Agricultural Economics Research Review 2021, 34 (2), 253-265
DOI: 10.5958/0974-0279.2021.00027.6

An introduction to randomized evaluations and its application in agricultural economics

A G Adeeth Cariappa[1]

ICAR – National Dairy Research Institute, Karnal, Haryana 132 001

Email: adeeth07@gmail.com

Abstract Large-scale randomized evaluations or randomized control trials (RCT) are useful in evidence-based policymaking, but these are seldom used by agricultural economists in India. This paper discusses in detail the application of RCTs—from agricultural technology adoption to nudging farmers to price sensitivity. It is necessary to estimate the sample size, level, and type of randomization to design and implement a randomized evaluation properly. Agricultural economists at various institutes can collaborate in conducting large-scale RCTs and evaluate the effect of interventions in different contexts. The paper also lists potential applications for RCTs in agriculture.

Keywords Randomized control trials (RCT), randomized evaluation, impact evaluation, cost-effectiveness, random assignment

JEL Codes Q10, Q12, Q14, C93

Randomized evaluations—also known as randomized control trials (RCT), random assignments, random experiments, and social experiments—are conducted to support policy with rigorous evidence. Over the past 15–20 years, RCTs have become popular in the social sciences, especially development studies, and many researchers at international research institutes, universities, nongovernmental organizations (NGO), etc. use RCTs to solve important problems. Randomized control trials are conducted by organizations like the Abdul Latif Jameel Poverty Action Lab[2] (J-PAL), Innovations for Poverty Action (IPA), International Initiative for Impact Evaluation (3ie), World Bank, and Centre for Global Development, to name a few. For pioneering the application of RCTs in the 21st century to answer potential questions in eradicating poverty, Abhijit Banerjee, Esther Duflo, and Michael Kremer won the Nobel Prize in Economics in 2019. The award of the prize to them is a testament to the effectiveness and popularity of RCTs.

Randomized control trials are not a new concept in agricultural research; experimental field-level research uses RCTs to evaluate the impact or effectiveness of a technology, or variants, like different doses of fertilizers, pesticides, weedicides, and animal feed. The use of large-scale RCTs in villages has helped to answer important questions in agriculture, and these can aid in policy design. However, RCTs are seldom used by

[1]The author implemented an RCT for his PhD dissertation titled "Impact evaluation of anionic mineral mixture supplementation on milk production and the milk fever: a randomized control trial". RCT ID: AEARCTR-0005108 (https://doi.org/10.1257/rct.5108-1.0).

[2]Abhijit Banerjee, Esther Duflo, and Sendhil Mullainathan founded the Poverty Action Lab (now Abdul Latif Jameel Poverty Action Lab, or J-PAL) in 2003. J-PAL is a global research centre that specializes in running randomized evaluations and has more than 181 affiliated researchers from 91 leading research universities of the world. In attempting to answer questions related to poverty, health, politics, agriculture, education, environment, finance, crime, gender, and labour markets, J-PAL and its affiliated researchers have 1109 ongoing and completed RCTs.

agricultural economists for policymaking in India. This article attempts to describe the potential uses of RCTs and motivate agricultural economic students and researchers of the ICAR to use it for good.

History of randomized evaluations

In 1747, James Lind demonstrated among sailors that eating lemons and oranges helped prevent scurvy. Lind was the first to introduce the concept of treatment and control in an experiment, and he is considered the father of clinical trials (Thomas 1997). In the 1920s, Neyman and Fisher separately used randomized trials in agriculture for treatment and control. Fisher's experimental work led to the famous book *The Design of Experiments* (Gibson and Sautmann 2017). Government-sponsored randomized trials were used to evaluate social policies in the 1960s to the 1990s. In small-scale experiments, the focus on subjects of interest shifted from plants and animals to human beings (Gibson and Sautmann 2017). Hundreds of randomized evaluations have been conducted since then, and the field is still evolving.

What is a randomized evaluation?

A randomized evaluation is a method used to estimate the impact of a programme or an intervention. The respondents are assigned randomly into two groups: the treatment group and the control group. If the sample size is large enough, all the respondents have an equal chance of being assigned to the treatment or control group; therefore, the groups are identical in both observable and unobservable characteristics, and the problem of selection bias is ruled out. The process of random assignment enables us to attribute any post-intervention difference in the control and treatment groups to treatment and not to any other factor (Gertler et al. 2016). The steps in programme evaluation are displayed in Table 1.

Randomization gives unbiased estimates

In an observational study, where respondents are randomly sampled from the treated and untreated (control) population (beneficiaries and non-beneficiaries of a programme), the difference in outcomes is the study's estimate of true impact. This estimate often differs from the true value due to systemic bias due to confounding variables, sampling variability, measurement errors, and spillover and monitoring effects (Ravallion 2018). The only error in an ideal RCT in which the treatment status is chosen randomly in addition to random sampling is the sampling variability. When the sample size increases, the sampling error approaches zero in expectation and the trial's estimate of impact gets closer to true value of impact. This is the sense in which an ideal RCT gives unbiased results (Ravallion 2018).

Bias is the product of correlation between explanatory-explained and explanatory-explanatory variables.

