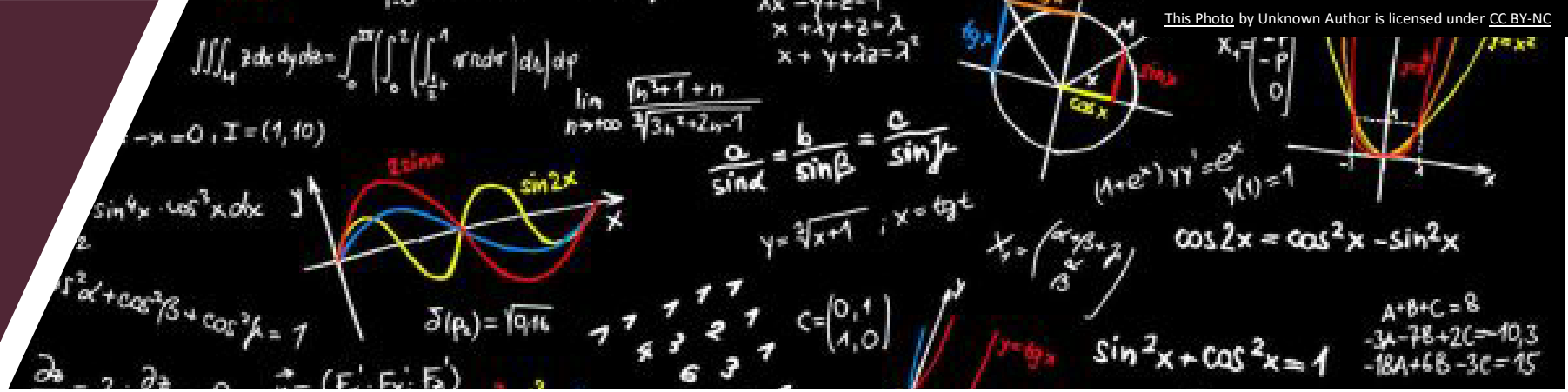


Research Design and Statistics

Laura A. Graham, Ph.D., MPH
lagraham@stanford.edu

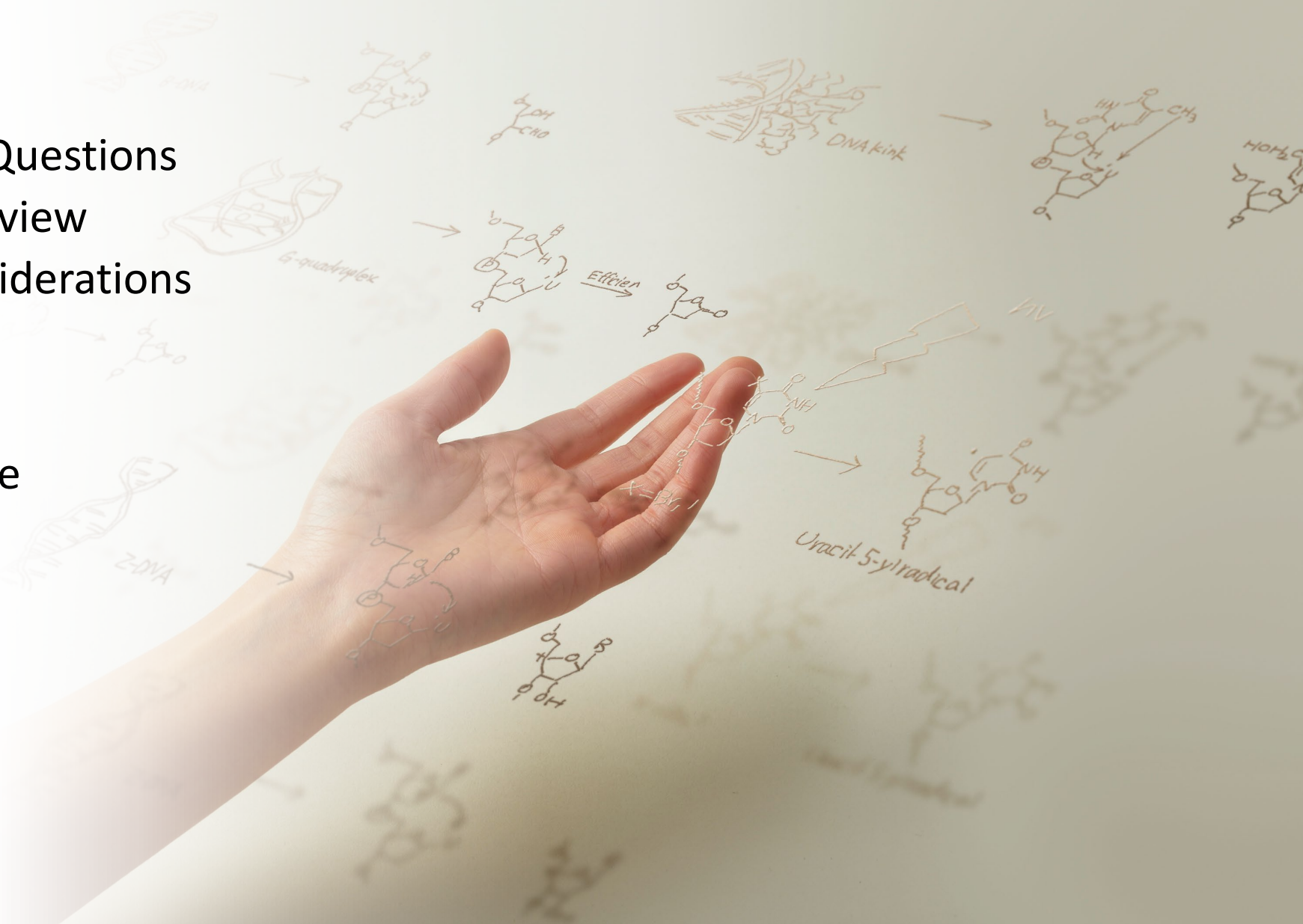
Many thanks to Amber Trickey for sharing her slides.

S-SPIRE Stanford–Surgery Policy Improvement Research & Education

Outline

- Research Design
  - Writing Research Questions
  - Study Design Overview
  - Study Design Considerations

- Statistics
  - Statistical Inference
  - Hypothesis testing
  - Statistical tests



First Step: The Research Question

P

- **Patient, Population or Problem**
  - Who are the relevant patients? Think about age, sex, geographic location, or specific characteristics that would be important to your question.

I

- **Intervention, Prognostic Factors, or Exposure**
  - What is the exposure, diagnostic test, or intervention that you are interested in?

C

- **Comparison**
  - What is the main alternative to compare with the intervention or exposure?

O

- **Outcome**
  - What can you hope to accomplish, measure, improve, or affect?
  - What are you trying to do for the patient?

T

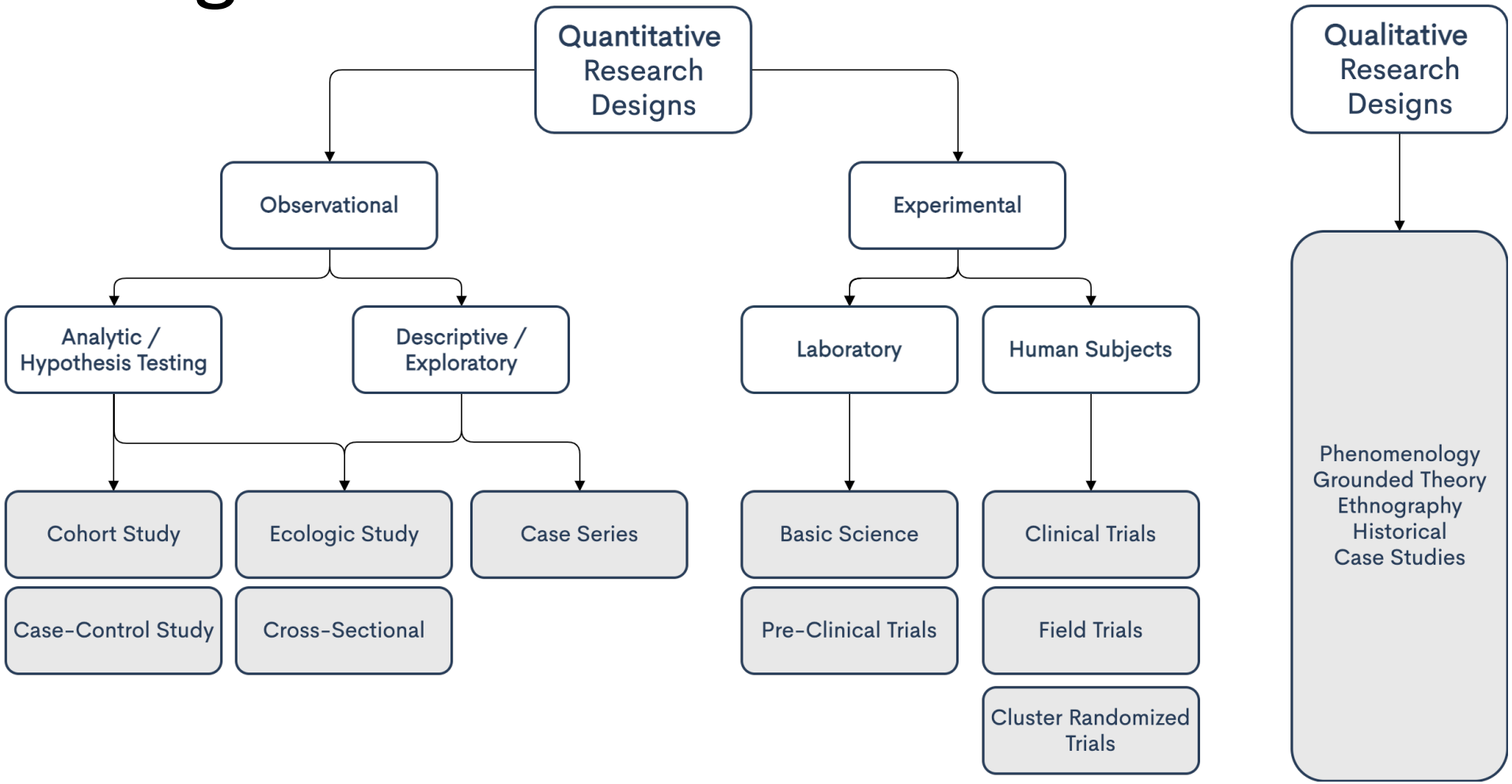
- **Type of Study**
  - What would be the best study design?

Research Question Examples

- Does a multimodal analgesia (I) administered intraoperatively in general surgery patients (P) reduce self-reported pain in the 24 hours after surgery (O) compared with opioid only analgesia (C)?

- Are 30- to 50- year old men (P) who have laparoscopic hernia surgery (I) compared with those undergoing open hernia surgery (C) at increased risk for readmission to the hospital (O) during the 30 days after surgery (T)?

