Ethics in Social/Political Science Research

POSC 3410 – Quantitative Methods in Political Science

Steven V. Miller

Department of Political Science

Goal for Today

Discuss some issues of ethics and replication in social science research.

Going Forward

This semester you learned:

- Concepts and measurements
- Descriptive infernece
- Random assignment/experimental design
- OLS regression
- What to do when your DV isn't continuous/interval
- Some advanced topics on top of that (e.g. simulation, Bayes)

Ethics and Replication

I can't make you do these things in good faith...

- Social science is rife with cases of academic misconduct.
- Publication incentives breed dishonesty; you are compelled to rise above it.

...but I can teach you some tools to help you be honest.

- i.e. this is academic workflow and replication.

Some Ethical Takeaways

Your theoretical model is causal. Your empirical model may not be.

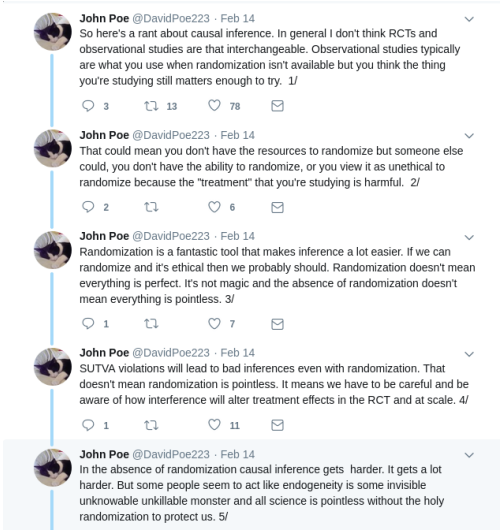
- Remember: everything is a "model."

But don't shirk from using causal language!

- Absent a causal drive, the aim of the research is directionless/vague.

There's an unnecessary tension between the RCT people and those doing observational analyses.

- Be forthright, but stand your ground.
- Again: your theoretical model is causal. Your empirical model may not be.



John Poe @DavidPoe223 · Feb 14

So here's a rant about causal inference. In general I don't think RCTs and observational studies are that interchangeable. Observational studies typically are what you use when randomization isn't available but you think the thing you're studying still matters enough to try. 1/

3 13 78

John Poe @DavidPoe223 · Feb 14

That could mean you don't have the resources to randomize but someone else could, you don't have the ability to randomize, or you view it as unethical to randomize because the "treatment" that you're studying is harmful. 2/

2 6

John Poe @DavidPoe223 · Feb 14

Randomization is a fantastic tool that makes inference a lot easier. If we can randomize and it's ethical then we probably should. Randomization doesn't mean everything is perfect. It's not magic and the absence of randomization doesn't mean everything is pointless. 3/

1 7

John Poe @DavidPoe223 · Feb 14

SUTVA violations will lead to bad inferences even with randomization. That doesn't mean randomization is pointless. It means we have to be careful and be aware of how interference will alter treatment effects in the RCT and at scale. 4/

1 11

John Poe @DavidPoe223 · Feb 14

In the absence of randomization causal inference gets harder. It gets a lot harder. But some people seem to act like endogeneity is some invisible unknowable unkillable monster and all science is pointless without the holy randomization to protect us. 5/

John Poe @DavidPoe223 · Feb 14

Endogeneity is not something that our minds can't comprehend without shattering. It's not Cthulhu. If we take it seriously and approach it logically we can break it down. We can build ways to deal with it. 6/

1 12

John Poe @DavidPoe223 · Feb 14

We've already broken endogeneity down into separate sources in the literature: relevant omitted variables, measurement error, self-selection, simultaneity, and dynamic effects modeled with lags or leads. These can all cause bias in treatment effects. 7/

1 18

John Poe @DavidPoe223 · Feb 14

We can approach each of these problems one at a time and try to solve them on a case by case basis. Is it easy? No. Is it always possible? No. But if it's possible even 1% of the time then that's better than if we threw our hands up and said that the truth is unknowable. 8/

1 13

John Poe
@DavidPoe223

Following

Would it be easier to randomize? Yes. But we probably end up learning less in the long run because randomization won't ever tell us what the sources of endogeneity were. If we end up knowing that we have a better model of the DGP as a whole instead of just the ATE or ATT 9/

4:23 PM - 14 Feb 2020

The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data

Causal inference is a core task of science. However, authors and editors often refrain from explicitly acknowledging the causal goal of research projects; they refer to causal effect estimates as associational estimates.