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$ - true equation

$\hat{y} = \beta_0 + \hat{\beta}_1 x_1$ - estimated equation (omitted variable)

Let us assume

$x_2 = \theta_0 + \theta_1 x_1 + \varepsilon$

Now substituting x_2 in the true equation, we get

$y = \beta_0 + \beta_2\theta_0 + (\beta_1 + \beta_2\theta_1)x_1 + \beta_2\varepsilon + \mu$

The estimation of the equation above can be written as

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \tilde{\mu}$

$\text{Bias} = \hat{\beta}_1 - \beta_1 = (\beta_1 + \beta_2\theta_1) - \beta_1 = \beta_2\theta_1$

No omitted variable bias exists when β_2 or θ_1 or both equal zero. The process of random assignment gives us unbiased estimates because it ensures that the intervention is independent of all the other variables

Table 1 Steps in programme evaluation

Needs assessment	What is the problem? Does the problem we propose to solve actually exist?
Theory of change	How in theory to we propose to solve these problems ? (Theoretical framework)
Process evaluation	Is the programme working as planned?
Impact evaluation	What is the effect of the programme?
Cost-effectiveness	Is the programme cost-effective?

Source Rossi, Lipsey, and Henry (2019)

($\theta_i = 0$). If the sample size is large enough—decided based on power calculation—random assignment gives us internally valid, unbiased, and precise estimates of impact.

How does RCT differ from other impact evaluation methods?

There are experimental and quasi-experimental methods to evaluate the impact of a programme (Table 2). Some quasi-experimental methods are assumption-heavy and are hard to test: difference-in-differences, propensity score matching, regression discontinuity design, and instrumental variable (particularly randomized promotion). The experimental method (RCT) makes very light (less) assumptions. These assumptions can be tested. The RCT method gives internally valid, unbiased, and precise estimates of impact—given enough sample size—and fits many operational contexts. Many researchers regard it as the gold standard of impact evaluation (Gertler et al. 2016).

Power calculation

The RCT method gives unbiased and precise estimates if the sample size is large enough. But what sample size is large enough to give precise estimates? We use the formula

$$MDE = (t_{(1-k)} + t_{\alpha})\sqrt{\frac{1}{P(1-P)}} * \sqrt{\left(\frac{\sigma^2}{N}\right)}$$

The formula is for individual-level randomization. We estimate sample size indirectly from the "minimum detectable effect size (MDE)" formula. The MDE formula helps to answer the question: "Given that I have the budget to sample only x households, what is the minimum effect size that I can distinguish from a null effect?" (World Bank 2019).

The MDE can be thought of as a hypothesized effect size that the intervention can generate, or a minimum level of impact below which the programme is considered unsuccessful; $t_{(1-k)}$ is the power, t_{α} is the significance level, P is the proportion of treatment in the sample, σ^2 is the variance, and N is the sample size (Duflo, Glennerster, and Kremer 2007). To calculate the sample size, we can use software packages like Optimal Design and STATA. The command in STATA is *sampsi (pre) (post), power(0.8) sd( )*.

If the randomization is done at the level of the cluster (village, block, school, district), we have to plug in intra-cluster correlation (ρ) and average cluster size (m) in the denominator of MDE in the formula.

$$\frac{MDE}{\sqrt{1+\rho(m-1)}} = (t_{(1-k)} + t_{\alpha})\sqrt{\frac{1}{P(1-P)}} * \sqrt{\left(\frac{\sigma^2}{N}\right)}$$

To randomize the example at the village level—with 80 farmers per village and an intra-cluster correlation of 0.3—we would use in STATA the command

sampsi 8.34 9.45, power(0.8) sd(3)

sampclus, obsclus (80) rho(0.3).

If is too high, there is no point in taking a large sample within the clusters; instead, increasing the number of clusters increases the power.

Data, randomization, and balance test

After deciding the optimum sample size required to detect a hypothesised effect size, we proceed to collect data. The RCT could be planned as a two-step or a one-step evaluation.

In the two-step process, a baseline survey of all the individuals in the study should be done before randomly assigning them to the treatment and control groups. The advantage of this method is that we can test if the two groups created are statistically similar on observable characteristics and the baseline information could be used to increase the precision of the impact estimates in the regression equation. The disadvantage is that it takes a lot of time and money to run the baseline survey.

In the one-step evaluation, we randomly assign some blocks/communities/villages to the treatment to control groups and randomly sample individuals from both groups. The survey is administered to randomly selected people; this method is simple and time-saving. The disadvantage is that we cannot test the pre-intervention similarity of the groups and improve the precision of the impact estimates.

How to randomize? What should be the level of randomization?

