

# Data Collection & Research the Process

- Some common data collection procedures & issues
- Choices & Combinations of research attributes
- Research Loop and its applications
- Research Process and what portions of the process give us what aspects of validity
- Study attributes that do and do not influence the causal interpretability of the results.

## Data collection methods

All data are collected using one of two major methods...

### Behavioral Observation Data

- When data are obtained by human recording and/or coding of participants' behaviors

### Self-Report Data (often called Survey Research)

- When data are obtained from participant's written or spoken responses to researcher's questions

Not all written or spoken responses are "self-report" – only those in response to questions from the researcher!

- Coding a response of the participant to a question asked by the researcher would give self-report data
- Coding the response of one participant to the same question asked by another participant would give observational data.

## Behavioral Observation Data Collection

It is useful to discriminate among different types of observation ...

### Naturalistic Observation

- Participants don't know that they are being observed
  - requires "camouflage" or "distance"
  - researchers can be VERY creative & committed !!!!

### Participant Observation (which has two types)

- Participants know "someone" is there – researcher is a participant in the situation
  - Undisguised
    - the "someone" is an observer who is in plain view
    - Maybe the participant knows they're collecting data...
  - Disguised
    - the observer looks like "someone who belongs there"

## Naturalistic Observation

### Advantages & Possibilities

- Probably offers the best external validity
- Participants don't know they are being observed, and so, "act naturally"
- Experimental or nonexperimental designs can be used
  - RA and Manip can require creativity – but are possible!

### Disadvantages & Challenges

- Limited to studying behavior
- Important ethical point → Limited to the observation of "public behaviors"
- Requires reliable/accurate coding to produce useful data

## Undisguised Participant Observation

### Advantages & Possibilities

- Behavior can be very "natural" after participants are "used to the observer"
  - Habituation -- observer shows up and waits until participant "gets used to" observer and then begins data collection
  - Desensitization -- observer slowly approaches so participant can gradually "get used to" them
- Experimental or nonexperimental designs can be used
  - RA and Manip can require creativity – but are possible!

### Disadvantages & Challenges

- Limited to studying behavior
- Important ethical point → Limited to the observation of "public behaviors"
- Some behaviors/participants don't habituate/desensitize
- Requires reliable/accurate coding to produce useful data

## Disguised Participant Observation

### Advantages & Possibilities

- The participant doesn't know they are being observed, and so, they should "act naturally"
  - access to "less public" behavior "among peers"
- Disguised researcher could change behavior of participants
- Experimental or nonexperimental designs can be used
  - RA and Manip can require creativity – but are possible!

### Disadvantages & Challenges

- Limited to studying behavior
- Important ethical point → Researcher is now "intruding" and must be careful about privacy/risk issues!
- Participation can cause loss of "objectivity" or be dangerous
- Requires reliable/accurate coding to produce useful data

How we collect the observational data ...

Although “written narratives” and “field notes” were the standards of observational data collection for decades, we now often use various kinds of instrumentation to record observational data, including...

- Audio recordings
- Pictures and video recordings
- Non-verbal behaviors – reaction time (RT), eye movements
- Medical/physiological recordings – EEG, EKG, EMG, GSR, MRI & PET scans, hormone levels

Instrumentation increases the availability & accuracy of the data...

- Audio/video recordings are more accurate than written records
- Computerized RT are more accurate than stopwatches
- Physiological recordings of “unseen” body processes



## Self-Report Data Collection

We need to discriminate among various self-report data collection procedures...

- Mail Questionnaire
- Computerized Questionnaire
- Group-administered Questionnaire
- Personal Interview
- Phone Interview
- Mail Questionnaire
- Group Interview (focus group)
- Journal/Diary

In each of these participants respond to a series of questions prepared by the researcher.

## Self-Report Data Collection

### Advantages & Possibilities

- can get data on “non-observables” or “mental behavior”
  - thoughts, opinions, attitudes, intentions, plans, etc.
- Experimental or nonexperimental designs can be used
  - RA and Manip are readily possible!

### Disadvantages & Challenges

- Dependent upon accuracy and honesty of the participant
- Ways to improve response honesty
  - Promises of anonymity and/or confidentiality
  - Rapport between researcher and participant
- Ways to improve response accuracy
  - Careful construction of questions and their sequence



One last thing about data collection (for both observational and self-report)

It is useful to discriminate between two kinds of data sources...

