



HAL
open science

Randomised Controlled Trials

Carlo Barone

► **To cite this version:**

| Carlo Barone. Randomised Controlled Trials. 2023, pp.4. hal-04021987

HAL Id: hal-04021987

<https://hal-sciencespo.archives-ouvertes.fr/hal-04021987>

Submitted on 9 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives | 4.0
International License

LIEPP METHODS BRIEF

This brief is part of a set of methods briefs published by LIEPP. As such, it is supported by the ANR and the State under the "Investissements d'avenir" programme within the framework of the LABEX LIEPP (ANR-11-LABX-0091, ANR-11-IDEX-0005-02) and the IdEx Université Paris Cité (ANR-18-IDEX-0001).

Randomised Controlled Trials

Carlo BARONE (Professor of sociology, Sciences Po, CRIS-LIEPP)

carlo.barone@sciencespo.fr



Distributed under a Creative Commons Paternité.
Attribution-NonCommercial-NoDerivatives | 4.0 International License

<https://www.sciencespo.fr/liepp/en.html>

How to cite this publication:

BARONE, Carlo, **Randomised Controlled Trials**,
Sciences Po LIEPP Methods Brief n°1, 2023-03

Cette publication est également disponible en français :

BARONE, Carlo, **Essais Contrôlés Randomisés**,
Sciences Po LIEPP Fiche méthodologique n°2, 2023-03

IN A NUTSHELL:

Randomised controlled trials (RCTs) aim at measuring the impact of a given intervention by comparing the outcomes of an experimental group (receiving the intervention) and a control group (not receiving it), to which individuals are randomly assigned. It is a useful quantitative method of *ex ante* evaluation, to test the impact of a program at a stage when it has not yet reached the totality of its target population (making the control group possible).

Keywords: Quantitative method, experimental method, experimental/treatment and control groups, random assignment, treatment, contamination

I. What does this method consist of?

Randomised Controlled Trials (RCTs) assess the impact of a policy by comparing two groups: one of them is given access to the policy (experimental group), while the other is temporarily excluded from the policy (control group). The researcher translates the goals of the policy into quantitative outcomes measures and assesses the efficacy of the policy by measuring these outcomes across these two groups. If the experimental group displays better values on these outcome measures, we conclude that the policy is effective. However, this conclusion is valid if, and only if, we can assume that the two groups were perfectly equivalent. This is why the assignment to the two groups must be done randomly: if the sample is sufficiently large, the random assignment ensures that the two groups are, on average, initially *equivalent on all characteristics*, known or unknown by the researcher, measured or unmeasured in the evaluation study. Hence, any difference in the outcomes observed after the implementation of the policy can be interpreted as an impact of the policy.

When conducting an RCT, the researcher draws a sample of individuals and invites them to participate in the study, explaining that they may be assigned to either the experimental or the control group. Among the participants who have accepted to participate, half of them will be randomly assigned to the treatment and half to the control group. This 50%-50% ratio is the most common one because it results in more precise estimates than unbalanced ratios (e.g., 70%-30%). Before delivering the intervention, we may carry out a baseline measurement of the outcomes. This is not strictly necessary, but it is often done for several reasons, for instance because it allows the researcher to study the impacts of the treatment in a more dynamic way by comparing variations in the outcomes across the two groups.

While the randomisation is a necessary condition to make plausible causal claims when comparing the two groups, it is not a sufficient condition. In particular, the control group must remain excluded from the policy during the entire period of implementation of the policy, that is, we must avoid any form of treatment contamination¹. This implies, for instance, that individuals of the two groups do not communicate about the treatment objectives and contents. Moreover, when individuals are assigned to the control group, they may react by trying to replace the treatment with a similar treatment. Treatment contamination and replacement can invalidate causal inferences if they happen on a large scale. Hence, the key requirement is that the control group acts 'as usual' and it is important that the researcher designs and presents the study in such a way to ensure that this is the case. Hence, while the randomisation is important, it is no less important to ensure the highest degree of control of these experimental conditions. The term 'randomised controlled trial' thus describes the two key requirements to make solid causal inferences: random assignment and control of the experimental conditions.

¹ The word "treatment", imported from clinical research, is often used to refer to the intervention under study.

II. How is this method useful for policy evaluation?

RCTs are aimed at estimating the causal impacts of policies, that is, at assessing whether policies produce changes in the outcomes reflecting their goals. The main challenge is that, even if a given policy is completely ineffective, these outcomes may change because of other policies, other economic or socio-cultural changes affecting these outcomes. For instance, we may deliver a training programme to unemployed individuals to improve their employability and observe the employment rates of individuals participating in this programme. However, it is unclear whether any observed change in this outcome can be attributed to the policy. For instance, it could be due to the economic cycle as well to any kind of other economic, labour or welfare policy (e.g., fiscal incentives to hire unemployed individuals, changes in eligibility rules for unemployment benefits, etc.). Hence, a simple pre-post comparison would be unable to isolate the genuine causal impact of this policy.

RCTs are not the only type of causal impact evaluation method, for instance regression discontinuity designs are another option (see separate brief). RCTs are a form of *ex ante* evaluation, that is, they must be carried out before the policy is delivered to the whole population of potential beneficiaries. This is because RCTs suppose that the policy is not delivered to some individuals, who constitute the control group. If the policy has already been generalised, RCTs are unfeasible. We may then resort to other types of causal impact evaluation methods to isolate the genuine causal impact of the policy.