Random assignment can be done at the field or the office. The most transparent method at the field to determine whether a farmer receives the treatment or

Table 2 Comparison of impact evaluation methods

Method	Treatment group	Comparison group	Core assumptions	Limitations
Difference-in-difference	Programme participants (beneficiaries)	Non-participants (non-beneficiaries)	The outcomes of the treatment and control groups will grow in parallel over time if the programme did not exist	Multiple waves of data before the intervention are required to test this assumption; not always possible
Propensity score matching	Programme participants (beneficiaries)	Non-participants (non-beneficiaries; based on observable characteristics, for each member in T, a member in C is predicted to have the same likelihood to have participated in the programme)	Only the observed characteristics used for matching affects programme participation	Assumes away any unobserved characteristics between the treatment and control groups
Regression discontinuity	Units above a particular cut-off(when units are ranked based on specific criteria, like a poverty index)	Units close to cut-off but not eligible to receive the intervention	Units above and below the cut-off are statistically identical Population close to the cut-off (above and below) is representative of the whole population	Comparability of units above and below the cut-off
Randomized control trial	Randomly assigned to treatment group before the intervention	Randomly assigned to control group before the intervention	Random assignment generates two identical groups The treatment and control groups are similar in both observable and unobservable characteristics	Limited external validity of an experiment conducted in a specific context

Source Compiled from various sources majorly from Gertler et al. (2016)

not is running a lottery or tossing a coin. On a computer, Stata, R, or any replicable software could be used to randomize (World Bank 2021).

Randomization can be done at the level of the individual or cluster (village, panchayat, or block) or it could be stratified. The level of randomization depends on factors such as the budget, spillover of treatment to control groups, and the target of the treatment; therefore, the choice of the level of randomization (individual or cluster) is context-specific (Duflo, Glennerster, and Kremer 2007).

Balance test

The balance test is the process of testing whether the randomized assignment has generated two similar groups. One way of checking the balance is to test whether the characteristics of the treatment and control groups differ statistically. If the difference in means of observable characteristics (age, land size, income, etc.) is not significant, it can be concluded that the randomization has generated two similar groups.

Although statistical similarities in individual variables are achieved, sometimes the differences in the characteristics of the treatment and control groups might be in the same direction. This is an indication of the inability of the random assignment to generate two statistically similar groups. A solution is to complement differences in means with a test for joint orthogonality (F-test) (McKenzie 2015): run a regression with treatment assignment[3] as the dependent variable with other observable characteristics as independent variable. If the null hypothesis of the F-test is accepted, random assignment to two groups has succeeded in generating balance. Under pure randomization, if balance is achieved in observed variables we can expect to have balance in unmeasured or unobserved variables (Bruhn and McKenzie, 2009).

Estimation

The basic impact evaluation formula (Gertler et al. 2016) is

$\Delta = (Y|P = 1) – (Y|P = 0)$

This formula states that the impact or causal effect (Δ) of a programme/intervention (P) on an outcome (Y) is the difference between the outcome with the programme (Y|P=1) and outcome without the programme (Y|P=0).

$$Y_i = \beta_i T_i + \sum_{j=1}^{J} \gamma_j x_{ij}$$

The average treatment effect can be estimated through this regression equation where

Y_i is the outcome variable of i^{th} individual;

T_i is the dummy dichotomous variable (1 if treated, 0 otherwise);

β_i is the individual treatment effect of treatment on i^{th} individual; and

x_{ij} is the observed or unobserved covariates (linear causes of the outcome), which we assume captures minimum set of causes of outcome Y_i sufficient to fix its value (Deaton and Cartwright 2018).

If we have baseline data, difference in differences or analysis of covariance (ANCOVA) approach could also be used to estimate intent to treat (ITT) estimates (Özler 2015).

Therefore, to summarize the steps in RCTs (Figure 1), we start with the evaluation question based on the needs assessment followed by designing the study in terms of deciding the study population, intervention, outcomes, and the sample size; after designing the study, we implement it by administering a baseline survey of the sample. This sample frame is then randomly assigned to the treatment and control groups; the balance test follows. If the groups are similar—randomization is successful in creating two equal groups—we can proceed with the intervention. The follow-up survey could be done after the intervention while monitoring the programme all along.

Overview of use of randomized evaluations

Randomized control trials are used widely in evaluating the

- adoption of agricultural technology (Ogutu et al. 2018; Jack et al. 2019); welfare and productivity

[3]If there are only two groups, 1 treatment and 1 control group, treatment indicator (dependent) variable will be a dichotomous variable taking value 1 for treatment group and 0 for control group.

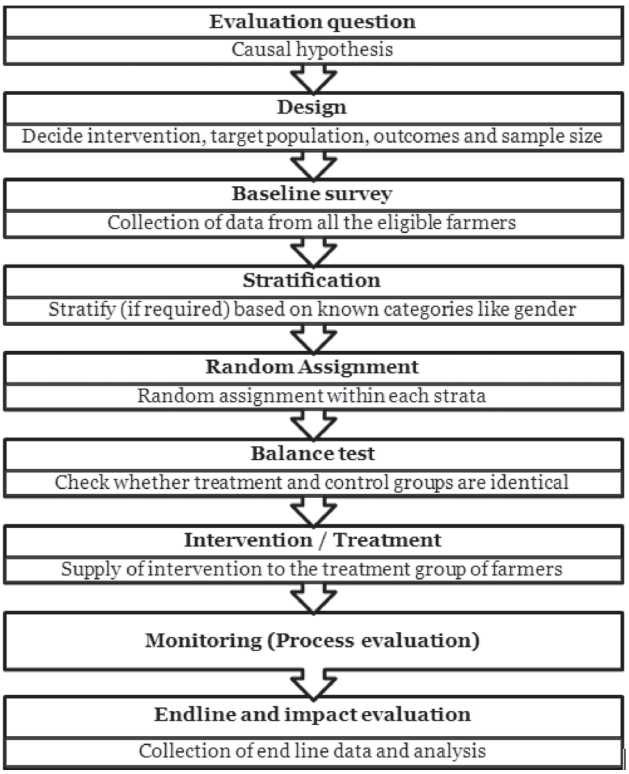

Figure 1 Steps in a randomized controlled trial

effects of contract farming (Arouna, Michler, and Lokossou 2021); digital extension advice on nutrient management (Arouna et al. 2021);

