

(Week 5) Causality and hypothesis testing
**(Week 6) Quant. analysis: strengths &
pitfalls**

Research design in comparative
political science

9 & 16 November, 2015

Prof. Andrew Eggers

A story about program evaluation

A story about program evaluation

National Supported Work Demonstration (1975-1979):
ex-offenders, drug addicts, etc.
receive 12-18 months of
subsidized employment in 10 US
cities.



MDRC implementing NSW in 1970s

Does it work? Of 6,600 eligible participants, some randomly assigned to **control group** (no subsidized employment).

	Treatment	Control
Avg earnings after program	\$4,670	\$3,819

Lalonde (1986): “Evaluating the econometric evaluations of training programs with experimental data”

Idea: Ignore the experimental control group; use standard economic surveys instead.

How close to the experimental benchmark do we get with standard econometric approaches? Not very close:



Robert
Lalonde,
University of
Chicago

“Policymakers should be aware that the available non-experimental evaluations of employment and training programs may contain large and unknown biases resulting from specification errors.” (p. 6 | 7)

Credibility revolution (?)

Major change throughout social sciences (though earlier in economics):

- Randomized experiments (RCTs) increasingly common (field, lab, survey)
- Observational studies are increasingly “design-based” & patterned after RCTs (“quasi-experiments”)
- Measuring effects of institutions, information, social movements etc. through mere cross-sectional regression is now **unacceptable**
- “What’s your identification strategy?”



Angrist and Pischke's book

Identification Taliban?

Does obsession with
“identification strategy”
limit us unnecessarily?

Cleverness over
substance?

What can/should we do
once we internalize the
“credibility revolution”?

Abandon explanation?



Angrist and Pischke

The research formula under the “old regime”

“Impact evaluation” paradigm

- A.** Identify an unresolved question about the effects of one or more independent variables in some setting (e.g. electoral system, social cleavages and number of parties in Europe)
- B.** Collect data from that setting
- C.** Regress outcome on independent variable(s) plus controls; interpret coefficients as “causal effects”

Explanatory paradigm

- A.** Develop an interest in some outcome (Y, e.g. civil war, turnout, voting for right-wing parties); read the literature about the topic
- B.** Identify a puzzle (theory vs empirics, empirics vs empirics, etc) and propose a resolution (i.e. theory)
- C.** Conduct an empirical test of your theory

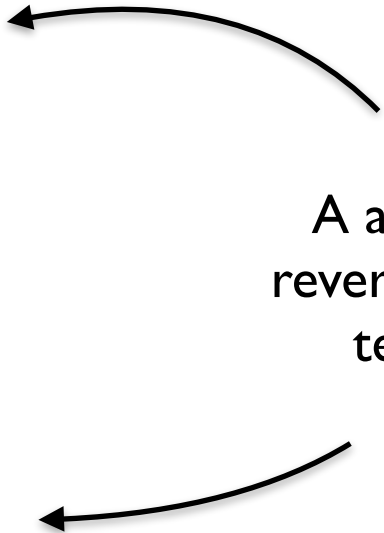
The research formula for “design-based” research

A. Learn about theoretical debates and empirical questions in a given area

B. Find/create setting where a **hypothesized determinant** differs between otherwise similar units in a favorable way:

- due to random assignment in your experiment, treatment and control are identical in expectation
- due to draft lottery, some young Americans sent to Vietnam and others not?
- due to arbitrary factors, some candidates elected to House of Commons and others not?
- due to an arbitrary population cutoff, some French villages have PR electoral system and others use plurality?

C. Collect data, estimate treatment effects using simple comparisons, show (in)sensitivity of results to specification



A and B can be reversed, but don't tell anyone!

