Internal Validity

- Measured & Manipulated Variables & Constants
- Causes, Effects, Controls & Confounds
- Components of Internal Validity
- Interrelationships between Internal Validity & External Validity
- "Creating" initial equivalence
- "Creating" & "Maintaining" ongoing equivalence

Some Review…

1. 3 kinds of knowledge?
2. 3 kinds of research hypotheses?
3. which two are dependent on which one?

1. 4 kinds of validity?
2. which three are dependent on which one?
3. which is often most important to "basic researchers" especially "lab researchers"?
4. which is often most important to "applied researchers" and "practitioners"?
5. which one is dependent on which three?
6. what are the four things that influence that one?

1. what are the four elements of external validity?
2. which of these are involved in what WE will call sampling?

Some Review (answers – no teasing here!) …

1. Descriptive - predictive - understanding
2. Attributive - associative - causal
3. Associative and causal dependent on attributive

1. Measurement - external - internal - statistical conclusion
2. External, internal & statistical conclusion dep on measurement
3. Internal (confounds change results)
4. External (what have we learned & how far can push it)
5. statistical conclusion dep on measurement, external & internal
6. measurement, external, internal & chance

1. Population, setting task/stim & temporal social
2. ALL FOUR → representation & inference

Before we can discuss Internal Validity, we have to discuss different types of variables and review causal RH:s and the evidence needed to support them…

Every behavior/charcteristic used in a research study is either a ...
Constant -- all the participants in the study have the same value on that behavior/characteristic

or a ...
Variable -- when at least some of the participants in the study have different values on that behavior/characteristic

and every behavior/characteristic is either …

Measured -- the value of that behavior/characteristic is obtained by observation or self-report of the participant (often called "subject constant/variable")

or it is …
Manipulated -- value of that behavior/characteristic is controlled, delivered, determined, etc., by the researcher (often called "procedural constant/variable")

So, every behavior/characteristic in any study is one of four types…

	constant	variable
measured	measured (subject) constant	measured (subject) variable
manipulated	manipulated (procedural) constant	Manipulated (procedural) variable

Identify each of the following (as one of the four above, duh!)…

- Participants reported practicing between 3 and 10 times
- All participants were given the same set of words to memorize
- Each participant reported they were a Psyc major
- Each participant was given either the "homicide" or the "self-defense" vignette to read

quick review of Causal Research Hypotheses:

From before...
- Causal RH: -- differences in the amount or kind of one behavior causes/produces/creates/changes/etc. differences in amount or kind of the other behavior

Using our newly acquired language…
- Causal RH: -- the value of the variable manipulated by the researcher causes the value of the variable measured from the participant

In a causal research hypothesis…

- the manipulated variable = the "causal variable"
- the measured variable = the "effect variable," the "response variable" or the "outcome variable"

Be sure to notice -- The "**causal variable**" absolutely must be **manipulated** in the study !!!!

Indeitify the manipulated/causal & underline measured/effect variable in @

- Practice improves performance.
- Treatment decreases depression.
- Schizophrenic symptomology is decreased by pharmacological intervention.
- Reading speed is improved by larger print size.

Try this one (you'll have to "figure out" what the manipulated variable is from the description of the different "conditions")

Completing the group therapy will lead to lower social anxiety scores than will completing the individual therapy.

manipulated variable --> Type of Therapy (group vs. individual)

measured variable --> Anxiety Score

Review of evidence required to support a causal research hypothesis …

Evidence needed to support a causal hypothesis...

- temporal precedence ("cause proceeds effect")
- demonstrate a statistical relationship
- elimination of alternative explanations (no other viable causes/explanations of the effect)

This identifies four different "roles" variables/constants might play in a study ….

Causal variable -- manipulated by the researcher -- the variable to which we want to attribute the effect

Effect variable -- measured from each participant after manipulation of causal variable by the researcher

Confounding variable(s) -- any variable (other than the one manipulated by the researcher) that might have caused the effect -- an alternative causal variable or explanation of the effect

Controls -- any constant/variable that can't have caused the effect because it is "equivalent" across conditions

One of those things about "how we use words oddly"

We often talk about two kinds of variables – like this…

"Variables" – behaviors or characteristics of interest in the study

Variables – behaviors or characteristics for which different participants have different values

Constants – behaviors or characteristics for which all participants have the same value

Control Constants vs. Control Variables

Control Constants

- any behavior or characteristic for which all participants have the same value
- "a constant can't be a confounding variable"

Control Variables

- any behavior or characteristic for which participants have different values, but for which the treatment or conditions are "balanced" or "equivalent" on that variable
- Examples
  - if ½ of the participants in each treatment/condition are Psyc majors and ½ Education majors, then major is a control variable (note – don't need a ½ - ½ split, only that the split is the same in each treatment/condition)
  - if the participants in each treatment/condition have the same average IQ, then IQ is a control variable

So, we have to be able to discriminate between these three things:
Constants vs. Control variables vs Confounding variables

So, we can tell these apart based on who is and isn't "different" !!!

