



Do and Don'ts when Implementing Randomized Controlled Trials in Developing Countries

Asad Islam

Director

Centre for Development Economics and Sustainability (CDES)

PIER, Bank of Thailand October 31 2024



Why RCTs?

- Policy makers want to know which policy works
- It is important to test policy before implementation as conventional wisdom may be wrong.
- RCTS have become popular in estimating the causal impacts of programs and policies
 - RCTs enable social scientists to draw valid inferences about cause and effect.
 - The essential ingredient of experimentation is random assignment of people into treatment and control groups
- Randomisation ensures that these groups differ solely due to chance.
- RCTs are widely used for programme and policy evaluations, simple and easier to explain the results to policymakers and practitioners.





Randomization

- When a randomized evaluation is correctly designed and implemented, it provides an **unbiased estimate** of the impact of the program in the sample under study—this estimate is **internally valid**.
- TODAY, we are going to learn how to correctly implement randomized evaluations to minimize failures and how to correctly analyze and interpret the results of such evaluations





Challenge in Conducting RCTS

Program Design Challenge:

- Do you have resources to extend the program to everyone?
- Do you need to think about offering the program to certain group based, such as using eligibility criteria?
- Should the program be available to all—is it an entitlement?
- Are you trying to run a small program to find out if it works- sample size small?
- For researchers, limited budget. How best we design a program to detect treatment effect, if there is any.

Implementation Challenge

- Identify logistical or political challenges: NGOs, government or private sector
- Thank if control group finds out about the treatment, or benefits from the treatment, or is harmed by the treatment





When to Randomise? – Common Examples

Opportunity	Description
New program design	A problem has identified, not sure what solution to implement
New programs	Program is new, pilot tested
New services	When an existing program offers new service
New people	Expansion of existing program to a new group of people
New locations	Expansion of existing program to a new area
Oversubscription	More people are interested than capacity
Undersubscription	Not everyone who is eligible takes it up
Rotation	Benefits are to shared by rotation
Admission cutoffs	A cutoff, those just below the cutoff can be randomly admitted
Admission in phases	Resource constraints, random admission in phases over time





Is Randomization Unethical?

- How do we allocate resources when we have limited budget?
- Randomisation is often criticized as unethical since one group receives the program while the other group does not.
- When you are resource constrained and know that you can not offer to everyone even following some strict rules or eligibility criteria, how should we allocate funds?
- A general ethical issue is whether at some point, the control group should be offered the treatment or be compensated in another way if the treatment turns out to be extremely effective.
- If the treatment or other compensation is offered, should the researcher inform the control group at the outset of these possible future benefits?





Avoid Testing the Control Group!

- This method is often used when testing health related interventions which includes medical treatments.
- In some cases, a way to adhere to these ethical guidelines and still not provide treatment to the control group may be found.
- An example (Miguel and Kremer 2004):
 - An evaluation of the impact of providing deworming medicine on educational outcomes in Kenya.
 - After about two years, the control group students were diagnosed and treated.
 - (But the analysis was complicated by lack of knowledge about whether those in the control group who were later diagnosed with having worms had also had them two years earlier).





Avoid Testing the Control Group

- Researchers sometimes argue giving placebo to the control group.
- In social science the issue can be tricky in many cases.
 - On the one hand, informing control group about the treatment raise the expectation.
 - Unlike drug trial, the intervention in the form of cash, in-kind, information etc. raise challenges and often need to 'hide' from the control group.
- Example: Mental Health intervention during Covid19: Offering information and psychological counselling to treatment groups, but do nothing for the control group.
 - Calling control group over phone but offering nothing substantive raise even more ethical concerns.





How to Randomize?

There are three general ways to randomly assign individuals or groups to a treatment group and others to control group:

- 1. Lotteries
- 2. Gradual phase-in
- 3. Encouragement design

These three methods have different benefits and drawbacks, and researchers should choose the most appropriate method depending on the intervention and context





Lotteries

- Some individuals or groups will be randomly assigned to the control group while the rest assigned to one or more treatment groups.
- Lotteries are the best method to obtain unbiased, consistent estimates of programme effects.
 - The main impediment: denying treatment indefinitely to the control group may seem unethical.
 - Especially if the program provides clear benefits to the disadvantage group.
- Solutions!: divide the target population into three groups:
 - the clearly deserving (all are treated),
 - the somewhat deserving (treatment randomly determined by lottery), and
 - the not deserving (excluded from the program).
 - Randomize only for the somewhat-deserving group,
 - the results of the RCT apply only to this group.