- effectiveness of social programmes like employment guarantee act (Bhatia et al. 2016) and integrated medical information and disease surveillance systems (Dhaliwal and Hanna 2017);
- impact of community participation in natural resource management (Karlan, Jamison, and O'Connor 2019); and the
- impact of nudging farmers to use fertilizers, save water, and store grains (Duflo, Kremer, and Robinson 2008; Chabé-Ferret et al. 2019; Aggarwal, Francis, and Robinson 2018).

Randomized control trials are used also in evaluating the impact of the use of

- inputs like drought-tolerant seeds (de Janvry et al. 2019) and of quality protein maize on nutrition (Tessema et al. 2016);
- the profitability of fertilizers (Beaman et al. 2013); and the
- impact of agricultural credit and insurance products (Karlan et al. 2014; Karlan et al. 2011) (Table 3).

Price sensitivity

Randomized evaluations can be designed to determine the price sensitivity or elasticity of a technology (a preventive health product or a new seed kit or biofortified food grains, etc) (Banerjee, Duflo, and Kremer 2016). These estimates can be used to decide the rate of subsidy of the technology. Randomized control trials can be used to determine the price at which a product should be sold or subsidized and inform researchers and policymakers planning on pricing (or subsidizing) a new agricultural technology

The most commonly used method to estimate the demand curve is willingness to pay: people are asked what they would be willing to pay for a product. But respondents may not think hard enough before answering because they are not incentivized. If the questions are not asked precisely, the respondents might interpret, and answer, the questions differently. For instance, if the respondents think that the enumerators are government or bank officials, and their answer might affect future subsidy policies, they might report a lower willingness to pay.

For this reason, researchers have moved towards field experiments in order to observe the "true" demand at each price point. One such measure is "take it or leave it". Experiments randomize the price and observe whether an individual purchases the product at that price. This is a straightforward revealed preference mechanism (Dupas and Miguel 2017).

Another widely used method is the BDM method, named after the theorists (Becker, DeGroot, and Marschak 1964). In the BDM method—a bidding process—respondents express their willingness to pay by quoting a price (or placing a bid). A random price is drawn from a known distribution of prices, and the item is sold if the bid exceeds the random price. The bidders pay only the random price even if their bid exceeds it; those who bid below the random price cannot purchase the item. This method allows for estimating the accurate demand of a product because the respondents who bid above the true value of a product risk buying the product when the price is higher than one would actually be willing to pay; and those

Table 3 Results of some RCTs on agriculture

Authors	Title	Main findings
Duflo, Kremer, and Robinson (2008)	How high are rates of return to fertilizer? evidence from field experiments in Kenya	Increased crop yield (28–91%) 15% increase in net income
Karlan et al. (2011)	Crop price indemnified loans for farmers: a pilot experiment in rural Ghana	86–92% take-up of loans No overwhelming evidence of change in the investment behaviour of farmers
Beaman et al. (2013)	Profitability of fertilizer: experimental evidence from female rice farmers in Mali	Increased likelihood of fertilizer use Increased output No increase in profit
Karlan et al. (2014)	Agricultural decisions after relaxing credit and risk constraints	Increased expenditure on agricultural inputs Farmers' trust in insurance schemes has a large effect on the take-up of rainfall insurance
Aggarwal, Francis, and Robinson (2018)	Grain today, gain tomorrow: evidence from a storage experiment with savings clubs in Kenya	Increased storage of grains Higher price realized Profitability approximately doubled
Shenoy and Rao (2021)	Got (clean) milk? transparency, governance, and incentives for cleanliness in Indian dairy cooperatives	Group incentives increased milk production quality in village dairy cooperatives Private, targeted incentives increase efficiency

who bid below the product's true value risk not buying it when the price is lower (Dupas and Miguel 2017).

In an experiment with households of school-age children in Kenya, Dupas (2014) found that the demand for long-lasting antimalarial bed-nets (LLINs) became less price-sensitive if subsidies were provided in the form of vouchers that households had to redeem at local retail shops within three months. At USD 0.60, the demand for LLINs was found to be 73%, but the demand dropped to 33% at USD 1.50 (an 80% subsidized price) and to 6% at USD 3.50 (a 50% subsidy). The primary driver of demand was the overall price of the LLINs, Dupas confirmed; different marketing strategies failed to change the slope of the demand curve. In an experiment by the author, the demand for a dairy animal health product was found to be highly sensitive to prices; 1% fall in prices led to a 18% increase in demand (Cariappa et al. 2021).

Similarly, various experiments were conducted in pricing preventive health products:

- deworming pills, in Western Kenya (Kremer and Miguel 2007);
- wearing rubber shoes to prevent worm infections, in Kenya (Meredith et al. 2013);
- water purifying solution (chlorine) in Zambia (Ashraf, Berry, and Shapiro 2010);
- hand washing soap in India (Spears 2009); and
- animal health product in India (Cariappa et al. 2021)

Dupas and Miguel (2017) review these studies and conclude that higher prices seem to have created too many errors of exclusion (leaves too many people out), and subsidies might be justified if these prevent the transmission of disease.

