Reproducibility to Robustness: Proposals for new tools to increase the policy utility of reproducibility practices

Jake Bowers
Senior Scientist, The Policy Lab
Methods Director, Evidence in Governance and Politics network (EGAP)
Political Science and Statistics, University of Illinois @ Urbana-Champaign

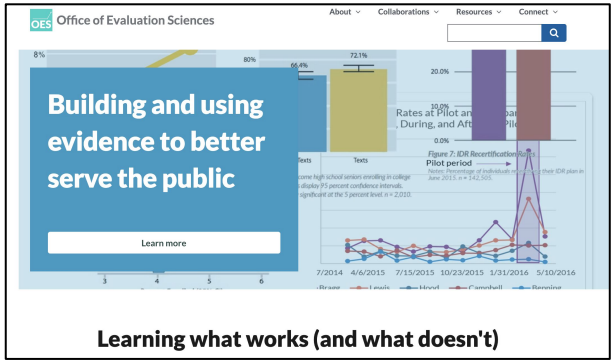
Kevin H. Wilson
Head of Data Science, The Policy Lab

Toronto Workshop on Reproducibility
February 25, 2021

BROWN
The Policy Lab

Context: Some Policy Labs Connecting Academics to Policy Problems

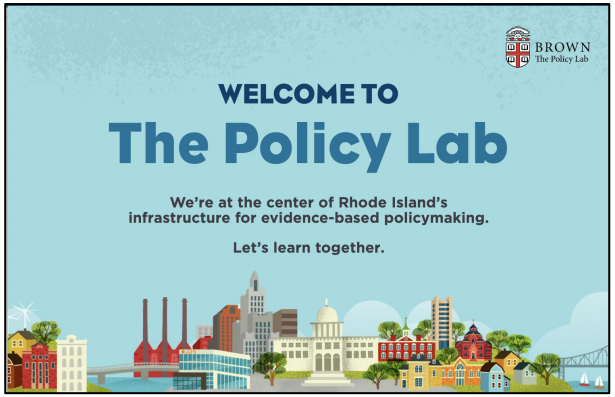
The US Office of Evaluation Sciences
(est 2014)



TThe Lab@DC
(est 2016)



The Policy Lab
(est 2018)