Gradual Phase-in

- ALL groups are eventually given the opportunity to participate in the program, but SOME are allowed to participate BEFORE others.
- Example: Covid19 Vaccine promotion in Bangladesh and India
- This method is used frequently and perhaps the most commonly used method.
- Randomization by gradual phase-in involves two distinct randomizations.
 - 1. The evaluation sample is randomly assigned to two or more groups, which ensures that the individuals in these groups are essentially the same.
 - 2. The order in which these groups are offered the treatment is randomized.





Encouragement Design

- Treatment and control groups have access to the program, but some individuals or groups are randomly assigned to receive encouragement to take up the program
- Implemented by randomly selecting individuals to be provided information on, or an incentive to participate in, a program.
- Effective encouragement leads to higher take-up of the program in the treatment group than in the control group.
- NOTE: In this case we can evaluate only the impact of receiving an encouragement to take up the program that is evaluated (and its indirect effect on program take-up), rather than the direct impact of the program itself.
- Example: Educational intervention by government during school closures, reminder about the program, timing, etc. to some pupil.
- Check: <u>https://users.monash.edu/~asaduli/covid19research.html</u>





Level of Randomization

- Randomization could be performed at individual level or group level.
- Randomizing at the individual level is usually preferred because it provides a larger effective sample size.
- Factors to consider when deciding the level of randomization:
 - How many groups to have in the sample and how many individuals to sample in each group must also be determined
 - If spillovers occur at the individual level, but not at the group level, then randomisation is better at group level.





Other Recommendations on Randomization

Whether randomisation is performed at the individual or group level, several other practices should, in general be followed

 Conducting multiple randomized experiments on the same area/sample is sometimes convenient (and cost reducing).

Example:

- During Covid19, we ran educational intervention- remote learning for children, and mental health intervention for women- in the same geographic area, different households, different sample.
- It reduced the cost of running surveys, could use within village contact who helped running the surveys and intervention





Randomization Bias

- Sometimes participants in the treatment group exert more effort because they want to support the program.
- When the program staff behave differently when they know the program is being monitored.
- Participants in the control group may exert more effort because they do not want to appear weak.
- Following steps can be taken in order minimize randomization bias:
 - Use blinding where the participants do not know whether they are in the treatment group or control group.
 - Separate evaluation team from the program team.
 - Evaluation team does not know if a participant is under treatment or control.
 - Minimize interaction between treatment and control groups to reduce feelings of competition, anticipation, or demoralization in the two groups.
 - Assure staffs that their jobs or incomes will not be affected by performance of the program.





Social Desirability Bias

- Being treated could alter the behaviour, might induce respondents to answer strategically to continue receiving the benefits or not losing so!
- Social desirability bias occurs when a participant is influenced to deny undesirable traits and ascribe to themselves traits that are socially desirable.
- This often occurs when collecting sensitive and personal information such political/ religious views, intimate partner violence, personal finances, and sexual activity.





Example

- Forced Displacement, Mental Health, and Child Development: Evidence from the Rohingya Refugees, By Asad Islam, Abu Siddique, Tabassum Rahman, Tanvir Mojumder and Tanvir Shatil, 2022
- Intervention provided psychosocial support for mothers and play-based activities with their children
- Respondents that received support for a year might feel more inclined to provide favorable responses to enumerators relative to control group respondents.
 - However, in this study, control group participants also participated in social gatherings (pre-pandemic) organized and invited by BRAC Bangladesh.
- Use the Marlon-Crowne social desirability scale, which is survey module developed by social psychologists to measure a person's propensity to give socially desirable answers (Crowne and Marlow, 1960).
 - Developed by psychologists and has been validated in various contexts.
 - The questionnaire asks whether respondents have various too-good-to-betrue personality traits such as whether respondents are excellent listeners or never hurting anyone's feelings on purpose to create an SDB.





