Causation when Experiments are Not Possible

The search for truth is like looking for Elvis … on any given day there will be many sightings --- most will be impersonators!

Confounds, Confounds, Confounds

Review

- Experiments manipulate the independent variable and measure changes in the dependent variable
- Major concern—confounding variables
  - Variables correlated with the independent variable that may be causes of the dependent variable
  - Subject confounds: differences between subjects
  - Procedural confounds: differences in way experimental and non-experimental groups are treated

Clicker Question

An experimenter sets out to conduct an experiment in which all subject confounds have been eliminated. What strategy should she adopt?

A. Randomly assign subjects to the experimental and control group
B. Match subjects in the experimental and control groups on all possible confounds
C. Match on the most likely confounds and then random assign subjects
D. Forget it. You can never eliminate all confounds

Clicker Question

Why cannot an experiment eliminate all subject confounds by randomization and matching?

A. The experimenter is likely to slip and not match correctly
B. There are too many variables and by chance some will end up correlated with the independent variable
C. The placebo effect will cause subjects to respond even in the control group
D. Experimenter bias will lead the experimenter to see an effect where there isn't one

Review - 2

- Strategies for removing risk of confounds:
  - Randomization: attempt to neutralize effects of subject confounds, known and unknown
  - Matching: make subjects equivalent in terms of possible known confounds
  - Locking: fixing a value of a procedural variable
  - Measuring: treat a possible confound as additional independent variable and then measure for correlations
- Two strategies for controlling subject confounds
  - Randomization when using between-subject designs
    - Major risk: non-equivalent groups of subjects
  - Counterbalancing when using within-subject designs
    - Major risk: contamination of subjects

Review - 3

- **Internal validity**: are the effects on the dependent variable due solely to the manipulation of the independent variable
  - Was there a confounding subject variable?
  - Was there a confounding procedural variable?
    - Demand characteristics (reactivity)
    - Experimenter bias (observer bias)
- **External validity**: do the results of the study generalize to the population, setting, and manipulation of interest

Clicker Question

When someone raises a question of internal validity to an experiment they are

A. Raising doubts about whether the effect produced is due to the purported cause
B. Questioning whether the effect would always occur when the cause is present
C. Raising doubts about whether the results are applicable to a different context
D. Questioning whether the p-value that was used was too small

External Validity

To what extent can you ***generalize*** the results of your study?

Population Generalization

Will a study using one population generalize to another population?

- Will a study of college sophomores generalize to middle-aged adults?
- Will a study of chronically depressed patients generalize to patients who are acutely depressed?
- Will a study of captive raised dolphins generalize to wild dolphins?
- Will a study on mice generalize to humans?

Are there differences between the studied and target populations that are causally related to the effect of interest?

Setting Generalization

Will a study conducted in one laboratory or clinical setting generalize to the setting of interest?

- Will results obtained in a flight simulator generalize to an actual cockpit?
- Will results obtained in an outpatient setting generalize to a psychiatric hospital?
- Will results obtained in a laboratory generalize to customers in a store?

Are there differences between the studied setting and the target setting that are causally related to the effect of interest?

Manipulation generalization

Will a result obtained with one task generalize to other tasks or stimuli?

- Will studies of perceiving visual illusions presented on a computer screen generalize to perception of ordinary objects?
- Will a survey of consumer attitudes generalize to consumer behavior?

Are there differences in the manipulation that are causally related to the effect of interest?

Assessing External Validity

Must make a plausibility judgment in assessing external validity (or do a separate study!)

- Is the target population different from the studied population in ways that are *likely to matter* for the causal claim?
- Is the target setting different from the studied setting in ways that are *likely to matter* for the causal claim?
- Is the manipulation used in the experiment different from the target process in nature in ways that are *likely to matter* for the causal claim?

Example: Rats and Saccharine

1977 Canadian study which fed pregnant rats up to 20% of their body weight per day in saccharine showed an increase in bladder tumors

Saccharine was banned in Canada and the FDA was about to ban its use in the US
when Congress intervened

Assessing external validity:
- Are rats relevantly like humans?
- Is living in the laboratory like living at home, etc.?
- Is feeding up to 20% of body weight like eating as part of regular diet?

Clicker Question

Which of the following concerns are about external validity?

