“A Revolution in Economics? It’s Just Getting Started…”

Shawn Cole[1] (Harvard Business School and J-PAL)
William Parienté (Université Catholique de Louvain and J-PAL)
Anja Sautmann (J-PAL)

Abstract: We have each experienced thrills and pain while supporting the mission of the Poverty Action Lab, which has facilitated many of the experiments described in the 2019 Nobel prize citation, and which, in many ways, seeks to fulfill what Angrist and Pischke called the “Credibility Revolution in Empirical Economics.” Even though (or perhaps because) we have conducted many RCTs, we share many of the concerns that critics have highlighted: high cost, long time lags, and limits to generalizability. Yet, we are quite optimistic that the power and importance of experimental work will only grow. We see two complementary developments which will make RCTs cheaper, faster, larger, and ultimately substantially more insightful. First, a new research literature seeks to improve the design of experiments, and what we can learn from them, through improved methodologies, meta-analyses, and improved understanding of heterogeneity. Second, the rise of administrative data, and in particular the promise of ‘closed-loop’ data environments, in which interventions can be delivered and evaluated digitally, often on very large samples, is rapidly opens new frontiers for investigation.

Randomized control trials (RCTs) are incredibly powerful and versatile, particularly well-suited to understanding causal relationships and evaluating policy interventions - but they are not new. Edward Chamberlin conducted economic classroom experiments as early as 1948, and Vernon Smith won the 2002 Nobel prize for his experiments on the workings of markets (in fact, randomization experiments were used in the field of “psychophysics” in 1883). This year’s Nobel Prize was not awarded for the invention of RCTs, but for putting them to systematic use, at a previously unprecedented scale - transforming development economics in the process, as the prize committee noted. The use of randomization in the field at large has exploded over the last 15 years, and become a standard tool in labor, finance, health, and many other subfields of economics.

But even though (or perhaps because) we, the authors, have conducted many RCTs, we are acutely aware of some of their drawbacks, which many critics have also highlighted. The data collection and intervention delivery for an RCT can create substantial costs; RCTs can take months or years (funding applications, baseline data collection, roll-out, follow-up data collection, and analysis). Individual studies may hold limited insights for other settings or populations. While many of these challenges are not particular to RCTs, they limit the speed and scale whereby systematic bodies evidence can be assembled. We therefore see great

[1] *Disclosures:* Shawn Cole, Anja Sautmann, and William Parienté receive compensation from the Jameel Poverty Action Lab by virtue of their board role or employment. Shawn Cole is co-founder and board chair of Precision Agriculture for Development (PAD), a non-profit which evaluates and scales agricultural extension; Cole and Killeen (2019) and Kasy and Sautmann (2019) were conducted in cooperation with PAD. The authors affirm they have no further competing interests.

promise in several complementary developments in the field that have the potential to address some of these challenges.

A Wealth of Administrative Data

The first of these developments is the increasing digitization of infrastructure and service delivery in developing countries, both in the private sector and in government.

From utility billing to food-assistance delivery, more and more publicly provided services are administered or monitored using electronic systems. Citizens can file their taxes electronically and contact their representatives via text message. Linking such data via reliable identifiers, such as the Aadhar ID in India, dramatically reduces study costs, while radically expanding the scope of outcomes that development researchers can measure.

Increasingly, academics negotiate access to this data for their work. Such data brings several important advantages. A comprehensive administrative dataset can eliminate costly data collection; for example, a representative database of citizens or clients can be used for selecting and stratifying an experimental sample without the need for a prior census. In many instances, administrative data is of higher quality than survey data, for example when it comes to information that is difficult to recall (such as exact work hours), hard to understand (such as medical test results), or unpleasant or risky to share (such as criminal records). It can be an effective tool for independent monitoring: Dhaliwal and Hanna (2017) use biometric records to measure attendance for public health workers, and Callen and Long (2015) compare administrative records at different aggregation levels to detect election fraud in Afghanistan.

Government data is only the beginning - the depth and breadth of information collected by the private sector is stunning. Private-sector data in development has been particularly helpful in understanding consumer financial services. Karlan and Zinman (2009) study the efficacy of messages encouraging saving; Schaner (2018) examines whether time-limited high interest rates have long term impacts on savings, and Bursztyn et al. (2019) work with a bank to evaluate whether moral messages encourage repayment by delinquent borrowers. We expect experiments to grow substantially more ambitious as researchers develop increasingly deep relationships with firms.

In parallel, satellite or street photography data, combined with machine learning algorithms, provide detailed information on crop choice, deforestation, and soil erosion; electrification and night lights; transport networks and property values. For example, Cole and Killeen (2019) use satellite data to evaluate an RCT which tests whether customized fertilizer advice affects smallholder farmer yields in Gujarat. The advantages of remote-sensing data relative to surveys are numerous: there is no attrition, additional years of post data, as well as pre-data, are virtually free to collect, and satellite yield data does not suffer from well-known biases such as over-estimation by the smallest plot holders. The net effect is to reduce the required sample size (holding power constant) by at least fifty percent.

Improved Analysis and Design

A second important development is the surge of research on improving the analysis and design of experiments.

Many critics have pointed out the problem of external validity: a single study in a specific context provides only incomplete information about its potential impact elsewhere. However, by now, hundreds of RCTs have been completed, and much of the data collected associated to these RCTs is published and accessible (due to the concerted efforts of many, including J-PAL). At a minimum, this allows us to qualitatively compare results across contexts (for example, if impacts of similar interventions replicate across areas and populations). Moreover, recent years have seen meta-analysis techniques applied to RCT data in social science, e.g. Bayesian hierarchical models (Bandiera et al., 2017; Meager, 2019) that combine the available evidence in a structured way. Those models disentangle heterogeneity of treatment effects *across* studies from sampling variation *within* each study. For example, Meager (2019) combines data from seven microfinance experiments and shows that the average effects of better access to microcredit are small to negligible across the board.

