
 - Parameter values for a power analysis
 - Examples for common designs
 - How to use the literature to inform specifications
- How to write-up (or read) a power analysis

Why is power important?

- EXAMPLE: You want to assess the impact of your program on program beneficiaries, but tracking the circumstances of all of your beneficiaries is cost prohibitive. If instead you tracked outcomes for some, but not all of your program participants, how many participants will you need to conduct a credible study?
- In some cases, you will need to conduct a CENSUS (population) and include every program participant in your study. But sometimes it is possible to use a SAMPLE of program participants and still capture evidence of impacts that can be GENERALIZED to the wider population (all program beneficiaries).

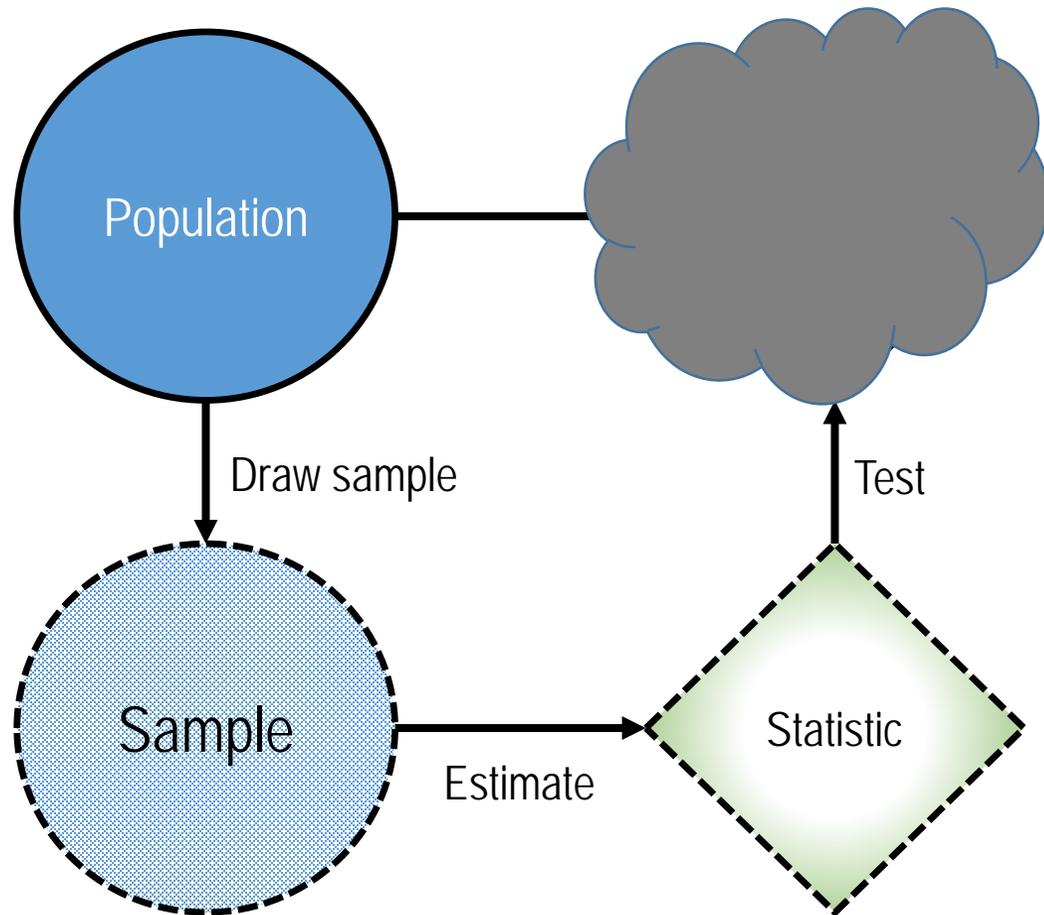
Power helps determine how large a sample size you need in order to obtain reliable evidence of program impacts.

Statistics for Impact Studies



- Impact studies want to infer the existence and size of the impact for the **population** based on **statistics** produced from **sample** data
- We will not know for sure whether a program will be successful in the future or for non-study participants
- We make an informed decision based on our sample from which we infer to the population

Statistics for Impact Studies



- Impact studies want to infer the existence and size of the impact for the **population** based on **statistics** produced from **sample** data
- We will not know for sure whether a program will be successful in the future or for non-study participants
- We make an informed decision based on our sample from which we infer to the population

Studies have few of many samples

Your study is only one of many possible samples

- Your margin of error is a range of possible outcomes from 95 percent of many other samples
- Given a true impact, will your sample happen to produce a result that is statistically significant?



You only get one jelly bean...will it be grape?