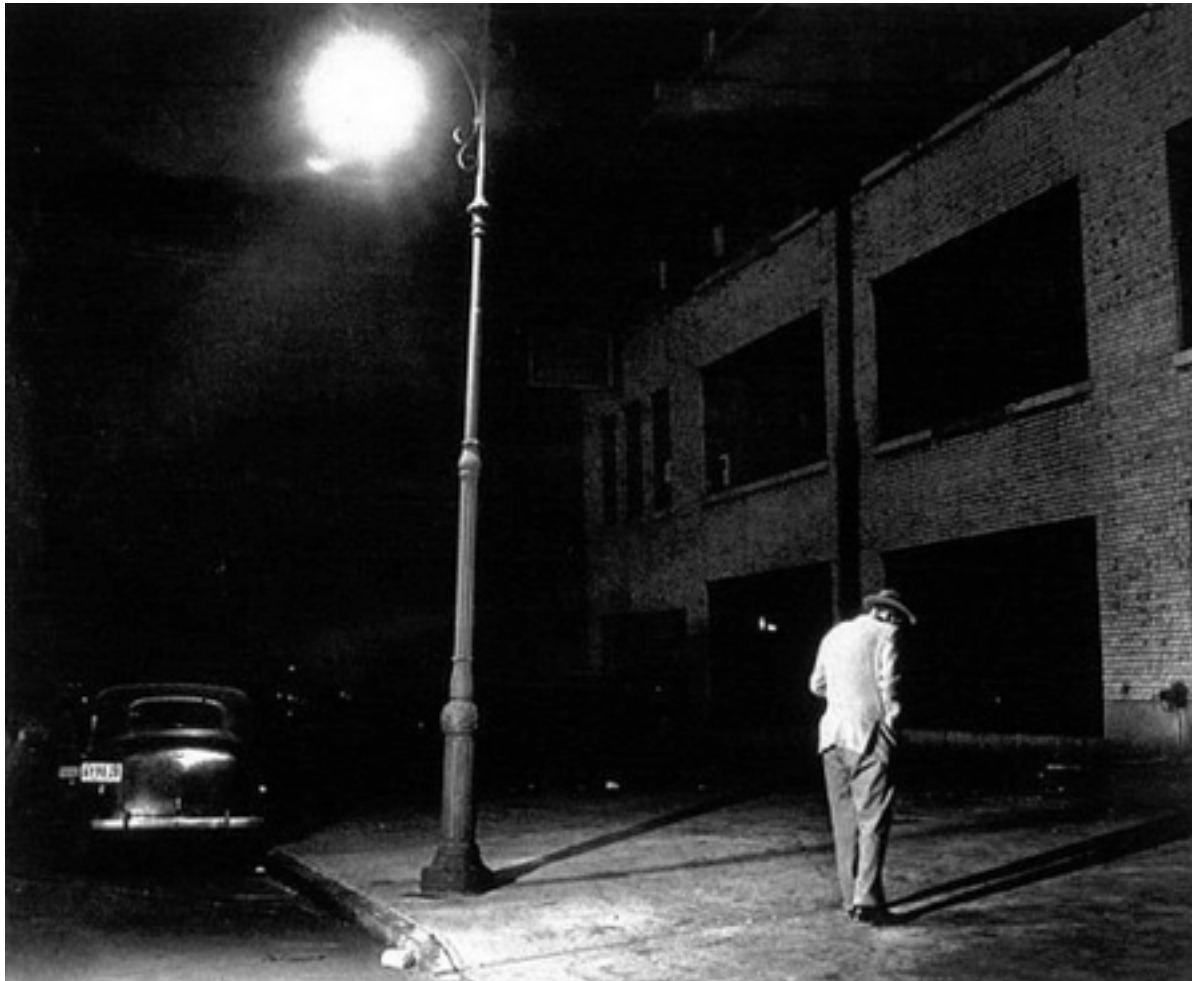
Design vs statistical control

Key feature of design-based approach: choosing settings where statistical control is less necessary.

Research question	Statistical control approach	(Non-experimental) design approach
What is effect of job training program?	Gather data on a bunch of people including participants and non-participants. Regress wages on participation indicator and controls.	Compare participants with non-participants who were barely not eligible.
What is effect of PR (compared to plurality) on turnout?	Gather data on turnout from various countries. Regress turnout on electoral system indicator and controls.	Compare French cities just above and below population cutoff that determines electoral system.

Design vs statistical control

Key feature of design-based approach: choosing settings where statistical control is less necessary.