### Primary Data Sources

- Sampling, questions and data collection completed for the purpose of this specific research
- Researcher has maximal control of planning and completion of the study – substantial time and costs

### Archival Data Sources (AKA secondary analysis)

- Sampling, questions and data collection completed for some previous research, or as standard practice
- Data later made available to the researcher for secondary analysis
- Often quicker and less expensive, but not always the data you would have collected if you had greater control.

Is each primary or archival data?

- Collect data to compare the outcome of those patients I've treated using Behavior vs. using Cognitive interventions
- Go through past patient records to compare Behavior vs. Cognitive interventions
- Purchase copies of sales receipts from a store to explore shopping patterns
- Ask shoppers what they bought to explore shopping patterns
- Using the data from some else's research to conduct a pilot study for your own research
- Using a database available from the WWW to perform your own research analyses
- Collecting new survey data using the WWW

### Data collection Settings

Same thing we discussed as an element of external validity...

Any time we collect data, we have to collect it somewhere – there are three general categories of settings

#### Field

- Usually defined as “where the participants naturally behave”
- Helps external validity, but can make control (internal validity) more difficult (RA and Manip possible with some creativity)

#### Laboratory

- Helps with control (internal validity) but can make external validity more difficult (remember ecological validity?)

#### Structured Setting

- A “natural appearing” setting that promotes “natural behavior” while increasing opportunity for “control”
- An attempt to blend the best attributes of Field and Laboratory settings !!!



Data collection Settings identify each as laboratory, field or structured...

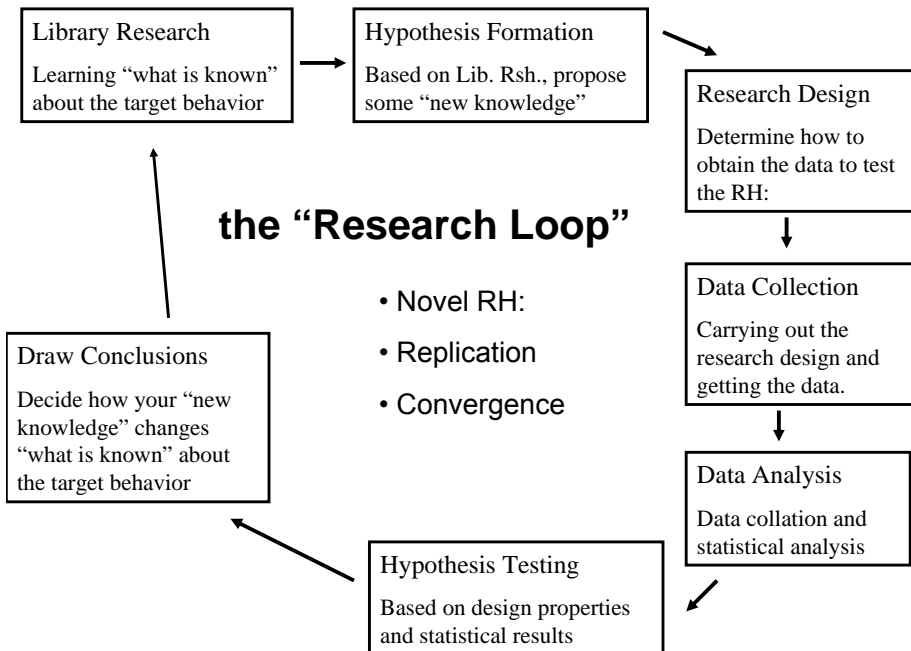
- Study of turtle food preference conducted in Salt Creek.
- Study of turtle food preference conducted with turtles in 10 gallon tanks.
- Study of turtle food preference conducted in a 13,000 gallon “cement pond” with natural plants, soil, rocks, etc.
- Study of jury decision making conducted in 74 Burnett, having participants read a trial transcript.
- Study of jury decision making with mock juries conducted in the mock trial room at the Law College.
- Study of jury decision making conducted with real jurors at the Court Building.



Now might be a good time to review the decisions made when conducting any research project.