Kind of thing…	Differences among individual people?	Differences among groups (on average)?
constant	no	no
Control variable	Yes (makes it a variable)	No (makes it a control)
Confounding variable	Yes (making it variable)	Yes (Making it a confound)

Control Constants, Control Variables & Confounds – some practice

80% of treatment group participants have prior experience with the task and 20% of the control group participants have prior task experience — confound

60% of treatment group participants have prior experience with the task and 60% of the control group participants have prior task experience — control variable

None of the participants in either group have prior task experience — control constants

All participants are 6 years old — control constants

The average age of the treatment group is 7 and the average age of the control group is 45. — confound

The average ate of the treatment group is 7.1 and the average age of the control group is 7.2, — control variable

So, to summarize ...

Before the study begins...		After the study is over ...
Causal Variable	→	Causal Variable
Effect Variable	→	Effect Variable
Potential Confounds	→	(Control) Constants
	→	Control Variables
	→	Confounding Variables

Let's try using these terms …

RH: Computerized spelling practice leads to better performance than does paper & pencil practice.

Twenty Spanish-English bilingual 4th grade students were given 10 words and practiced them 5 times each on the computer. Twenty Spanish-English bilingual speaking 2nd grade students were given the same 10 words and practiced them 3 times each using paper & pencil. When tested, the "computer practice" students did better than the "paper & pencil practice" students

What's the intended causal variable? Type of practice (comp.vs. pap&pen)

What's the intended effect variable? Test performance

Any control variables/constants & is each measured or manipulated?
- S-E speaking – meas. const
- same words -- manip. const

Any confounds & is each measured or manipulated ?
- grade -- measured
- # practices -- manipulated

So, can these results be used to support the causal RH: why or why not?

NO! We have temporal precedence, we have a statistical relationship, but we also have **confounds**, so we can't be sure what caused the effect

Here's another...

RH: Group therapy will lead to lower dep. scores than individual therapy

Five employed & five unemployed patients with no prior therapy completed a 24-session course of group therapy, meeting each time at the university psychiatric clinic. A different group of five employed & five unemployed patients, each of whom had previously received therapy for depression, completed a 10-session series of individual therapy, meeting at the same clinic. After the respective therapies, the group therapy patients had lower depression scores.

What's the intended causal variable? Type of therapy (grp vs. ind.)

What's the intended effect variable? Depression score

Any control variables/constants & is each measured or manipulated?
- Tx location -- manipulated const.
- Employment -- measured var.

Any confounds & is each measured or manipulated ?
- # sessions -- manipulated
- prior therapy -- measured

So, can these results be used to support the causal RH: why or why not?

NO! We have temporal precedence, we have a statistical relationship, but we also have **confounds**, so we can't be sure what caused the effect

Notice that the RH: determines what's a causal variable and a confound !

RH: More therapy sessions will lead to lower dep. scores.

Five employed & five unemployed patients with no prior therapy completed a 24-session course of group therapy, meeting each time at the university psychiatric clinic. A different group of five employed & five unemployed patients, each of whom had previously received therapy for depression, completed a 10-session series of individual therapy, meeting at the same clinic. After the respective therapies, the group therapy patients had lower depression scores.

What's the intended causal variable? # therapy sessions (24 vs. 10)

What's the intended effect variable? Depression score

Any control variables/constants & is each measured or manipulated?
- Tx location -- manipulated const.
- employment -- measured const.

Any confounds & is each measured or manipulated ?
- Type of Tx -- manipulated
- prior therapy -- measured

So, can these results be used to support the causal RH: why or why not?

NO! We have temporal precedence, we have a statistical relationship, but we also have **confounds**, so we can't be sure what caused the effect

Quick review … then on to Internal Validity...