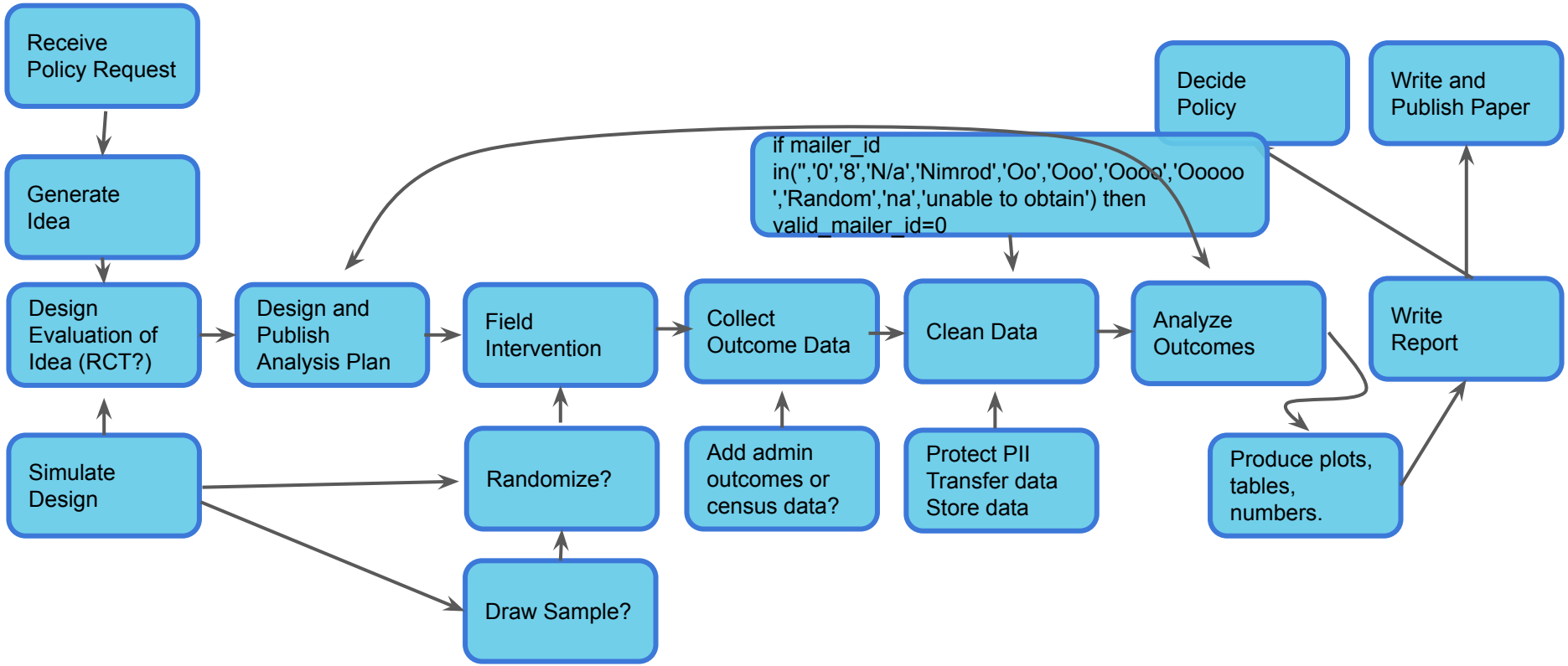


Foreshadowing: What do we need to enhance robust evidence-based policy making?

- Tools and norms for **participatory pre-analysis planning** with non-technical participants.
- Tools for **blind re-analysis** while protecting privacy with large administrative data sets and/or **unit testing of data tasks.**
- **Build systems** better than our current combo (Something better than make+poetry(for python)+renv(for R)+unix shell tools)
- Tools to **automate data cleaning** with complex administrative data.