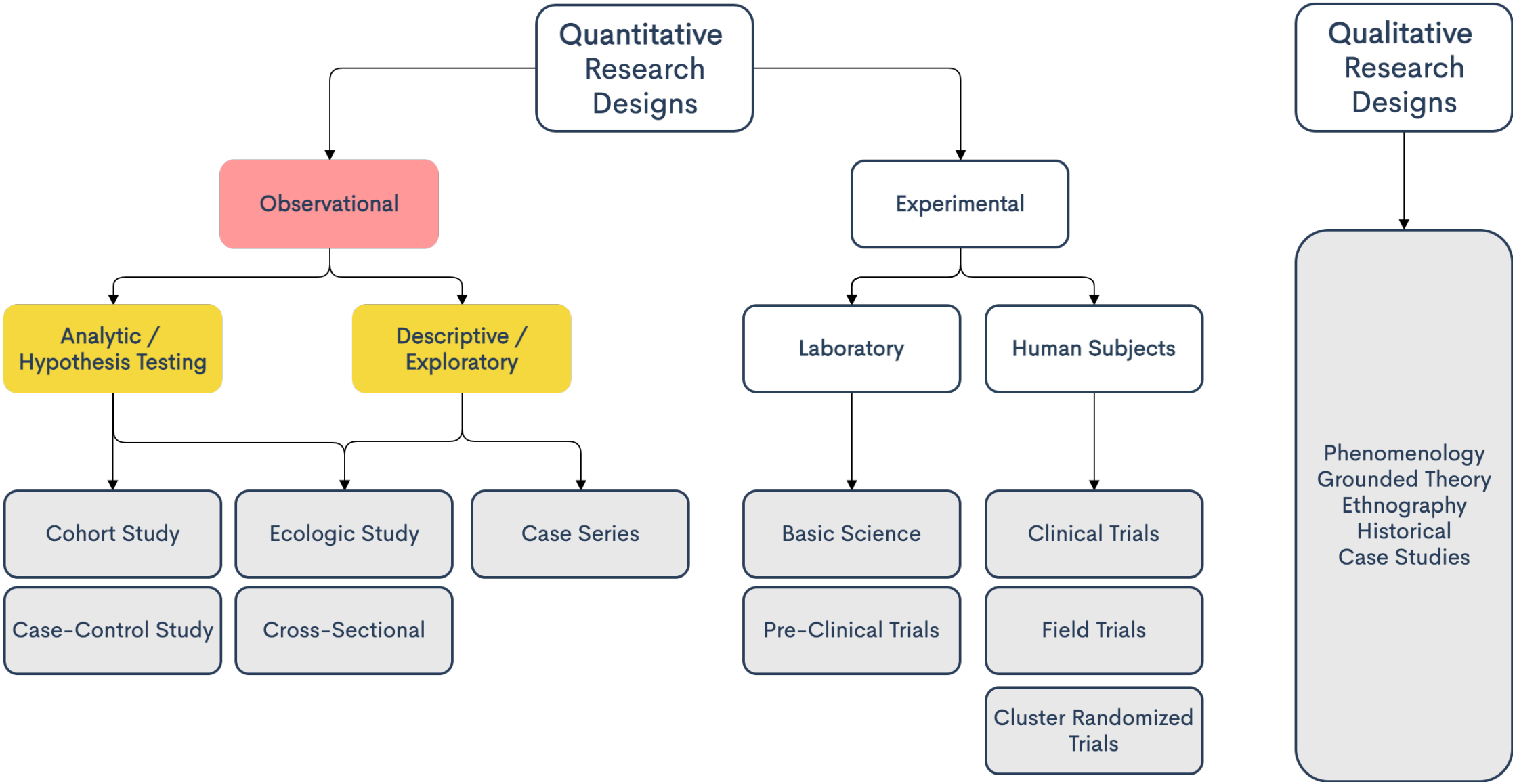
Study Designs



Experimental vs Observational Designs

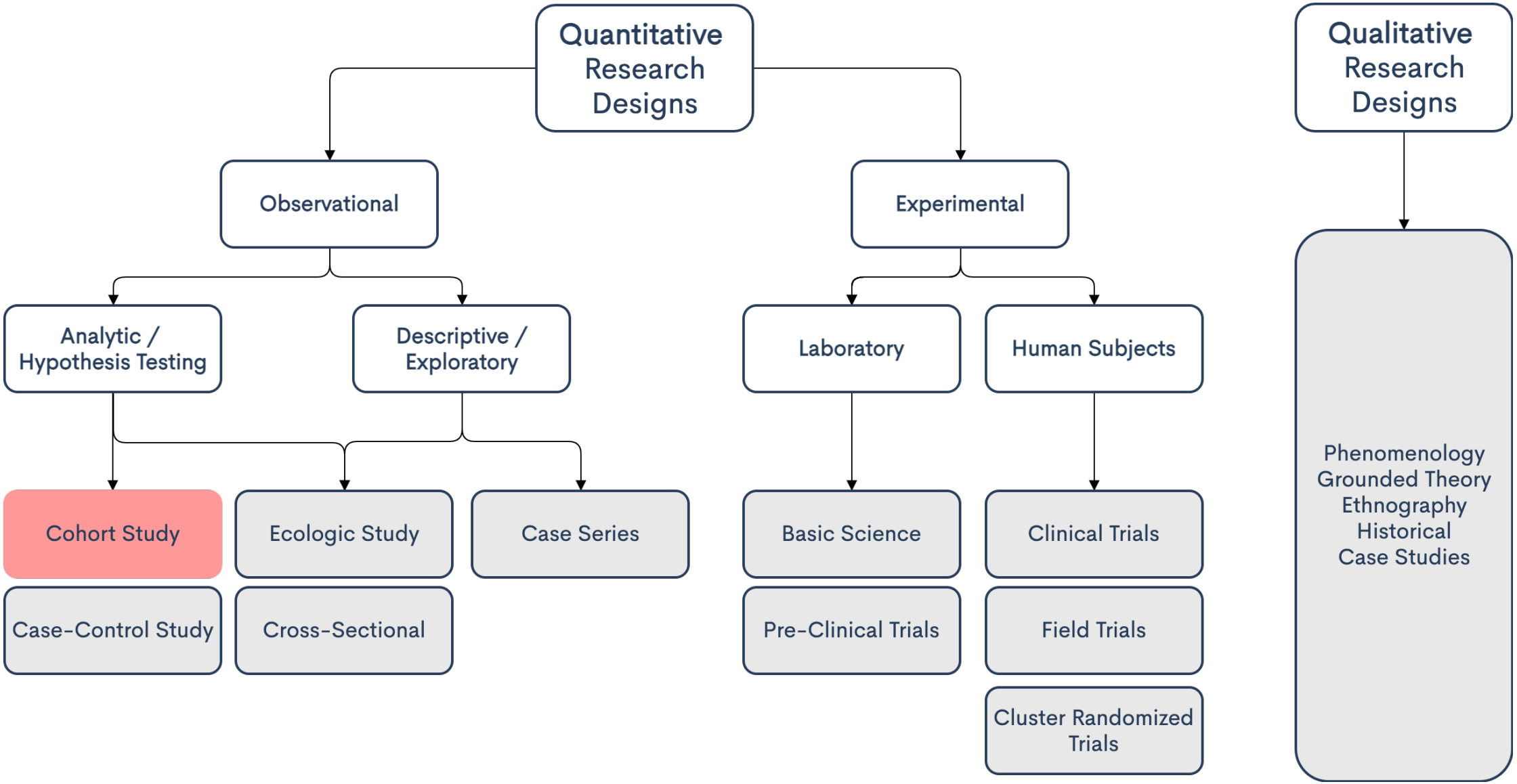
- **Experimental / Interventional**
  - Higher quality evidence
  - Investigator manipulates the conditions (i.e., Assigns treatment groups)
  - Experimental studies are only ethically permissible when "adherence to the protocol does not conflict with the subject's best interest."
    - Example. It is unethical to force some patients to smoke and others not to smoke

- **Non-experimental / Observational**
  - Came about due to ethical and cost restrictions of experimental studies
  - The investigator **does not assign exposure status**
  - Rely heavily on understanding the **selection of subjects** into treatment groups
    - Source of A LOT of our research design concerns.
  - **Less valid** than experimental designs but also **less resource-intensive** (time, money, data, etc.)
  - May be better for **rare outcomes**


Quantitative Research Designs
Observational
Analytic / Hypothesis Testing
Descriptive / Exploratory
Cohort Study
Ecologic Study
Case Series
Case-Control Study
Cross-Sectional
Experimental
Laboratory
Human Subjects
Basic Science
Pre-Clinical Trials
Clinical Trials
Field Trials
Cluster Randomized Trials
Qualitative Research Designs
Phenomenology
Grounded Theory
Ethnography
Historical
Case Studies

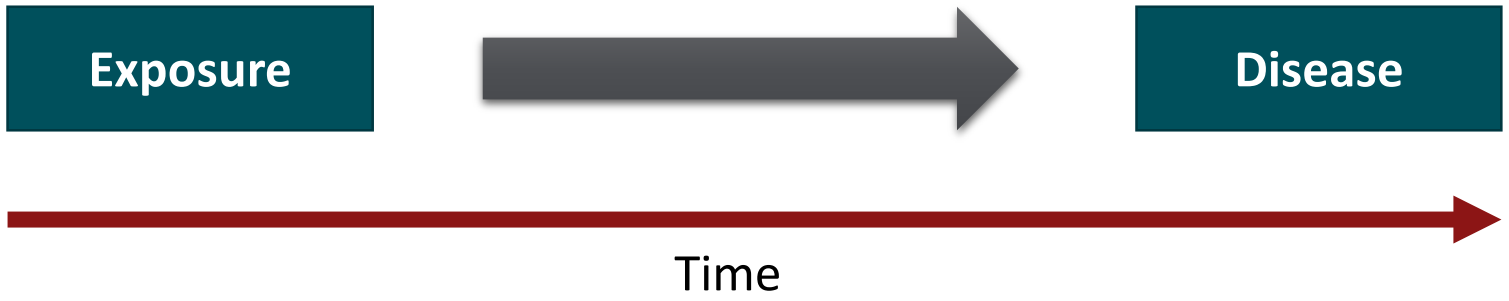
Analytic vs. Descriptive

Analytic	Descriptive
Test hypotheses	**Generate hypotheses**
Quantify the direction and magnitude of associations.	Identifies and describes patterns by place, time, and/or person in a population
	Lacks a comparison group!


Quantitative Research Designs
Observational
Experimental
Analytic / Hypothesis Testing
Descriptive / Exploratory
Laboratory
Human Subjects
Cohort Study
Ecologic Study
Case Series
Basic Science
Clinical Trials
Case-Control Study
Cross-Sectional
Pre-Clinical Trials
Field Trials
Cluster Randomized Trials
Qualitative Research Designs
Phenomenology
Grounded Theory
Ethnography
Historical
Case Studies

Cohort Studies

- Well-defined group of subjects that are followed over time for an outcome of interest.
- Research subjects are identified by their **exposure status**.



Cohort Studies

- **Prospective**
  - Exposure is assessed before the disease develops

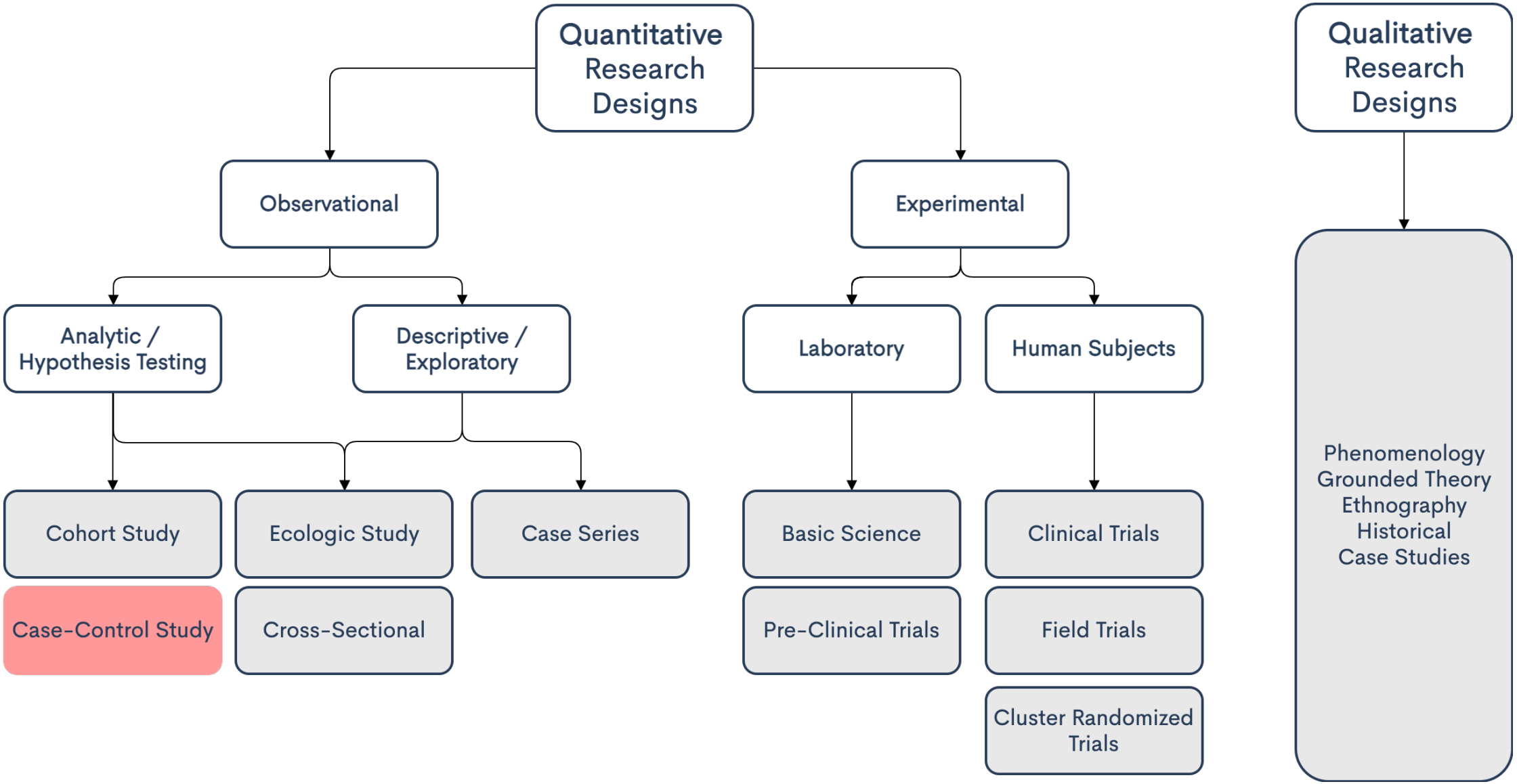
- **Retrospective**
  - Exposure is assessed after some people have already developed disease

Cohort Studies

- **Strengths**
  - Establishes a **temporal association** between exposure and disease
  - Can measure **incidence**
  - Good for **rare exposures** and common diseases
  - Can look at **multiple outcomes**
  - Prospective studies allow better control over sampling and **better quality assessments** over time.
    - Existing data may be incomplete, inaccurate, or measured in ways that are not ideal for answering the research question.