What Should We Do When Implementing RCTs?

- Researchers often face numerous obstacles when implementing RCTs which in return lead to biased estimates of program effects.
- Common issues to deal with:
 - Attrition,
 - spillover effects,
 - heterogenous effects
- Use of PAP and pre-registration

Reference:

Asad Islam, Michael Vlassopoulos, Abu Siddique, Tabassum Rahman, Deb Pakrashi, and Firoz Ahmed, 'Improving Women's Mental Health During a Pandemic'





Spillover Effects

- Treatment impact spills over onto control group members especially if the control group is in close proximity.
- Spillovers can lead to biased estimates of programme effects, despite 100 percent compliance with randomization.
- Random assignment should be done at a level that is high enough or large enough to eliminate, or at least minimize the possibility of spillovers.
- Consider whether spillovers are positive or negative.
 - If <u>positive</u>, the true effect will be underestimated, providing a <u>lower</u> bound of the true impact.
 - If <u>negative</u>, the true effect will be overestimated, providing an <u>upper</u> bound of the true impact.
- In many cases, estimating the extent of spillovers could be useful because it may have implications for when to implement , or how to design the program in question.
- References:
- The Value of Information in Technology Adoption: Theory and Evidence from Bangladesh | PAPER





Replication and Scalability

- Experimental studies are often conducted at a significantly smaller scale.
- A program that was successful in a relatively small area may not have the same effect when it is scaled up to a national level even if the small area is representative of the country as a whole.
- If the target population and area is too large, for example a country-wide program, then we can replicate the intervention under different samples and contexts.
- At the very least, show who are included in your sample, how they compare with the population to be scaled up. How your sample is 'representative' of population such as national sample survey.
 - Example: Asad Islam, Abu Siddique, Tabassum Rahman, Deb Pakrashi, and Firoz Ahmed <u>Awareness Campaigns during COVID-19 and Health</u> <u>Behavior: Evidence from Two Rapid Randomized Experiments in</u> <u>Bangladesh and India</u>, *Review of Economics and Statistics*, forthcoming





General Equilibrium Effects

- Implementing the program on a national scale may have economy-wide implications, which are often referred to as general equilibrium effects.
 - Consider a program that increases years of completed schooling among school-age children or an agricultural training program.
 - Additional schooling should provide them with higher incomes when they become adults.
 - But the program may not have as large an effect on income when it is scaled up to the national level because the increase in the educated labour force could reduce the wages of educated labour.
 - In other words, once it becomes more common to be well educated, the relative advantage in wages for well-educated people is likely to decrease.
- Little can be done about this potential threat to external validity, but at a minimum such possibilities should be kept in mind when extrapolating the results from a small program to a nationwide program.



MONASH



Scaling Up Is Not Always the Answer

- Although researchers find significant treatment effects from an intervention, it does not necessarily mean the treatment should be scaled up to a higher level such as national level.
- A few reasons why scaling up could potentially fail:
 - General equilibrium effects
 - Lack of resources and capacity to deliver intensive treatments at scale.
 - Weak incentives to government and other parties involved.
 - Especially in developing countries, improving management quality in public bureaucracies at scale is likely to be much more challenging than in private sectors.





Multifaceted Programs

- Recently researchers often test the effectiveness of multifaceted programs, especially in developing countries.
- Multifaceted programs provide a number of treatments simultaneously and observe the aggregate treatment effects.

Example:

- 1. Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa by Oriana Bandiera, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman, 2020
 - Intervention attempting to jump-start adolescent women's empowerment in Uganda by simultaneously providing girls vocational training and information on sex, reproduction, and marriage.
- 2. <u>Maternal Knowledge and Practices in Early Childhood Development: A Cluster-Experimental</u> <u>Evidence from Bangladesh</u>, Asad Islam, Umair Khalil and Tabassum Rahman
 - Implement a maternal skill training intervention focused on three key childcare areas: hygiene, nutrition, and engagement





Example

- A multifaceted program causes lasting progress for the very poor: Evidence from six countries, By Abhijit Banerjee, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, Christopher Udry, 2015
- They conducted RCTs in six countries to improve livelihoods among the very poor.
- The programs were implemented by six different organisations in six different countries (Ethiopia, Ghana, Honduras, India, Pakistan, and Peru).
- Although the objective of the program was the same, the intervention was not identical in all six settings.
- The core of the program is a productive asset transfer.
 - Plus training and support, life skills coaching, temporary cash consumption support, and access to savings accounts and health information or services.
- Although the asset type differed across countries, the principle in choosing the asset was consistent.
- The sample size used in the analysis varies from 925 households (Ethiopia) to 2,606 households (Ghana) from site to site.