A. A study did not achieve statistical significance
B. A non-representative group of participants was studied and one can't anticipate what sorts of effects the independent variable will have on other participants
C. Participants might have responded to the novelty of the experiment more than the specific effect of the independent variable that was manipulated
D. The participants figured out which treatment they were on

The main advantage of experiments

Experiments manipulate the independent variable
- Unless there are confounds, any change in the dependent variable can be attributed to the independent variable

When the independent variable is not being manipulated
- You have much less confidence that the independent variable is what is responsible for the change in the dependent variable
- There is increased risk that it is due to other factors —confounding variables

Sometimes you cannot manipulate the independent variable

You want to study whether sex affects income

You know there is a correlation

- Women earn $0.69 for every $1.00 men earn

Is the causal link between being female and income or between some correlated confounding variable?

Let's do an experiment:

We will randomly assign people to be men or women . . .

Impossibility of experiments

Sometimes experiments are physically impossible

- Cannot randomly make metal gold or silver
- Cannot randomly assign people to IQ
- Cannot randomly assign genes to mice—but we are getting there

Sometimes experiments are *ethically* problematic

- Immoral to simply give people HIV or cut out parts of their brains
- Immoral to randomly assign people to the values *College* and *No College*
- Sometimes immoral to have proper control groups (withholding treatment)

Settle for *controlled* correlations

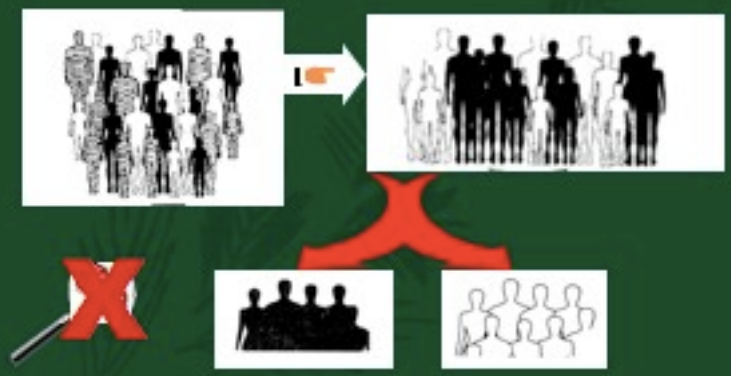
Measure pre-existing value of the independent variable and select subjects based on their existing value rather than manipulating the variable

- Control as much as possible for confounds
- Draw tentative causal conclusions based on correlation

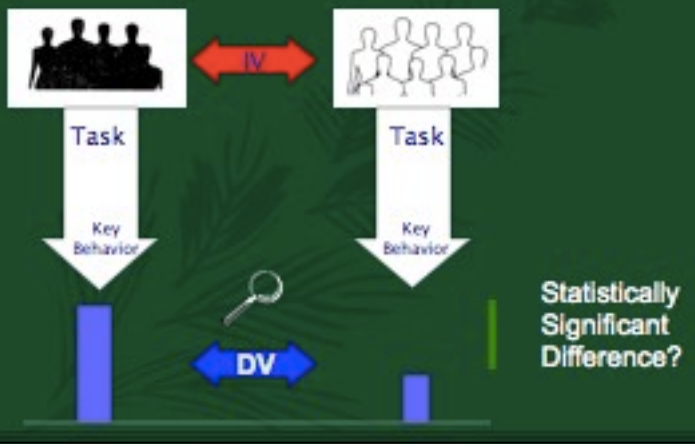
Two strategies:

- **Prospective studies**: identify groups in terms of possible *cause* variables and measure possible *effect* variables
- **Retrospective studies**: identify groups in terms of possible *effect* variables and measure possible *cause* variables

Prospective studies



Prospective studies



Benzypyrene and lung cancer

Benzypyrene, an ingredient in coal tar pitch and asphalt, is known to cause skin cancer.
It is also present in cigarettes.
Could it be a factor in lung cancer?

Roofers are constantly in contact with coal tar pitch and asphalt—exposed to the amount of benzypyrene equal to smoking 35 packs a day!