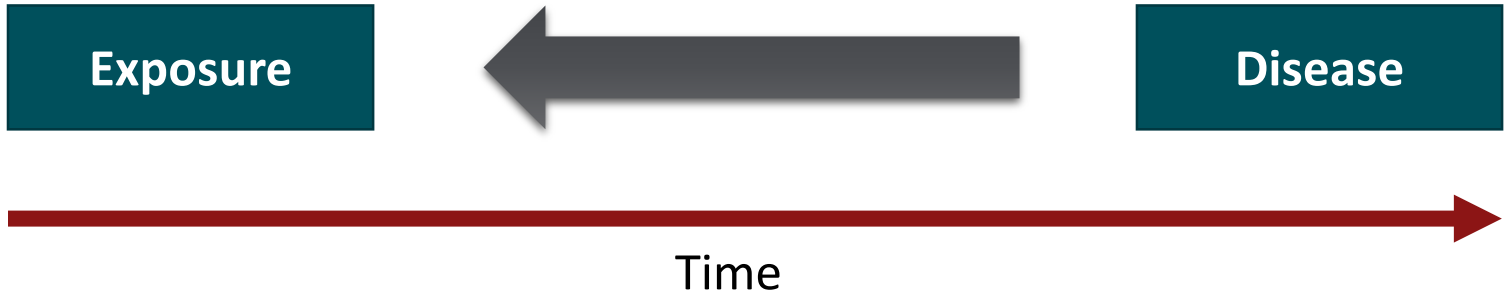
Cohort Studies

- **Weaknesses**
  - **Recall bias** can be an issue for retrospective studies
  - **Loss-to-follow-up** can also become an issue in long prospective studies
  - Prospective cohort studies can be **resource-intensive** (large sample size, long follow-up)
  - Not good for rare diseases/outcomes


Quantitative Research Designs
Observational
Experimental
Analytic / Hypothesis Testing
Descriptive / Exploratory
Laboratory
Human Subjects
Cohort Study
Ecologic Study
Case Series
Basic Science
Clinical Trials
Case-Control Study
Cross-Sectional
Pre-Clinical Trials
Field Trials
Cluster Randomized Trials
Qualitative Research Designs
Phenomenology
Grounded Theory
Ethnography
Historical
Case Studies

Case-Control Studies

- Research subjects are identified by their **disease status**
- **Always retrospective**



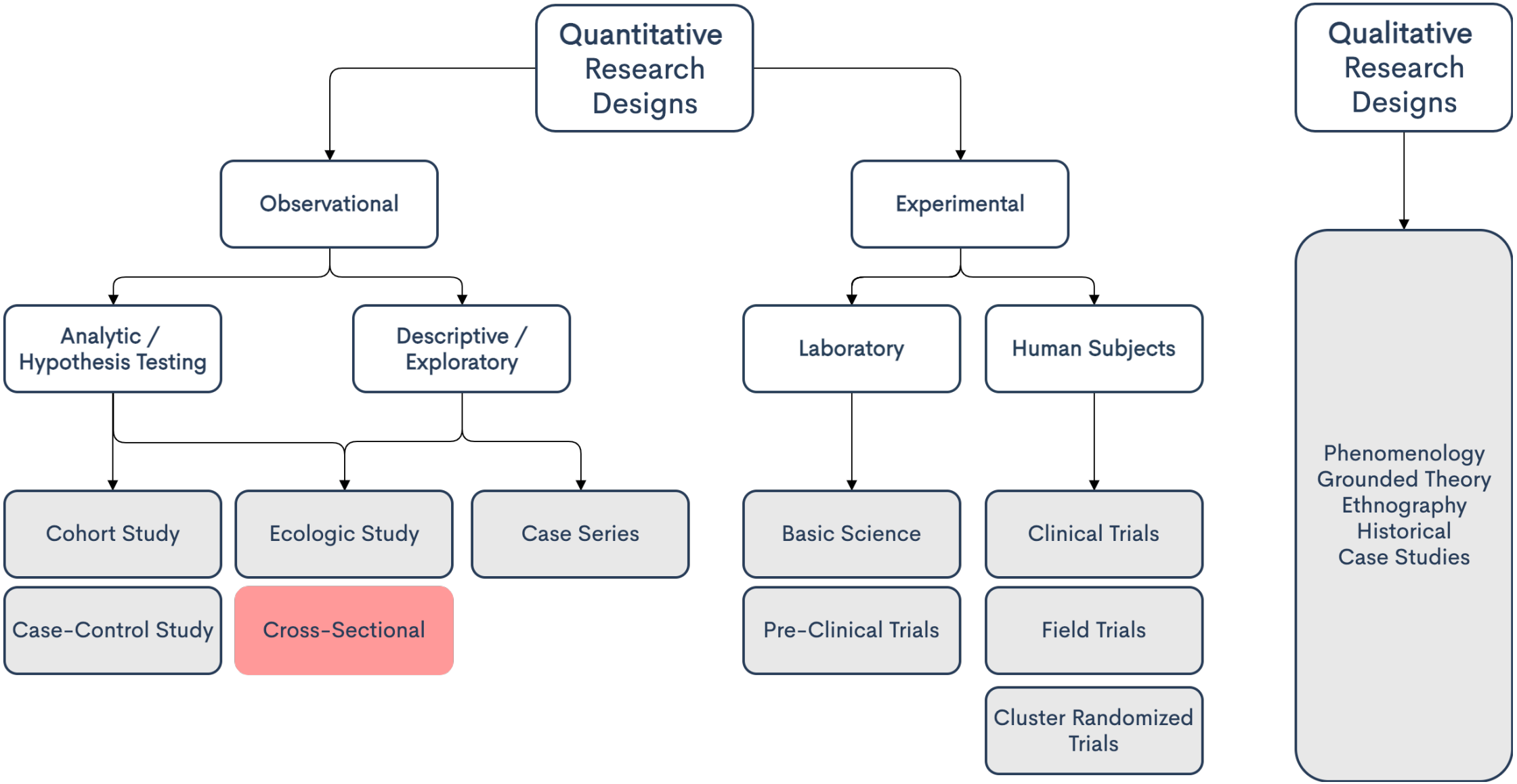
Case-Control Studies

- Key considerations
  - Case selection
    - Cases should be **representative of all of diseased subjects** in the community
  - Control selection
    - Controls should be similar to the cases in all respects other than the disease in question
    - Should be representative of all persons without the disease in the population from which the cases are selected
    - **Should have the potential to become cases**

Case-Control Studies

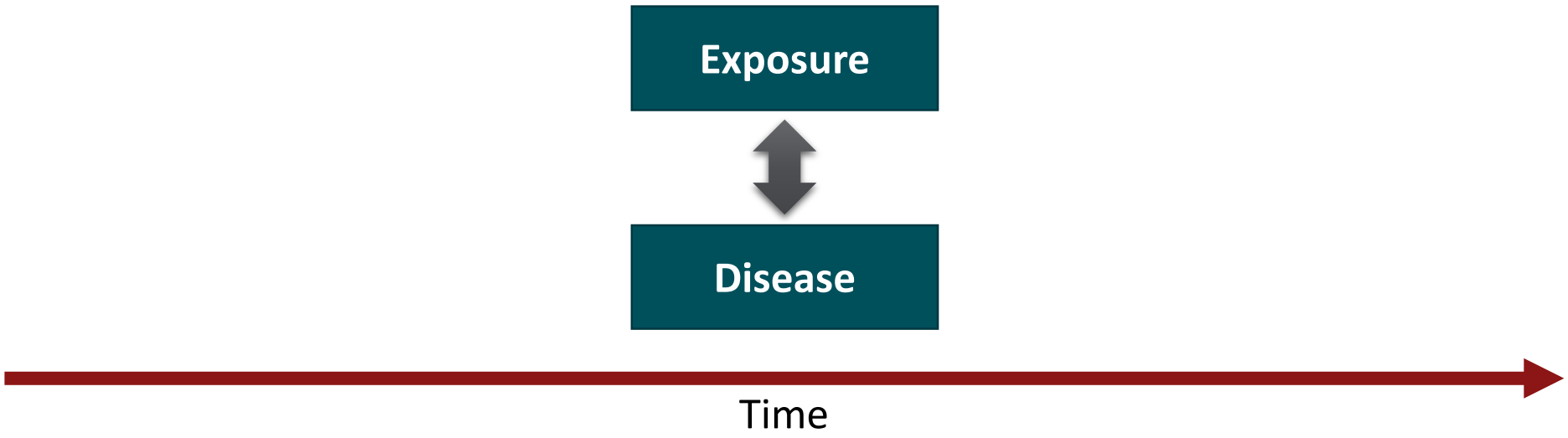
- **Strengths**
  - Good for **rare outcomes**
  - Can be less resource-intensive
  - Can assess **multiple exposures**
    - Case-control studies are useful for generating hypotheses about the causes of an outcome variable.

- **Weaknesses**
  - **More prone to bias (recall bias, selection bias, etc.)**
  - Do not estimate incidence or prevalence
  - Examine only one outcome


Quantitative Research Designs
Observational
Experimental
Analytic / Hypothesis Testing
Descriptive / Exploratory
Laboratory
Human Subjects
Cohort Study
Ecologic Study
Case Series
Basic Science
Clinical Trials
Case-Control Study
Cross-Sectional
Pre-Clinical Trials
Field Trials
Cluster Randomized Trials
Qualitative Research Designs
Phenomenology
Grounded Theory
Ethnography
Historical
Case Studies

Cross-Sectional Studies

- Both the exposure and outcome are assessed at the **same point in time or over a short period of time**.



Cross-sectional Studies

- **Strengths**
  - Provide a point-in-time **prevalence** estimate
  - Require less time to complete and **avoids the problem of loss to follow-up**
  - Can be used at the beginning of a cohort or clinical trial to provide baseline characteristics

- **Weaknesses**
  - Does not estimate incidence
  - Provides **less evidence of a causal relationship** because temporality cannot be confirmed

Ecological Studies

- Unit of analysis is a **group**, not the individual.
- Result in aggregate measures that are reported (descriptive) or compared (analytic).
- Also, good for rare diseases or to study the effect of large-scale public health interventions.
- Should always consider the potential **ecologic fallacy**
  - When the relationship observed at the group level does not represent the relationship at the individual level (ex., relationship may differ based on grouping levels)

Case Series

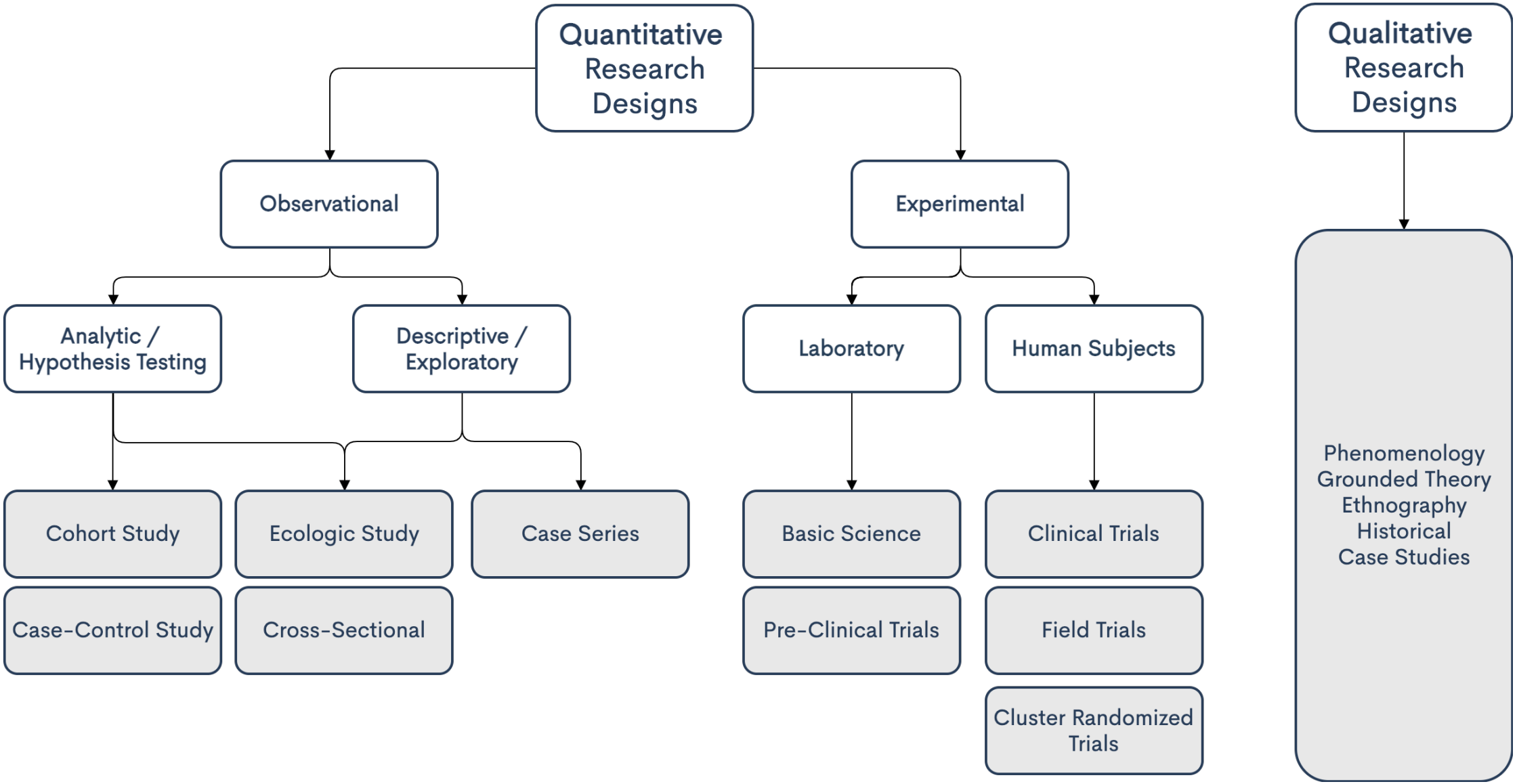
- Useful for:

1. Describing a **new disease** process
2. Identifying and describing **rare manifestations**
3. Identifying **emerging** health conditions

- Example. A case series of the **first 1000 patients with AIDS**. 72.7% were homosexual or bisexual males and 23.6% were injection drug users. It did not require a formal control group to conclude that these groups were at higher risk.

