American Economic Review: Papers & Proceedings 2017, 107(5): 462–465
https://doi.org/10.1257/aer.p20171060

Challenges to Replication and Iteration in Field Experiments: Evidence from Two Direct Mail Shots†

By Jake Bowers, Nathaniel Higgins, Dean Karlan, Sarah Tulman, and Jonathan Zinman*

We conducted an experiment marketing microloans to farmers in the United States during spring 2015, and found that a simple direct mail letter increased borrowing from a government program. The subsequent spring, we built on this finding and enriched the design to test for information spillovers. The second experiment used a different sample frame and slightly different letter, due to operational considerations. The direct effect result did not replicate in the second year, thus lowering the likelihood that spillovers would be present and detectable.

These results add to recent evidence on how (seemingly subtle) differences in context and treatment content affect consumer responses (e.g., Allcott 2015; Bertrand et al. 2010). At the same time, many treatment effects do seem to hold in multiple contexts (e.g., Banerjee et al. 2015).

So what is an experimenter to do given the constraints on the number of factors that can be randomized? We offer several considerations and potential solutions.

I. USDA Microloans

The United States Department of Agriculture's Farm Service Agency (FSA) provides loans to farmers through several programs. Farmers can use the funds for working capital or fixed asset purchases including real estate.

The FSA Microloan program[1] launched in 2013 to meet the "financing needs of small, beginning farmers, niche and non-traditional farm operations." The maximum loan size is $50,000, with a relatively short application and relaxed underwriting criteria for managerial experience, production history, and collateral (Tulman et al. 2016).

II. Experiments: Sampling, Design, Data

Providing credit access to nontraditional operations means conducting outreach to producers who may not have regular contact with FSA. FSA has conducted prior mailing outreach experiments in other programs, with almost uniformly positive treatment effects thus far on the targeted measures of farmer engagement (see, e.g., Wallander, Ferraro, and Higgins forthcoming).

The objectives of the experiment in the projects reported here are to (i) measure whether direct mail increases take-up of the Microloan program (Experiments 1 *and* 2); (ii) compare direct mail effectiveness across different sample frames—different databases that can be tapped for outreach—(Experiment 1 *versus* 2); and (iii) estimate whether word of mouth amplifies

* Bowers: University of Illinois, 420 David Kinley Hall, 1407 W Gregory Drive, Urbana, IL 61801, and SBST (e-mail: jbowers@illinois.edu); Higgins: Social and Behavioral Sciences Team, 1650 Pennsylvania Avenue NW, Washington, DC 20502 (e-mail: nathaniel.higgins@gsa.gov); Karlan: Yale University, 27 Hillhouse Avenue, New Haven, CT 06511 (e-mail: dean.karlan@yale.edu); Tulman: US Department of Agriculture, Economic Research Service, 1400 Independence Avenue SW, Mail Stop 1800, Washington, DC 200250-0002 (e-mail: sarah.tulman@ers.usda.gov); Zinman: Dartmouth College, HB6106, Hanover, NH 03755 (e-mail: jzinman@dartmouth.edu). The authors thank Hunt Allcott for constructive and helpful comments that improved the paper, and Michael Gerling for invaluable assistance. Views expressed herein are the authors' and do not represent the views or policies of the US Department of Agriculture, Economics Research Service, the Social and Behavioral Sciences Team, or the General Services Administration. The research here was funded by the Economic Research Service, USDA.

† Go to https://doi.org/10.1257/aer.p20171060 to visit the article page for additional materials and author disclosure statement(s).

[1] The program name has since changed to the Direct Farm Operating Microloan program.

outreach (Experiment 2). It turns out that objectives (ii) and (iii) are at odds in our context.

Each of our experiments randomizes a single direct mailer sent in the spring, the busiest borrowing season. The mailers provide basic information on loan purposes, terms, and the application process (online Appendix Figure 1). The outcome of interest is the same in both experiments: a successful loan application.[2] The note to Table 1 provides additional details on the sample frames and randomizations.

