POLS 205
Political Science as a Social Science

Experiments & Observation

Christopher Adolph

University of Washington, Seattle

April 14, 2010

Outline

What is a research design?

A research design is a plan to answer your research question, and includes:

1. A (causal) theory and implied hypotheses
2. A unit of analysis on which the hypotheses operate
3. A set of variables, including a dependent variable & covariates
4. A plan to collect these data
5. A plan to analyze these data

Today we focus on steps 4 and 5, and a single powerful strategy: **experiments**.

Key concepts for experiments

Treatment A variable under the control of experimenters

Key concepts for experiments

Treatment A variable under the control of experimenters

Treatment Group Subjects given the treatment

Key concepts for experiments

Treatment A variable under the control of experimenters

Treatment Group Subjects given the treatment

Control Group Subjects not given the treatment

Key concepts for experiments

Treatment A variable under the control of experimenters

Treatment Group Subjects given the treatment

Control Group Subjects not given the treatment

Randomization Subjects are assigned *purely* by chance to either a treatment group or a control group

Key concepts for experiments

Treatment A variable under the control of experimenters

Treatment Group Subjects given the treatment

Control Group Subjects not given the treatment

Randomization Subjects are assigned *purely* by chance to either a treatment group or a control group

Testing Subjects are measured on some dependent variable

Key concepts for experiments

Treatment A variable under the control of experimenters

Treatment Group Subjects given the treatment

Control Group Subjects not given the treatment

Randomization Subjects are assigned *purely* by chance to either a treatment group or a control group

Testing Subjects are measured on some dependent variable

Internal validity A properly designed experiment will correctly estimate the effect of treatment under laboratory conditions

Key concepts for experiments

Treatment A variable under the control of experimenters

Treatment Group Subjects given the treatment

Control Group Subjects not given the treatment

Randomization Subjects are assigned *purely* by chance to either a treatment group or a control group

Testing Subjects are measured on some dependent variable

Internal validity A properly designed experiment will correctly estimate the effect of treatment under laboratory conditions

External validity Even properly designed experiments may not capture a particular real world situation

Key concepts for experiments

Treatment A variable under the control of experimenters

Treatment Group Subjects given the treatment

Control Group Subjects not given the treatment

Randomization Subjects are assigned *purely* by chance to either a treatment group or a control group

Testing Subjects are measured on some dependent variable

Internal validity A properly designed experiment will correctly estimate the effect of treatment under laboratory conditions

External validity Even properly designed experiments may not capture a particular real world situation

Intent to Treat In some experiments, subjects assigned to treatment group may evade treatment

Advantages & Disadvantages of Experiments

Well-designed experiments have three major advantages:

Maximize internal validity If correctly designed, experiments provide best possible estimates of a causal effect under ideal (lab) conditions

Advantages & Disadvantages of Experiments

Well-designed experiments have three major advantages:

Maximize internal validity If correctly designed, experiments provide best possible estimates of a causal effect under ideal (lab) conditions

Avoid measurement quandaries Clever experimental design can control for hard or impossible to measure confounders

Advantages & Disadvantages of Experiments

Well-designed experiments have three major advantages:

Maximize internal validity If correctly designed, experiments provide best possible estimates of a causal effect under ideal (lab) conditions

Avoid measurement quandaries Clever experimental design can control for hard or impossible to measure confounders

Isolate causal effects Experiments can isolate the effect of covariates than tend to covary in the real world

Advantages & Disadvantages of Experiments

Unfortunately, experiments have three major downsides:

Ethical qualms Experiments on human might harm either treatment or control subjects

Advantages & Disadvantages of Experiments

Unfortunately, experiments have three major downsides:

Ethical qualms Experiments on human might harm either treatment or control subjects

Poor feasibility Many causal variables resist manipulation, especially when people are involved

Advantages & Disadvantages of Experiments

Unfortunately, experiments have three major downsides:

Ethical qualms Experiments on human might harm either treatment or control subjects

Poor feasibility Many causal variables resist manipulation, especially when people are involved

Low external validity Human awareness of the experimental environment often invalidates lab findings

A Closer Look at Internal Validity

Experiments are uniquely suited to make causal inferences:

A Closer Look at Internal Validity

Experiments are uniquely suited to make causal inferences:

No chance of reverse causation Because random assignment precedes treatment, it cannot cause it

Note this assumes subjects stay in the study.

