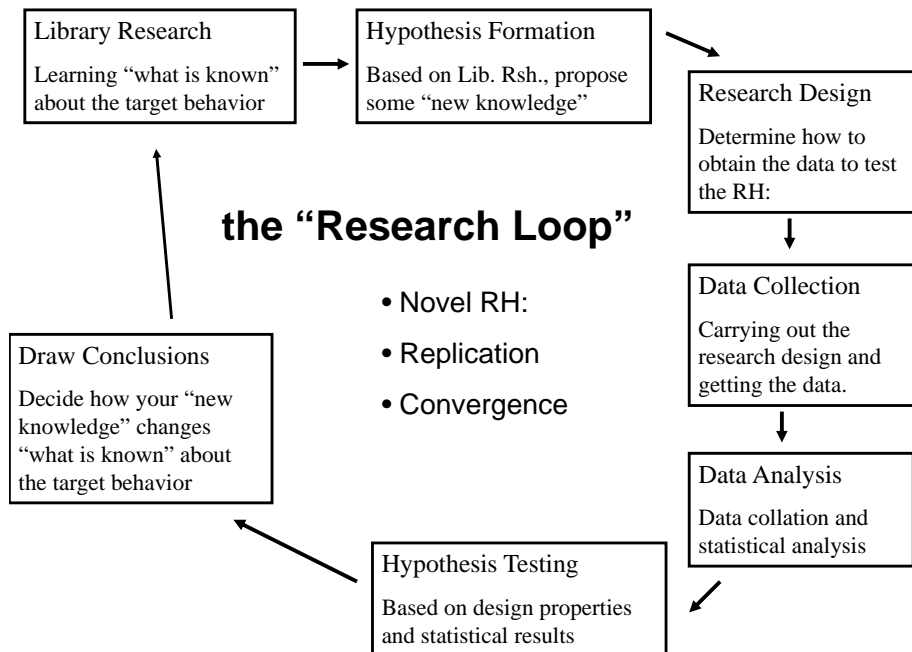# Research Process & Data Collection

- Choices & Combinations of research attributes
- Research Loop and its Applications
- Research Process and what portions of the process give us what aspects of validity
- Observational, Self-report & Trace Data Collection
- Research Settings
- Primary & Archival data sources
- Study attributes that do and do not influence the causal interpretability of the results.
- Initial equivalence for Within-Groups designs
- Revisiting External and Internal Validity – Attrition
- Troubles with researchers & participants

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")
- Data collection (observational vs. self-report)

Considering these choices, any one study could be run 1536 different ways !!! (2x4x8x3x2x2x2 = 1536)

---

the "Research Loop"

- Novel RH:
- Replication
- Convergence

Library Research

Learning "what is known" about the target behavior

Hypothesis Formation

Based on Lib. Rsh., propose some "new knowledge"

Research Design

Determine how to obtain the data to test the RH:

Data Collection

Carrying out the research design and getting the data.

Data Analysis

Data collation and statistical analysis

Hypothesis Testing

Based on design properties and statistical results

Draw Conclusions

Decide how your "new knowledge" changes "what is known" about the target behavior

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?

Types of Validity

Measurement Validity
  - do our variables/data accurately represent the characteristics & behaviors we intend to study ?

External Validity
  - to what extent can our results can be accurately generalized to other participants, situations, activities, and times ?

Internal Validity
  - is it correct to give a causal interpretation to the relationship we found between the variables/behaviors ?

Statistical Conclusion Validity
  - have we reached the correct conclusion about whether or not there is a relationship between the variables/behaviors we are studying ?

External Validity

Do the who, where, what & when of our study represent what we intended want to study?

Measurement Validity

Do the measures/data of our study represent the characteristics & behaviors we intended to study?

Internal Validity

Are there confounds or 3rd variables that interfere with the characteristic & behavior relationships we intend to study?

Statistical Conclusion Validity

Do our results represent the relationships between characteristics and behaviors that we intended to study?
- did we get non-representative results "by chance" ?
- did we get non-representative results because of external, measurement or internal validity flaws in our study?

Research process ...

Statement of RH: ⟶
- tells associative vs. causal intent
- tells variables involved
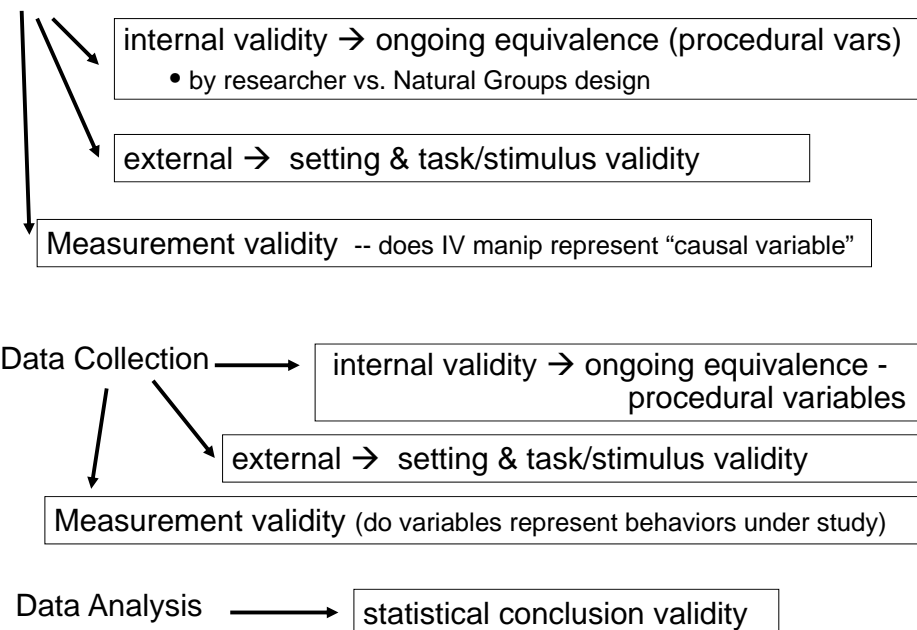- tells target population