Case Series

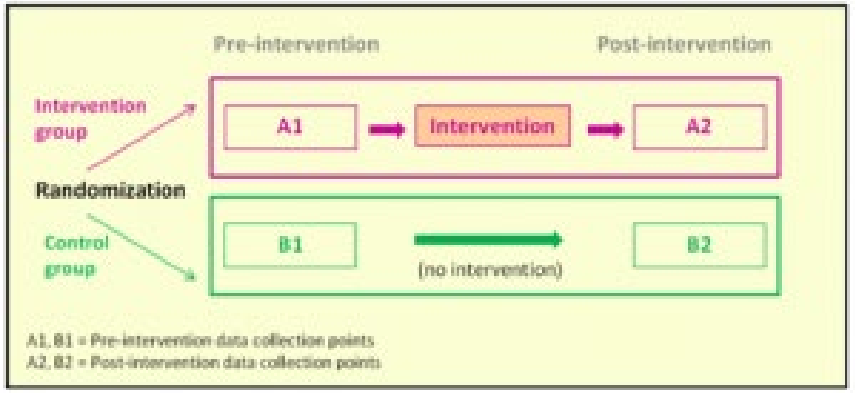
- **Strengths**
  - Cost-effective method to describe rare manifestations and new/emerging diseases
- **Weaknesses**
  - Purely descriptive
  - Weakest form of evidence
  - Misleading and may suggest a plausible causal relationship where none exists in real population


Quantitative Research Designs
Observational
Experimental
Analytic / Hypothesis Testing
Descriptive / Exploratory
Laboratory
Human Subjects
Cohort Study
Ecologic Study
Case Series
Basic Science
Clinical Trials
Case-Control Study
Cross-Sectional
Pre-Clinical Trials
Field Trials
Cluster Randomized Trials
Qualitative Research Designs
Phenomenology
Grounded Theory
Ethnography
Historical
Case Studies

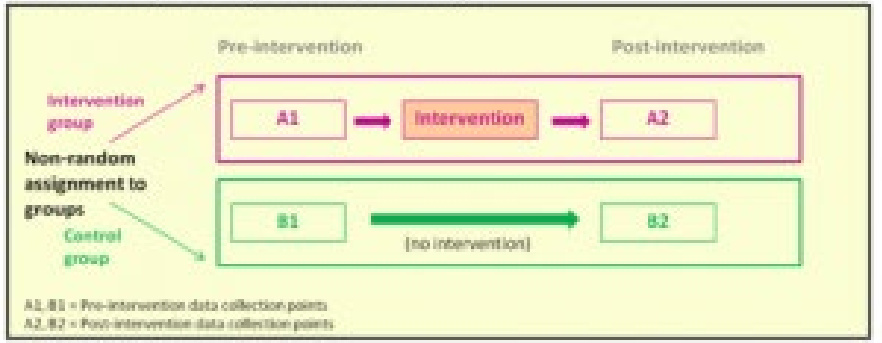
Experimental Designs

- Investigator manipulates the independent variable (experimental variable)
- Better quality than observational designs
- True experimental designs involve randomization
- Quasi-experimental designs do not use randomization
  - Regression Discontinuity
  - Difference-in-Differences
  - Instrumental Variable Analysis

Classical Design of Randomized Experiments

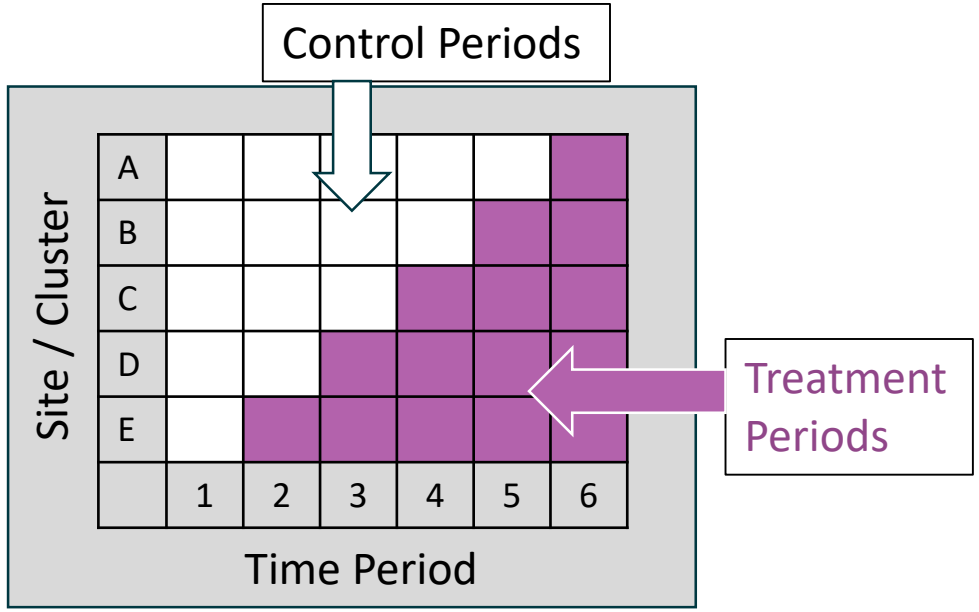


Classical Quasi-Experimental Design



Randomization in Experimental Designs

- Individual randomization
  - Pros: Best power per n (optimal efficiency)
  - Considerations: Contamination within sites, may be infeasible
- Cluster randomization
  - Pros: Minimize contamination
  - Considerations: Individuals not independent, intracluster correlation coefficient (ICC), larger n needed to achieve power
- Stepped-wedge design
  - Pros: All sites get intervention, random timing/order
  - Considerations: Individuals not independent (ICC), much larger n needed to achieve power



[Rutterford 2015]

Hybrid Study Designs

- Combine elements of **different designs**
  - A nested case control study within a cohort study
  - A study that incorporates both a qualitative and quantitative design (Mixed Methods Study)

- Can be used to address some of issues of a single study design

Hybrid Study Designs

Design Concern	Hybrid Study Suggestion
Underlying hypothesis is not well-supported	Use a qualitative design to support and guide findings in a quantitative study
Retrospective cohort data does not include **detailed disease information**	Nested case-control or case-cohort to get more granular data that is not already collected
Concern about **case and control selection**	Nested case-control design can ensure all cases and controls come from the same population

Study Design Considerations

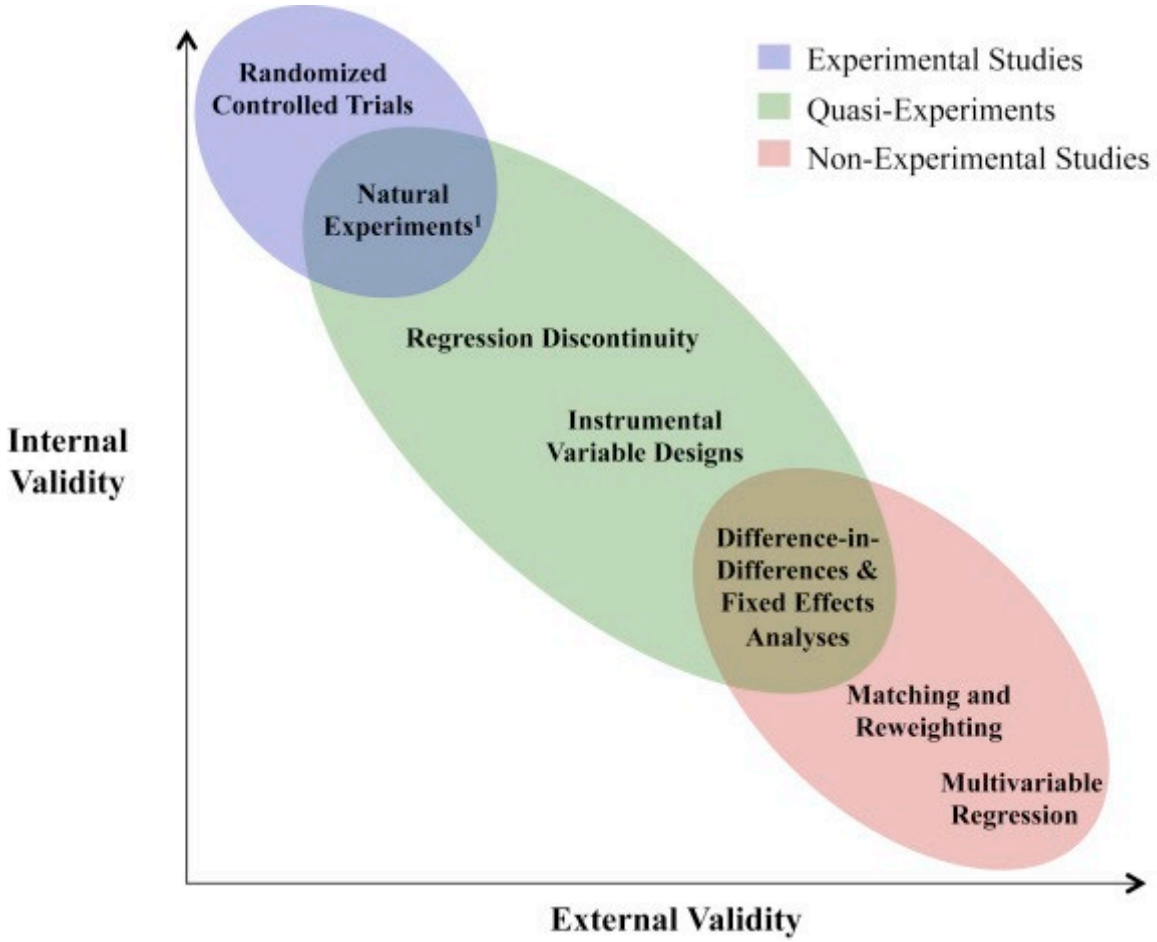
Validity

Internal validity: The extent to which the observed results represent the truth in the population we are studying and, thus, are not due to methodological errors

- Study design + analysis
  - Randomized treatment assignment
  - Specific information collected (or not)
  - Data analysis methods

External validity: Generalizability to other settings and populations

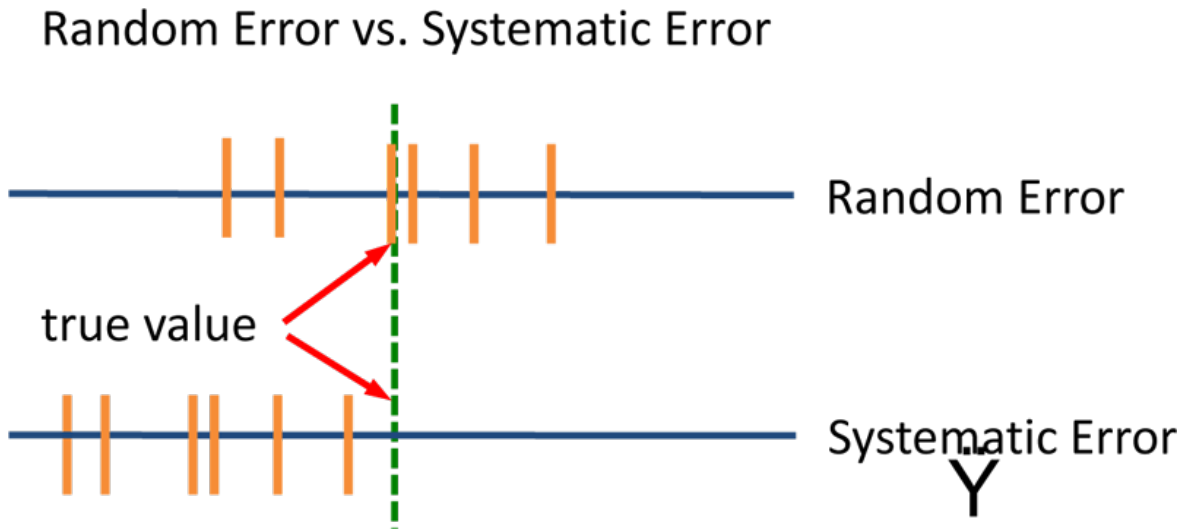
- Study design
  - Which patients are included
  - How the treatment is implemented



Measurement Error

- **Error**: difference between the observed result and the truth
- The goal of a good research design is to **minimize error**

- Random Error
- Systematic Error



Measurement Error

- **Random Error** (Precision / Reliability)
  - The degree to which our research methods produce consistent results
  - Example. Blood pressure measurements when there is not standardized protocol
  - Exists in ALL Research Designs

- **Systematic Error** (Accuracy / Validity)
  - Closeness of a measured value to the truth
  - The degree to which we are measuring what it is supposed to measure
  - Example. Taking weights with a scale that consistently reads 5 pounds light.

