Testing Social Policy Using Experiments: Lessons from the United States

Judith M. Gueron, MDRC

Colloque sur les Expérimentations pour les Politiques Publiques de l'Emploi et de la Formation, 23 May 2008, Paris

Experimentation

- A research method

- A vision of how to make government more rational and effective

The Welfare Story: 40 Years of Experiments

1. How did this happen?
2. What does this teach us about research methods?
3. What does it tell us about what works?
4. Did these studies affect policy?
5. What are the big lessons?

Evaluation Questions

1. Was the program well implemented?
2. **Did the program achieve its goals?**
3. Are costs reasonable in relation to achievements?
4. Do the answers vary for subgroups, approaches, and local conditions?
5. What are the lessons for policy and practice?

Defining Outcomes & Impacts

- **Outcomes**: status of people at a specific time
  - E.g., percent working, percent not poor

- **Impacts**: difference between the outcomes which did occur and the outcomes which would have occurred without the program

Key Methodological Challenge

- Getting a reliable estimate of what people would have done on their own (called the "counterfactual") in order to determine what the program really accomplished

- The counterfactual is a moving target

What Causes Change?
Avoiding Chantecler's Reasoning

- Most reliable method: a lottery (random assignment) to program and control groups

- "Experiment" ≡ Random assignment study

U.S. Context

1. Desire for change

2. Multiple sources of innovation and money

3. Demand for proof

Demand for Proof

1970s: Dark Ages, no answers to basic questions
1. Do welfare-to-work programs have any effect?
2. Do impacts vary by subgroups or approaches?
3. Do goals trade off: e.g., ↑ work vs. ↓ poverty?
4. Are results replicable?
5. Can you get believable answers/do experiments?
6. Can you do this on a large scale?
7. Will high quality information matter?

Now have some answers to all of these questions

Example #1: National Supported Work Demonstration (1974)

- First random assignment test of multi-site employment program

- Offered up to 18 months of paid work to long-term welfare recipients, former drug addicts, people leaving prison, and young school drop outs

- Goal: Long-term change in behavior

How Did This Happen?

- Prior studies ended in dispute on methods

- Learn whether it would work and what it would cost at small scale, before proposing costly national program

- Goal: build a record so that knowledge could cumulate

Lessons on Methods

1. Feasibility of random assignment

2. Feasibility of survey strategy

3. Value of transparent, relevant impact measures

Lessons About What Works

- Surprise: positive impacts for women on welfare, not primarily-male other groups

- Employment programs *can* work, but all "good ideas" don't and may even do harm

- People with high outcomes may have low impacts

Did the Study Affect Policy or Practice?

- Null findings → avoid spending on ineffective programs. Confirmed value of testing before passing a law

- Positive findings for welfare recipients did not lead to expansion

Lessons for Other Fields

- Value of centralized control of funding and design

- Reanalysis showed nonexperimental research designs would have yielded incorrect conclusions

Example #2: Study of State Work/Welfare Initiatives (1980s)

President Reagan's election a turning point

- Welfare policy → more conservative
- States given freedom to test work requirements
- Federal government stopped funding most social policy research

Result

- Prospects for experiments looked bleak
- Surprise: Experiments flowered, led to new decentralized approach that had greater policy impact

How Did This Happen?

Ford Foundation funded the study

Controversial issues → used random assignment

- Anticipated modest impacts so needed large sample (35,000 people)

- Used existing administrative records

Experiments had never been done before at this scale, in operating welfare offices, independent of Washington, offering no special funds

Why Did States Participate?

- Sought states with large programs, useable data, representative of nation, willing to risk backlash and negative findings

- Reasons states joined
  - ➤ Answer their questions
  - ➤ Get assistance on program design
  - ➤ Gain visibility

Lessons on Methods

1. Feasible to conduct experiments in regular welfare offices and not disrupt operations

2. Feasible to use existing administrative records to follow people and get accurate estimates of impacts

Lessons About Mandatory Welfare-to-Work Programs

- Increased work and reduced welfare

- Did not hurt children

- Average impacts small to modest

- Did not reduce poverty (program rules and wages in U.S.)

- Saved money

Did the Studies Affect Policy or Practice?

Widespread agreement: studies changed opinions, program design, and laws because of

1. Technical strength of random assignment
2. Replication across multiple sites
3. Timing and relevance of programs tested
4. Convincing scale
5. Active communication; shared good & bad news
6. Less partisan context

Example #3: The Next 10 Years

Many experiments. U.S. government insisted on random assignment to prove budget neutrality

Findings

1. Outcome standards can → inefficient decisions

2. No policy did best on all goals. "Work First" strategies increased work, saved money, did not harm children, but did not reduce poverty. Earnings supplements increased work, reduced poverty, benefited children, but cost more

3. Other methods failed to replicate experiments

12 Lessons from the U.S. Experience

1. Address important issues

2. Have a reasonable treatment

3. Design a real-world test

Lessons, continued

4. Address the questions people care about

5. Fight for random assignment

6. No single experiment is definitive

Lessons, continued

7. Do not define success as working miracles or you are likely to fail

8. Simplify

9. Actively communicate the results

Lessons, continued

10. Do not confuse dissemination with advocacy

11. Be honest about failures

12. Get partners and buy-in from the beginning