- H_0 Null Hypothesis– No difference between study groups
- H_1 Alternative Hypothesis– Difference between study groups (program and control/comparison)

In hypothesis testing we hope to be able to **reject the null hypothesis** of no differences and **accept the alternative hypothesis** that there is a difference between study groups by showing a very low chance that the null hypothesis is correct.

*The level of power calculated for your study is the **probability** of rejecting the null hypothesis (i.e., claim a program impact)*

- Power is about planning
- Power analysis sections are estimates about the future
- Power is about using evaluation resources wisely
- Power should be estimated prior to conducting a study (i.e., power analysis)

Why is power important?

High powered studies have many benefits:

- For grantees: Helps increase the chance of detecting true effects (obtaining credible evidence of program impact)
- For funders: Helps ensure you are resourcing an evaluation that will produce reliable evidence of impacts
- For researchers: Helps increase reliability and trust in study findings

Why power is important for program evaluations

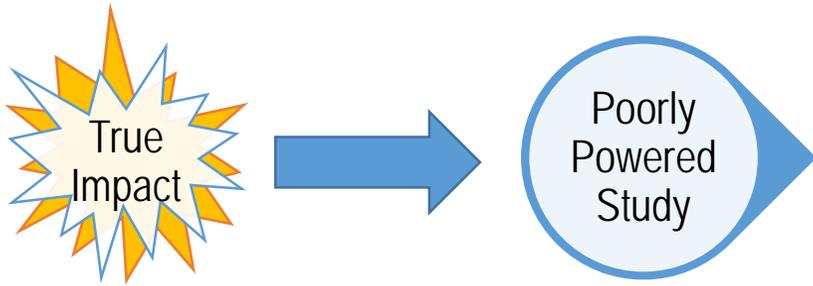
Results of a poorly powered study (i.e., Type II Error)



Results of a well powered study

Why power is important for program evaluations

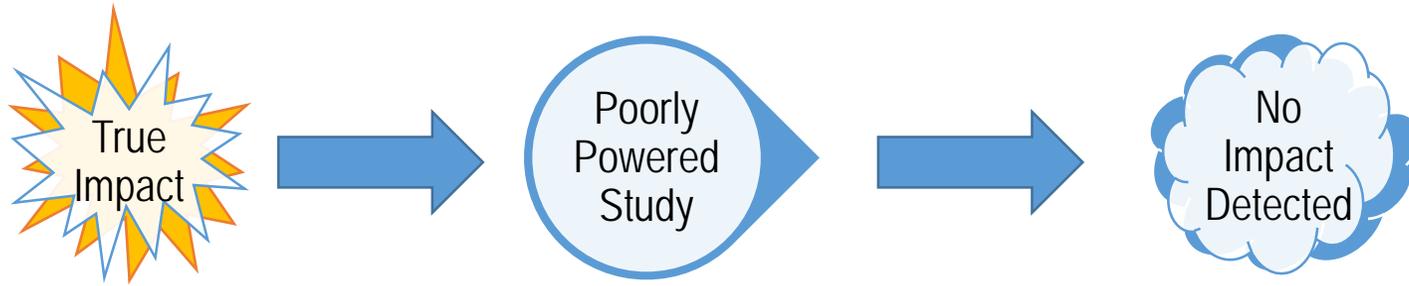
Results of a poorly powered study (i.e., Type II Error)



Results of a well powered study

Why power is important for program evaluations

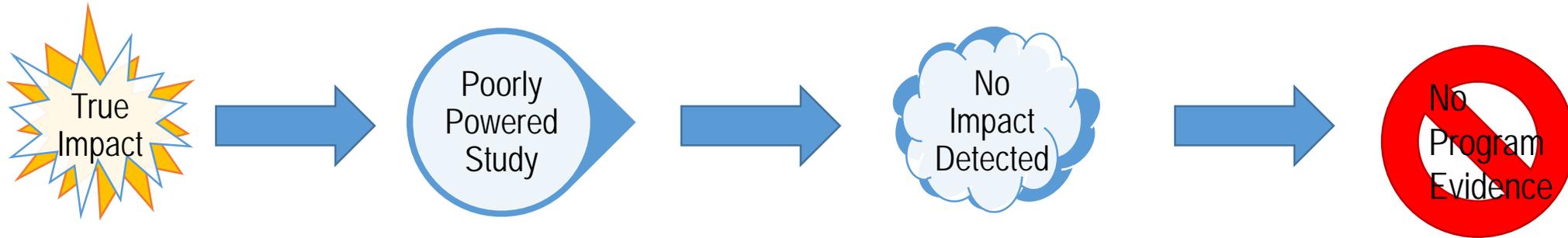
Results of a poorly powered study (i.e., Type II Error)