Bias

- Bias is a **systematic error** in the design, conduct or analysis of a study that results in a mistaken estimate of an exposure's effect on the risk of disease — (Schlesselman and Stolley, 1982)

    - Selection bias
    - Information bias
    - Confounding

Selection Bias

- Method of **participant selection** distorts the exposure-outcome relationship from that present in the target population

  - Surveying by phone may systematically exclude patients without phones (**non-response bias**)
  - Patients without the exposure may be more likely to not complete the study (**loss-to-follow-up bias**)
  - Healthier patients may be more likely to get a certain risky treatment (**confounding by indication**)
  - Patients affected by the disease may be more likely to participate (**volunteer bias**)

Information bias

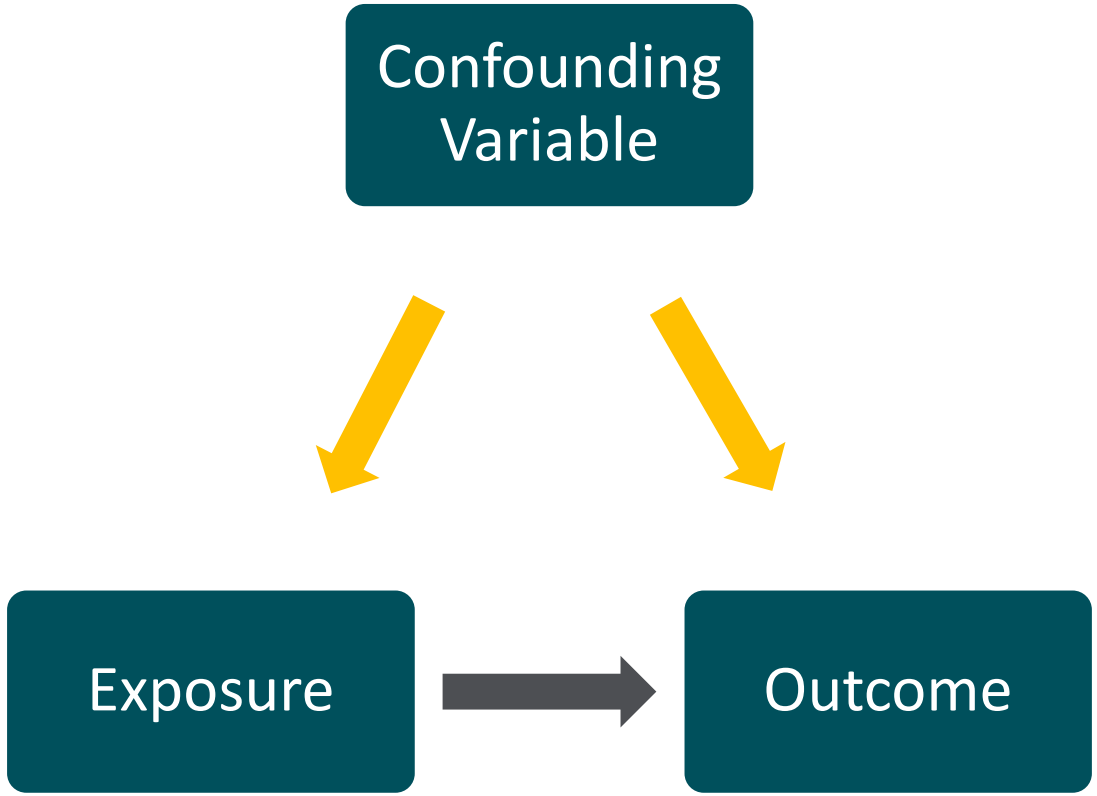
- Information bias occurs when **information is collected differently between two groups (misclassification)**
  - **Non-differential** misclassification occurs when the level of misclassification does not differ between the two groups
  - **Differential** misclassification occurs when the level of misclassification differs between the two groups
- Differential misclassification may lead to a distortion of the effect.

Confounding

- Confounding occurs when the **observed result** between exposure and disease **differs from the truth** because of the influence of the third variable

- In contrast, effect modification is when the effect of the exposure is different among subgroups – not a distortion of the effect due to a systematic error.

Confounding



- Associated with both exposure and outcome
- Distributed **unequally** among comparison groups
- **NOT in the causal pathway** from exposure to outcome

Confounding

- Research Design Solutions
  - **Restrict** the cohort
  - **Instrumental variables**
  - **Match** comparison groups
  - Covariate **adjustment** (statistical control)
  - **Randomize** subjects (experimental design)

Statistics

Talk to a biostatistician before collecting any data!

Statistical Inference

- The process of forming judgments about the parameters of a target population using a sample drawn from that population

1. Estimation
   - Point estimation: Summarize the sample by a single value
   - Interval Estimation: Defining a range of values within which the true population parameter exists

2. Hypothesis Testing

Hypothesis Testing Steps

State the hypothesis

Identify all relevant alternative hypotheses

Consider the statistical assumptions

Variable type

Distributional characteristics

Determine the appropriate statistical test and significance level

Conduct the statistical test

Variable Types

Independent Variable (primary IV)

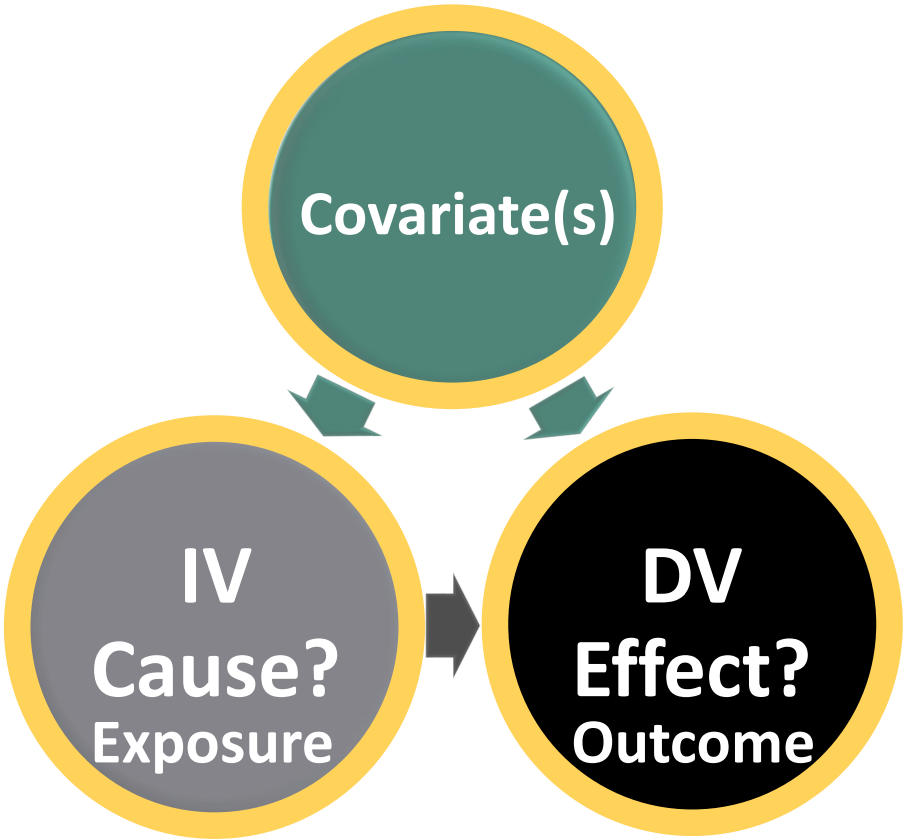
- Exposure (Intervention)
- Occurring first
- Causal relationship (?)

Dependent Variable (DV)

- Outcome
- Response variable
- Occurring after predictors

Covariate(s)

- Related to both outcome and exposure
- Must be included for internal validity

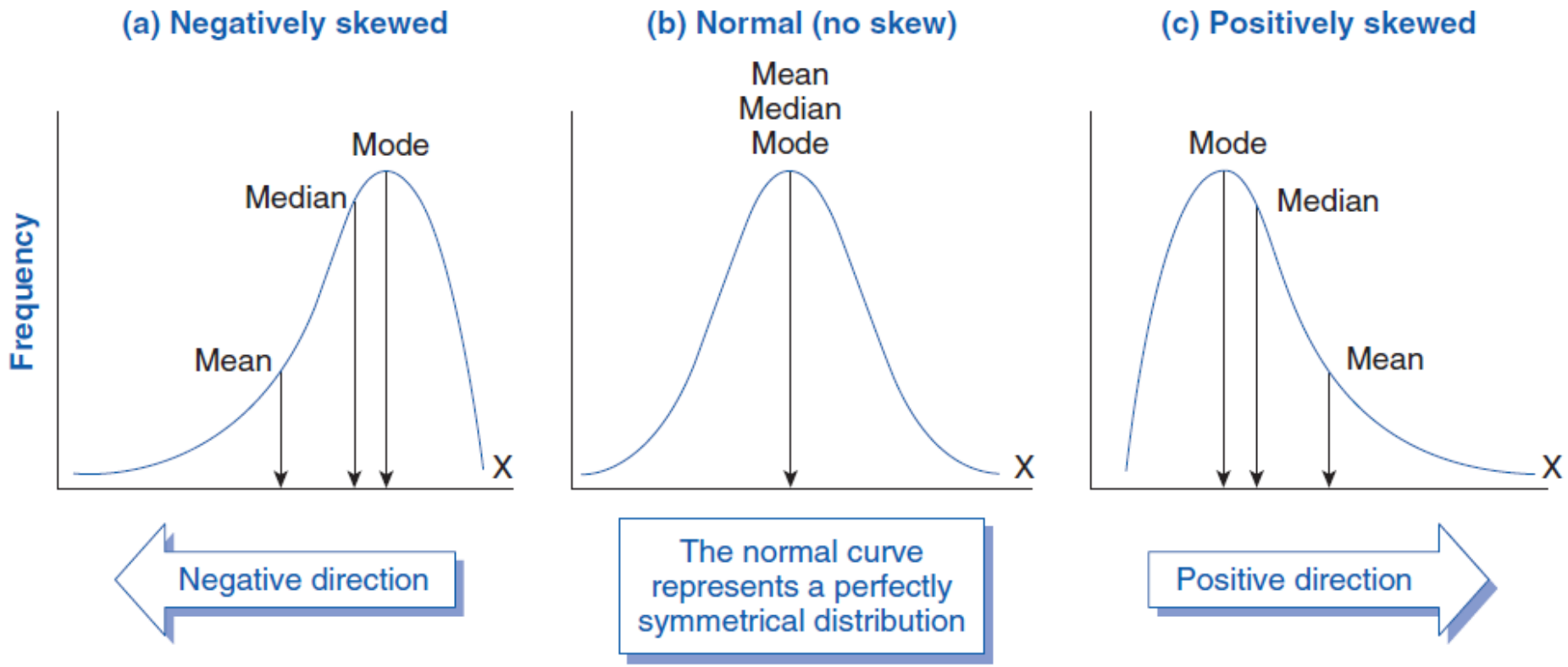


Variable Measurement Scales

Type of Measurement	Characteristics	Examples	Descriptive Stats	Information Content
Continuous	Ranked spectrum; quantifiable intervals	Weight, BMI	Mean (SD) + all below	Highest
Ordered Discrete		Number of cigs / day	Mean (SD) + all below	High
Categorical Ordinal (Polychotomous)	Ordered categories	ASA Physical Status Classification	Median	Intermediate
Categorical Nominal (Polychotomous)	Unordered Categories	Blood Type, Facility	Counts, Proportions	Lower
Categorical Binary (Dichotomous)	Two categories	Sex (M/F), Obese (Y/N)	Counts, Proportions	Low

[Hulley 2007]

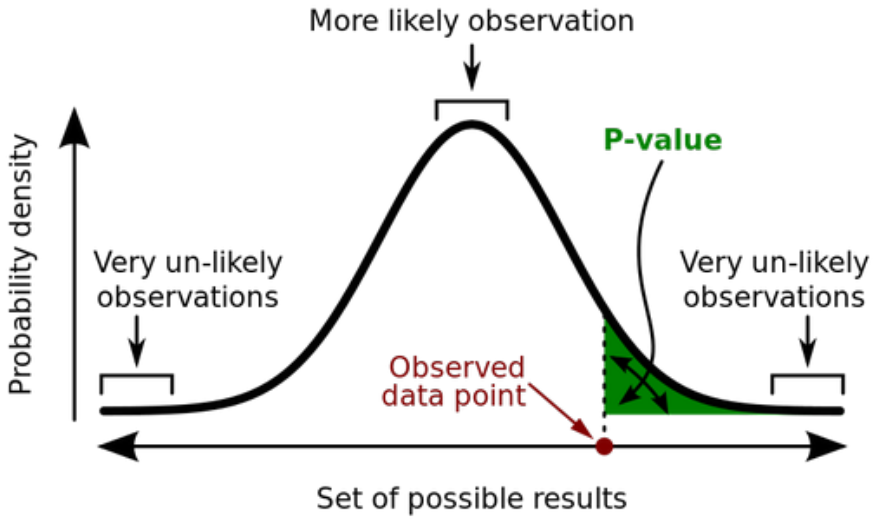
Variable Distributional Characteristics



P-values

The p-value represents the probability of finding the observed **or more extreme** test statistics **if the null hypothesis is true.**

- Measures evidence against H_0
- Smaller p-value, larger evidence against H_0
- Reject H_0 if p-value ≤ α