- Research hypothesis (associative or causal)
- Research design (true vs. nonexp & BG vs. WG)
- Sampling procedure (complete vs. purposive, researcher vs. self selected & simple vs. stratified)
- Setting (laboratory vs. structured vs. field)
- Task (while there are thousands of possible tasks, they generally divide into “natural, familiar tasks” and “contrived, novel & artificial tasks”)
- Stimuli (the thousands of possible stimuli are generally divided into “natural, familiar stimuli” and “contrived, novel & artificial stimuli”)
- Data collection (observational vs. self-report)

Considering these choices, any one study could be run 1536 different ways !!! ( $2 \times 4 \times 8 \times 3 \times 2 \times 2 \times 2 = 1536$ )



## Applying the Research Loop

The “research loop” is applied over and over, in three ways...

- Initial test of a RH:
  - The first test of a research hypothesis -- using the “best” design you can
- Replication Study
  - being sure your conclusions about a particular RH: are correct by repeating exactly the same research design
  - the main purpose of replication is to acquire confidence in our methods, data and resulting conclusions
- Convergence (Converging Operations) Study
  - testing “variations” of the RH: using “variations” of the research design (varying population, setting, task, measures and sometimes the data analyses)
  - the main purpose of convergence is to test the limits of the “generalizability” of our results
    - what design/analysis changes lead to different results?



Research process ...

Statement of RH: →

- tells associative vs. causal intent
- tells variables involved
- tells target population

Participant Selection (Sampling) →

- external → population validity
  - Complete vs. Purposive
  - Researcher- vs. Self-selection
  - Simple vs. Stratified

\*Participant Assignment (necessary only for Causal RH:)

↙

internal validity → initial equivalence (subj vars)

- random assignment of individuals by the researcher ☺
- random assignment of groups ☺
- random assignment – arbitrary conditions by researcher ☺
- random assignment – arbitrary conditions by “administrator” ☺
- self assignment ☺
- non-assignment (e.g., natural or pre-existing groups) ☺

\*Manipulation of IV (necessary only for Causal RH:)

↙ ↘

internal validity → ongoing equivalence (procedural vars)

- by researcher vs. Natural Groups design

↘

external → setting & task/stimulus validity

↘

Measurement validity -- does IV manip represent “causal variable”

Data Collection →

internal validity → ongoing equivalence - procedural variables

↘ ↘

external → setting & task/stimulus validity

↘

Measurement validity (do variables represent behaviors under study)

Data Analysis →

statistical conclusion validity



Study attributes that do and don't directly influence the causal interpretability of the results & a couple that make it harder

Attributes that DON'T directly influence causal interpretability...

- Participant Selection (population part of external validity)
- Setting (setting part of external validity)
- Data collection (measurement validity)
- Statistical model (statistical conclusion validity)

Attributes that DO directly influence causal interpretability...

- Participant Assignment (initial eq. part of internal validity)
- Manipulation of the IV (ongoing eq. part of internal validity)

Attributes that make it harder to causally interpret the results ...

- Field experiments (harder to maintain ongoing equivalence)
- Longer studies (harder to maintain ongoing equivalence)



Something else to remember...

There are certain combinations of data collection, design, setting and/or statistics that co-occur often enough that they have been given names.

- But, the names don't always accurately convey the causal interpretability of the resulting data.
- Remember, the causal interpretability of the results is determined by the design & the presence/absence of confounds
- You have to check the type of design that was used (experimental or non-experimental) and whether or not you can identify any confounds !!!

Some of those combinations ...

Research "Types" named for the data collection used

- "Survey research"
- "Observational research" **Usually implies a non-experiment conducted in the field**
- "Trace research"

**Remember: Any data collection method can be used to obtain causally interpretable data it is part of a properly conducted true experiment.**

Research "Types" named for the research setting used

- "Field research" **usually implies a non-experiment**
- "Laboratory research" **usually implies an experiment**
- "Trace research"













**Remember: Any research setting can be used to obtain causally interpretable data it is part of a properly conducted true experiment.**

Research "Type" seemingly named for the statistical analysis used

- "Correlational research" **usually implied a non-experiment**

**Remember: Any data collection method can be used to obtain causally interpretable data it is part of a properly conducted true experiment.**

So there's a lot of possible combinations of data collection, setting and design (even if we simplify things as below)...

Data collection	Experimental Design w/o confounds Setting		Non-experimental Design & Exp Design w/ confounds Setting	
	Laboratory	Field	Laboratory	Field
Observation				
Self-report				
Trace				

All three attributes are important when describing the study!

But only the design type and confound control actually determine the causal interpretability of the results!!!!  