Prospective study traced 5,788 roofers for 12 years

Benzypyrene and lung cancer

Rather than following an explicit control population, researchers used US mortality rates for the general population as the comparison

- Roofers with less than 20 years experience showed no increase in rates of lung cancer
- Roofers with 20-30 years experience showed 1.5 times the usual rate of lung cancer
- Roofers with 30-40 years experience showed 2.47 times the usual rate of lung cancer

With selection comes confounds

- Many other variables may correlate with both the independent and dependent variables and one of these may be responsible for the observed group differences
  - What might correlate with Benzypyrene exposure in roofers?
- The nature and number of potential confounding variables may not even be known
- Without randomization, have no way of countering possible effects of ***unknown*** confounding variables
- There are strategies for dealing with ***known*** possible confounding variables

Matching to control confounds

As long as we know what might be the possible confounds, we can control for them by **matching** the different groups.

Two strategies for matching
- Match ***each subject*** in the different treatment groups on each confounding variable
- Match ***means*** for confounding variables across treatment groups

Limits: there may be many other variables (or too many to match on) that differ between groups that might have causal effects on the dependent variable

Clicker Question

When matching is used to control possible confounds

A. Researchers match individuals on the dependent and independent variables
B. Researchers match individuals on the dependent and potentially confounding variables
C. Researches match individuals on all variables
D. Researchers match individuals on the independent and potentially confounding variables

Measuring to control confounds

Sometimes it is not practical to match the groups on all suspected confounds

But if you can measure values on these variables, you can investigate whether they correlate with the dependent variable

If they do, they become possible causes

Multi-factor studies examine the contributions of multiple independent variables on the dependent variable

Example of measuring confounds

Condition known as failure to thrive

- infant's weight gain (relative to height) is in the bottom 3% of the distribution

What is the effect of failure to thrive (independent variable) on mental development?

Operationally define mental development as score on Bailey Mental Scale of Infant Development (dependent variable)

Measure several other possible confounds and evaluate whether they correlated with dependent variable.

Two found:

- Education level of parents
- Time placed with alternative care-giver

Clicker Question

If you were to do a study of the effect of gender on job performance and suspected child care responsibilities as a confound, how would you try to investigate it?

A. Determine whether there is a correlation between time spent in child care and gender
B. Determine whether there is a correlation between time spent in child care and job performance
C. Measure time spent on child care and eliminate those who spend too much time from the study
D. Measure time spent on child care and study only those who did a lot of child care

From Prospective to Retrospective Studies

- To do a prospective study you must identify groups based on the relevant independent variable, then wait until you can measure the dependent variable
  - In some cases of interest, that may mean waiting years
- Alternative strategy is to start with the effect and look backwards to isolate the possible cause
  - This is what a **retrospective study** attempts to do

Retrospective studies

Both experiments and prospective studies begin with the groups identified in terms of the ***independent*** variable (suspected cause)

- Either assign or select subjects
- Measure the ***dependent variable*** (suspected effect)

Retrospective studies work the other way around

- Begin with subjects who show the value on the ***dependent variable*** (suspected effect)
- Match them with others who lack the value on the dependent variable
- Measure the presence or absence of the ***independent variable*** (suspected cause)

Why retrospective studies?

An effect occurs but we are lacking in good hypotheses as to what might cause it
- making it hard to do either an experiment or a prospective study

The effect (dependent variable) of interest occurs very infrequently
- which would require enormously large samples to get enough cases with the effect
- but we still want to know why it occurs

There is not time for a prospective or an experimental study
- but we need answers NOW

Birth control pills and blood clots

In the 1960s a surprising number of fatal blood clots started appearing among relatively young women
Most of these women had started taking birth control pills within the last year

Was the pill the culprit?
It would take years to design and run a proper study
meanwhile, women were dying

Search for women who had been treated for nonfatal clots (legs or lungs) within previous two years—58 such women found

Birth control pills and blood clots

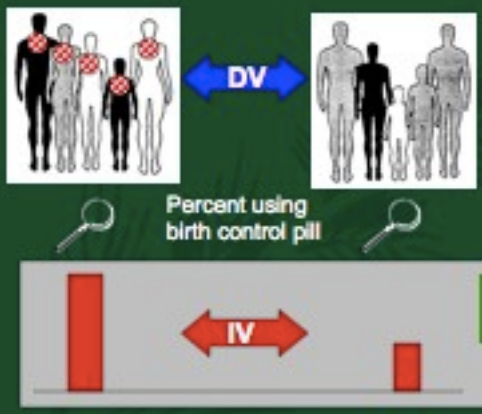
Need a comparison group: 116 married women who had been admitted to the same hospitals for serious surgery or other medical condition than blood clotting.
Matched on age, number of children, etc. (the likely confound variables)

Of the 58 admitted for blood clots, 26 (45%) had taken oral contraceptives in the preceding month

Of the 116 matched individuals, 10 (9%) had taken oral contraceptives in the preceding month

This difference is statistically significant
But NOTE: you cannot judge how much the risk is!