If the effects of the treatment or control cause selective dropout, this *does not hold*

A Closer Look at Internal Validity

Experiments are uniquely suited to make causal inferences:

No chance of reverse causation Because random assignment precedes treatment, it cannot cause it

Note this assumes subjects stay in the study.

If the effects of the treatment or control cause selective dropout, this *does not hold*

Limited effects of confounding variables As sample size increases, treatment and control should have similar distributions on all confounding variables

Note this only holds on *average*.

In a given experiment, control and treatment may be unbalanced by chance

More likely if sample size is small

A Closer Look at Internal Validity

Other threats to interval validity can be solved through careful design

Measurement Error The tests given to subjects may measure their results with error

If this error is random, results will still be valid on average

If this error is *correlated* with the treatment or confouders, the experiment will be invalidated

A Closer Look at Internal Validity

Other threats to interval validity can be solved through careful design

Measurement Error The tests given to subjects may measure their results with error

If this error is random, results will still be valid on average

If this error is *correlated* with the treatment or confouders, the experiment will be invalidated

Testing Effects Human subjects learn—and may get better at the "test" on their own!

This can include simply learning how to answer test questions quickly and efficiently

FYI: This is almost all SAT test prep does: give you tests until you get a "testing effect"

A Closer Look at External Validity

Experiments often have low external validity:

1. Lab conditions may not fully replicate real world situations:

A Closer Look at External Validity

Experiments often have low external validity:

1. Lab conditions may not fully replicate real world situations:
    - Real world social enviroments often involve interaction over years

A Closer Look at External Validity

Experiments often have low external validity:

1. Lab conditions may not fully replicate real world situations:
   - Real world social enviroments often involve interaction over years
   - . . . in richly layered social networks . . .

A Closer Look at External Validity

Experiments often have low external validity:

1. Lab conditions may not fully replicate real world situations:
    - Real world social enviroments often involve interaction over years
    - . . . in richly layered social networks . . .
    - E.g., could we ever replicate a “legislature” in a lab? Or a “campaign”?

A Closer Look at External Validity

Experiments often have low external validity:

1. Lab conditions may not fully replicate real world situations:
   - Real world social enviroments often involve interaction over years
   - . . . in richly layered social networks . . .
   - E.g., could we ever replicate a “legislature” in a lab? Or a “campaign”?

2. Compliance rates in labs may differ from real world settings:

A Closer Look at External Validity

Experiments often have low external validity:

1. Lab conditions may not fully replicate real world situations:
   - Real world social enviroments often involve interaction over years
   - . . . in richly layered social networks . . .
   - E.g., could we ever replicate a “legislature” in a lab? Or a “campaign”?

2. Compliance rates in labs may differ from real world settings:
   - Intent To Treat (ITT) effects differ from pure treatment effects

A Closer Look at External Validity

Experiments often have low external validity:

1. Lab conditions may not fully replicate real world situations:
   - Real world social enviroments often involve interaction over years
   - ... in richly layered social networks ...
   - E.g., could we ever replicate a “legislature” in a lab? Or a “campaign”?
2. Compliance rates in labs may differ from real world settings:
   - Intent To Treat (ITT) effects differ from pure treatment effects
   - We may be interested in either or neither (what about real world ITT?)