Results of a well powered study

Why power is important for program evaluations

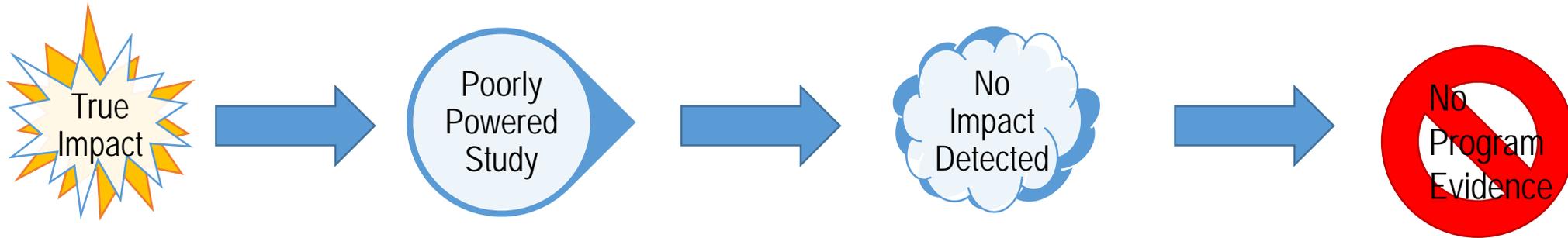
Results of a poorly powered study (i.e., Type II Error)



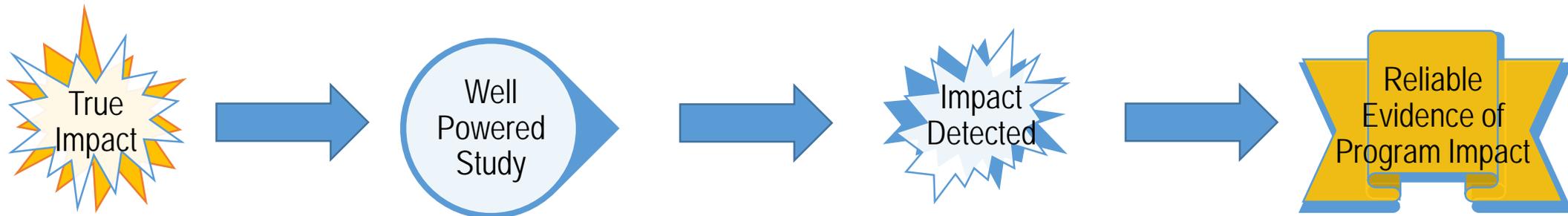
Results of a well powered study

Why power is important for program evaluations

Results of a poorly powered study (i.e., Type II Error)