This commentary argues that using the term "causal" is necessary to improve the quality of observational research.

Specifically, being explicit about the causal objective of a study reduces ambiguity in the scientific question, errors in the data analysis, and excesses in the interpretation of the results. (*Am J Public Health*. 2018;108: 616–619. doi:10.2105/AJPH.2018.304337)

Miguel A. Hernán, MD, DrPH

See also Galea and Vaughan, p. 602; Begg and March, p. 620; Ahern, p. 621; Chiolero, p. 622; Glymour and Hamad, p. 623; Jones and Schooling, p. 624; and Hernán, p. 625.

You know the story:

> Dear author: Your observational study cannot prove causation. Please replace all references to causal effects by references to associations.

Many journal editors request authors to avoid causal language,[1] and many observational researchers, trained in a scientific environment that frowns upon causality claims, spontaneously refrain from mentioning the C-word ("causal") in their work. As a result, "causal effect" and terms with similar meaning ("impact," "benefit," etc.) are routinely avoided in scientific publications that describe nonrandomized studies. Instead, we use terms like

Confusion then ensues at the most basic levels of the scientific process and, inevitably, errors are made.

We need to stop treating "causal" as a dirty word that respectable investigators do not say in public or put in print. It is true that observational studies cannot definitely prove causation, but this statement misses the point, as discussed in this commentary.

OF COURSE "ASSOCIATION IS NOT CAUSATION"

glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.

The risk ratio of 0.8 is a measure of the association between wine intake and heart disease. Strictly speaking, it means that drinkers of one glass of wine have, on average, a 20% lower risk of heart disease than individuals who do not drink. The risk ratio of 0.8 does not imply that drinking a glass of wine every day lowers the risk of heart dis-

Some Ethical Takeaways

The proliferation of machine learning/AI/"algorithms" creates more ethical issues.

1. "Treat", don't manipulate.
2. There's no bias-free model; *you are the bias.*
3. Evil is evil, whether intentional or unintentional.

Andrew Heiss (Georgia State), re: third point: "don't let stupidity transform into evil."

- Good academic workflow can help.

THE WALL STREET JOURNAL.

Blue Feed, Red Feed

See Liberal Facebook and Conservative Facebook, Side by Side

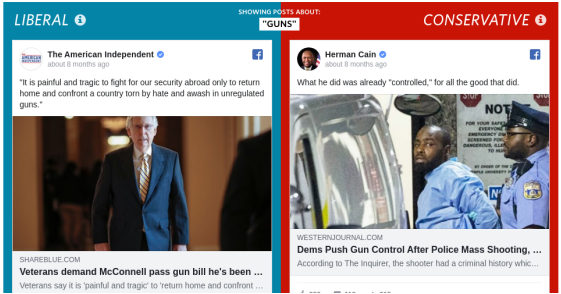
By **Jon Keegan**

Published May 18, 2016 at 8:00 a.m. ET | Last Updated Aug. 19, 2019

This page has been archived and will no longer be updated.

FILTER ARCHIVED FEEDS BY TOPIC:

PRESIDENT TRUMP | HEALTH CARE | GUNS | ABORTION | ISIS | BUDGET | EXECUTIVE ORDER | IMMIGRATION



NEWS · 24 OCTOBER 2019 · UPDATE 26 OCTOBER 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals — and highlights ways to correct it.