A Closer Look at External Validity

Experiments often have low external validity:

1. Lab conditions may not fully replicate real world situations:
   - Real world social enviroments often involve interaction over years
   - . . . in richly layered social networks . . .
   - E.g., could we ever replicate a “legislature” in a lab? Or a “campaign”?
2. Compliance rates in labs may differ from real world settings:
   - Intent To Treat (ITT) effects differ from pure treatment effects
   - We may be interested in either or neither (what about real world ITT?)

Thoughts on how to solve these problems?

A Closer Look at External Validity

Experiments often have low external validity:

1. Lab conditions may not fully replicate real world situations:
   - Real world social enviroments often involve interaction over years
   - ...in richly layered social networks ...
   - E.g., could we ever replicate a “legislature” in a lab? Or a “campaign”?
2. Compliance rates in labs may differ from real world settings:
   - Intent To Treat (ITT) effects differ from pure treatment effects
   - We may be interested in either or neither (what about real world ITT?)

Thoughts on how to solve these problems?

Field experiments or natural experiments, perhaps?

Experimental design

To run an experiment properly, we must combine random assignment, treatment, and testing to ensure accurate causal inference

But in the real world, these steps are expensive, so we also want to choose the design that maximizes interval validity subject to our budget constraint

Consider the following designs:

1. Pre-test, post-test
2. Post-test
3. Multi-group
4. Case-control

Pre-test, Post-test Experimental Design

Classic experimental method is to test before and after treatment

By comparing in two directions:

1. Before and after the test for treated group
2. Between the treated group and the control

we can calculate the Average Treatment Effect, or ATE

(Why *Average* Treatment Effect?

Pre-test, Post-test Experimental Design

Classic experimental method is to test before and after treatment

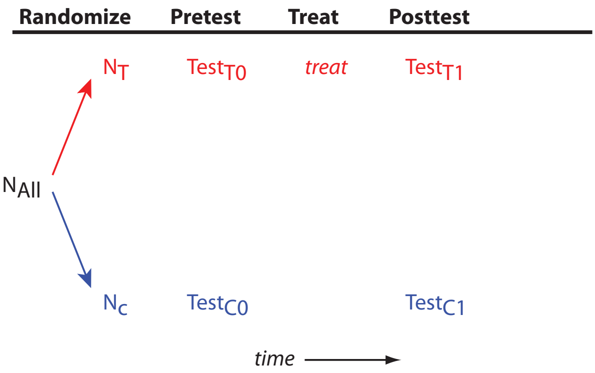
By comparing in two directions:

1. Before and after the test for treated group
2. Between the treated group and the control

we can calculate the Average Treatment Effect, or ATE

(Why *Average* Treatment Effect? Because we want to remove random noise)

Pre-test, Post-test Experimental Design



$$\text{Average Treatment Effect} = \frac{1}{N_T}\sum_{i=1}^{N_T}(\text{Test}_{T1} - \text{Test}_{T0}) - \frac{1}{N_C}\sum_{j=1}^{N_C}(\text{Test}_{C1} - \text{Test}_{C0})$$

Post-test Experimental Design

- What if we can't afford two waves of tests?
- What if we fear a very strong testing effect?
- What if we don't want subjects to even know what we are testing until after the treatment?

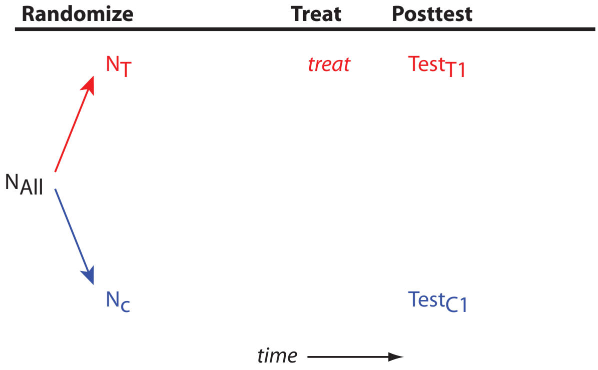
Then we can just drop the pre-test altogether!