Similarly, a rapidly growing literature applies machine learning methods in RCTs. Machine learning can be used to identify heterogeneity in treatment effects across subgroups without "mining" the data for significant effects (Athey and Imbens, 2017; Chernozhukov et al., 2018). An innovative application is Hussam et al (2018), who compare the results from predicting which entrepreneurs are most likely to benefit from a cash grant in India through machine learning and through community information. Those new methods could be used in theory to predict effects in other contexts (without necessarily bearing the cost conducting another RCT). Another set of studies have begun to incorporate adaptive sampling designs, familiar from multi arm bandit problems and many computer science applications, to improve learning (e.g. Dimakopoulou et al. (2017), Kasy and Sautmann (2019)). These new designs have the potential to make experiments with many treatment arms more efficient and less expensive.

Conducting Experiments in Closed-Loop Data Environments

The power of A/B tests in e-commerce settings, where firms digitally implement treatments and instantly measure outcomes (such as purchase decisions or user engagement), has transformed software and services. Academic economists have partnered with companies to answer questions of strategic interest to companies, such as Ho et al. (2019), who randomize customer service messages (and rebates) to over one million potentially dissatisfied Uber passengers, seeking to understand how firm behavior affects customers propensity to continue to purchase rides via Uber.

Public and social services are increasingly delivered via internet or mobile phone, opening the door to closed-loop experimentation for the purposes of better policy. Fabregas et al. (2019) report on a series of experiments in which farmers in Kenya were sent text messages to encourage adoption of lime; by partnering with One Acre Fund, the research team was able to obtain outcome data (lime purchase) at almost no marginal cost, allowing several rounds of iterations in message content. Similarly, Kasy and Sautmann (2019) apply their adaptive sampling algorithm (above) to help an NGO to quickly identify the most effective way to enroll rice farmers into a free phone information service using a voice-based, automated profiling system. Six treatments were tested in 17 experimental waves with 10,000 subjects in less than 40 days; and, thanks to the adaptive algorithm, nearly 40% of the sample benefitted from the treatment with the highest enrollment rate.

Conclusion

Over the past 15 years, RCTs have become an integral part of empirical economics. Hundreds, if not thousands, of RCTs have helped shape a collective understanding of the methodology's advantages and limitations. As the field is matures, it is transforming again. Researchers are refining its methods and beginning to tackle complex problems such as external validity. Innovations in data use and acquisition, sampling and design methods promise to make RCTs more nimble and efficient. The parallel digitization of data and information collection - as well as service delivery itself - allow experimentation at unprecedented speed and nearly zero marginal cost. These developments increase the scope for research even further, but they also hold particular promise for governments or NGOs who are most interested in quickly and cheaply identifying policies that work.

REFERENCES

Athey, S., & Imbens, G. (2017). The econometrics of randomized experiments. *Handbook of Economic Field Experiments*, 1:73–140

Bandiera, O., Fischer, G., Prat, A., & Ytsma, E.(2017). Do Women Respond Less to Performance Pay? Building Evidence from Multiple Experiments (Working Paper)

Bursztyn, L., Fiorin, S., Gottlieb, D., & Kanz, M. (2019). Moral Incentives in Credit Card Debt Repayment: Evidence from a Field Experiment. Journal of Political Economy, 127(4), 1641–1683. doi: 10.1086/701605.

Callen, M, & Long, J. (2015). Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan. *American Economic Review*, 105(1): 354-381.

Casaburi, L, Willis, J. (2018). Time vs. State in Insurance: Experimental Evidence from Contract Farming in Kenya. *American Economic Review*, 108 (12): 3778-3813.

Chernozhukov, V., Demirer, M., Duflo, E.,& Fernandez-Val, I.(2018).Generic Machine Learning Inference on Heterogenous Treatment Effects In Randomized Experiments. (Working Paper)

Cole, S, & Killeen, G. (2019). The promise of satellite yield measurements: Evidence from a randomized controlled trial in Gujarat (Working Paper).

Dimakopoulou, M., Athey, S., &Imbens, G. (2017). Estimation considerations in contextual bandits. arXiv preprint arXiv:1711.07077.

Fabregas, R., Kremer, M., & Schilbach, F. (in press). Realizing the Potential of Digital Development: The Case of Agricultural Advice. *Science*.

Halperin, B., Ho, B., List, J., & Muir, I. (2019). Toward an Understanding of the Economics of Apologies: Evidence from a Large-Scale Natural Field Experiment (NBER Working Paper No.25676). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w25676

Hussam, R., Rigol, N., & Roth, B. (2018). Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design In The Field. *Working Paper*

Karlan, D., & Zinman, J. (2009), Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment. Econometrica, 77, 1993-2008. doi:10.3982/ECTA5781

Kasy, M., & Sautmann, A. (2019). Adaptive treatment assignment in experiments for policy choice. Manuscript submitted for publication.

Meager, R.(2019). Understanding the Average Impacts of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments. *American Economic Journal: Applied Economics.* 11(1), 57-91

Schaner, S. (2015). Do Opposites Detract? Intrahousehold Preference Heterogeneity and Inefficient Strategic Savings. American Economic Journal: Applied Economics, 7(2), 135–174. doi: 10.1257/app.20130271