Heidi Ledford

Numbers don’t always tell the truth

Mark J. Girouard, an employment attorney at Nilan Johnson Lewis, says one of his clients was vetting a company selling a resume screening tool, but didn’t want to make the decision until they knew what the algorithm was prioritizing in a person’s CV.

After an audit of the algorithm, the resume screening company found that the algorithm found two factors to be most indicative of job performance: their name was Jared, and whether they played high school lacrosse. Girouard’s client did not use the tool.

Q. If machine learning is so smart, how come AI models are such racist, sexist homophobes? A. Humans really suck

Our prejudices rub off on our computer pals, sadly

By Katyanna Quach 5 Sep 2019 at 07:02 64 SHARE ▼

Academic Workflow and Replication

Replication crises/academic misconduct are proliferating in social science. Examples:

- Economics: Reinhart and Rogoff's (2010) Excel error
- Psychology: too many to list
  - Recurring themes: small-*n*, *p*-hacked experiments, or even fabricated data
- Sociology/criminology: Stewart retractions
- Political science: Lacour and Green (2014) scandal

I'm not going to assign motives (naiveté or something worse) to these scandals and those involved.

- But, assuming honesty, you can avoid a similar pitfall with good workflow.

Some Tips on Good Workflow/Replication

"Kondo" your projects into sub-directories.

- Keep things tidy/de-cluttered in your project.
- I have my recommendations, but tweak for what works for you.

"Launder" your data; never overwrite them.

- Never overwrite original columns. Recode into new columns/objects.
- *Definitely* never overwrite raw data.

Related: invest in cloud storage (e.g. Box, Dropbox).

- Create separate folders for raw data (`data`) and your individual projects (`projects`).
- Tongue in cheek: think of "my laptop broke/fried/got stolen" as the 21st century equivalent of "the dog ate my homework."

An Example of Sub-Directories

```
my_project_name
+-- _cache
+-- _dross
+-- abstract
+-- appendix
+-- cover-letter
+-- data
|   +-- data.rds
|   +-- models.rds
|   +-- sims.rds
+-- doc
+-- figs
+-- presentation
+-- readings
+-- src
|   +-- 1-load.R
|   +-- 2-clean.R
|   +-- 3-analysis.R
|   +-- 4-sims.R
|   +-- 5-create-tabs-figs.R
+-- my_project_name.Rproj
+-- my_project_name.Rmd
+-- .gitignore
+-- README.Rmd
```

Some Tips on Good Workflow/Replication

- Surprise! You're a computer programmer now. Embrace it.
  - Make comments to yourself in code.
  - Make your steps apparent, something you can't do in Excel.
- Make your document "dynamic."
  - Learn R Markdown. You'll thank me later.
- REQUIRED: Share your work.
  - I recommend Github. Will also help with version control.
  - Don't be surprised when you get asked to share what you did.

How to improve your relationship with your future self*

Cómo mejorar su relación con su futuro yo

JAKE BOWERS
Universidad de Illinois

MAARTEN VOORS
Wageningen University

ABSTRACT

This essay provides practical advice about how to do transparent and reproducible data analysis and writing. We note that doing research in this way today will not only improve the cumulation of knowledge within a discipline, but it will also improve the life of the researcher tomorrow. We organize the argument around a series of homilies that lead to concrete actions. (1) Data analysis is computer programming. (2) No data analyst is an island for long. (3) The territory of data analysis requires maps. (4) Version control prevents clobbering, reconciles history, and helps organize work. (5) Testing minimizes error. (6) Work *can* be reproducible. (7) Research ought to be credible communication.

Key words: research transparency, reproducible research, workflow, methodology

Conclusion

I hope you learned a fair bit semester.

- Everything from concepts to regression and beyond.

You must act in good faith, no matter perverse incentives.

- Be mindful of emerging ethical issues (esp. in machine learning context).
- Be prepared to 100% show the world how you did what you did.

Don't make future you begrudge past you's incompetence or dishonesty.

Table of Contents