A. *Experiment* 1

The sample frame for Experiment 1 is the National Agricultural Statistics Service's (NASS) "list frame," which is used to conduct the Census of Agriculture. This "statistical" frame, to use government data management parlance, has the advantage of including farmers who had not previously interacted with FSA. But it has the disadvantage of being divorced from FSA "administrative" program data in an operational sense, making data merges (e.g., between treatment assignment and program take-up) cumbersome and costly. The list frame also has usage restrictions, designed to protect individual identifying information, that preclude us from randomizing at the individual level.

The zip code is the lowest permissible level of aggregation for the list frame and, in spring 2015, we block-randomized all 3,683 US zip codes with at least 1 likely-eligible farmer and at least 20 percent of farmers in the zip code likely-eligible, in nine target states, 50–50 into treatment (every farmer sent one mailer) versus control (no farmer sent mailer), with likely eligible farmers being defined as those targeted by the Microloan program. We end up with 143,981 treatment and 144,924 control farmers and then measure loan take-up at the individual level.

B. *Experiment* 2

In year two (spring 2016), we prioritized testing whether a similar direct mailer would increase take-up among women farmers. The list frame is ill-suited for this purpose, so we use the Service Center Information Management System (SCIMS) maintained by the FSA. This frame has the advantages of being administrative data that is easier to merge with data on program take-up, and of having lighter usage restrictions that permit targeting and randomization at the individual level. But it only includes farmers who have previously interacted with FSA.

Targeting women motivated us to make (seemingly) minor changes to the mailer as well: a photo of a female instead of a male farmer, and several text modifications—including changes from "farmers" to "women" (online Appendix Figure 1).

Another design change sought to build on the anticipated replication of the finding that mailers would increase take-up, by adding a test for spillovers between letter recipients and other nearby female farmers. We randomly assigned 524 noncontiguous US counties, 50–50, to either 0 percent treated ("pure control") or to 50 percent of female farmers treated. We can then estimate spillovers by comparing *untreated* farmers in treated counties to the pure control group. All told we include the universe of 548,546 female farmers in the SCIMS in the 524 experimental counties.

III. Results

Table 1 presents the experimental results, which we summarize here. In year one we find that farmers in treated zip codes (those assigned to get a mailer) are 0.06 percentage points(pp) (se $= 0.02$) more likely to take-up than control farmers (those assigned to no mailer), a large increase on the base take-up rate of 0.22 percent in the control group. In year two this finding does not replicate, with a point estimate of 0.0001pp, (se $= 0.007$pp), on a control group take-up rate of 0.04pp. The main treatment effects are statistically different across the two years (p-value < 0.01). Unsurprisingly, given the lack of a main effect in year two, we do not find a spillover effect either (0.005pp, se $= 0.007$pp).

IV. Discussion

We conducted two seemingly similar direct mail outreach experiments one year apart and found different results. A mailer increased take-up of a government agricultural microloan

[2] The difference between actual applications and successful applications is so small that we ignore it. To apply, a borrower needs to work with a local loan officer, who is able to provide guidance as to the likely success of an application, and as such, very few rejections are actually recorded.

Table 1—Regression Results

Dependent variable = 1 if borrowed under program	Experiment 1[a]	Experiment 2
Direct effect of mailer	0.00056404	−0.00006748
	(0.000167)	(0.0001183)
No mailer in mailer county (spillover effect)		−0.0001437
		(0.0001184)
Constant	2.21E-16	0.001323
	(5.39E-17)	(0.0001699)
Observations	288,905	515,048