"Kinds of behaviors/measures" -- need to be able to think simultaneously with two "systems"

First, any behavior/characteristic in a study is one of four kinds
- measured (subject) constant
- measured (subject) variable
- manipulated (procedural) constant
- manipulated (procedural) variable

Second, each behavior/characteristic has one of 4 "roles" in the study
- Causal variable
- Effect (response, outcome) variable
- Control variable/constant -- for causal interpretation, every behavior/characteristic not the causal or effect variable need to be "controlled"
- Confounding variable -- anything other than the causal variable that might be causing "the effect'

Components of Internal Validity

-- remember, Int. Val. Primarily applies when testing causal RH:
-- but "cleaner" studies of associative RH: are easier to interpret

Initial Equivalence

– Prior to manipulation of the causal variable, participants in the different conditions are the same (on the average) on all measured/subject variables

Ongoing Equivalence

– during manipulation of the causal variable, completion of the task, and measurement of the effect variable, participants in the different conditions are the same (on the average) on all manipulated/procedural variables

Practice with Types of Variables & Types of Equivalence

Tell the confounding variable, whether it is sub/msr or manip/proc and tell the type equivalence that is at "risk" ...

I'm concerned that before the treatment began, those in the Drug Treatment group were more depressed than were those in the Therapy Treatment group.

Depression:
- Subject/Measured Variable
- Initial Equivalence

Are you sure that there was no problem allowing those in the Drug Treatment group to attend an extra 5 sessions ? Those in the Therapy Treatment group didn't have the extra sessions.

sessions:
- Manip./Procedural Variable
- Ongoing Equivalence

More practice ...

Tell the confounding variable, whether it is sub/msr or manip/proc and tell the type equivalence that is at "risk" ...

To save time, only those who are familiar with computers were included in the Computer Training Condition, and everybody else was put in the Lecture Condition.

Familiarity:
- Subject Variable
- Initial Equivalence

Because of the class schedule, those in the Computer Training Condition only had 20 minutes to take the test, while those in the Lecture Condition had 30 minutes.

Training time:
- Procedural Variable
- Ongoing Equivalence

From before -- using our new language

RH: Computerized spelling practice leads to better performance than does paper & pencil practice.

Twenty Spanish-English bilingual 4th grade students were given 10 words and practiced them 5 times each on the computer. Twenty Spanish-English bilingual speaking 2nd grade students were given the same 10 words and practiced them 3 times each using paper & pencil. When tested, the "computer practice" students did better than the "paper & pencil practice" students

We identified "grade" as a confound.

Does it mess up initial or ongoing equivalence & how do you know ??

initial equivalence -- it is a subject/measured variable

We identified "number of practices" as a confound.

Does it mess up initial or ongoing equivalence & how do you know ??

ongoing equivalence -- it is a manipulated/procedural variable

Another from before -- using our new language

RH: Group therapy will lead to lower dep. scores than individual therapy

Five employed & five unemployed patients with no prior therapy completed a 24-session course of group therapy, meeting each time at the university psychiatric clinic. A different group of five employed & five unemployed patients, each of whom had previously received therapy for depression, completed a 10-session series of individual therapy, meeting at the same clinic. After the respective therapies, the group therapy patients had lower depression scores.

We identified "# sessions" as a confound.

Does it mess up initial or ongoing equivalence & how do you know ??

ongoing equivalence -- it is a manipulated/procedural variable

We identified "prior therapy" as a confound.

Does it mess up initial or ongoing equivalence & how do you know ??

initial equivalence -- it is a subject/measured variable

Just one more -- this one has changed -- find all the confounds and tell what part of internal validity each "screws up"

RH: More therapy sessions will lead to lower dep. scores.

Ten employed patients with no prior therapy completed a 24-session course of group therapy, meeting each time at the university psychiatric clinic. Ten other unemployed patients, each of whom had previously received therapy for depression, completed a 10-session series of individual therapy, meeting at a local church. After the respective therapies, the group therapy patients had lower depression scores.

Initial equivalence confounds?

- Employment -- msr/sub var
- Prior Therapy -- msr/sub var

Ongoing equivalence confounds?

- # sessions -- manip/proc var
- meeting location -- manip/proc var

Traditionally.. Internal Validity is about Causal Interpretability

Why? Causal Interpretability is about eliminating confounds (so that a statistical relationship between two variables can be properly interpreted as a causal relationship between the two !!

Why? Because we know that **confounds change results**! If both the causal variable and the confound are "operating" then the results will be changed from what they would be if there is no confounds and only the causal variable is operating!!

With this in mind.. I'd like to expand how we think about Internal validity a bit..

The better the internal validity, the more accurate the results – both Causal results and Associative results !!!

