

Data for Research on Science Funding

Funders Forum &
RoRI Consortium meeting
October 2021

Adam B Jaffe
Professor Emeritus, Brandeis University



Research on Research Funding

- Program design choices affect the outcomes of program activities.
- We'd like to be able to choose those design options that best achieve program objectives.
- To make such choices systematically, we need evidence on how outcomes differ under different program design choices.
- Social scientists have thought a lot about what kind of data is needed to develop this evidence.

Evidence

- Understanding how and to what extent a program generates outcomes inherently involves comparing observed outcomes with some **counterfactual**.
- A U.S. National Science Foundation program officer once told me, “We know our program is crucial for U.S. science because 80% of the Nobel Prize winners in XXXX were funded by our program.”
 - If you were presented with a research proposal that displayed this kind of reasoning, would you fund it?
- Just as we judge a new drug by comparing outcomes for patients who get it to outcomes for patients who don't get it, we need some kind of comparison group to judge how researchers who are **'treated'** by our science program are affected by that program.

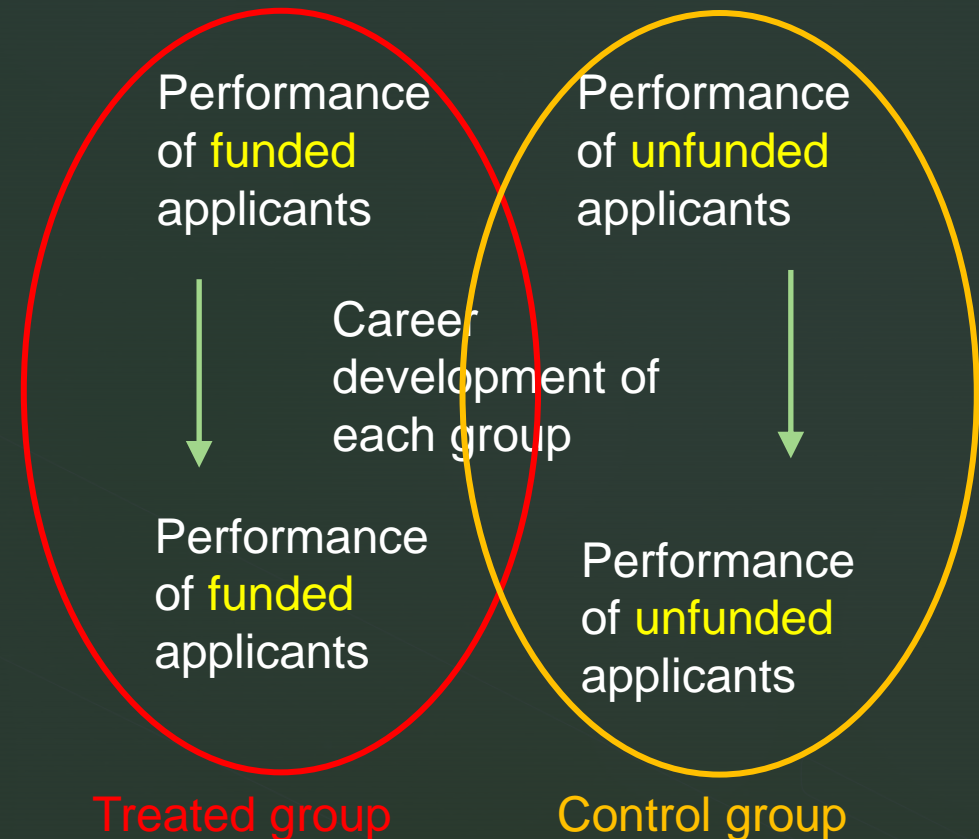
Treated and Control Groups

- The 'Gold Standard' for collecting data on treated and control groups is an experiment, or Randomized Controlled Trial ("RCT"). (more on these in a minute)
- But ongoing program activities can often be used as 'quasi-experiments', **if appropriate data on their operation is collected and retained.**

Counterfactual analysis in a quasi-experiment

Before program participation

After program participation



Implication: data is needed on four different performance observations

Data implications

- Data is needed on all applicants regarding their performance **prior** to their interaction with the program
 - Can be collected as part of application, or can be merged in from external sources, if appropriate identifying information (e.g. ORCID number) is collected
- Data on **unfunded** applications must be retained to provide the control group needed for counterfactual analysis
- If a numerical score or ranking is used as part of the selection process, retention of these scores allows statistical analysis to deal with the non-random assignment of applicants to the control and treated groups.



Other considerations

- **Privacy:** all applicant data must be collected, maintained, shared and analyzed in ways that respect legal and personal concerns about confidentiality and privacy
- This is a solvable problem. Health research routinely involves statistical analysis of individuals' health conditions and health outcomes. These are among the most legally protected and personally sensitive data people have.
 - appropriate permissions
 - appropriate structures for confidentiality and potentially anonymization
- **Other data on processes:** e.g. understanding how referee bias affects selection requires retention of demographic information on referees



The need for pre-planning

- Many of these identified data needs can be fulfilled only if the data needs are built into program design in advance.
- Ongoing program operation can provide quasi-experiments for program evaluation, but only if that evaluation is anticipated so that needed data are collected, retained, and enveloped in permissions and confidentiality structures to protect privacy.



From quasi- to real experiments

- Suppose your reading of research on unconscious bias leads you to decide to switch to blind reviewing of applications.
- This is a potentially good idea, but you don't really know. Maybe with unblind reviewing referees consciously compensate for unconscious biases they know to be in the system, and when you blind the process this goes away leaving behind unconscious biases that are still triggered by implicit differences in the language of different genders or different groups.
- So instead of switching wholesale to the new approach, conduct an experiment. Evaluate some applications with the new procedure and some with the old procedure.

Objections to experiments

“It’s unfair to evaluate some applications with one method, and some with a different method”

- How is this unfair? Some applicants are exposed to a procedure that you believe to be better (though you don’t really know), and others are exposed to what was until very recently the standard procedure. Surely this is no more problematic than giving some people the new vaccine and some people a placebo?

“One program experimenting in this way will generate too little data to be meaningful”

- Some data has to be better than no data.
- If the data do not yield statistically reliable conclusions, they will still allow for a qualitative picture of what is going on, particularly if the whole effort is structured to yield as much insight as possible.

“It will be embarrassing if we learn that we went to a lot of trouble to make a change that has no effect”

- Isn’t it more embarrassing to make a change and not be able to demonstrate if it is valuable?

Programmatic experiments create multiple treated/control comparisons

Performance of **funded** applicants reviewed blind

Performance of **unfunded** applicants reviewed blind

Performance of **funded** applicants reviewed non-blind

Performance of **unfunded** applicants reviewed non-blind



Career development of each group



Do the funded applicants from the blind process look different than the funded applicants from the non-blind process?

Is the difference between the funded and non-funded applicants bigger or smaller for the blinded versus non-blinded process?

Parting Thoughts



- The perfect is the enemy of the good.
- Data needs for evaluation should be built into program design up front
- Don't miss opportunities to use program improvements/changes as experiments
- Develop a long-term culture of evaluation
 - all information is 'data'
 - privacy protocols become routine
 - think about sharing data across organizations