Notes: Unit of observation is the individual in column 1, and the household (potentially multiple female farmers residing at the same address) in column 2. Each column presents results from a single OLS regression on the treatment variable(s)—based on random assignment of getting a direct mailer—and randomization strata. The 2015 sample frame is the US Census of Agriculture in Alabama, Arkansas, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, and Tennessee. The 2015 experiment was randomized at the zip-code level, stratified by using a block randomization on number and density of farmers. Specifically, using the 2012 Census of Agriculture, we calculated the deciles of the estimated number and density (as a proportion of all farmers) of likely eligible farmers by zip code within each state, with likely eligible farmers being defined as those targeted by the Microloan program (minorities, women, beginning farmers, etc.). We dropped zip codes with less than 1 targeted farmer, and zip codes such that more than 80 percent of farmers were non-targeted farmers. We then combined these deciles, creating one hundred blocks, and then within each block assigned, 50 percent of zip codes to treatment and 50 percent to control. Standard errors in Experiment 1 are clustered at the zip code level. For Experiment 2 we constructed pairs of potential-to-treat counties (rather than zip codes) and pure-control counties based on the number of farmers within each county, prioritizing the best matches between counties that were feasible given the constraints on the number of letters we could send out. We identified adjacent counties and removed them from consideration of paired counties for sampling. This was done to improve the likelihood that the farmers in pure control counties were not in contact with any farmers that received the marketing. We then ordered counties by number of eligible farmers, to create pairs of counties, and then randomly assigned one from each pair to treatment and one to control. Within treatment counties, we then randomly assigned 50% of female farmers to receive a marketing letter, and 50 percent to not.

program in year one. Anticipating that the result would replicate in year two, we enriched the experimental design to test for spillovers. But the direct effect did not replicate.

What caused the different treatment effects? It may be that effects vary from one year to the next. It may be that treatment density matters: our first experiment randomized at the zip code level and had a denser concentration of treated farmers. It may be that subtle differences in mailer content matter. It may be that sample frame matters because treatment effects are heterogeneous across farmer characteristics: after all, agriculture differs between the 9 southern states in experiment 1 and the 47 states included in experiment 2.

What should an experimenter who is interested in testing for spillovers and has seen our year one results do? An obvious strategy is to hold the sample frame and everything about the treatment fixed, and send another mail shot as soon as possible. But in our case the practical interest in testing whether direct mail would be effective on a specific (and different) target market precluded us from fixing the sample frame, and strong seasonality in demand for the marketed program (coupled with sample size constraints) forced us to wait a year. In many settings difficulty in merging data from different sources—e.g., demographic or transaction data with contact information; treatment assignment with downstream outcomes—can also slow iterative testing.

Backing up, one could in principle have learned more about several of the open questions with a richer design in year one; e.g., by randomizing content within or across mailers, and/or by testing whether treatment effects for the same mailer vary across different sample frames or across baseline farmer characteristics within frame. But, as is often the case, sample size and capacity constraints permitted only the simplest of designs in year one.

Our results highlight the value and obstacles to using iterative field experimentation for basic and/or applied research. Seemingly subtle research design differences can change results. Yet the number of permutations required to run a full and contemporaneous set of tests far exceeds the sample size in a typical setting, even one with “big data.” This statistical constraint, coupled with the time dimension, dictates that iteration is key to identifying heterogeneous treatment effects and any limits to external validity.

But to iterate quickly, “rapid-fire testing” and re-testing across many dimensions, research stakeholders must overcome several obstacles. We will have to increase sample sizes, make the implementation of randomizations more nimble, hasten the capture of outcome data and its merging onto baseline and randomization data, quicken the analytic turnaround from test to result, and speed up the feedback loop from results to design changes to implementation.

REFERENCES

Allcott, Hunt. 2015. “Site Selection Bias in Program Evaluation.” *Quarterly Journal of Economics* 130 (3): 1117–65.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Pariente, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. “A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries.” *Science* 348 (6236): 1260799.

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman. 2010. “What’s Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment.” *Quarterly Journal of Economics* 125 (1): 263–306.

Tulman, Sarah, Nathaniel Higgins, Robert Williams, Michael Gerling, Charles Dodson, and Bruce McWilliams. 2016. *USDA Microloans for Farmers: Participation Patterns and Effects of Outreach.* US Department of Agriculture Economic Research Report 222.

Wallander, Steven, Paul Ferraro, and Nathaniel Higgins. Forthcoming. “Addressing Participant Inattention in Federal Programs: A Field Experiment with the Conservation Reserve Program.” *American Journal of Agricultural Economics.*