Additionally, it is unethical to use RCTs in many settings like estimating the losses due to pests where some farmer's field is infested with pests and some are not, or evaluating the impact of unemployment by terminating some employees. In many settings it is not feasible to run an RCT, like deliberately changing the tax rates for some states. In general, it is difficult to experimentally study institutions like property rights, legal system, democracy, religion, and marriage.

But researchers have intelligently used the experimental set-up to answer these kinds of questions. In Liberia, for instance, where a customary legal system, administered by the village chief, co-exists with a formal legal system, a legal empowerment intervention—advocacy services provided by community paralegals trained in formal law—was experimentally studied among who wanted to resolve a legal dispute (Sandefur and Siddiqi 2013).

In another experiment, Karlan and Zinman (2009) identify the level of information asymmetries (adverse selection and moral hazard) in credit markets and find that moral hazard is important. Also they find that about 13–21% of the loan defaults are due to moral hazard. Thus, RCTs are used to answer some questions thought to be untestable empirically.

Cost-effectiveness analysis (CEA)

Cost-effectiveness analysis (CEA) expresses the impact of the costs of implementing a complex programme on an outcome in terms of a simple ratio, or it measures an intervention's effectiveness at a given cost. Unlike cost-benefit analysis, CEA makes no judgement of the value of outcome variables. Cost-effectiveness analysis is widely used to compare various interventions used to solve a common problem and find the most cost-effective option (Dhaliwal et al. 2015).

$$\text{CE ratio} = \frac{\text{Total impact of programme/ intervention on specific outcome}}{\text{Total cost of implementing the programme}}$$

$$\text{Effect per INR 1,000 spent} = \frac{\text{Impact on specific outcome}}{\text{Cost of implementing}} * \text{INR 1,000}$$

Any programme—or any RCT, in particular—aims to provide solutions to a problem, explain why the intervention did or did not work, and scale up the most cost-effective solution among different alternative solutions to solve the particular problem. Thus, CEA lets researchers weigh the impact estimates based on the cost incurred and compare alternative solutions to a problem (Dhaliwal et al. 2015). Comparing the cost-effectiveness of programmes in pilot and at scale lets researchers test whether the programme works when scaled up.

Comparing interventions targeted at improving educational outcomes, it was found that school-based deworming of children, iron and vitamin A supplementation, and information campaigns were much more cost-effective than conditional cash transfers (J-PAL 2017). For instance, for every USD 100 spent on school-based deworming, children spent an additional 11.91 years in school, but the equivalent increase in school years was 2.61 for iron and vitamin A supplementation and 0.01-0.09 for conditional cash transfers (J-PAL 2018).

Bhula, Mahoney, and Murphy (2018) conducted an experiment of deworming children of primary and secondary school age; it was found that for every USD 100 spent (the cost-effectiveness), the additional years of education improved from 11.91 years in the pilot to 22.50 years at scale. Kremer, Miguel, and Thornton (2009) conducted an experiment to evaluate the effects of merit scholarships on girls' test scores in Kenya. They compared their impact estimates with that of six other programmes—teacher incentives, textbook provision, flip chart, deworming, and child sponsorship—and concluded that providing merit scholarships was the most cost-effective intervention.

A CEA tested and compared the effectiveness of cash and non-cash incentives in conditional cash transfer (CCT) programmes in many contexts. Nicaragua runs a cash transfer programme for women and girls; every USD 100 spent increases the number of years of education by 0.13 (Barham, Macours, and Maluccio 2018). In Morocco, every USD 100 spent on both conditional and unconditional cash transfers led increased the number of years of education by 0.02 (Benhassine et al. 2015).

Critical arguments against RCT

First, randomization does not always ensure orthogonality; to show that the estimates are unbiased, the post-randomization correlates with the treatment should be insignificant (balance test).

Second, what is true for the sample may not hold true for the population (Deaton and Cartwright 2018). This is the inference, or generalizability, problem. If there is considerable heterogeneity in the study sample, its average treatment effect will differ from that of the population.

One RCT is a piece of a big puzzle. Trying the same intervention in different contexts and understanding the behaviour of people lets us scale successful experiments.

Critics argue that the results of RCTs are obvious: using insecticide-treated bed nets reduces malaria; wearing eyeglasses improves the test scores of visually challenged students; fertilizer use enhances yield. But identifying problems (needs), deciding interventions, and designing the study so that it can be scaled is critical—not only for RCT but also for any methodology. Cost-effective projects form the basis of policy interventions.

The effectiveness of randomized evaluations may be threatened by attrition; partial compliance, contamination, or diffusion; and spillover. The solutions are, respectively, double sampling and buffer; placebo control design; and reducing interaction between the treatment and control groups spatially or temporally (Table 4).

There are other serious concerns regarding the ethics of running an experiment for evaluating policy options, especially among the poor when some get the intervention and others do not (Deaton and Cartwright 2018; Deaton 2020).

Conclusions

The 2019 Nobel Prize in Economics was awarded not only for the experimental approach used, but because of the impact this approach has had on converting evidence into policy. The evidence from randomized evaluations influenced many policies (Table 5).

