

THE ADVANTAGES OF RANDOMIZED EVALUATIONS

In a world with limited resources, decision-makers in government and other organizations must decide which social programs to invest in. A good place to start is by measuring impact: How well do these programs work?

Randomized evaluations make it possible to understand the causal impact of an intervention: In other words, what specific changes to participants' lives can be directly attributed to the program itself? They also allow researchers and policymakers to tailor their research designs to answer specific questions about the effectiveness of a program and its underlying economic theory.

In recent years, randomized evaluations, also called randomized controlled trials (RCTs), have gained global prominence as a tool for measuring impact in policy research. The 2019 Nobel Memorial Prize in Economics was awarded to J-PAL co-founders Abhijit Banerjee and Esther Duflo, and longtime J-PAL affiliate Michael Kremer, in recognition of how this research method has transformed the field of social policy and economic development.

RANDOMIZED EVALUATIONS IN PRACTICE

Randomized evaluations can answer policy-relevant questions with the potential to affect millions of lives. In Indonesia, for example, a government program called Raskin provided subsidized rice to low-income households. It reached 17.5 million people in 2012, but researchers estimated that eligible households saw only about one-third of the benefits they were entitled to receive due to issues related to corruption and inefficiency.

A randomized evaluation by Banerjee et al. (2018)¹ tested whether giving households identification cards, referred to as Raskin cards, that included information about their eligibility and program rules would help them access the program. If the Raskin cards worked,

more benefits would reach households in need. But if not, it would be a waste of money. In this case, how could we study its impact?

In an ideal world, we would distribute Raskin cards and measure the benefits that each household received—then go back in time to see what would have happened to those same households without the cards. This way, we would know exactly what effect the cards had.

We can't time travel—but we can achieve a similar goal using **randomization**. In this study, the researchers randomly assigned a large number of villages to one of two versions of the intervention (the **treatment groups**) or a **comparison group**.



Treatment group 1

In the first treatment group, eligible households received cards with information about the quantity of subsidized rice they were entitled to purchase.



Treatment group 2

The second treatment group received cards with the same information, plus the official price of Raskin rice.



Comparison group

In comparison villages, the government continued to run the program under the status quo with no cards.



Outcomes of interest

Researchers measured how much rice households in each group received, the price they paid for the rice, and individuals' satisfaction with the program.

CREDIBLE ESTIMATES OF IMPACT

An important advantage of randomized evaluations is that they help ensure that **systematic differences** between groups do not drive differences in outcomes. In other words, we can more confidently attribute the difference in outcomes to the intervention, rather than to other factors.

In our example, suppose instead that program officials had been able to choose which households received cards. Officials who weren't following the program's rules might not want their citizens to receive information about the program and may therefore be less likely to enroll their community. In this case, the communities with the most irregularities in rice distribution would not receive the cards, while communities where Raskin was already working well would get them. As a result, we might *overestimate* the impact of the intervention.

Alternatively, suppose the government had rolled out the cards to the villages with the most citizen complaints first—i.e., villages with irregularities—and we had measured impact by comparing them to villages not yet enrolled, where Raskin was performing better to begin with. This might lead us to *underestimate* the effect of the program.

Random assignment solves this selection problem by ensuring the treatment and comparison groups are comparable. Of course, the two groups will not be exactly the same. Any two individuals assigned to treatment and comparison groups are likely to be quite different in all sorts of ways, leading to some chance variation between the different groups.

However, as we add more individuals (or in our example, villages), the groups as a whole tend to become more similar to each other, and more similar to the population from which they are drawn. Statistical methods allow us to gauge how likely it is that differences in outcomes we observe are due to the program being evaluated. With large enough samples, we can learn the true effect of the intervention with a high degree of confidence.

ANSWERING SPECIFIC QUESTIONS

Another important benefit of randomized evaluations is that the intervention and data collection can be designed to answer specific questions. For example, it may be important to understand the individual impact of different components of a program and the channels through which they work.

In the Raskin example, researchers created a second treatment group to test whether providing information about the official subsidized price had an additional impact on the quantity of rice eligible households received and the price they paid.

They found that households that received price information did not pay less—but they did receive higher quantities of subsidized rice. This suggests that the price information increased households' bargaining power with local officials. In part based on these results, the government decided to scale up social assistance identification cards to 15.5 million households across the country, reaching over 65 million people in 2013.

A randomized evaluation can also be designed to credibly measure program effects on different groups, such as men and women. Researchers who conduct randomized evaluations often collect novel data as well, such as detailed location or GPS data, or information on preferences, language, quality of goods, and more. This can be combined with administrative data, such as hospital or tax records, enabling researchers to answer questions about a host of subjects at relatively low cost.

IMPORTANT CONSIDERATIONS

Randomized evaluations are excellent tools for learning about the impact of a policy or program, but they are not appropriate for all situations. For example, a randomized evaluation is not the right choice under the following circumstances.

- If the sample size is too small to statistically detect an impact, even if one may exist.
- When a policy cannot be randomly assigned, such as a macroeconomic policy.
- When it is unethical to provide the intervention to a treatment group or withhold it from a comparison group.
- If a randomized evaluation is not cost-effective, too time-consuming, or politically infeasible.
- If threats to the integrity of the research design, such as spillover effects, are too difficult to control.

J-PAL affiliated researchers consider these factors carefully when determining whether an evaluation is ethical and feasible, and every J-PAL evaluation is reviewed by a formal Institutional Review Board whose mission is to protect the rights and welfare of study participants.

When appropriately designed and implemented, randomized evaluations are powerful tools for measuring the impact of policies and programs.

To learn more about our work, visit povertyactionlab.org.

¹ Banerjee, Abhijit, Rema Hanna, Jordan Kyle, Benjamin A. Olken, and Sudarno Sumarto. 2018. "Tangible Information and Citizen Empowerment: Identification Cards and Food Subsidy Programs in Indonesia." *Journal of Political Economy* 126 (no. 2): 451-491.

Authors: Eliza Keller, Sarah Kopper, Anja Sautmann