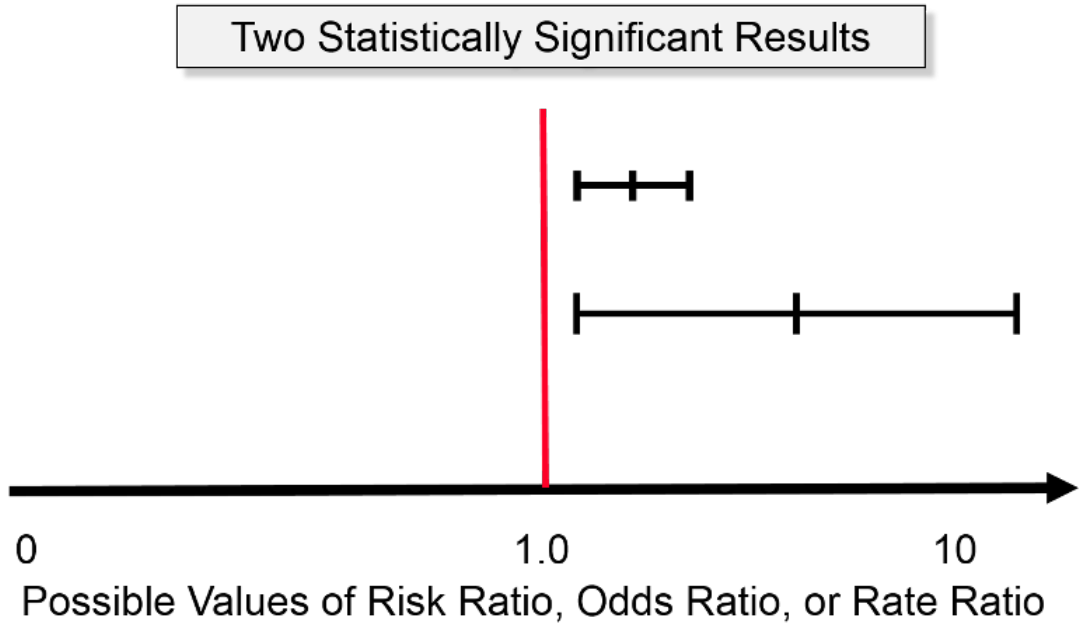


P-Value Pitfalls

- P is highly dependent on sample size
- The *statistical* significance …
  - ➢ does not equal *clinical* significance
  - ➢ does not equal *the magnitude of the effect*
  - ★ Report descriptive statistics with p: n1, n2, %'s, means, SD…
- P is not dichotomous yes/no, but a continuum, <0.001 to >0.99

Confidence Intervals

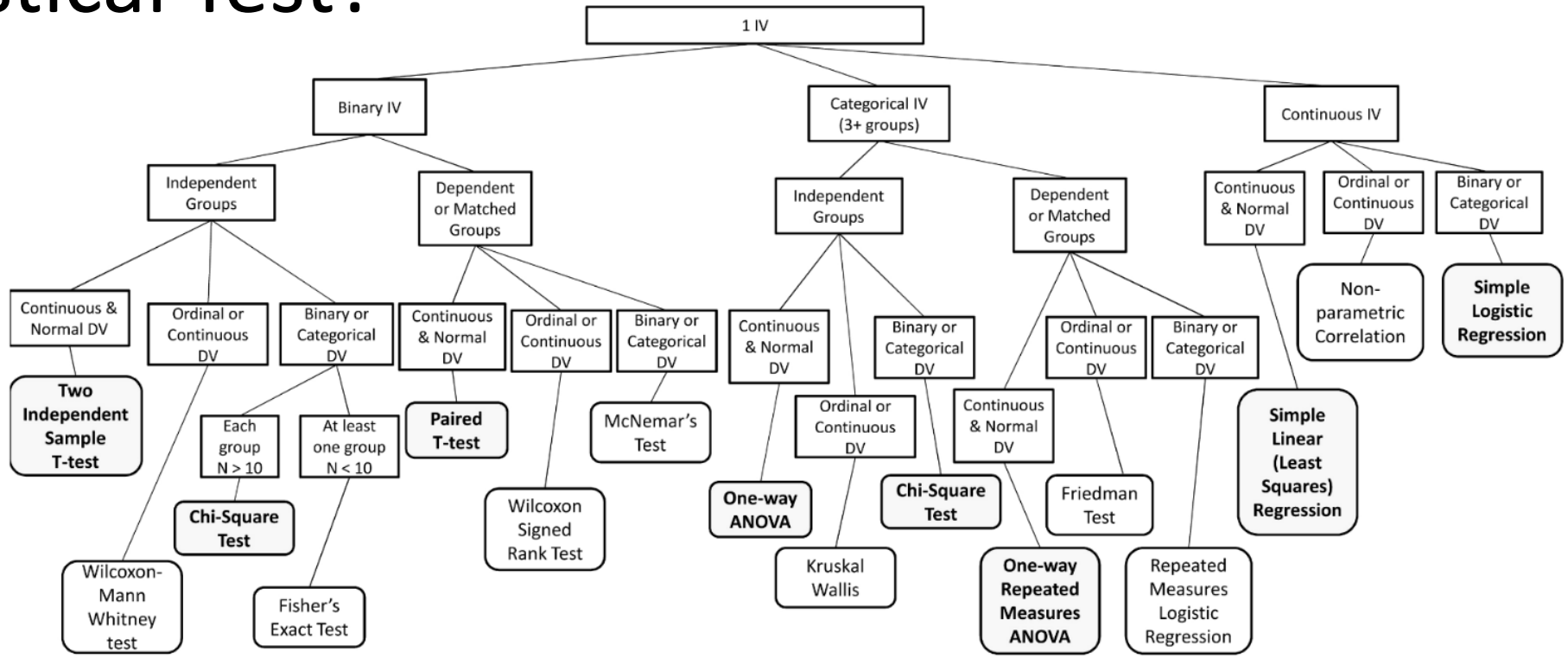
- Range of values within which the true population parameter exists
  - Width of the range of values is determined by a function of sample size and sample variability
- Provides more information than just the p-value alone



Determining the Statistical Test to Use

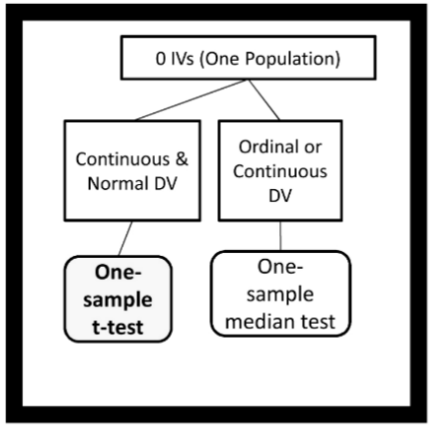
Which Statistical Test?

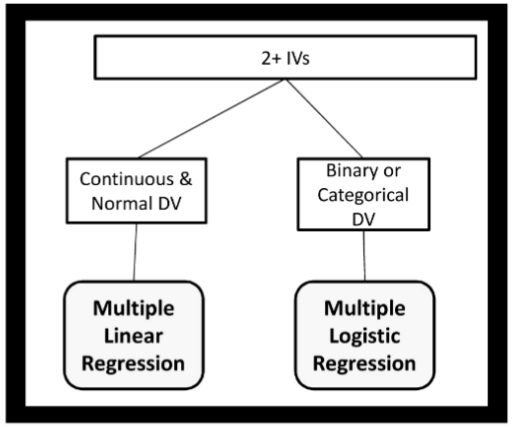
1. **Number of IVs**
2. **IV Measurement Scale**
3. **Independent vs. Matched Groups**
4. **DV Measurement Scale**



LEGEND:

IV = Independent Variable
(i.e. predictor, exposure)
DV = Dependent Variable
(i.e. response, outcome)





Common Regression Models

OUTCOME VARIABLE	APPROPRIATE REGRESSION	MODEL COEFFICIENT
Continuous AND Normal	Linear Regression	Slope (β): How much the outcome increases for every 1-unit increase in the predictor
Binary / Categorical	Logistic Regression	Odds Ratio (OR): How much the **odds** for the outcome increases for every 1-unit increase in the predictor
Time-to-Event	Cox Proportional-Hazards Regression	Hazard Ratio (HR): How much the **rate** of the outcome increases for every 1-unit increase in the predictor

Hierarchical / Mixed Effects Models

Correlated Data

- Grouping of subjects
- Repeated measures over time
- Multiple related outcomes

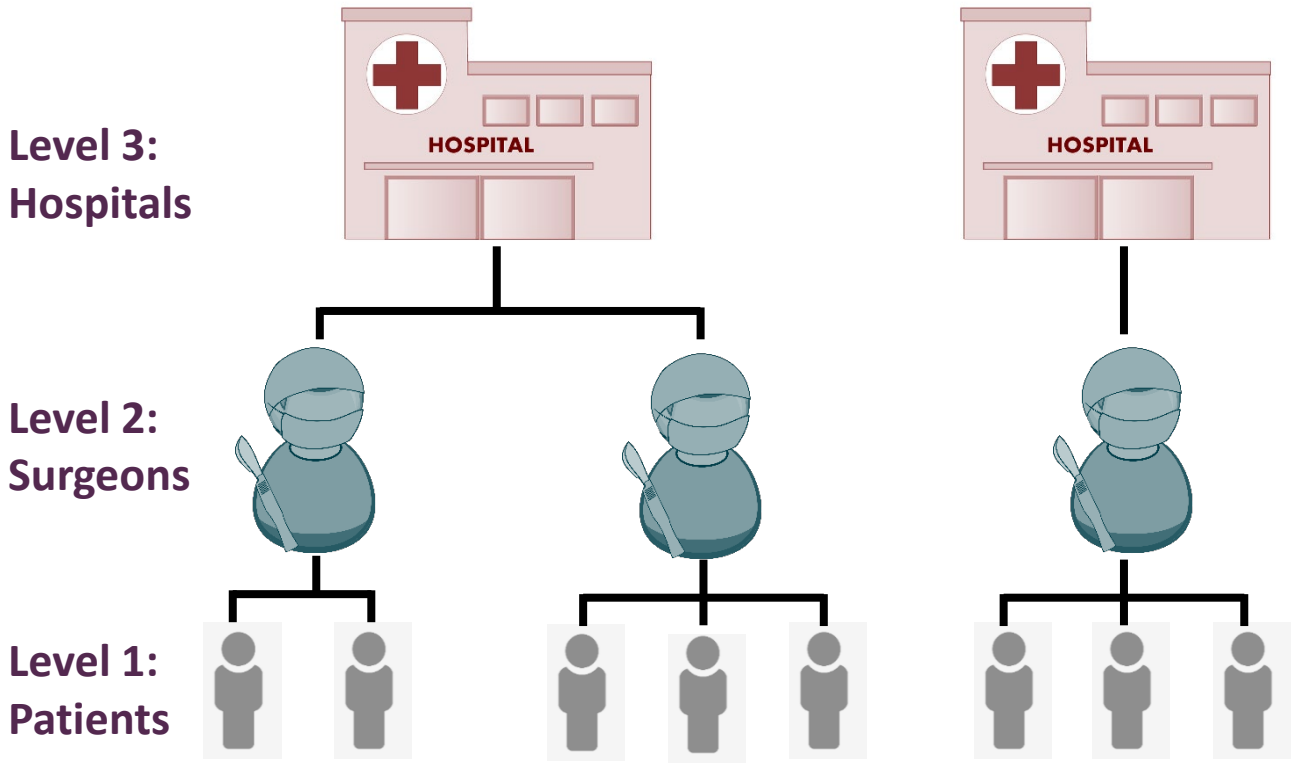
Can handle

- Missing data
- Nonuniform measures

Outcome Variable(s):

- Categorical
- Continuous
- Counts

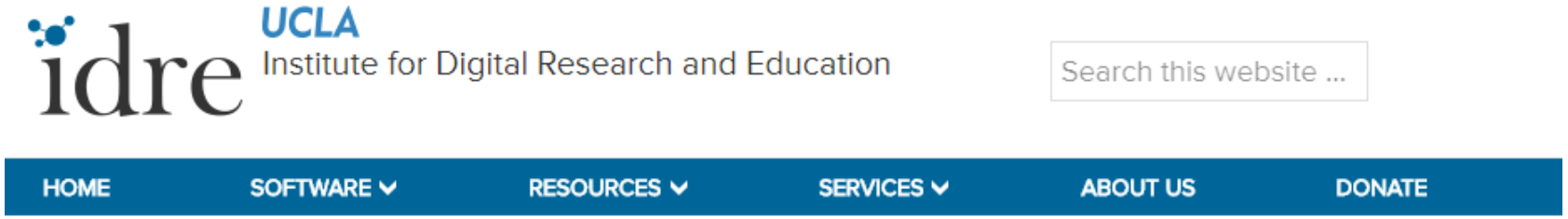
Nested Data



S-SPIRE

Resource

- https://stats.idre.ucla.edu/other/dae/



DATA ANALYSIS EXAMPLES

Regression Models					
Robust Regression	Stata	SAS			R
Models for Binary and Categorical Outcomes					
Logistic Regression	Stata	SAS	SPSS	Mplus	R
Exact Logistic Regression	Stata	SAS			R
Multinomial Logistic Regression	Stata	SAS	SPSS	Mplus	R
Ordinal Logistic Regression	Stata	SAS	SPSS	Mplus	R
Probit Regression	Stata	SAS	SPSS	Mplus	R

Count Models					
Poisson Regression	Stata	SAS	SPSS	Mplus	R
Negative Binomial Regression	Stata	SAS	SPSS	Mplus	R
Zero-inflated Poisson Regression	Stata	SAS		Mplus	R
Zero-inflated Negative Binomial Regression	Stata	SAS		Mplus	R
Zero-truncated Poisson	Stata	SAS			R
Zero-truncated Negative Binomial	Stata	SAS		Mplus	R
Censored and Truncated Regression					
Tobit Regression	Stata	SAS		Mplus	R
Truncated Regression	Stata	SAS			R
Interval Regression	Stata	SAS			R

Power Analysis / Sample Size						
Single-sample t-test	Stata	SAS			R	G*Power
Paired-sample t-test	Stata	SAS			R	G*Power
Independent-sample t-test	Stata	SAS			R	G*Power
Two Independent Proportions	Stata	SAS				G*Power
One-way ANOVA	Stata	SAS				G*Power
Multiple Regression	Stata	SAS				G*Power
Accuracy in Parameter Estimation	Stata					

Thanks!
Questions?