Randomized evaluations can be used in agriculture to evaluate the impact of new technologies and of different incentives on technology adoption. These can nudge farmers to reduce stubble burning and export their products. Randomized evaluations can determine the differential impact of subsidies and direct cash transfers and which crop insurance model works. Researchers can use RCTs to design the most effective incentive to encourage farmers to participate in various schemes. The gossip theory (Banerjee et al. 2019) can be tried to disseminate information to farmers.

Randomized evaluations can help agricultural researchers answer questions like why farmers overuse inputs or do not participate in welfare programmes/ adopt new technology and why productivity is not increasing. More randomized evaluations should be taken up by agricultural researchers, and students should be encouraged to learn and conduct RCTs. Resources—like manuals, toolkits, teaching resources,

Table 4 Threats to random evaluation and its solutions

Threat	Solution
Attrition	Double sampling
Respondents drop out of either the treatment or control group during the evaluation	Random sampling is done again, and the missing sample is replaced with the new one if the new sample is identical to the missing sample
	Buffer
	Planning early and having an extra sample to replace a missing sample
Partial compliance/contamination/diffusion	Placebo control design
The respondents do not comply with the assignment to the treatment or control group	Comparing compliers in the treatment and control groups
	Estimating complier average causal effect instead of average treatment effect
Spillover	Design the study to reduce interaction between the treatment and control groups spatially or temporally, probably by selecting villages which are at least 10 km away
Indirect benefit or cost (positive or negative externality) incurred by the respondents in the control group	
(who did not directly receive the treatment)	If the study aims to estimate spillovers, inverse probability weights should be used to get consistent estimates

Source Gerber and Green (2012)

Table 5 Coverage of some of the successful randomized evaluations

Project	Country	Coverage
Teaching at the right level (TaRL)	India and Zambia, Nigeria and Côte d'Ivoire	60 million students
Deworming in schools	India, Ethiopia, Kenya, Nigeria, and Vietnam	292 million children
Give Directly – cash transfer programme	Kenya, Rwanda, and Uganda	125,000 households
Graduation approach – targeting ultra-poor households	Bangladesh and India	2.2 million
Community chlorine dispensers for better health	Kenya, Malawi, and Uganda	4 million
Phone-based technology for agricultural information delivery	India, Kenya, Pakistan, Rwanda, Ethiopia, Bangladesh, Uganda, Zambia, Nigeria	3.56 million farmers

Source J-PAL (2019) (https://www.povertyactionlab.org/evidence-to-policy/adapting-and-scaling-program)

and online courses on designing and running randomized evaluations—are available aplenty.[4]

Agricultural economists at various institutes can collaborate and solve each part of the larger puzzle and see how the intervention works in different contexts. For instance, RCTs could be used to test zero budget natural farming in different parts of the country. Similarly, the effectiveness of biofortified food can be tested using clinical trials and the incentives to increase adoption (like providing information, subsidy, or cash transfer) could be tested experimentally.

Inter- and intra-disciplinary experimental research conducted in a variety of contexts will provide policymakers evidence. That might be the biggest contribution of agricultural economists to the RCT revolution. As they say, better late than never.

References

Aggarwal, S, E Francis, and J Robinson. 2018. Grain today, gain tomorrow: evidence from a storage experiment with savings clubs in Kenya. *Journal of Development Economics* 134: 1–15. https://dx.doi.org/ 10.1016/ j.jdeveco.2018.04.001

Arouna A, J D Michler, and J C Lokossou. 2021. Contract farming and rural transformation: Evidence from a field experiment in Benin. *Journal of Development Economics* 151: 102626. https://doi.org/10.1016/ j.jdeveco.2021.102626

Arouna A, J D Michler, W G Yergo, and K Saito. 2021. One size fits all? Experimental evidence on the digital delivery of personalized extension advice in Nigeria. *American Journal of Agricultural* Economics 103: 596–619. https://doi.org/10.1111/ajae.12151

Ashraf, N, J Berry, and J Shapiro. 2010. Can higher prices stimulate product use? evidence from a field experiment in Zambia. *American Economic Review* 100 (5): 2383–413. https://dx.doi.org/ 10.1257/ aer.100.5.2383

Banerjee, A V, A G Chandrasekhar, E Duflo, and O M Jackson. 2019. Using gossips to spread information: theory and evidence from two randomized controlled trials. *The Review of Economic Studies*, 86(6), 2453–490. https://academic.oup.com/restud/article/86/6/ 2453/5345571

Banerjee, A V, E Duflo, and M Kremer. 2016. The influence of randomized control trials on development research and on development policy, in *The state of economics, the state of the world*, 1–76. MIT Press. https:// dx.doi.org/10.7551/mitpress/11130.003.0015

Barham, T, K Macours, and J A Maluccio. 2018. Experimental evidence of exposure to a conditional cash transfer during early teenage years: young women's fertility and labor market outcomes. Working paper. https://www.povertyactionlab.org/evaluation/ long-term-effects-conditional-cash-transfer-program-nicaragua

Beaman, L, D Karlan, B Thuysbaert, and C Udry. 2013. Profitability of fertilizer: experimental evidence from female rice farmers in Mali. *American Economic Review* 103 (3): 381–86. https://dx.doi.org/ 10.1257/ aer.103.3.381