That is, if N is large enough, the treatment and control should have the same pretest:

$$\text{Test}_{T0} - \text{Test}_{C0} \rightarrow 0 \text{ as } N \rightarrow \infty$$

This leads to a simplified research design

Post-test Experimental Design



$$\text{Average Treatment Effect} = \frac{1}{N_T}\sum_{i=1}^{N_T}\text{Test}_{T1} - \frac{1}{N_C}\sum_{j=1}^{N_C}\text{Test}_{C1}$$

Longitudinal Experimental Design

Other extensions:

- What if we want to measure effects of treatment over time?
- Or effects of different treatments?

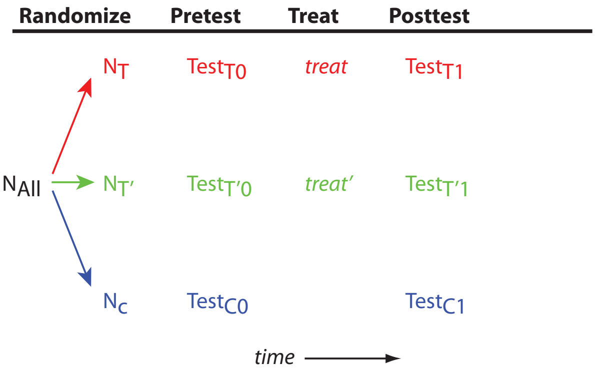
Longitudinal Experimental Design

Randomize	Pretest	Treat	Posttest	Posttest
N_T	Test_{T0}	*treat*	Test_{T1}	Test_{T2}
N_{All}				
N_C	Test_{C0}		Test_{C1}	Test_{C2}

time $\longrightarrow$

$$\text{Average Treatment Effect} = \frac{1}{N_T}\sum_{i=1}^{N_T}(\text{Test}_{T2} - \text{Test}_{T0}) - \frac{1}{N_C}\sum_{j=1}^{N_C}(\text{Test}_{C2} - \text{Test}_{C0})$$

Multigroup Experimental Design



$$\text{Average Treatment Effect} = \frac{1}{N'_T} \sum_{i=1}^{N'_T} (\text{Test}_{T'1} - \text{Test}_{T'0}) - \frac{1}{N_C} \sum_{j=1}^{N_C} (\text{Test}_{C1} - \text{Test}_{C0})$$

Case-Control Experimental Design

My favorite experimental design (strongly recommended over previous):

1. Gather all controls and confounders as you would for an observational study

Case-Control Experimental Design

My favorite experimental design (strongly recommended over previous):

1. Gather all controls and confounders as you would for an observational study
2. Randomly draw a case from the subject pool

Case-Control Experimental Design

My favorite experimental design (strongly recommended over previous):

1. Gather all controls and confounders as you would for an observational study
2. Randomly draw a case from the subject pool
3. Find a matching control from the subject pool (match on observables)

Case-Control Experimental Design

My favorite experimental design (strongly recommended over previous):

1. Gather all controls and confounders as you would for an observational study
2. Randomly draw a case from the subject pool
3. Find a matching control from the subject pool (match on observables)
4. Randomly assign the treatment to one of the pair and not the other

Case-Control Experimental Design

My favorite experimental design (strongly recommended over previous):

1. Gather all controls and confounders as you would for an observational study
2. Randomly draw a case from the subject pool
3. Find a matching control from the subject pool (match on observables)
4. Randomly assign the treatment to one of the pair and not the other
5. Measure the difference in Y for the pair

Case-Control Experimental Design

My favorite experimental design (strongly recommended over previous):