Retrospective Design



Confounds in retrospective studies

Must match on the Dependent Variable

Must be able to *detect differences* in the Independent Variable

Only look for those differences you suspect are relevant

Often this requires relying on *memory* of the participants

- Memory may be different between those who exhibit the value of interest on the dependent variable, especially if it is negative, and those who don't
  - Those who are ill may be more attuned to what they have done or what has happened to them

What predicts or causes Alzheimer's?

Start with population, some of whom have developed Alzheimer's and some who haven't

Study of the School Sisters of Notre Dame, an order of nuns

Examined 678 nuns from Minnesota Texas, Wisconsin, Connecticut, Maryland, Missouri, and Illinois

Look back into the records of those who developed Alzheimer's and those who didn't

Look for differences earlier in their lives

Nun study

- Taking folic acid negatively correlated with Alzheimer's
- Occurrence of small strokes a predictor of Alzheimer's
- The more ideas nuns packed into the sentences of their early autobiographies, the less likely they were to get Alzheimer's disease six decades later
- Maybe also the prevalence of positive emotions in early writing predicts less Alzheimer's

Nun study

Contrast:

> "My father, Mr. L.M. Hallacher, was born in the city of Ross, County Cork, Ireland, and is now a sheet-metal worker in Eau Claire"

with:

> "My father is an all-around man of trades, but his principal occupation is carpentry, which trade he had already begun before his marriage with my mother"

From Retrospective Study to Prospective Study to Experiment

Growing phenomenon of childhood obesity (dependent variable)

- Hypothesis: Hours reading is a cause (independent variable)

Retrospective Study of Childhood Obesity

- Begin with group of obese children
  - Need operational definition of obesity!
- Find non-obese matches on possible confound variables
  - Obese parents
  - Foods in diet
  - Grades in school

From Retrospective Study to Prospective Study to Experiment

- If, after matching on these other variables, there is a statistically significant difference in hours spent reading
  - Then hours spent reading is a candidate cause of childhood obesity
  - But, despite care in matching, many variables will not be matched
- Follow up retrospective study with a prospective study
  - Identify groups of children who are readers and non-readers
  - Match the two groups on all known potential confounds

From Retrospective Study to Prospective Study to Experiment

- Prospective Study of Reading and Obesity using Pretest-Posttest Design
  - Measure participants degree of obesity at outset
  - After test period, measure participants degree of obesity
  - Determine the change in obesity
- If there is a statistically significant difference in the increase in obesity in the readers versus non-readers, it is highly plausible that reading is a cause of obesity
  - But there may well be unsuspected confounds
  - Unknown confounds can only be controlled in an experiment

Clicker Question

What is the major advantage of a randomized experiment over a prospective study?

A. In an experiment one manipulates the independent variable
B. Randomizing can control for unknown subject confounds
C. Randomizing can control for unknown procedural confounds
D. There is no experimenter bias in a randomized experiment

From Retrospective Study to Prospective Study to Experiment

Set up a controlled experiment

Choose a sample of children

- Randomly assign some of them to a reading enticement program

Still need to control for confounding procedural variables such as time spent reading

- What do those not in reading enticement program do with their time?
- Perhaps create a crafts enticement program

If correlation between participation in reading enticement program and increase in obesity holds up

- You have the best possible evidence for a causal link between reading and obesity

A word on reporting results

With experiments and prospective studies, one can ask not only if the result is statistically significant, but what is the effect size

But be careful! Often reports of dependent variable are made in terms of percentage increases

- An increase from 1/1000 to 5/1000
- An increase from 10/1000 to 50/1000
- An increase from 100/1000 to 500/1000
  - Are all 5 fold increases (500% increases) but one is an increase of 4/1000 while the last is an increase of 400/1000