[4] Some are listed in the Appendix

Becker, G M, M H Degroot and J Marschak. 1964. Measuring utility by a single-response sequential method. *Behavioral Science* 9: 226-232. https://doi.org/10.1002/bs.3830090304

Benhassine, N, D Florencia, E Duflo, P Dupas, and V Pouliquen. 2015. Turning a shove into a nudge? a "labeled cash transfer" for education. *American Economic Journal. Economic Policy* 7 (3): 86–125. https://dx.doi.org/ 10.1257/pol.20130225

Bhatia, R, S L Chinoy, B Kaushish, J Puri, V S Chahar, and H Waddington. 2016. Examining the evidence on the effectiveness of India's rural employment guarantee act. 3ie Working Paper 27. https://www.3ieimpact.org/file/8086/download?token=uX30WC3U

Bhula, R, M Mahoney, and K Murphy. 2018. Conducting cost-effectiveness analysis (CEA). https://www.povertyactionlab.org/research-resources/cost-effectiveness

Bruhn, M and D McKenzie. 2009. In pursuit of balance: randomization in practice in development field experiments. *American Economic Journal: Applied Economics* 1 (4): 200–232. https://dx.doi.org/10.1257/app.1.4.200

Cariappa, A G A, B Chandel, R Sendhil, A K Dixit, G Sankhala, V Mani and B Meena. 2021. Do the prices of a preventive animal health product affect dairy farmers' willingness to pay and product use? Evidence from an experimental study. Working paper. https://doi.org/10.21203/rs.3.rs-731711/v1

Chabé-Ferret, S, P Le Coent, A Reynaud, J Subervie, and D Lepercq. 2019. Can we nudge farmers into saving water? evidence from a randomized experiment. *European Review of Agriculture Economics* 46 (3): 393–416. https://dx.doi.org/ 10.1093/erae/jbz022

Cohen, J and P Dupas. 2010. Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics* 125 (1): 1–45. https://dx.doi.org/ 10.1162/qjec.2010.125.1.1

de Janvry, A, E Sadoulet, K Emerick, and M H Dar. 2019. The impact of drought-tolerant rice on local labor markets in India. https://www.povertyactionlab.org/evaluation/impact-drought-tolerant-rice-local-labor-markets-india

Deaton, A. 2020. Randomization in the tropics revisited: a theme and eleven variations. NBER Working paper 27600. https://dx.doi.org/10.3386/w27600

Deaton, A and N Cartwright. 2018. Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine* 210: 2–21. https://dx.doi.org/10.1016/j.socscimed.2017.12.005

Dhaliwal, I, E Duflo, R Glennerster and C Tulloch. 2015. Comparative cost-effectiveness analysis to inform policy in developing countries, in *Education policy in developing countries*, 285–338. https://doi.org/10.7208/chicago/9780226078854.003.0008.

Dhaliwal, I and R Hanna. 2017. The devil is in the details: the successes and limitations of bureaucratic reform in India. *Journal of Development Economics* 124: 1–21. https://dx.doi.org/ 10.1016/j.jdeveco.2016.08.008

Duflo, E, M Kremer, and J Robinson. 2008. How high are rates of return to fertilizer? evidence from field experiments in Kenya. *American Economic Review* 98 (2): 482–88. https://dx.doi.org/ 10.1257/aer.98.2.482

Duflo, E, R Glennerster, and M Kremer. 2007. Using randomization in development economics research: a toolkit. Centre for Economic Policy Research Discussion Paper Series No 6059, London. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=6059

Dupas, P and E Miguel. 2017. Impacts and determinants of health levels in low-income experiments, *Handbook of Economic Field Experiments, vol 2, chapter 1*, 3–95. Elsevier BV. https://doi.org/10.1016/bs.hefe.2016.09.003

Dupas, P. 2014. Short-run subsidies and long-run adoption of new health products: evidence from a field experiment. *Econometrica* 82 (1): 197–228. https://dx.doi.org/ 10.3982/ECTA9508

Gerber, A S and D P Green. 2012. *Field experiments: design, analysis, and interpretation*. W. W. Norton & Company. https://wwnorton.com/books/9780393979954

Gertler, P J, S Martinez, P Premand, L B Rawlings, and C M J Vermeersch. 2016. *Impact evaluation in practice*. 2nd ed. Washington, DC: Inter-American Development and World Bank. https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464807794.pdf?sequence=2&isAllowed=y

Gibson, M and A Sautmann. 2017. Introduction to randomized evaluations. https://www.povertyactionlab.org/research-resources/introduction-evaluations

Jack, W, M Kremer, J de Laat, and T Suri. 2019. Borrowing requirements, credit access, and adverse selection: evidence from Kenya. NBER Working Paper 22686. https://dx.doi.org/10.3386/w22686

J-PAL. 2017. Roll call: getting children into school. https://www.povertyactionlab.org/publication/roll-call-getting-children-school

J-PAL. 2018. Conducting cost-effectiveness analysis (CEA). https://www.povertyactionlab.org/resource/conducting-cost-effectiveness-analysis-cea