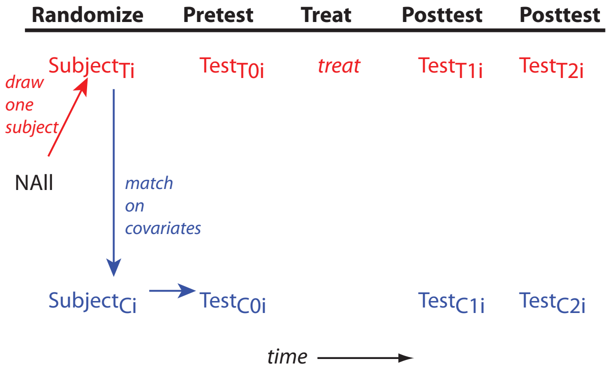
1. Gather all controls and confounders as you would for an observational study
2. Randomly draw a case from the subject pool
3. Find a matching control from the subject pool (match on observables)
4. Randomly assign the treatment to one of the pair and not the other
5. Measure the difference in Y for the pair
6. Repeat many times, and average the result

Case-Control Experimental Design

My favorite experimental design (strongly recommended over previous):

1. Gather all controls and confounders as you would for an observational study
2. Randomly draw a case from the subject pool
3. Find a matching control from the subject pool (match on observables)
4. Randomly assign the treatment to one of the pair and not the other
5. Measure the difference in Y for the pair
6. Repeat many times, and average the result
7. Report Average Treatment Effect

Case-Control Experimental Design



$$\text{Average Treatment Effect} = \frac{1}{N_T} \sum_{i=1}^{N_T} [(\text{Test}_{T1i} - \text{Test}_{T0i}) - (\text{Test}_{C1i} - \text{Test}_{C0i})]$$

Case-Control Experimental Design

Case-control solves three problems at little cost:

1. Impossible to be unbalanced on observables, even with random assignment!

Case-Control Experimental Design

Case-control solves three problems at little cost:

1. Impossible to be unbalanced on observables, even with random assignment!
2. If a subject drops out of the study, discard its pair, not the whole experiment!

Case-Control Experimental Design

Case-control solves three problems at little cost:

1. Impossible to be unbalanced on observables, even with random assignment!
2. If a subject drops out of the study, discard its pair, not the whole experiment!
3. If you think treatment effect is heterogenous as a function of observables, can report *Local* Average Treatment Effect by looking at a subset of cases.

Case-Control Experimental Design

Case-control shares with other experimental designs these problems:

1. If random assignment on unobservables if unbalanced, could be biased

Case-Control Experimental Design

Case-control shares with other experimental designs these problems:

1. If random assignment on unobservables if unbalanced, could be biased
2. If drop out is non-random with respect to unobservables, could be bias

Solution to both is to expand sample size!

Alternatives to Experiments

Suppose we can't do a lab experiment or want higher external validity

Alternative research designs:

1. **Natural Experiment**: Find a case where nature assigned a random treatment

Alternatives to Experiments

Suppose we can't do a lab experiment or want higher external validity

Alternative research designs:

1. **Natural Experiment**: Find a case where nature assigned a random treatment

2. **Field Experiment**: Assign a treatment to people in a real-world environment

Alternatives to Experiments

Suppose we can't do a lab experiment or want higher external validity

Alternative research designs:

1. **Natural Experiment**: Find a case where nature assigned a random treatment
2. **Field Experiment**: Assign a treatment to people in a real-world environment
3. **Matching**: Use case-control matching with (non-random) naturally assigned variation

Alternatives to Experiments

Suppose we can't do a lab experiment or want higher external validity

Alternative research designs:

1. **Natural Experiment**: Find a case where nature assigned a random treatment
2. **Field Experiment**: Assign a treatment to people in a real-world environment
3. **Matching**: Use case-control matching with (non-random) naturally assigned variation
4. **Regression**: Use observational data and control for *every* confounder

Natural Experiments

Find a case where nature assigned a random treatment, measure average treatment effect

Key: Convincing other scientists treatment assignment is unrelated to all confouders

Examples:

1. **Snow's cholera map**
2. **2000 Presidential election, Palm Beach ballot**

Natural Experiment 1: Snow's cholera map

The pipes of each Company go down all the street... A few houses are supplied by one Company and a few by the other, according to the decision of the owner or occupier at that time when the Water Companies were in active competition. In many cases a single house has a supply different from that on either side. Each company supplies both rich and poor, both large houses and small; there is no difference either in the condition or occupation of the persons receiving the water of either company...