Results of a well powered study

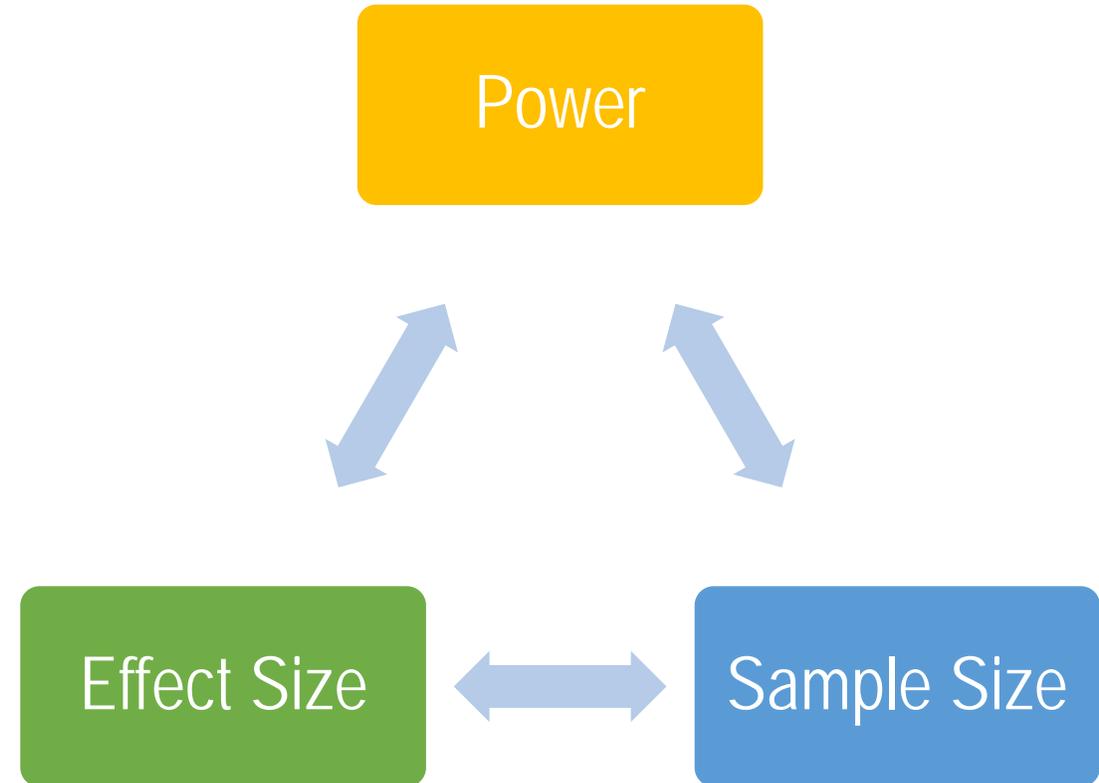


The probability of “statistically significant results” is based on

- The design of the study
- The amount/structure of the data
 - Generally speaking: more information=more power
- How the data are analyzed
- The size of the intervention/program’s impact
- The selected significance level

What is a power analysis?

- A calculation that helps determine if a study has an adequate chance to detect a statistically significant effect (if one truly exists)
- Power analysis is based on the relationship between power, sample size, and effect size (assume two of the elements and calculate the third)



Power increases with impact size

Holding sample size/design constant:

- Bigger impacts have more power
- Smaller impacts have less power



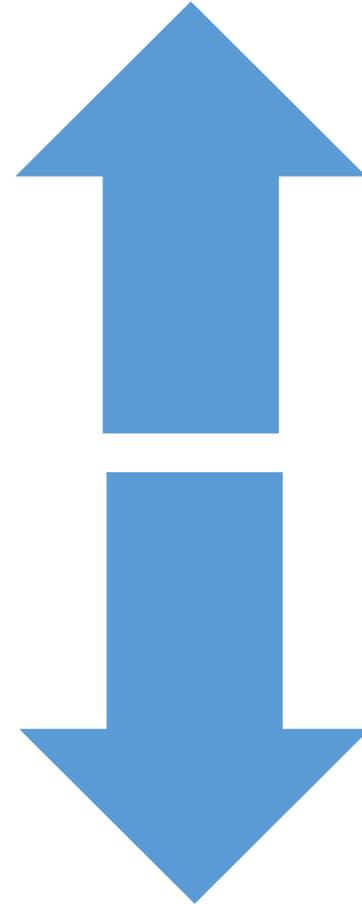
Big impacts
have more
power

Small impacts
have less power

Power increases with sample size

Holding impact constant:

- Bigger samples, depending on design, have more power
- Smaller samples, depending on design, have less power



Big samples have more power

Small samples have less power

Power, impact size and sample size

- Data are like a microphone trying to detect a sound
- Small samples (microphone) can only detect large impacts (loud sounds)
- Big samples (microphone) can detect impacts both large and small (loud sounds and pin drops)



Why conduct a power analysis for your study?



- Power analysis gives researchers a chance to determine if their sample design is adequate to detect the expected impact before the study
- Power analysis encourages research teams to think critically about and explore
 - Expected Impacts of their intervention
 - The design of their study sample and how it impacts the analysis plan
 - Whether the analysis plan is feasible (Is there enough data?)
 - Evaluation budget (ensure using resources wisely)
- A power analysis can help programs effectively use and target resources

What does a power analysis tell you?



- KEY QUESTION #1: How many students are needed to detect an effect size of X ?
ANSWER: Y students are needed in the treatment group and Y students are needed in the control group.
- KEY QUESTION #2: What effect size must the program achieve to detect a statistically significant effect, given the known sample sizes?
ANSWER: An effect size of X standard deviations must be produced by the intervention.

BEFORE DATA COLLECTION!

- Power analyses are only informative and helpful prior to data collection
- When there is no power analysis and the results are not statistically significant:

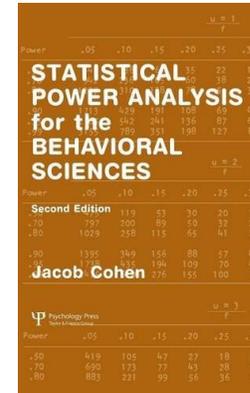
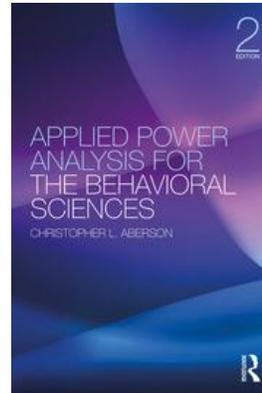
Perhaps there is no impact of the intervention
OR