Unbundling Multifaceted Program

- Generally, multifaceted program effects are unbundled into its components to check which components are more effective and if there are any components with no effects.
- Due to limited resources, providing multiple treatments simultaneously can be impractical, especially when scaling up.
- Thus, policymakers often try to implement only one or two components of the multifaceted program.
- However, implementing only component may not yield as much effect as the multifaceted program. It may even have no effect without the remaining components.





Example

- Unpacking a multi-faceted program to build sustainable income for the very poor, Banerjee et al. 2021
- Examines whether for the population targeted by the graduation program, it is possible to get similar results with just one of the main components of the program.
- The savings-only program has statistically significant positive effects on financial inclusion and consumption at two years, but both effects are much weaker by the three-year mark.
- The asset-only treatment has no evidence of any positive welfare effects after either two years or three years.
- It can be argued that there are complementarities between the initial program's components.





Pre-analysis Plan

- RCTs are generally less prone to issues such as data mining and publication bias due to PAPs and pre-registration of RCTs.
- PAPs are designed to avoid data mining by committing researchers, in writing, to a limited set of analyses before they begin to analyse the data.
- The most rigorous PAP would be written, and registered, before any data are collected and before the programme has been implemented.
- PAPs should specify:
 - (1) the population subgroups that will be examined
 - (2) the outcomes that will be considered
 - (3) the statistical and econometric specifications to be used





Conducting a Pilot Program

- Sometimes it is recommended to conduct a pilot program prior to program implementation.
- Pilot programs allow the researcher to learn about complications and problems that may be impossible to foresee.
- Despite the low statistical power due to small sample size, a pilot test will serve as a basis for a more successful large-scale intervention.
- Benefits of a pilot programme:
 - 1. The evaluation team will learn about how to implement the program and how to collect data.
 - 2. Will convince the program administrators and local govt. that the RCT is neither harmful nor politically threatening and thus could produce useful results.





Researcher's Involvement

- Involvement of researchers is crucial when conducting RCTs.
- Researchers must monitor how the program is being implemented (including the randomization) and how the data are being collected (including baseline data).
- Many randomized trials have been aborted or have provided useless results when researchers delegated monitoring to local collaborators.
- Researchers should be prepared to go to the field frequently to ensure that the randomized trial is being correctly implemented.
- After more experience has been gained by local collaborators, monitoring could be reduced, but it is better to continue monitoring since even competent local collaborators could do a poor job if they are not interested.





Cost Effectiveness

- It is recommended to include a cost effectiveness analysis
- Unlike researchers, policymakers have to justify their actions to the public and these interventions have to be effective as well as cost effective.
- Cost-effectiveness analysis shows the effect a program achieves on one outcome measure for a given cost incurred.
 - Cost effectiveness is calculated by summing up all the program costs and then divided by the total impact the program had on a single outcome measure.
 - On the other hand, cost benefit analysis translates all the different benefits and costs of programs onto one scale (usually a monetary scale).
- This can be used to compare different programs or to assess whether a program is worth investing in.
- Example: Asad Islam, Vy nguyen, Russell Smyth and Zabid Iqbal "<u>Financial</u> <u>Diaries and Women's Money Management Behavior: Evidence from an RCT</u>"





Rethinking of doing RCTs?

- When there is insufficient observations
- If it is unethical
- Are you just interested in testing theory? Do you see any policy relevance?
- When the research cannot implement the ideal research design due to government restrictions
- If the intervention is not scalable, and is not policy relevant
- When the local community will end up in disadvantage or no benefits after the intervention

