External and Internal Validity

Political Analysis, Week 7 (MT 2015)
Oxford Q-Step Centre
Prof. David Kirk
Key Questions to Ask about Research

1) Do the data support the authors’ conclusions with respect to the population studied? (Internal validity)

2) If the conclusions are valid, do they generalize beyond the population sampled and the setting studied? (External validity)
Validity in Experimental Design

- **Validity**: degree of support for an inference
  - “the extent to which relevant evidence supports that inference as being true or correct” (Shadish et al. 2002, p. 34)

- **External validity** refers to the generalizability of results (in this case, experimental results); are results applicable to other settings and persons (i.e., to people and places outside the laboratory)?

- **Internal validity** refers to the validity with which one can conclude that the observed relationship (covariation) between an independent and dependent variable reflects a causal relationship (as opposed to spurious).
  - Prior to determining whether there is a causal association between an independent and dependent variable, we must establish that there is even a correlation. In this sense, internal validity relates to the validity of the correlation and then the validity of the causal effect
“Threats” to validity

- Threats are specific reasons why our inferences about the relationship (correlation, causal) between an independent variable and dependent variable may be wrong.

- Researchers should consider the potential threats to the validity of their studies, and describe these possibilities or limitations of their studies.

- But in the event that threats and limitations are not adequately described by authors, readers of research studies should also think about possible threats to validity when evaluating the strength of a study.
External Validity and Generalizability
Generalization Example, Gerber et al (2008)

• In Week 5 we discussed the experiment by Gerber, Green, and Larimer in Michigan on the effect of different mailings – with different levels of social pressure (i.e., neighbours see voting record) – on the likelihood to vote

• But the context was a primary election with very few contested political races (i.e., a lot of people did not likely care about the election). Would the treatment effect be nearly as large (8 percentage point effect) in an election that more people actually cared about?

• Many social scientists argue that given the advent of social media (e.g., Facebook, online dating), that the nature of our social networks is fundamentally different relative to even just 10 years ago. So if our social networks are online, do we even care anymore what our neighbours might think about us?
Recall the Treatment (Gerber et al, 2008)

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

This chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

<table>
<thead>
<tr>
<th>MAPLE DR</th>
<th>Aug 04</th>
<th>Nov 04</th>
<th>Aug 06</th>
</tr>
</thead>
<tbody>
<tr>
<td>9905 JOSEPH JAMES SMITH</td>
<td>Voted</td>
<td>Voted</td>
<td></td>
</tr>
<tr>
<td>9905 JENNIFER KAY SMITH</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9907 RICHARD B JACKSON</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9909 KATHY MARIE JACKSON</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9909 BRIAN JOSEPH JACKSON</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9910 JENNIFER KAY THOMPSON</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9911 BOB R THOMPSON</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9913 BILL S SMITH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9918 WILLIAM LUKE CASPER</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9919 JENNIFER SUE CASPER</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9917 MARIA S JOHNSON</td>
<td>Voted</td>
<td>Voted</td>
<td></td>
</tr>
<tr>
<td>9918 TOM JACK JOHNSON</td>
<td>Voted</td>
<td>Voted</td>
<td></td>
</tr>
<tr>
<td>9918 RICHARD TOM JOHNSON</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9916 ROSEMARY S SUE</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9916 KATHRYN L SUE</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9915 HOWARD BEN SUE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9915 NATHAN CHAD BERG</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9916 CARRIE ANN BERG</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9917 EARL JOEL SMITH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9916 DEBORAH KAY WAYNE</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9917 JOEL R WAYNE</td>
<td>Voted</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Ladd and Lenz study discussed last week examined the effect of media endorsement switches on voting, concluding that readers of papers that switched endorsements to Labour were then 8.6 percentage points more likely to vote for Labour.

The context was British elections from the 1990s, and switched endorsements by just four newspapers, all to Labour. Only 1 of the 4 had a large circulation.

We don’t know if the same findings would hold...

- If a different newspaper had switched endorsement
- If the endorsement switch was from Labour to Conservative instead of Conservative to Labour
- In a different country
- In a different time period; newspaper circulation has declined by nearly 50% in the UK since 2000 and the rise of digital media has transformed the media. Do newspaper endorsements still have the same effect in 2015 as they did in 1997 given this transformed media context?
Recall the Results:

- The red line is the counterfactual: what would have happened with the treatment group if their newspapers had not switched endorsements (slope of red line is the same as Untreated line).

- In the absence of any kind of change in media endorsements, the temporal change in % voting for Labour would be the same for Treated and Untreated groups.

This is the treatment effect: $15.2 - 6.6 = 8.6$
Diverging Findings: Gerber, Karlan, and Bergan (2009)


- Research Design: one month prior to gubernatorial election in Virginia, researchers randomly assigned non-newspaper subscribers to receive a free Washington Post subscription (a liberal-leaning paper), a free Washington Times subscription (a conservative paper), or no newspaper (the controls).