So, you always want to run the “cleanest” & “most controlled” study possible – whether you are testing Causal or Associative research hypotheses.

This is a VERY high standard for us to meet!

It is very difficult to run a completely controlled study – with no confounds!!

So we have to constantly worry about how we can “clean up” or “control” our research and data collection, so that we get the most valid results possible!

Internal validity is always important, because Confounds Change Results !

From which study will you learn more???

Study #1 Those who got more practices were also more motivated and were run during a different semester than those who got fewer practices

Study #2 Those who got more practices were equally motivated and were run during the same semester than those who got fewer practices

Whether you are testing a causal or an associative RH, the data from Study #2 is going to be easier to interpret!

The fewer confounds you have, the more you learn from their being a statistical association between two variables, whether what you are trying to learn is associative or causal !!!

Internal validity is always important, because Confounds Change Results !

The Relationship between Internal & External Validity

There are two different ways to think about the relationship between these two types of validity

• logically they are mutually exclusive, but we seem to alternate between using them both

- “Trade-off” characterization
  - it is impossible to promote both internal and external validity within a single study
  - the researcher must choose which will be emphasized in a particular study
    - internal validity (control)
    - external validity (representativeness)
- “Precursor” characterization
  - without causal interpretability (from having internal validity), what is there to generalize ???
  - focuses on causal information - suggesting associative information is not valuable

Participant Assignment – "creating" initial equivalence

- "Who will be in what condition of the study when?"
- goal is to for participants in each condition of the study to be equivalent, on the average, before the manipulation begins
- related type of validity is Internal validity - initial equivalence
- Note: participant assignment has nothing to do with the External Validity of the study -- only the internal validity component of internal validity (causal interpretability)

How this works for each type of design …

In Between Groups Designs

- each participant will complete only one condition -- randomly determine which condition for each participant

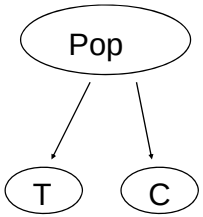
In Within-Groups Designs

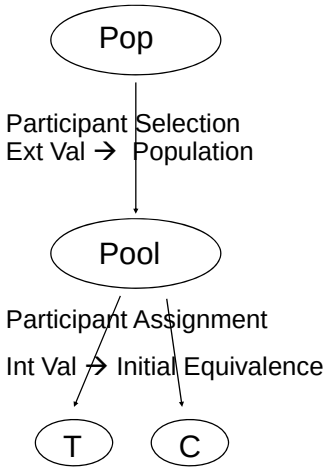
- each participant will complete all conditions -- randomly determine the condition order for each participant

Separating "Selection" & "Assignment"

A common representation of the participant acquisition process is shown below.

Folx are randomly chosen from the pop and placed into one of 2 groups.





What usually happens is shown above: First participants are selected into a "pool" and then are assigned into groups. Different aspects of validity are influenced by each step!!!

Acceptable Participant Assign. Procedure for Causal RH:

- Random Assignment of individuals by the researcher
  - each participant has an equal chance of being in each condition of the study (BG) or each condition order (WG)
  - thus, all subject variables are "balanced" or "averaged out" across the conditions before manipulation begins
  - this what gives us "initial equivalence" in a true experiment

Random assignment for Between Groups Designs

- Each participant will complete one condition (Tx1 or Tx2)
- 1st participant -- flip a coin assign Tx1 if heads or Tx2 if tails
- 2nd participant -- gets opposite of 1st participant
- 3rd participant -- flip coin again & assign Tx1 or Tx2
- 4th gets opposite condition of 3rd participant

Remember …

- random assignment doesn't guarantee initial equivalence (though we act like it does)
- random assignment is more likely to produce initial equivalence the larger the sample -- "better chance for chance to work"

Random assignment for Within-Groups Designs

- Each participant will complete both conditions (Tx1 & Tx2)
- For WG designs, RA is called "counterbalancing"
- 1st participant -- flip a coin assign the order Tx1-Tx2 if heads or the order Tx2-Tx2 if tails
- 2nd participant -- gets opposite order of 1st participant
- 3rd participant -- flip coin again & assign the condition order
- 4th gets opposite order of 3rd participant

Remember …

- random assignment doesn't guarantee initial eq.
- random assignment "works better" the larger the sample

Two important things about RA for WG designs…

- Not all studies can be run with a WG design
  - e.g. can't run gender as a WG design (or other subject variables)
- Can't counterbalance all sets of conditions
  - e.g., can't counterbalance "0 vs. 10 practices" or "before-after"

Unacceptable -- procedures that thwart testing Casual RH:

- Random assignment of groups (rather than individuals)
  - don't know that the groups were equivalent
- Arbitrary Assignment by the researcher
  - anything not using a "probabilistic" process -- might even be based on a "good idea" -- but isn't random
- Self Assignment by the participant
  - participant chooses what condition/order they will be in
- Administrative Assignment
  - non-random assignment determined by someone else
- Non-Assignment or "Natural Assignment"
  - participant is already "in" conditions before they arrive at the study -- "causal variable" is really a subject variable

Problem with all of these?