Karlan, D, E. Kutsoati, M. U. McMillan, and Chris Udry. 2011. Crop Price Indemnified Loans for Farmers : A Pilot Experiment in Rural Ghana. *Journal of Risk and Insurance* 78 (1) : 37–55. https://dx.doi.org/ 10.1111/j.1539-6975.2010.01406.x

Karlan, D, J Jamison, and T O'Connor. 2019. Community based rangeland management in Namibia. https://www.povertyactionlab.org/evaluation/community-based-rangeland-management-namibia

Karlan, D, R. Osei, I. Osei-Akoto, and C. Udry. 2014. Agricultural decisions after relaxing credit and risk constraints. *Quarterly Journal of Economics* 129 (2): 597–652. https://dx.doi.org/ 10.1093/qje/qju002

Karlan, D and J Zinman. 2009. Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica* 77: 1993-2008. https://doi.org/10.3982/ECTA5781

Kremer, M, and E. Miguel. 2007. The illusion of sustainability. *Quarterly Journal of Economics* 122 (3): 1007–65. https://dx.doi.org/ 10.1162/qjec.122.3.1007

Kremer, M, E. Miguel, and R. Thornton. 2009. Incentives to learn. *Review of Economics and Statistics* 91 (3): 437–56. https://dx.doi.org/ 10.1162/rest.91.3.437

McKenzie, D. 2015. Tools of the trade: a joint test of orthogonality when testing for balance. Development Impact Blog. https://blogs.worldbank.org/impactevaluations/tools-trade-joint-test-orthogonality-when-testing-balance

Meredith, J, J Robinson, S Walker, and B Wydick. 2013. Keeping the doctor away: experimental evidence on investment in preventative health products. *Journal of Development Economics* 105: 196–210. https://dx.doi.org/10.1016/j.jdeveco.2013.08.003

Ogutu, S O, A Fongar, T Gödecke, L Jäckering, H Mwololo, M Njuguna, M Wollini, and M Qaim. 2018. How to make farming and agricultural extension more nutrition-sensitive: evidence from a randomised controlled trial in Kenya. *European Review of Agricultural Economics* 47 (1): 95–118. https://dx.doi.org/10.1093/erae/jby049

Özler, B. 2015. Why is difference-in-difference estimation still so popular in experimental analysis? Development Impact Blog. https://blogs.worldbank.org/impactevaluations/why-difference-difference-estimation-still-so-popular-experimental-analysis

Rao, M and A Shenoy. 2021. Got (clean) milk? governance, incentives, and collective action in Indian dairy cooperatives. https://www.povertyactionlab.org/sites/default/files/research-paper/working-paper_863_Got-Clean-Milk-Dairy-Cooperatives_India_Jan2021_0.pdf

Ravallion, M. 2018. Should the randomistas (continue to) rule? CGD Working Paper 492. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/should-randomistas-continue-rule

Rossi, P H, M W Lipsey, and G T Henry. 2019. *Evaluation: a systematic approach, 8th edn*. SAGE Publications. https://us.sagepub.com/hi/nam/evaluation/book243885

Sandefur, J and B Siddiqi. 2013. Delivering justice to the poor: theory and experimental evidence from Liberia, in *World Bank Workshop on African Political Economy, Washington, DC, May*, Vol. 20. 2013. http://cega.berkeley.edu/assets/cega_events/61/5D_Political_Economy-_Violence.pdf

Spears, D. 2009. Bounded rationality as deliberation costs: theory and evidence from a pricing field experiment in India. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1444912.

Tessema, M, N S Gunaratna, K Donato, J L Cohen, M McConnell, D Belyaneh, I D Brouwer, T Belachew, and H De Groote. 2016. Translating the impact of quality protein maize into improved nutritional status for Ethiopian children: study protocol for a randomized controlled trial. *BMC Nutrition* 2 (54). https://bmcnutr.biomedcentral.com/articles/10.1186/s40795-016-0089-z

Thomas, D P. 1997. Sailors, scurvy and science. *Journal of the Royal Society of Medicine* 90 (1): 50–4. https://dx.doi.org/10.1177%2F014107689709000118

World Bank. 2019. Minimum detectable effect. DIME Wiki. https://dimewiki.worldbank.org/Minimum_Detectable_Effect

World Bank. 2021. Randomization in Stata. DIME Wiki. https://dimewiki.worldbank.org/Randomization_in_Stata

Received 6 March 2020 Accepted 10 May 2020

Appendix

Readings and teaching resources on RCTs

1. J-PAL (https://www.povertyactionlab.org/research-resources?view=toc)
2. Running randomized evaluations (http://runningres.com/)
3. Data and statistical codes of published studies (https://dataverse.harvard.edu/dataverse/DFEEP)
4. Impact Evaluation in Practice (https://www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice)
5. Handbook of Economic Field Experiments, Volume 1 and 2 (http://dx.doi.org/10.1016/bs.hefe.2016.09.005)
6. de Janvry Alain, Kyle Emerick, Elisabeth Sadoulet and Manzoor Dar. 2016. The agricultural technology adoption puzzle: what can we learn from field experiments? Ferdi Working paper 178.

   https://are.berkeley.edu/esadoulet/wp-content/uploads/2018/10/Agricultural-Technology-Adoption-Puzzle-Ferdi-WP.pdf
7. Teaching and learning resources for impact evaluation (mostly RCT) (https://adeeth07.github.io/more/)