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 constraints by researcher ☹
- random assignment – arbitrary constraints 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 validty)
- 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)

All data are collected using one of three major methods…

Behavioral Observation Data
- Studies actual behavior of participants
- Can require elaborate data collection & coding techniques
- Quality of data can depend upon secrecy (naturalistic, disguised participant) or rapport (habituation or desensitization)

Self-Report Data
- Allows us to learn about non-public "behavior" – thoughts, feelings, intentions, personality, etc.
- Added structure/completeness of prepared set of ?s
- Participation & data quality/honesty dependent upon rapport

Trace Data
- Limited to studying behaviors that do leave a "trace"
- Least susceptible to participant dishonesty
- Can require elaborate data collection & coding techniques

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"

Observational data collection can be part of Experiments (w/ RA & IV manip) or of Non-experiments !!!!!

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
- 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 can be part of Experiments (w/ RA & IV manip) or of Non-experiments !!!!!

Trace data are data collected from the "marks & remains left behind" by the behavior we are trying to measure.
There are two major types of trace data…
Accretion – when behavior "adds something" to the environment
• trash, noseprints, graffiti
Deletion – when behaviors "wears away" the environment
• wear of steps or walkways, "shiny places"

3. **Garbageology** – the scientific study of society based on what it discards -- its garbage !!!

• Researchers looking at family eating habits collected data from several thousand families about eating take-out food
• Self-reports were that people ate take-out food about 1.3 times per week
• These data seemed "at odds" with economic data obtained from fast food restaurants, suggesting 2 times per week
• The Solution – they dug through the trash of several hundred families' garbage cans before pick-up for 3 weeks – estimated about 2.8 take-out meals eaten each week

---

Data collection Methods – identify each as Observation, Self-report or Trace…

Gave 3rd grade students an arithmetic test and…

• Watched them to see if they used counted on their fingers — Observation

• Use the test % score as my DV — Self-report

• Counted the number of erasures they made — Trace

Did a interview with each lab manager candidate in my office, and …

• Had a series of 6 questions each was asked — Self-report

• I counted the copies of the "What we do in this lab" flyer that was available in the waiting room before and after each candidate waited there — Trace

• I recorded whether or not they took notes during the interview — Observation

---

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

---

Data Sources …
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 that are 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 — primary
- Go through past patient records to compare Behavior vs. Cognitive interventions — archival
- Purchase copies of sales receipts from a store to explore shopping patterns — archival
- Ask shoppers what they bought to explore shopping patterns — primary
- Using the data from some else's research to conduct a pilot study for your own research — archival
- Using a database available from the web to perform your own research analyses — archival
- Collecting new survey data using the web — primary

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 lot of possible combinations of data collection, setting and design (even if we simplify things as below)…

	Experimental Design w/o confounds Setting		Non-experimental Design & Exp Design w/ confounds Setting	
Data collection	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!!!!!   ☺ ☹

Attrition – also known as drop-out, data loss, response refusal, & experimental mortality – we're gonna have 2 kinds

Attrition endangers the population portion of externa validity
- After carefully obtaining a representative sample, some people drop out
- that attrition could render the sample non-representative
- E.g., the study is "harder" so participant drop out – producing a sample that doesn't represent the motivation of the pop

Differential Attrition endangers initial equivalence part of internal validity (of subject variables)

• random assignment is intended to produce initial equivalence of subject variables – so that the groups (IV conditions) have equivalent means on all subject variables

• e.g., If one condition is "harder" and so more participants drop out of that condition, producing a "motivation" difference between the two conditions.

So, "attrition" works much like "self assignment" to trash initial equivalence

Both involve a non-random determination of who provides data for what condition of the study!

Imagine a study that involves a "standard treatment" and an "experimental treatment"…

• random assignment would be used to ensure that the participants in the two groups are equivalent

• self-assignment is likely to produce non-equivalence (different "kinds" of folks likely to elect the different treatments)

• attrition (i.e., rejecting the randomly assigned condition) is similarly likely to produce non-equivalence (different "kinds" of folks likely to remain in the different treatments)

"Counterbalancing failure" causes Initial Equivalence problems!