For each of these there is a "reason" for why participants are in a particular condition/order -- that reason, and anything associated with it produces a confounding of initial equivalence

Tell whether each is random, arbitrary, self, administrative or involves no assignment (were in "natural groups" before arriving to participate in the study...

- after being presented with the options, each patient chose whether they would receive the "standard" or the "experimental" operation — Self
- the researcher decided that the first 20 participants would be assigned to the treatment condition, the rest would be assigned to the control — Arbitrary
- the Hospital Executive Committee determined that people who were over 60 years old would all receive the "standard" operation and all others would be randomly assigned to which operation they would receive — Admin
- medical records were examined to determine if the each participant had received the "standard" or "experimental" operation — None
- whether each patient would receive the "standard" or "experimental" operation was determined by a coin-flip — RA
- the researcher flipped a coin to decide which dormitory would receive in-room internet access and which would continue with common-room access — RA- groups

How do we “produce” the 2 components of internal validity????

Important point -- we use different techniques to produce initial equivalence (of subject variables) and to produce ongoing equivalence (of procedural variables).

Initial equivalence of subject variables

Random assignment of individual participants to treatment conditions before treatment begins

Ongoing equivalence of procedural variables

Random assignment of individual participants to procedural alternatives before treatment begins

Procedural standardization of manipulation, instructions, interaction, task completion and performance measurement

Random Assignment to Control Initial vs. Ongoing Equivalence

Randomly assigning individual participants to the conditions of a study (which condition for BG or condition order for WG) is used to control initial equivalence of subject variables.

- RA “ensures” that, on average, participants in the different conditions (BG) or different condition orders (WG) are the same “on average” on all subject variables

We also use random assignment to help control the ongoing equivalence of some procedural variables, for example…

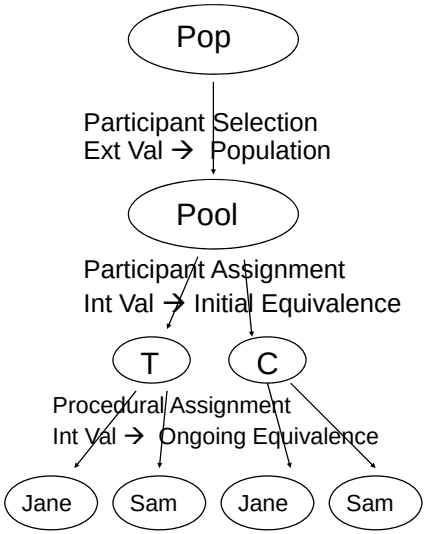
- if we have multiple research assistants – we should RA which research assistant runs each participant
- researcher gender, age, appearance, race/ethnic & perceived comfort are all known to influence participant motivation, attention & performance !!!
- if we have multiple sets of instrumentation – we should RA which set is used for each participant

Separating Assignment for Initial & Ongoing Equivalence

So, the whole process often looks like this…

Multiple Procedural Assignment steps may be necessary:

Data collector, room, equipment, stimulus set, data coder, etc.



Tell whether each random assignment controls subject variables or procedural variables and whether the RA improves initial eq. or ongoing eq. …

IV is type of operation

- whether each patient would receive the "standard" or "experimental" operation was determined by a coin-flip → SV → initial
- we flipped another coin to decide which of four surgeons would perform the operation → PV → ongoing

IV is vision vs. touch

- ½ the participants were assigned to use the old stimulus set we've been using for years and ½ were assigned to use the new stimulus set we just had made this semester → PV → ongoing
- ½ the participants were randomly assigned to the visual condition, while the other ½ completed the touch condition → SV → initial

IV is treatment vs. control

- Jane ran a random ½ of the participants and Sam ran the other ½ → PV → ongoing
- whether the participant was run in the treatment or control condition was based the roll of a 6-sided die. → SV → initial