- A Google Docs-alike or Overleaf-alike **editing environment** for **easy writing** of text and **collaboration** across big diverse teams where figures and tables and numbers can be linked **without any copy and pasting**. (Not quite Overleaf).

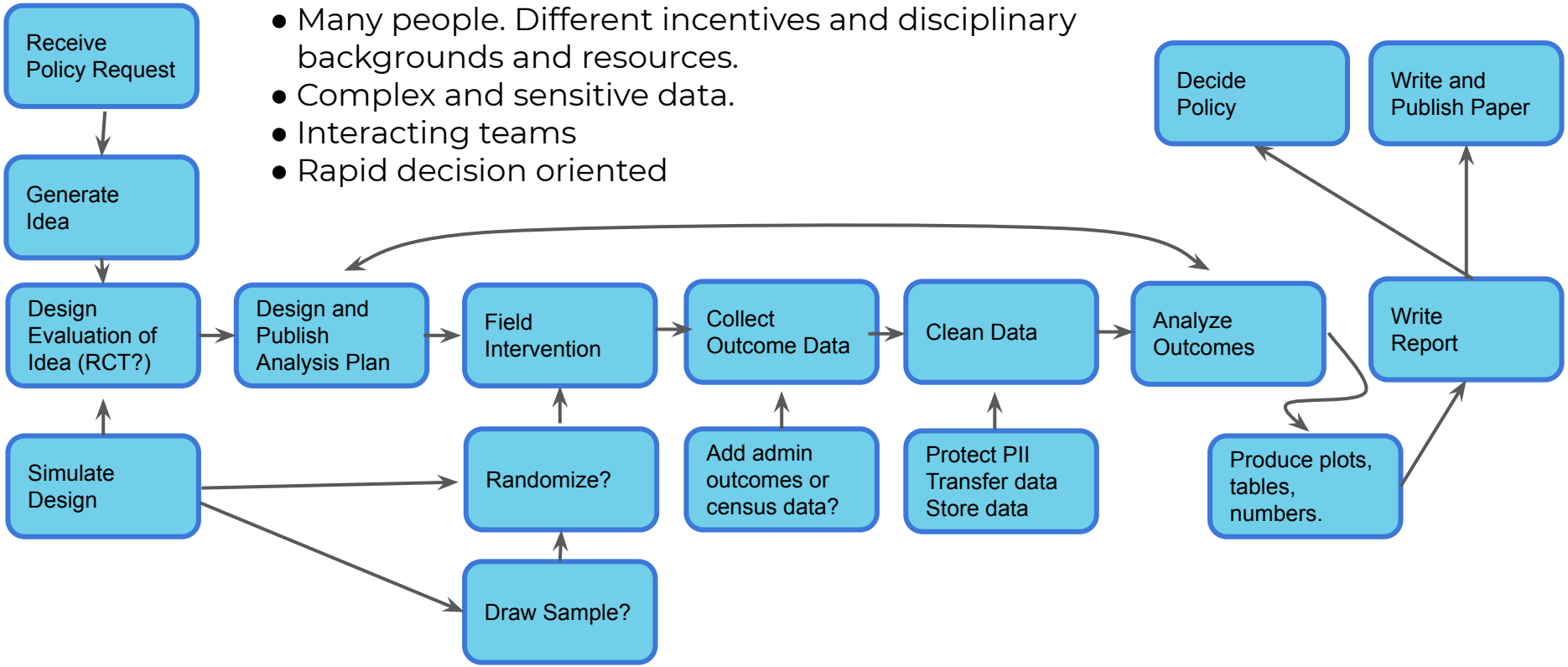
Overall: We want something like “robustness” --- a decision made based on one of our plots or numbers or tables should do not harm OR at least, we should know quickly that it is doing harm so that we can minimize harm..