Huh?? "counterbalancing failure" happens "during the procedure" – so why isn't it ongoing equivalence!

Initial equivalence – before manipulation of the IV, participants in the conditions are equivalent (on average) on all subject variables.

Ongoing equivalence – during manipulation of the IV, completion of the task, and measurement of the DV, participants in the conditions are equivalent (on average) on all procedural variables (except for the IV).

Two things to notice:
- Initial equivalence is about subject variables
- Ongoing is not "during the study" but during manip of IV, task, measurement of DV – for a particular condition

Keep going…

In a WG design, we are counting on counterbalancing to ensure that the set of participants are "equivalent on average on all subject variables" just before they begin the manipulation of the Control condition, as they are just before they begin the manipulation of the Treatment condition.

If the participants are different on any subject variables before they begin manip of the Control condition than they are before manipulation of the Treatment condition – then those difference on those subject variables is an Initial Equivalence problem!

So, if we don't counter balance (or if counterbalancing fails) then the differences produced are subject variable differences – a problem of Initial Equivalence

Experimenter Expectancy Effects

A kind of "self-fulfilling prophesy" during which researchers unintentionally "produce the results they want".   Two kinds…

Modifying Participants' Behavior
- Subtle differences in treatment of participants in different conditions can change their behavior…
- Inadvertently conveying response expectancies/research hypotheses
- Difference in performance due to differential quality of instruction or friendliness of the interaction

Data Collection Bias (much like observer bias)
- Many types of observational and self-report data need to be "coded" or "interpreted" before they can be analyzed
- Subjectivity and error can creep into these interpretations – usually leading to data that are biased toward expectations

Data Collection Bias: Observer Bias & Interviewer Bias

Both of these are versions of "seeing what you want to see"

Observer Bias is the term commonly used when talking about observational data collection
- Both observational data collection and data coding need to be done objectively and accurately
- Automation & instrumentation help – so does using multiple observers/coders and looking for consistency

Interviewer Bias is the term commonly used when talking about self-report data collection
- How questions are asked by interviewers or the interviewers' reactions to answers can drive response bias
- More of a challenge with face-to-face interviews
- Computerized and paper-based procedures help limit this

Participant Expectancy Effects

A kind of "demand characteristic" during which participants modify their behavior to respond/conform to "how they should act".

Social Desirability
– When participants intentionally or unintentionally modify their behavior to match "how they are expected to behave"
– Well-known social psychological phenomenon that usually happens between individual's and their "peer group"
– Can also happen between researcher and participants

Acquiescence/Rejection Response
– If participant thinks they know the research hypothesis or know the behavior that is expected of them they can "try to play along" (acquiescence) or "try to mess things up" (rejection response)
– Particularly important during within-groups designs – if participants think study is "trying to change their behavior"

Participant Expectancy Effects: Reactivity & Response Bias

Both of these refer to getting "less than accurate" data from the participants

Reactivity is the term commonly used when talking about observational data collection
– the participant may behave "not naturally" if they know they are being observed or are part of a study
– Naturalistic & disguised participant observation methods are intended to avoid this
– Habituation and desensitization help when using undisguised participant observation

Response Bias is the term commonly used when talking about self-report data collection and describes a situation in which the participant responds how they think they "should"
– The response might be a reaction to cues the researcher provides
– Social Desirability is when participants describe their character, opinions or behavior as they think they "should" or to present a certain impression of themselves
– Protecting participants' anonymity and participant-researcher rapport are intended to increase the honesty of participant responses

Data collection biases & inaccuracies -- summary

Type of Data Collection

	Observational	Self-report
Researcher Expectancy	Observer Bias "inaccurate data recording/coding"	Interviewer Bias "coaching" or "inaccurate recording/coding"
Participant Expectancy	Reactivity "reacting" to being observed	Response Bias "dishonest" responding

Single & Double-blind Procedures

One way to limit or minimize the various biasing effects we've discussed is to limit the information everybody involved has

In Single Blind Procedures the participant doesn't know the hypotheses, the other conditions in the study, and ideally, the particular condition they are in (i.e., we don't tell how the task or manipulation is designed to change their behavior)

In Double-blind Procedures neither the participant nor the data collector/data coder knows the hypotheses or other information that could bias the interaction/reporting/coding of the researcher or the responses of the participants

Sometimes this simply can't be done (especially the researcher-blind part) because of the nature of the variables or the hypotheses involved (e.g., hard to hide the gender of a participant from the researcher who is coding the video tape)

Let's say we are studying weightlifting…

We want to see if, for some particular movement, there is a difference in the number of reps a person can do with a "wide grip" vs. a "narrow grip". We decide to use a WG design.

We have every body do the wide grip trial first and then immediately do the narrow grip trial. See a problem?

What is the problem? The variable of concern is "fatigue" – a subject variable! Participants are more fatigued just before the second narrow grip condition than just before the first wide grip condition!

Doing the study with this improper counterbalancing leads to a "fatigue diference" between the conditions – a subject variable problem → an initial equivalence problem! ;)