It is obvious no experiment could have been designed which would more thoroughly test the effect of water supply on the progress of cholera than this.

John Snow (1885: 74-75)

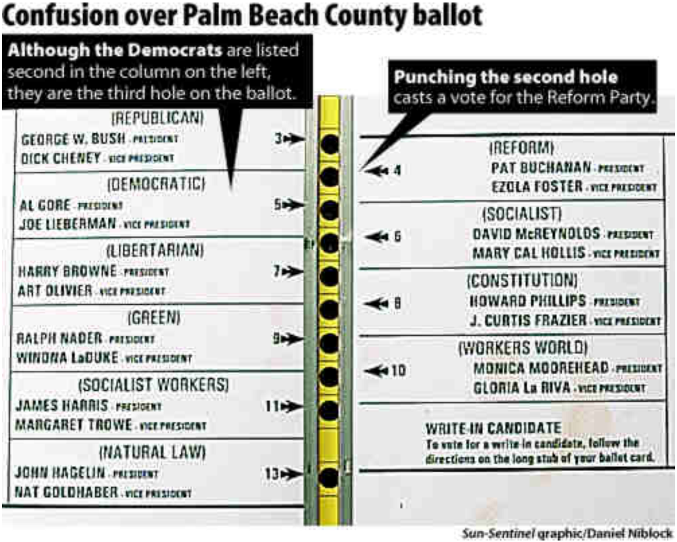
Natural Experiment 2: Palm Beach ballot

2000 Presidential election between Bush & Gore came down to Florida

Florida was achingly close:
537 votes in the last tally before the US Supreme Court called the election

But in populous, Democratic Palm Beach, at least 3400 voters picked conservative third party candidate Pat Buchanan, 3 to 8 times more than he expected

What happened?



One explanation: poor ballot design led many Gore voters to accidently punch Buchanan

Natural Experiment 2: Palm Beach ballot

Is this a natural experiment?

Natural Experiment 2: Palm Beach ballot

Is this a natural experiment?

Different counties used different ballots

Natural Experiment 2: Palm Beach ballot

Is this a natural experiment?

Different counties used different ballots

Could ballot design be correlated with partisan support? A Reform or Republican party "plot"?

Natural Experiment 2: Palm Beach ballot

Is this a natural experiment?

Different counties used different ballots

Could ballot design be correlated with partisan support? A Reform or Republican party “plot”?

Ballot designed by a Democratic county official, Theresa LaPore!

Natural Experiment 2: Palm Beach ballot

Is this a natural experiment?

Different counties used different ballots

Could ballot design be correlated with partisan support? A Reform or Republican party "plot"?

Ballot designed by a Democratic county official, Theresa LaPore!

Thought it would *help* elderly voters by magnifying text size

Natural Experiment 2: Palm Beach ballot

Is this a natural experiment?

Different counties used different ballots

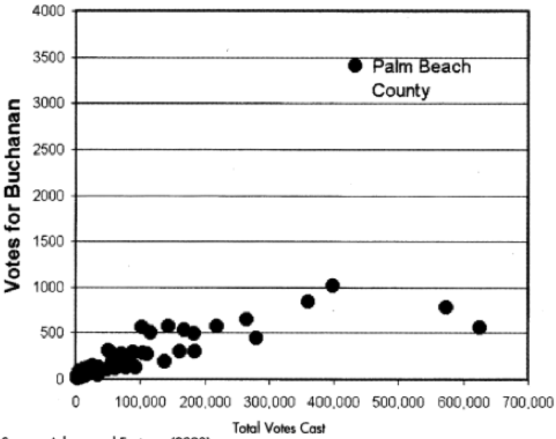
Could ballot design be correlated with partisan support? A Reform or Republican party "plot"?

