Overview

1. Overview of Research Design
2. Pros & Cons of Commonly Used Study Designs
3. Measurement Error and Bias Considerations

Focus will be on human subjects research and quantitative designs.
Poll 1

- What is your background? Check all that apply
  - Clinical
  - Biostatistics
  - Epidemiology
  - Economics
  - Data Science
  - Other Mathematics or Science Background
  - Other non-Mathematics or non-Science Background
Poll 2

- How many years have you been working in research?
  - <2 years
  - 2-5 years
  - 5-10 years
  - >10 years
What is Research Design?

- **Framework** or **strategy** to conduct research

- Study Methods
What is Research Design?

- Equator network: Enhancing the Quality and Transparency Of health Research (https://www.equator-network.org/)
Consolidated Health Economic Evaluation Reporting Standards (CHEERS)

- Target Population
- Setting and Location
- Study Perspective
- Comparators
- Time Horizon
- Discount Rate
- Choice of Health Outcomes
- Measurement of Effectiveness
- Preference Based Outcomes
- Estimating Resources and Costs
- Current, Price Date, and Conversion
- Choice of Model
- Assumptions
- Analytic Methods

https://www.equator-network.org/reporting-guidelines/cheers/
Consolidated Health Economic Evaluation Reporting Standards (CHEERS)

- Study Design
  - Target Population
  - Setting and Location
  - Study Perspective
  - Comparators
  - Time Horizon
  - Discount Rate
  - Choice of Health Outcomes
  - Measurement of Effectiveness

- Analyses
  - Preference Based Outcomes
  - Estimating Resources and Costs
  - Current, Price Date, and Conversion
  - Choice of Model
  - Assumptions
  - Analytic Methods

https://www.equator-network.org/reporting-guidelines/cheers/
* Note that Mixed Methods Study Designs incorporate both quantitative and qualitative research designs to answer their research question
Quantitative vs. Qualitative

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal, <strong>objective</strong>, systematic process for</td>
<td>Systematic <strong>subjective</strong> approach used to</td>
</tr>
<tr>
<td>obtaining information about the world</td>
<td>describe life experiences and give them meaning</td>
</tr>
<tr>
<td><strong>Test</strong> relationships and describes or <strong>examine</strong></td>
<td>Gain insight, <strong>discover</strong> frameworks, or <strong>explore</strong> a</td>
</tr>
<tr>
<td>causal associations</td>
<td>particular phenomenon</td>
</tr>
<tr>
<td><strong>Tests</strong> theory</td>
<td><strong>Develops</strong> theory</td>
</tr>
</tbody>
</table>

* Note that Mixed Methods Study Designs incorporate both quantitative and qualitative research designs to answer their research question.
Experimental vs. Observational

- **Experimental / Interventional**
  - Higher quality evidence $\rightarrow$ Better validity
  - 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.”

- **Non-experimental / Observational**
  - Came about due to ethical and cost restrictions of experimental studies
    - i.e. It is unethical to force some patients to smoke and others not to smoke
  - Investigator does not control the exposure (i.e. Subjects self-select into group)
Observational Research Design
Observational Research Design

- Unlike experimental designs ....
  - The investigator **does not assign exposure status**
  - Rely heavily 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**
# Analytic vs. Descriptive

<table>
<thead>
<tr>
<th>Analytic</th>
<th>Descriptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test hypotheses</td>
<td>Generate hypotheses</td>
</tr>
<tr>
<td>Quantify the direction and magnitude of associations.</td>
<td>Identifies and describes patterns by place, time, and/or person in a population</td>
</tr>
<tr>
<td></td>
<td><strong>Lacks a comparison group!</strong></td>
</tr>
</tbody>
</table>
Quantitative Research Designs

Observational

- Analytic / Hypothesis Testing
  - Cohort Study
  - Case-Control Study

- Descriptive / Exploratory
  - Ecologic Study
  - Cross-Sectional
  - Case Series

Experimental

- Laboratory
  - Basic Science
    - Pre-Clinical Trials
    - Cluster Randomized Trials

- Human Subjects
  - Clinical Trials
  - Field 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
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 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
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 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 processes
  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
* Note that Mixed Methods Study Designs incorporate both quantitative and qualitative research designs to answer their research question.
Guess the Study Design

- I want to know if aspirin is associated with postoperative bleeding. I ask patients on the day of surgery if they took an aspirin that morning or the day before. Later, I query the medical records for postoperative bleeding events in those patients.

- What type of study is this?
Guess the Study Design

- I want to know if aspirin reduces your risk of becoming infected with SARS-CoV-2. I send out a survey that asks about daily aspirin use and also asks about history of SARS-CoV-2 infection.

- What type of study is this?
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

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

Diagram:
- Random Error vs. Systematic Error
- True value
- 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 Design

- **Systematic Error** (Accuracy / Validity)
  - Closeness of a measured value to the truth
  - The degree to which an method/study actually measures what it is supposed to measure
Systematic Error

- 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
  - Endogeneity
Selection Bias

- Method of **participant selection** that 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)**, leading to an error in the conclusion of the association.

- **Differential** misclassification occurs when the level of misclassification differs between the two groups.
- **Non-differential** misclassification occurs when the level of misclassification does not differ between the two groups.
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 & Endogeneity

- Not the same

- **Endogeneity** occurs when a variable in a multiple regression model is **correlated with the error term**

- May be due to:
  - An omitted variable/residual confounding
  - Measurement error of collected variables
  - Simultaneity
    - X causes Y but Y also causes X
Confounding & Endogeneity

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

- Visual representation of causal assumptions of your research question
  - A conceptual framework unique to your research question
- Illustrate sources of bias

- **Directed**: Factors are connected with arrows, the arrows represent the direction of the causal relationship
- **Acyclic**: No directed path can form a closed loop, a factor cannot cause itself
Directed Acyclic Graphs (DAGs)

**DAG**

```
E -> C -> D
```

**Not a DAG**

```
E -> C -> E
```
Directed Acyclic Graphs

- **Directed acyclic graphs (DAGs)** can help to identify confounding and endogeneity during the study design phase.
References & Resources


Thank you!

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

@lagrahamepi

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