

Replication Issues in Social Experiments

Burt S. Barnow

George Washington University

David Greenberg

University of Maryland Baltimore County

**Prepared for the Institute for Employment Research International
Conference on Field Experiments in Policy Evaluation,
Nuremberg, Germany October 18 and 19, 2012**

Topics Covered

- Why replicate?
- Why replications are often unsuccessful
- Examples of replications
 - Income maintenance experiments
 - Job clubs
 - Unemployment insurance bonus experiments
 - Job search experiments
 - Center for Employment Training
- Improving replication efforts

Why Replicate?

- Larger sample size
- Additional target groups
- Additional geographic locations
- Different intervention parameters
- Additional related treatments

Why Replications Are Often Unsuccessful

- Insufficient understanding of what made the original intervention successful
- Insufficient care and resources devoted to the quality of implementation and the process of scaling up
- Insufficient attention to the culture within the helping organization and the regulatory and systems context surrounding it
- Insufficient attention to local capacity and the organizational environment within which the intervention must be sustained

Why Replications Are Often Unsuccessful

- Failure to understand that what works for most children and families may not change outcomes for the children and families who are most at risk
- Failure to understand the “uptake problem” among local front-line personnel and supervisors
- Funders’ reluctance to devote significant sums to the substantial operational costs of scaling up.
(Schorr & Farrow)

Examples of Replications

Income Maintenance Experiments

- Original IME was NJ IME used to test “negative income tax,” then popular welfare approach
- Operated from 1968-1972 in 5 cities
- Included 1,357 families and included 4 guarantee levels and 3 implicit tax rates
- Replications
 - Rural areas in 2 states 1969-1972
 - Gary, Indiana 1971-1975
 - Seattle and Denver 1971-1978

Income Maintenance Experiments

- Findings were similar among sites and as expected: 7% and 17% reductions in labor supply for husbands and wives
- Replications conducted before initial experiment completed—why?
 - HEW wanted to run own experiments
 - HEW staff thought they could “do it better”
 - Replications were larger and tested additional items (training vouchers, child care)

Job Clubs

- Prior to a series of experiments by Azrin, it was believed that job search should be done on an individual basis
- In first experiment, 120 job seekers randomly assigned with 90% of treatment group employed in 2 months compared to 55% of controls
- Azrin & Philip (1979) tested approach on disabled population with stronger results: 95% of T employed at 6 months compared to 28% of C
- With funding from DOL, Azrin et al. (1980) had 1,000 welfare recipients randomly assigned to job club or control status with 87% of T and 59% of C employed at 12 months
- DOL replicated the welfare population study in Louisville with similar findings
- Job clubs are now the established approach to job search assistance—not only are they more effective, they are cheaper

Unemployment Insurance Bonus Experiments

- Premise was that offering cash bonus to UI claimants who found job quickly would shorten claims and save money
- First experiment in IL offered claimants \$500 if they found job within 11 weeks and remained employed 4 months
- Experiment also had employer experiment where employers received the bonus, but it had low participation and no statistically significant impact and was not replicated
- IL claimant experiment had 4,186 in treatment group and 3,952 in control group

Unemployment Insurance Bonus Experiments

- IL experiment found duration reduced by 1.15 weeks and benefits reduced by \$194 per claimant, with no significant reduction in earnings
 - Reductions are large in aggregate with 9M claimants annually projected through 2017
- IL program saved lots of money, but in part due to low take-up rate
- Replications conducted in 3 states: NJ, PA, WA
- Replications had more treatment variations, and NJ included job search assistance

Unemployment Insurance Bonus Experiments

- Replications had similar sample sizes and structure as in IL
- Results in replication states not as encouraging
 - Marginal impact of bonus in NJ was .4 weeks
 - Results in PA and WA generally not statistically significant except for most generous plans
- Bonus concept lost favor due to disappointing results in replication
 - Meyer (1995) notes that large savings depended in large part on low take-up rate, unlikely to be sustained in permanent program

Job Search Experiments

- These experiments started in the 1970s and 1980s and combined stricter enforcement of the work test with job search assistance
- The Charleston, SC experiment had 4,247 claimants assigned to 3 treatment groups and 1,428 in the control group
- The treatments reduced UI by .5 to .75 weeks and saved \$46-\$56/week

Job Search Experiments

- Similar experiments were conducted in New Jersey and Washington with similar findings
- Three other states conducted their own experiments, with similar findings but weaker designs (WI, NV, MN)
- The Department of Labor began funding grants to provide these Reemployment and Eligibility Assessment (REA) services in 2005
 - DOL currently spends \$65.5M in 40 states
 - States supposed to develop random control group, but done with limited success
 - Evaluation in 2012 found reductions in weeks claimed of 1.1-3.0 weeks in 3 states but no impact in one state that did not implement the program well
- The mix of “carrot” and “stick” appears to work better than either alone
 - Bonus experiments have limited effectiveness
 - Ashenfelter et al. (2004) evaluation of stick only experiments found no benefits to states or workers

Center for Employment Training

- CET established in 1967 in San Jose, CA and provided workforce services and training to welfare recipients, agricultural workers, dropouts, etc.
- In the 1980s, CET participated in two separate multiple site RCTs with very impressive findings:
 - JOBSTART was a workforce demo for disadvantaged youth implemented in 13 sites
 - The Minority Female Single Parent demo provided comprehensive employment and training services to female parents in 4 sites
 - In both projects CET had very strong, statistically significant impacts on employment and earnings, and none of the other sites had positive, significant findings
 - In JOBSTART, earnings for the T group were \$14,271 higher for yrs 3 and 4
 - In MFSP, earnings for the T group were \$2,000 higher for 30 months

Center for Employment Training

- Based on the strong findings, DOL replicated the CET program in 12 sites from 1995-1999
- Replications conducted in 6 existing CET sites and 6 newly established sites
- Follow-up conducted at 30 and 54 months
- Fidelity was high only in 4 established sites, but never in other sites
- In the first follow-up, positive impacts on earnings for women only in high-fidelity sites
- By the second follow-up, so positive impacts

Center for Employment Training

- Miller et al. (2005) offer 3 hypotheses on lack of positive impact in replication
 - Treatment group may not have needed services because economy was better
 - Treatment group failed to take advantage of credentials received
 - Programs used by control group offered services similar to CET during replication
- Successful youth programs very rare in US, so should CET be abandoned?
 - There are many important features of CET, and replications may not have implemented them all
 - Worth exploring in current project to identify and test strategies for disconnected youth

Improving Replication Efforts

- Give more systematic on *when* to replicate and *how* to replicate
 - If program successful, do we need replication?
 - Are there ever cases where replication of unsuccessful program desirable?
- Think about replicating for other target groups
- Give more thought to assuring fidelity: Do not replicate on the cheap
- Be more systematic in deciding number of sites and T and C sizes