Perhaps the study was underpowered to detect the actual effect

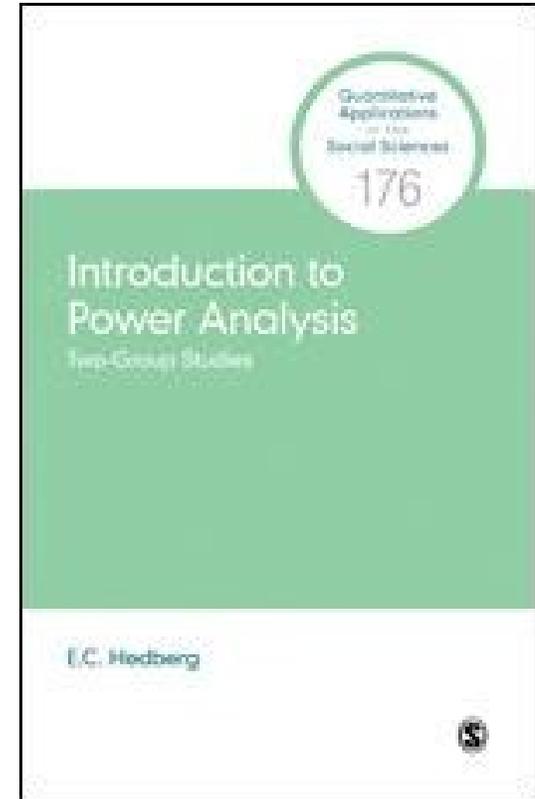
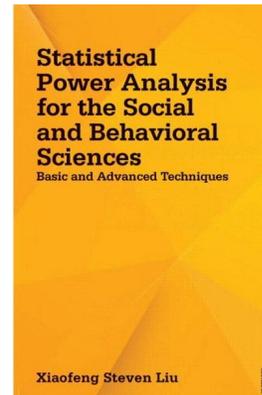
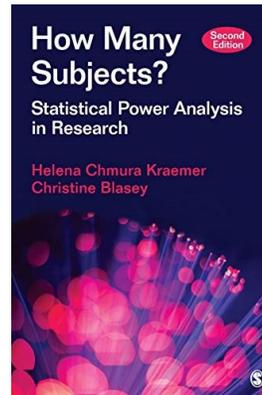
Yuan, K. H., & Maxwell, S. (2005). On the post hoc power in testing mean differences. *Journal of educational and behavioral statistics*, 30(2), 141-167.

- Power is the chance to find an estimated impact that is “statistically significant”
- Poorly powered studies can reduce the chance of detecting the true effect of programs
- Power is the result of a process involving study design, analysis, and expected impacts
 - For every design and analysis, there is a unique power analysis procedure
- Larger samples can detect smaller or larger impacts (are more sensitive) compared to smaller samples that can only detect larger impacts (are less sensitive)
- Power analysis must occur before data collection in a study

- Aberson, C. L. (2011). *Applied power analysis for the behavioral sciences*. Routledge.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- Hedberg, E. C. (2017). *Introduction to Power Analysis: Two-group Studies* (Vol. 176). Sage Publications.
- Kraemer, H. C., & Blasey, C. (2015). *How many subjects?: Statistical power analysis in research*. Sage Publications.
- Liu, X. S. (2013). *Statistical power analysis for the social and behavioral sciences: Basic and advanced techniques*. Routledge.
- Many more papers and blogs exist, likely specific to your outcome



Power	$d = 1$					Power	$d = 2$					Power	$d = 3$						
	.05	.10	.15	.20	.25		.30	.05	.10	.15	.20		.25	.30	.05	.10	.15	.20	.25
.80	17	27	35	42	48	55	11	18	23	28	33	38	44	50	56	62	68	74	80
.70	27	42	55	68	80	93	18	28	35	42	50	58	66	74	83	92	101	110	119
.60	42	68	93	119	145	171	28	42	55	68	80	93	106	120	134	148	162	176	190



Closing Remarks

Power Analysis for Program Evaluation I: An Introduction



Dr. Carrie Markovitz

Principal Research Scientist

NORC

Q&A

Power Analysis for Program Evaluation I: An Introduction



OFFICE OF
RESEARCH AND
EVALUATION



AmeriCorps

Questions?

Dr. Andrea Robles

*Research and Evaluation Manager,
Office of Research and Evaluation, AmeriCorps*

Visit nationalservice.gov/ORE for past webinars, research and evaluation resources, reports, and more!