It looks so easy

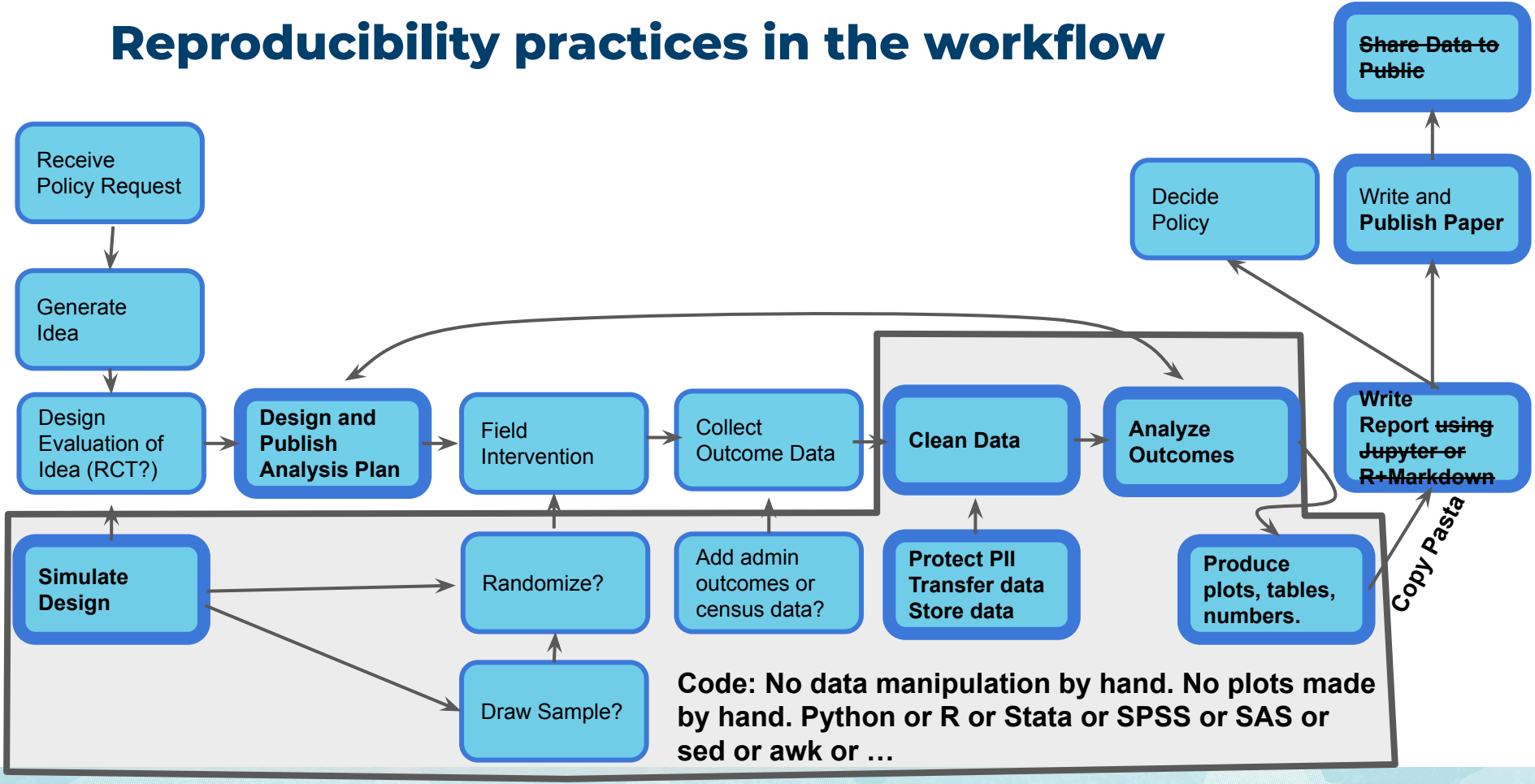


It looks so easy and there are so many meetings!

- Many people. Different incentives and disciplinary backgrounds and resources.
- Complex and sensitive data.
- Interacting teams
- Rapid decision oriented



Reproducibility practices in the workflow



Public pre-analysis plans in science vs policy

Academia	Policy Labs
Ease interpretation of statistical results.	Ease interpretation of statistical results.
Address file-drawer problem	Address file-drawer problem
	Stakeholder involvement can increase actionability and quality of debate; bring "politics" back into policy. (Yokum)
	Forces interdisciplinary decisions about stats.
	Addresses criticisms of policy-based-evidence. (No hiding work. Public commitment to analyses.)

PAP Questions: How to involve the public and stakeholders in our data work? How to manage the politics?

Reproducibility practices in science help over the long run.

Academia	Policy Labs
Over years, the community corrects errors via **peer review** and professional incentives and **public data**.	Time to decision less than 1 year. Data private and protected and sensitive. Professional incentives to move fast.
Consequences of errors limited to a few careers (if at all).	Errors can affect entire populations.
Publications increasingly enforce reproducibility practices: data, code, pre-analysis plans.	No external incentives to reduce errors.

Code Questions: What workflow is better than executing Jupyter/R+Markdown code chunks by hand (and maybe out of order? Maybe using cruft from old projects in memory?) (Or running entire files by hand (and maybe out of order, maybe with old versions or old versions of data)? Or copying and pasting results into Google Docs?

Statistical inference as commonly practiced not designed for robust and rapid decisions.

Academia	Policy Labs
P-values and Posteriors with weak priors summarize evidence in this design about the theory of change. But never meant to guide direct decisions.	P-values and Posteriors must guide immediate decisions. Costs of errors rarely (if ever) built into α or priors.
Studies end by handing off to another person or team (or future versions of existing person or team).	Studies inform ongoing relationships between government and public.

Stats Question: Should we be a) including costs into our calculations? and/or b) monitoring the results for changes in context or results? (Phase 4 trials?) and/or c) attending more to priors (from policy experts, public, academics)?

What do we need to enhance robust evidence-based policy making?

- Tools and norms for **participatory pre-analysis planning** with non-technical participants.
- Tools for **blind re-analysis** while protecting privacy with large administrative data sets and/or **unit testing of data tasks.** (Not allowed to report a plot without a unit-test? A re-analysis?)
- **Build systems** better than our current combo (Something better than make+poetry(for python)+renv(for R)+unix shell tools)
- Tools to **automate data cleaning** with complex administrative data.
- A Google Docs-alike or Overleaf-alike **editing environment** for **easy writing** of text and **collaboration** across big diverse teams where figures and tables and numbers can be linked **without any copy and pasting**. (Not quite Overleaf).
- What is the role of **monitoring** in evidence-based public policy making?

Overall: We want something like “robustness” --- a decision made based on one of our plots or numbers or tables should do not harm OR at least, we should know quickly that it is doing harm so that we can minimize harm.

Thank you.

Interested in talking more?
Pitching in?
jake@jakebowers.org

thepolicylab.brown.edu

Questions?
Thoughts?

thepolicylab.brown.edu

A few other efforts to enhance evidence-based public policy

- Evidence in Governance and Politics (EGAP)
- Causal Inference for Social Impact Lab at CASBS
- Research4Impact (Org to spur cross-sectoral collaboration)
- The Office of Evaluation Sciences (OES) at the USA GSA
- The Lab@DC
- DeclareDesign (R package for design simulation)

See also:

- Bowers and Voors. 2016. “How to improve your relationship with your future self (V2.0)”