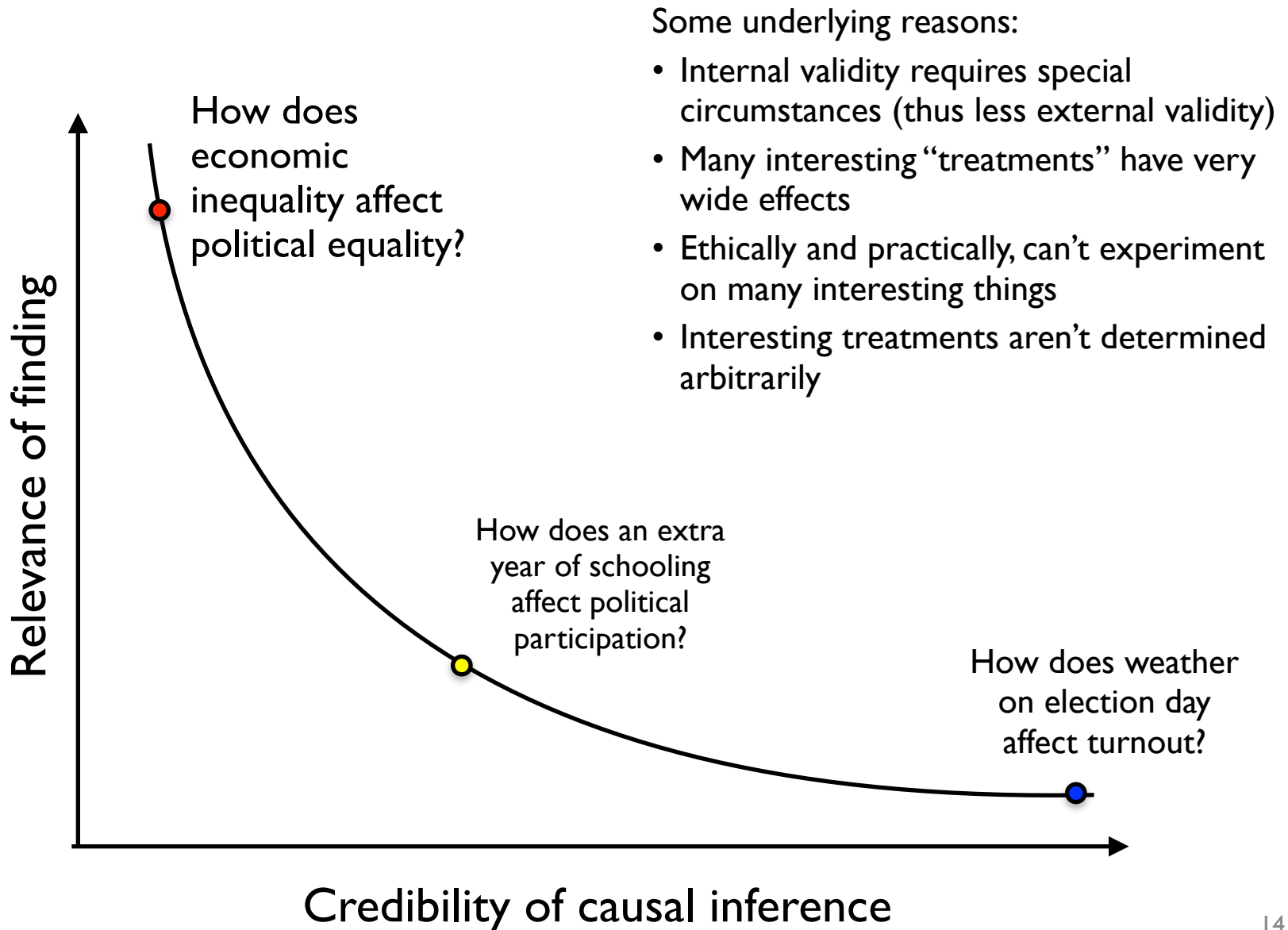


What questions are amenable to the “design-based” approach?

- **Effects of causes:** “what is effect of X”, not “what explains Y”
- Effects of causes that have **local effects:** e.g. “exposure to Fox News”, “PR vs plurality in municipal elections”, not “unipolar international system”
- Effects of causes that **vary in sizable populations** (e.g. “across individuals”, “across municipalities”, not “across hemispheres”)
- Effects of causes that are **binary** (not strictly necessary but helps!)

Most important question: What is the control group?

The credibility/relevance tradeoff



The research design check-list

Research design	Important?	Well-identified?	Data collection feasible?
Time-series cross-sectional regression to measure effect of economic development on democratization	Yes	No	Yes
Exploit randomized candidate order on California ballots to measure effect of ballot order on vote outcomes	No?	Yes!	Yes
Compare attitudes toward gov't just before French revolution in similar areas with different governing arrangements	Yes	Maybe?	Maybe

The big questions about design-based inference and the “credibility revolution”

- Does a study on elections in French villages tell us anything about national elections? (external validity)
- What about explanation? What about theory?
- What about “effects of causes” questions that can’t be answered this way: what is the effect of globalization?

Design-based research and hypothesis testing

Four kinds of “search” to worry about

Specification search: Having chosen an X and Y of interest and a setting, try various control variables, functional forms, etc until you find a significant relationship between X and Y

Treatment search: Having chosen a Y of interest and a setting, run a regression and choose your hypotheses based on what coefficients turn out to be significant/interesting

Outcome search: Having found a setting where X is quasi-randomly assigned, try various outcome variables Y until find a significant relationship

Subgroup search: Having found a setting where X is quasi-randomly assigned, try various subgroups (e.g. young Asian men) until find a significant relationship

Which of these is better with credibility revolution?
Which is worse?

Replication movement and DA-RT

<http://www.dartstatement.org/>



[Petition](#) to delay DA-RT implementation

Petition to Delay DA-RT Implementation

November 3, 2015 [list includes those who signed of November 8 5:15 pm EST]

Dear Colleagues,

We write as concerned members of the American Political Science Association to urge an important amendment to the statement, "Data Access and Research Transparency (DA-RT): A Joint Statement by Political Science Journal Editors." In the joint statement, dated October 6, 2014, journal editors committed their respective journals to a set of principles, to be implemented by January 15, 2016.

DA-RT organizers have made many efforts over the past five years to reach out to members of the profession through various symposia and meetings. However, these issues began to gain widespread attention only when the journal editors signed the statement of October 6, 2014 and panels at the 2015 annual meeting of the American Political Science Association brought the issue to the attention of many scholars who had not realized the possible implications of that statement for their own research, despite the previous outreach activities. Conversations at the panels, roundtables, section business meetings, and other venues at the recent annual meeting demonstrated that members of the Association have only just begun to grapple with the