Ballot designed by a Democratic county official, Theresa LaPore!

Thought it would *help* elderly voters by magnifying text size

Appears to be an excellent natural experiment: assignment of ballot uncorrelated with essentially every other political variable

FIGURE 2
Presidential Election Results for Florida, by County



Source: Adams and Fastnow (2000).

Palm beach Buchanan vote a huge outlier: can't be explained by any other variable

Field experiments

What if the “world” really was our laboratory?

Field experiments

What if the "world" really was our laboratory?

That is, we could randomly select some political actors, and subject them to a significant treatment, then compare their *real world* behavior to controls

Field experiments

What if the "world" really was our laboratory?

That is, we could randomly select some political actors, and subject them to a significant treatment, then compare their *real world* behavior to controls

A good field experiment combines the internal validity of lab experiments and the external validity of observational studies

Field experiments

What if the “world” really was our laboratory?

That is, we could randomly select some political actors, and subject them to a significant treatment, then compare their *real world* behavior to controls

A good field experiment combines the internal validity of lab experiments and the external validity of observational studies

Requires a field-manipulable variable

Field experiment example: New Haven voter turnout

Don Green & Alan Gerber (1999, PNAS) conducted a voter turnout experiment in New Haven

Field experiment example: New Haven voter turnout

Don Green & Alan Gerber (1999, PNAS) conducted a voter turnout experiment in New Haven

Randomly encouraged some voters to vote through door canvasing

Field experiment example: New Haven voter turnout

Don Green & Alan Gerber (1999, PNAS) conducted a voter turnout experiment in New Haven

Randomly encouraged some voters to vote through door canvasing

Compared to an unvisited control group

Field experiment example: New Haven voter turnout

Don Green & Alan Gerber (1999, PNAS) conducted a voter turnout experiment in New Haven

Randomly encouraged some voters to vote through door canvasing

Compared to an unvisited control group

Found a 6% increase in turnout!

Field experiment example: New Haven voter turnout

Don Green & Alan Gerber (1999, PNAS) conducted a voter turnout experiment in New Haven

Randomly encouraged some voters to vote through door canvasing

Compared to an unvisited control group

Found a 6% increase in turnout!

Potential problem: Told some voters election would be close (it wasn't)

Field experiment example: New Haven voter turnout

Don Green & Alan Gerber (1999, PNAS) conducted a voter turnout experiment in New Haven

Randomly encouraged some voters to vote through door canvasing

Compared to an unvisited control group

Found a 6% increase in turnout!

Potential problem: Told some voters election would be close (it wasn't)

Can political scientists use field experiments widely?

Field experiment example: New Haven voter turnout

Don Green & Alan Gerber (1999, PNAS) conducted a voter turnout experiment in New Haven

Randomly encouraged some voters to vote through door canvasing

Compared to an unvisited control group

Found a 6% increase in turnout!

Potential problem: Told some voters election would be close (it wasn't)

Can political scientists use field experiments widely?

Another problem (noticed by Kosuke Imai): Random assignement turned out to be very non-random—results change a lot if corrected

Field experiment example: New Haven voter turnout

Don Green & Alan Gerber (1999, PNAS) conducted a voter turnout experiment in New Haven

Randomly encouraged some voters to vote through door canvasing

Compared to an unvisited control group

Found a 6% increase in turnout!

Potential problem: Told some voters election would be close (it wasn't)

Can political scientists use field experiments widely?

Another problem (noticed by Kosuke Imai): Random assignement turned out to be very non-random—results change a lot if corrected

Case-control assignment would have solved this problem

Next time. . .

Non-experimental methods:

Matching (both qualitative & quantitative)

Regression