Laura.Graham@va.gov
lagraham@Stanford.edu

@lagrahamepi

"We are all apprentices in a craft where no one ever becomes a master." —Ernest Hemingway

Seven Habits of Highly Effective Data Users

1. **Check quality before quantity.** All data are not created equal; good statistics cannot salvage biased data.
2. **Describe before you analyze.** Special data require special tests; improper analysis gives deceptive results.
3. **Accept the uncertainty of all data.** All observations have some random error; interpretation requires estimating precision or confidence.
4. **Measure error with the right statistical test.** Positive results should be qualified by the chance of being wrong, negative results should be qualified by chance of having missed a true effect.
5. **Put clinical importance before statistical significance.** Statistical tests measure error, not importance; an appropriate measure of clinical importance must be checked.
6. **Seek the sample source.** Results from one dataset do not necessarily apply to others.
7. **View science as a cumulative process.** A single study is rarely definitive.

[Rosenfeld 2014]

Chi-Square or Fisher's Exact (if any cell N<10)

Outcome Variable:

- Binary

Predictor Variable(s):

- Categorical
- Independent groups
- 2+ groups
- 1 independent variable

2 x 2 Table:

		Outcome		
		MI +	**MI –**	**Tot**
Predictor	**GA**	324	37,178	37,502
	RA	20	4,743	4,763
	Tot	344	41,921	42,265

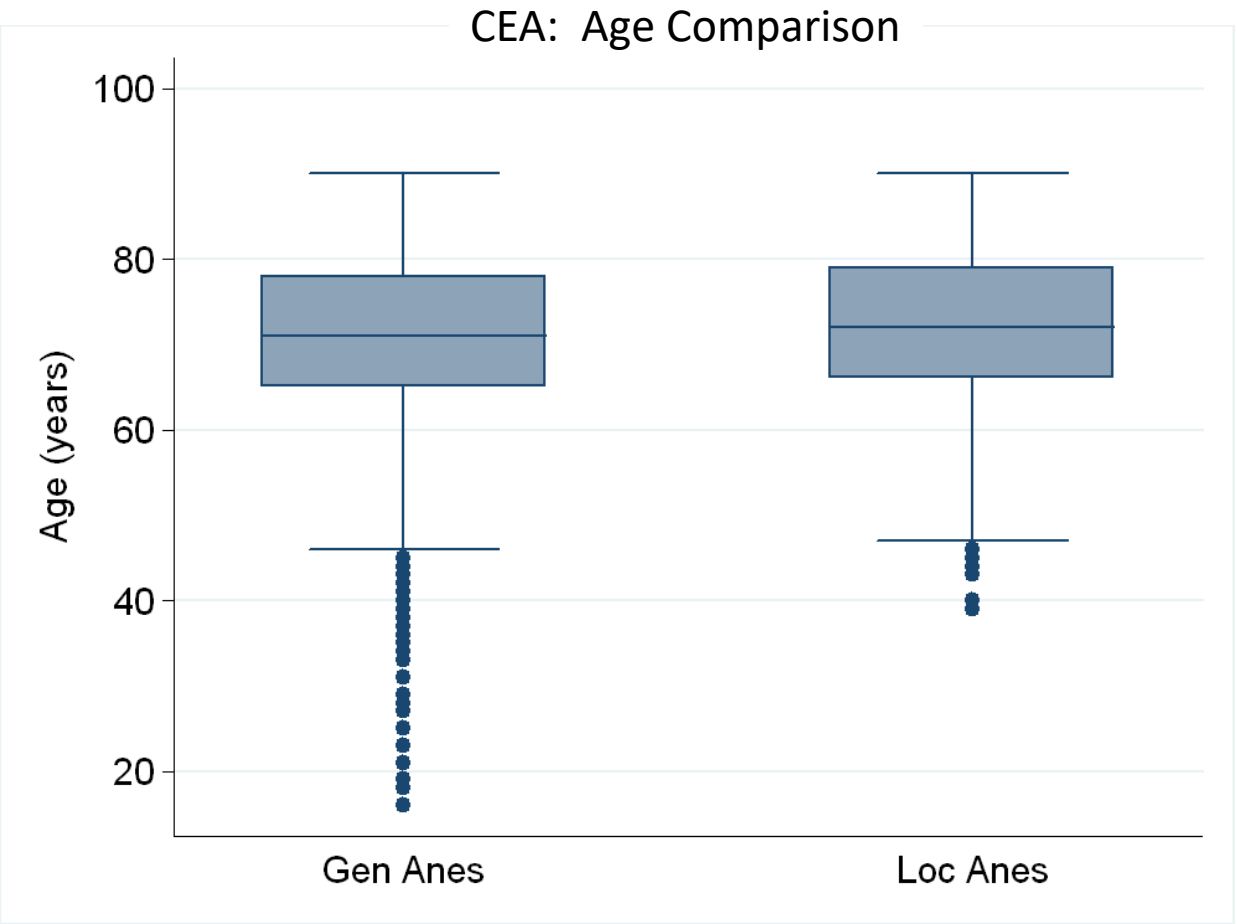
Student's T-test

Outcome Variable:

- Continuous
- Normally distributed

Predictor Variable(s):

- 1 Categorical IV
- 2 Independent groups



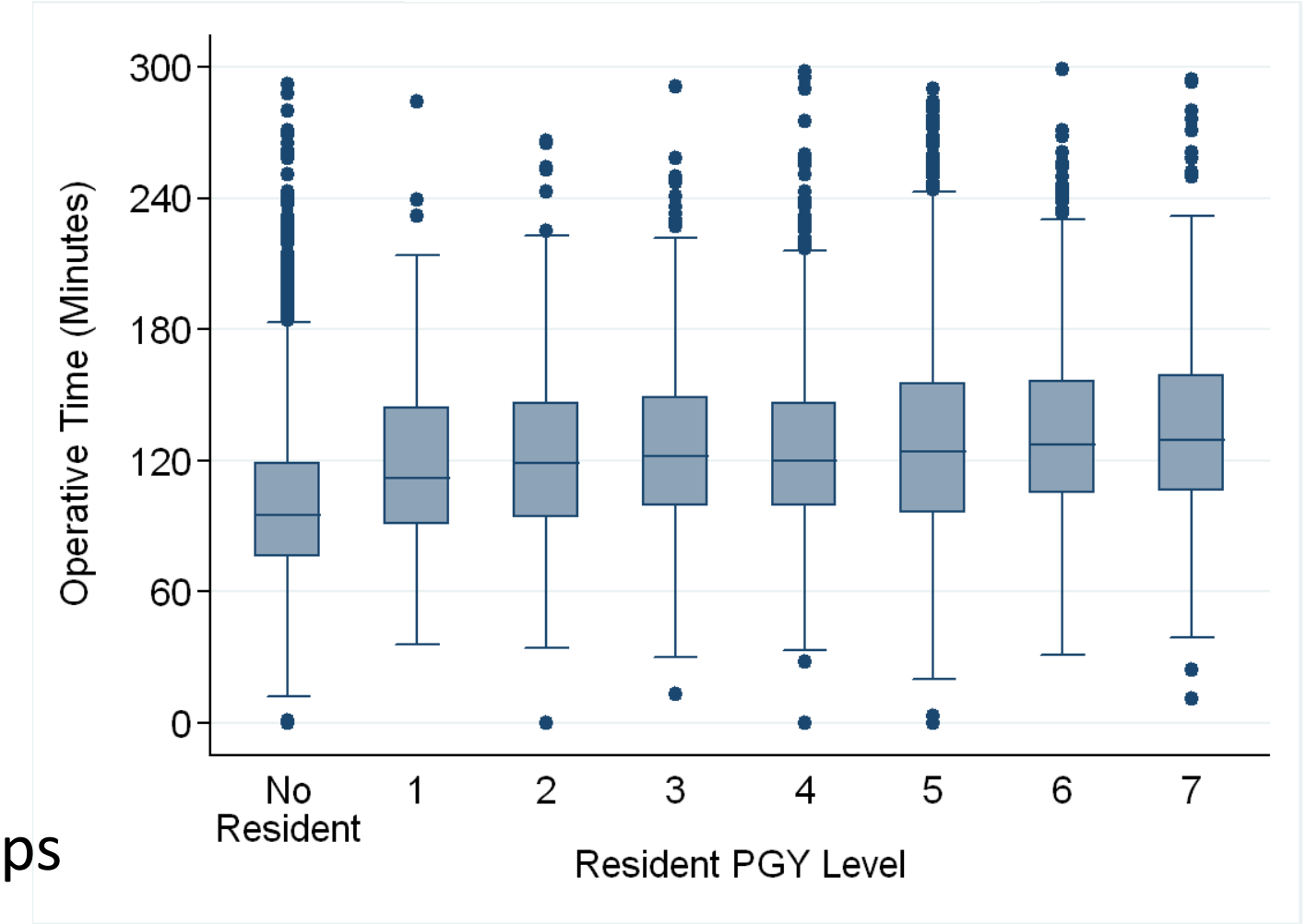
Analysis of Variance (ANOVA)

Outcome Variable:

- Continuous
- Normally distributed

Predictor Variable(s):

- 1 Categorical IV
- 3+ Independent groups



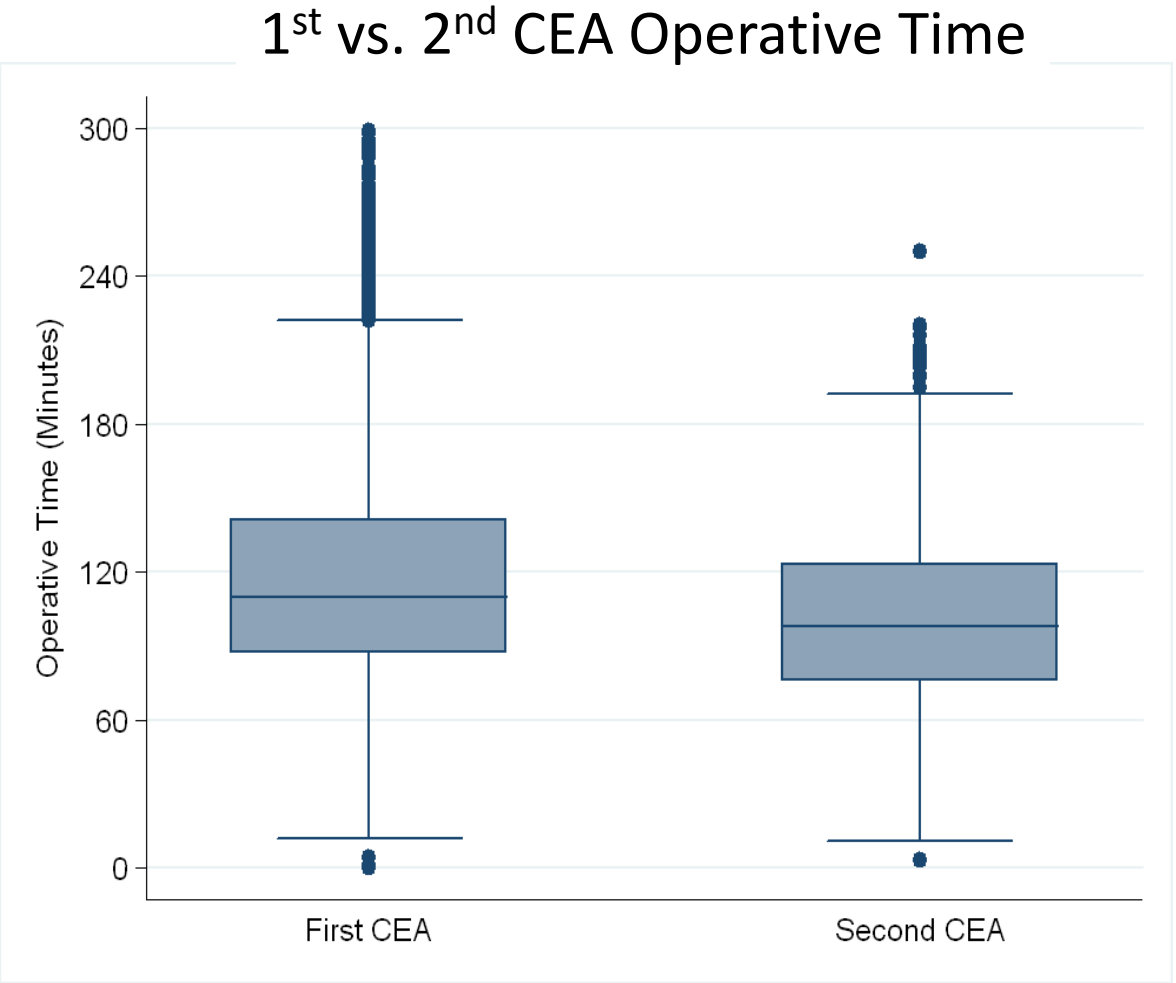
Paired T-test

Outcome Variable:

- Continuous
- Normally distributed

Predictor Variable(s):

- 1 Categorical IV
- 2 **matched / dependent** groups



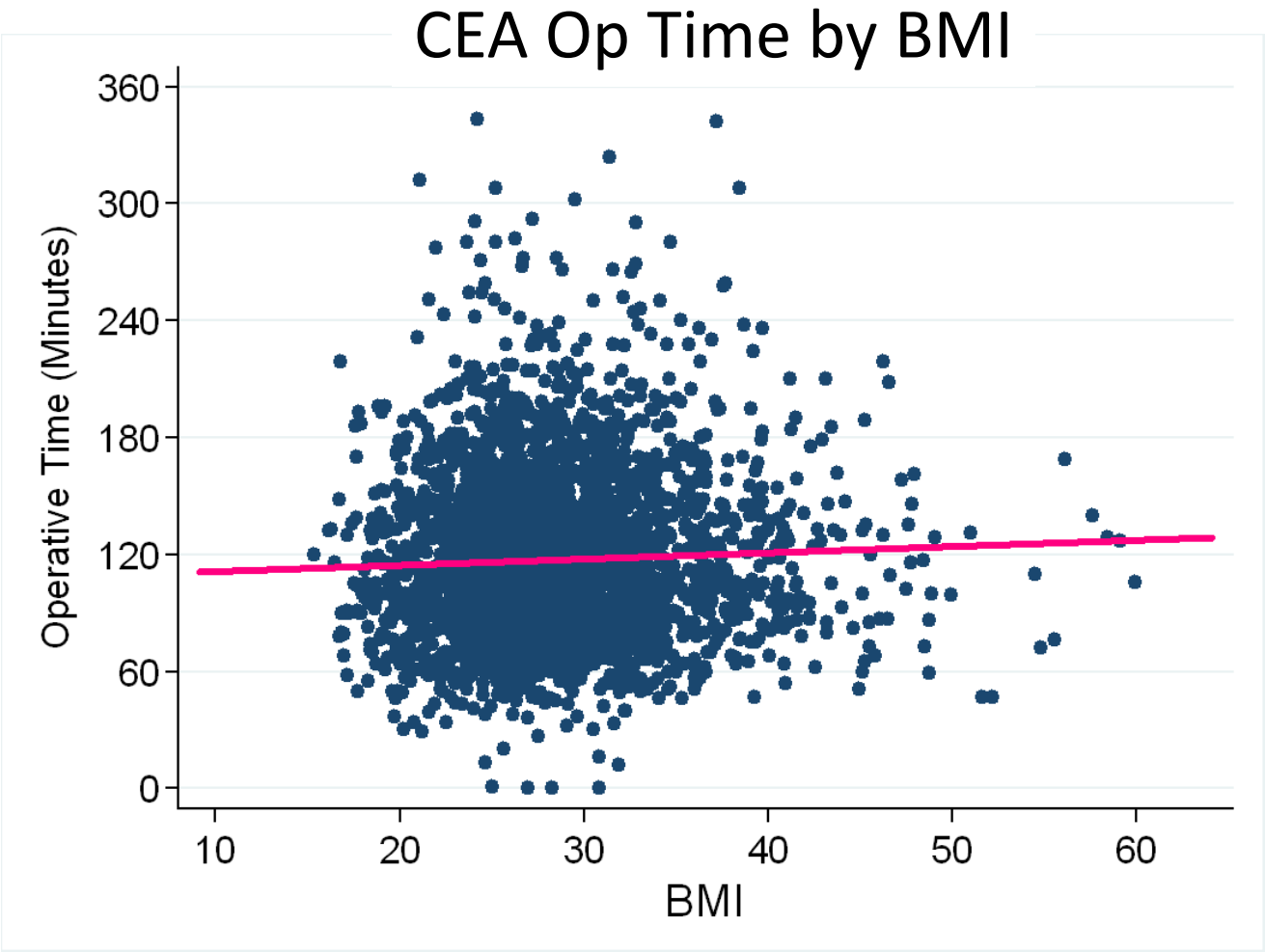
Pearson's Correlation

Outcome Variable:

- Continuous

Predictor Variable(s):

- 1 Continuous IV
- Normally distributed



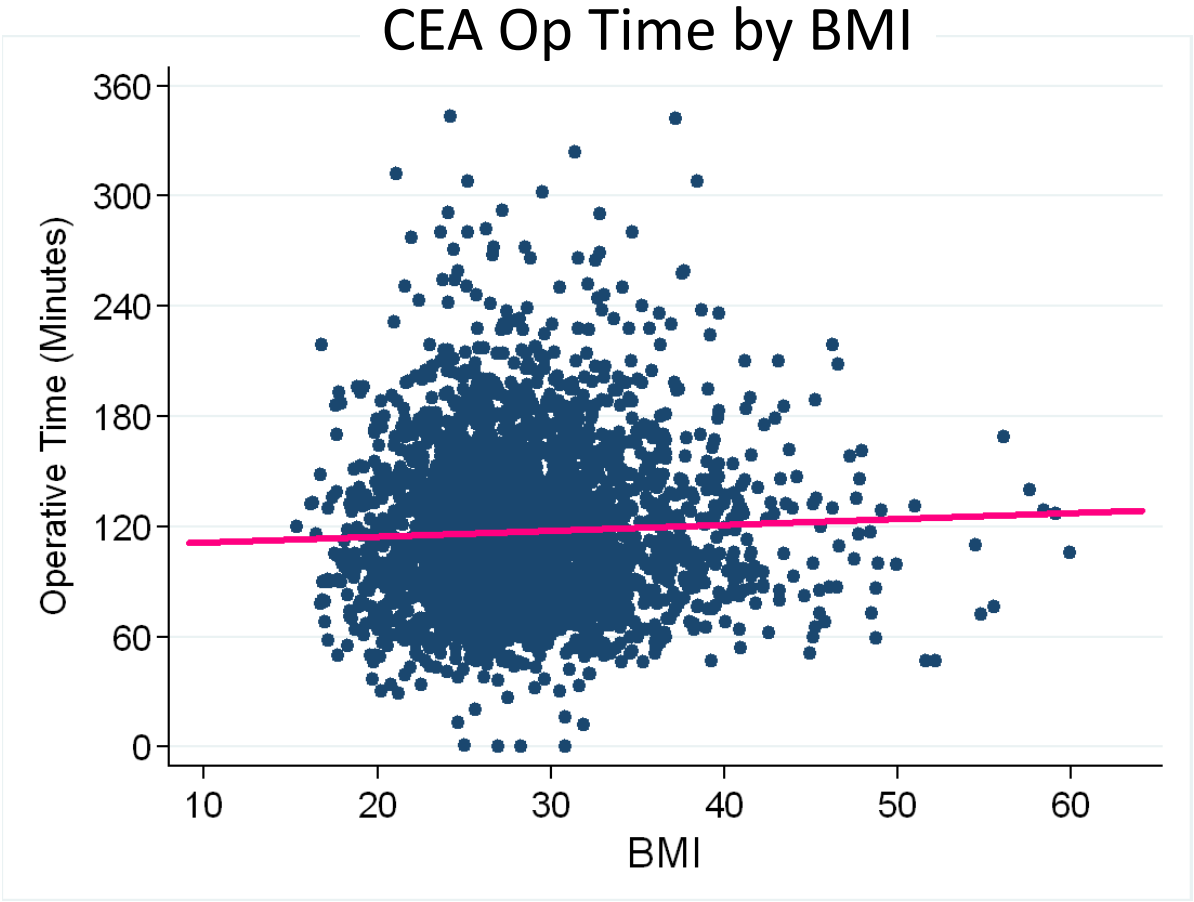
Linear Regression

Outcome Variable:

- Continuous

Predictor Variable(s):

- Categorical OR continuous
- Number of IVs
  - 1 = simple linear regression
  - 2+ = multiple linear regression



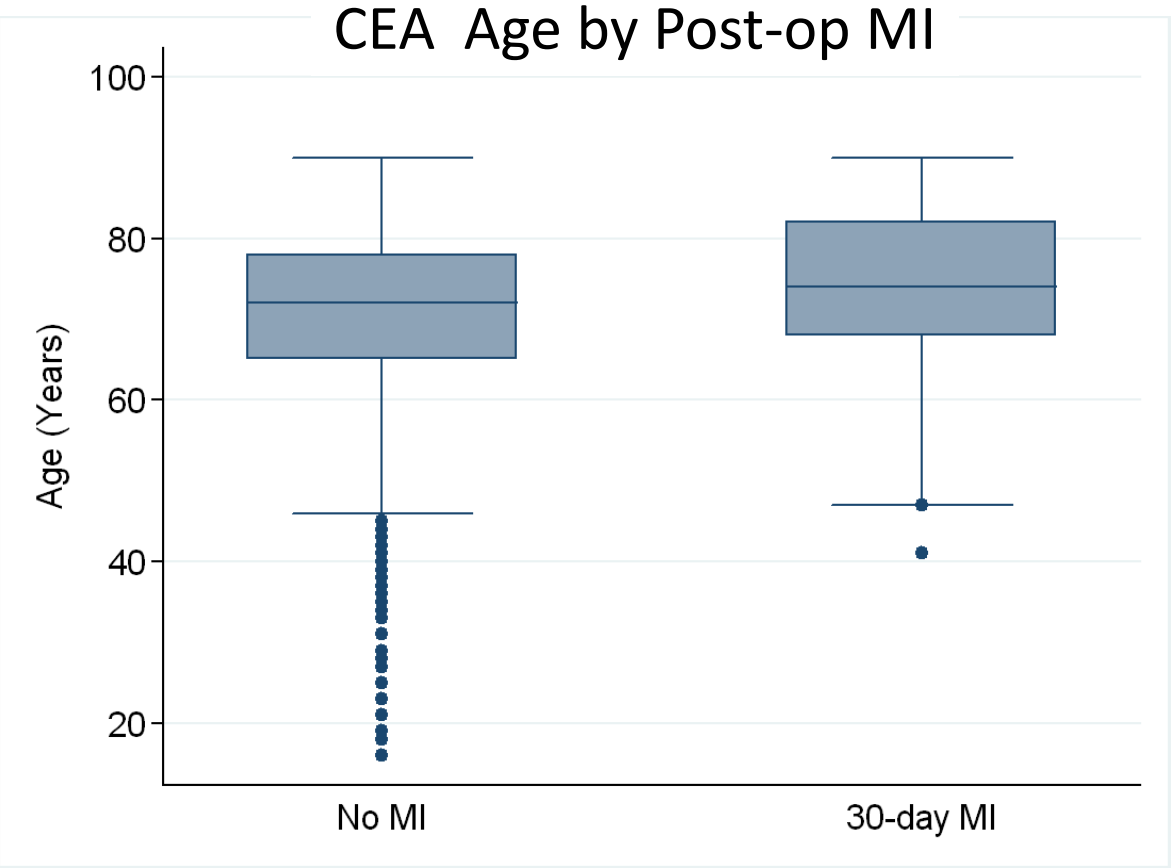
Logistic Regression

Outcome Variable:

- Dichotomous (binary)

Predictor Variable(s):

- Categorical OR continuous
- Number of IVs
  - 1 = simple logistic regression
  - 2+ = multiple logistic regression



P-values

$$n_i = 2\left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{Effect\ Size}\right)^2$$

		Null hypothesis is	
		TRUE	FALSE
The null hypothesis was	rejected ($P < \alpha$)	**Type I error**, false positive probability = α ✗	true positive probability = $1 - \beta$ (**power of the test**) ✓
	not rejected ($P \geq \alpha$)	true negative probability = $1 - \alpha$ ✓	**Type II error**, false negative probability = β ✗

References

1. Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. International journal of epidemiology. 2015 Jun 1;44(3):1051-67.
2. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. (2007). *Designing Clinical Research*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins.
3. Farrokhyar F, Reddy D, Poolman RW, Bhandari M. Practical Tips for Surgical Research: Why perform a priori sample size calculation?. *Canadian Journal of Surgery*. 2013 Jun;56(3):207.
4. Pye V, Taylor N, Clay-Williams R, Braithwaite J. When is enough, enough? Understanding and solving your sample size problems in health services research. BMC research notes. 2016 Dec;9(1):1-7.
5. Rosenfeld R. Chapter 2: Interpreting medical data. In: Flint, P.W., Haughey, B.H., Robbins, K.T., Thomas, J.R., Niparko, J.K., Lund, V.J. and Lesperance, M.M., 2014. *Cummings Otolaryngology-Head and Neck Surgery E-Book.* Elsevier Health Sciences.
6. https://stats.idre.ucla.edu/stata/whatstat/what-statistical-analysis-should-i-usestatistical-analyses-using-stata/
7. Rothman K, Greenland S, & Lash TL. (2008). Modern Epidemiology, 3rd Edition. Lippincott Williams & Wilkins.
8. Gunasekara FI, Carter K, & Blakely T. Glossary for econometrics and epidemiology. J Epidemiol Community Health 2008;62;858-861
9. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999:37-48.
10. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. https://www.equator-network.org/reporting-guidelines/strobe/

S-SPIRE