- Results: no effect of either paper on political knowledge, stated opinions, or turnout in election

- However, receiving either paper increased the likelihood of voting for the Democratic candidate, suggesting that media slant mattered less than media exposure.
# Table 4 — Effect of Treatment on Voting Behavior in Virginia Governors Race (OLS)

<table>
<thead>
<tr>
<th></th>
<th>Voted in 2005 election&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Voted in 2005 election&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Voted in 2006 election&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Voted for Democrat (set to missing if did not vote)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Voted for Democrat (set to zero if did not vote)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Separate treatment effects estimated for</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Washington Post</strong></td>
<td>0.001</td>
<td>0.011</td>
<td>0.025</td>
<td>0.112</td>
<td>0.072</td>
</tr>
<tr>
<td>treatment</td>
<td>(0.033)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.045)</td>
<td>(0.035)</td>
</tr>
<tr>
<td><strong>Washington Times</strong></td>
<td>0.005</td>
<td>-0.006</td>
<td>0.031</td>
<td>0.074</td>
<td>0.060</td>
</tr>
<tr>
<td>treatment</td>
<td>(0.033)</td>
<td>(0.019)</td>
<td>(0.020)</td>
<td>(0.045)</td>
<td>(0.035)</td>
</tr>
</tbody>
</table>

*Difference between .112 and .074 was not statistically significant*
While the Ladd and Lenz findings suggest that the endorsement or slant of a paper substantially influences voter preferences (i.e., voters switched from Conservative to Labour), Gerber et al. find a rather weak effect of media slant but a strong effect of media exposure:

- The likelihood of voting for the Democratic candidate increased regardless of whether someone received the liberal (*Post*) or conservative (*Times*) newspaper.
- There was no statistical difference between exposure to the *Post* vs. *Times* in the likelihood of voting for a Democrat.
Accounting for diverging results

- Differences in settings: different country, different election, different newspapers
- Differences in treatment: receipt of a free newspaper (distinguished by type of newspaper) vs. switched endorsement of newspaper
- Rigour of design: natural experiment (Ladd and Lenz) vs. randomized experiment (Gerber et al.)

So what should you believe?

- It is a judgment call
- One option is to think that media endorsements are indeed important for persuading voters, but perhaps only in the UK. If true, then further research should be done on why they matter in the UK but not US
- Another option is to conclude that the results of one study are biased
- Another option is to conclude that the Gerber et al study did not have enough “statistical power” to detect statistical differences in voting among *Times* vs. *Post* readers (more on “power” later…)
What about exposure to television media?


- The authors exploit the natural experiment induced by the timing of the entry of the Fox News Channel in local cable markets and consider the impact on voting.
- They compare the change in the Republican vote share between 1996 and 2000 for the towns that had adopted Fox News by 2000 with those that had not.
- Conditional on a set of control variables, the authors argue that availability of Fox News in a given town is random
- Findings: The entry of Fox News increased the Republican vote share in presidential elections by 0.4 to 0.7 percentage points
Representativeness: The Idea Behind Sampling

1) We seek information about our units of interest (i.e., a population)

2) We observe a selection of these units of analysis (i.e., a sample)

3) We make inferences about the population based on findings from the sample
Why Sample?

- There are vast differences in attributes between the various units (i.e., people, organizations, objects) we study.
- To make inferences about the population (i.e., generalize to the population), we need a representative sample of the population.
- Representative means that the range of variation in the units within the entire population is represented adequately in the sample.
- If a representative sample is used in an experiment, it means that the average causal effect observed in the sample would hold across others (i.e., those not sampled) in the population.
Sampling Designs and Generalizability

A probability sample is one in which each person in the population has a known non-zero probability of selection

- A simple random sample is one type of probability sample, in which each person in the population has an equal probability of selection
- Results can then generalize to the entire population since the sample is representative of the entire population
- Random sampling and random assignment are different: sampling involves selecting cases from the population to be in the study (preferably a representative sample); assignment involves then assigning those cases to different treatment

Two types of non-probability samples are convenience samples and purposive samples

- Convenience: individuals are selected because it is easy to access them (e.g., students)
- Purposive: subjects selected for a good reason tied to purposes of research.
- Results generalizable only to the sample; no way of determining if the results apply to anyone else.
Sampling in Experiments

- Representative samples of the population are generally less common than in observational studies (particularly surveys)

- Random and natural experiments often rely upon convenience samples
  - *Creaming*: sometimes those individuals most ready for a treatment are selected for the experiment (e.g., those unemployed individuals most job-ready are included in an employment program). Effect of the treatment may be far greater than if a representative sample has been used

- Accordingly, the researcher may be limited by the extent to which s/he can claim that inferences generalize to the wider population (or other countries).
Experiments are generally strong on internal validity, but weaker on external validity. In contrast, observational studies, particularly those done with representative samples, are generally weaker than randomized experiments on internal validity but stronger on external validity.

- **External validity** refers to the generalizability of results.
- **Internal validity** refers to the validity with which one can conclude that the observed relationship (covariation) between an independent and dependent variable reflects a causal relationship (as opposed to spurious).

Good researchers systematically consider the various threats to the validity of their results. Consumers of research should also weigh the threats when evaluating research evidence.