III. An example of application: what messages best favour tax compliance?

Tax compliance, that is the truthful reporting of taxable income and the timely payment of taxes, is essential to finance public services. Researchers partnered with the tax authority in Belgium to test the impact of different messages encouraging tax compliance (De Neve et al, 2019). Between 2014 and 2016, researchers randomly assigned around 2.5 million taxpayers to receive different messages: simplified messages presenting the key information in simpler terms, deterrence messages aimed at making the consequences of non-compliance explicit, and tax morale messages aimed at motivating taxpayers to appreciate the importance of compliance for the provision of public goods. The remaining 4 million taxpayers were assigned to a comparison group where taxpayer communication remained unchanged (this sample size is exceptional, most RCTs are based on a few hundreds or thousands cases). Using administrative data, researchers measured the impact of the intervention on the probability of making a payment or filing their taxes, and the amount of reported income. Simpler communication had the largest effect on tax compliance, inducing people to file and pay their taxes sooner. Adding deterrence messages further enhanced compliance, while tax morale messages were ineffective.

IV. What are the criteria for judging the quality of the mobilisation of this method?

In some contexts, experiments are unfeasible because the risks of treatment contamination or replacement are too high, for instance when treated and controlled, individuals can easily communicate on the contents of an information intervention and are highly motivated to do so. Some policies cannot be tested with an RCT because, by construction, they involve the whole population, therefore we cannot temporarily exclude the control group. For instance, this is the case of several macroeconomic, foreign or defense policies (for instance, a change in the military expenses).

Moreover, while most commonly we assign individuals to the treatment or control group, sometimes we may assign whole families, streets or villages to the treatment or to the control status. For instance, this is the case when a given intervention is more effectively delivered, or can only be delivered, at these supra-individual levels. These types of higher-level randomisations (cluster randomisation) can be necessary or extremely practical, but they demand large sample sizes and thus large budgets.

Finally, we should keep in mind that internal validity (i.e., the strength of causal inferences in the case under study) is only one of the quality criteria in evaluation research. Another important criterion is external validity, that is, the generalisability of conclusions beyond the sample under study. This second criterion, when applied to RCTs, demands that we draw large, random samples of the population under study and that participants do not drop out of the study. A third important criterion relates to the validity and reliability of the outcome measures, including the capability to observe the long-term outcomes of a policy, and the coverage of all potential (positive and negative) effects of the policy.

V. What are the strengths and limitations of this method compared to others?

As explained above, the main strength of the RCTs is that they allow assessing the genuine causal impact of a policy before delivering it to the whole population of beneficiaries. In clinical research, RCTs are the standard method to assess the efficacy of any kind of therapy or medicament and they are increasingly used for the evaluation of public policies, more so in educational, labour market, health and housing policies.

The most common applications of this method involve the randomisation between two groups of individuals. However, sometimes we may arrange three or more groups of individuals in order to compare qualitatively different variants of an intervention or different dosages of the intervention. For instance, in a study to promote the use of bike-sharing services, we may compare the control group to a first treatment group that has information about bike-sharing, a second treatment group receiving a monetary incentive and a third group receiving a larger monetary incentive.

RCTS are not always feasible. In particular, policymakers or potential participants may refuse the principle of randomisation. Indeed, some people argue that experiments are 'unethical' because they exclude the individuals of the control group from the benefits of the policy. This critique forgets that the exclusion is temporary, that is, it lasts only for the time needed to demonstrate that policy is effective. This temporary exclusion allows assessing if the policy is effective before generalising it to the whole population. Moreover, the resources available in ex ante evaluation studies allow treating only a small share of the total population, so treating everyone would anyway be impossible: the random assignment instead gives everyone the same chances of being treated.

It is critically important that researchers explain in simple terms why randomisation is ethical and why it is necessary to ensure the reliability of the comparisons between the two groups. Whenever it is possible, the social acceptability of the randomisation can be increased by creating a waiting-list, that is, the control group receives the policy at the end of the study, or a compensatory treatment (a treatment that is different from the one under study and that does not affect the outcome of the study). For instance, in a study providing information on childcare services to pregnant mothers to enhance the recourse to these services, the control group may receive this information at the end of the study or may receive some other type of information for instance on healthy practices during pregnancy. If a waiting list is created, it is not possible to observe long-term outcomes because the control group is no more excluded from the intervention. Waiting lists and compensatory treatments can be used also to reduce the risk that the individuals assigned to the control group drop out of the treatment. It is indeed important that the dropout rates of the two groups are similar in order to preserve their equivalence throughout the study.

Compared to laboratory experiments, RCTs have higher ecological validity, meaning that we are studying people in real life situations and in naturalistic contexts. Hence, the risk that their behaviour is influenced by the awareness of being part of a study is less important. At the same time, relative to laboratory experiments, RCTs allow a lower degree of control on the behaviour of participants. In clinical and psychological experiments, the awareness of being treated is often neutralised by administering placebos to the control group, that is, treatments that are specifically designed to have no effect. In social policies,

this practice is less common because we tend to regard the benefits deriving from the awareness of being treated as an integral part of the policy.

Most fundamentally, while RCTs are a reliable tool to assess the causal impacts of policies, they are not in a strong position to investigate the underlying processes. For instance, if an RCT concludes that a policy is ineffective or less effective than expected, this method is unable to explain what did not work and how we may improve this policy. For this reason, it is extremely useful to integrate RCTs with qualitative techniques of process evaluation. Moreover, the beliefs and perceptions of the policy that beneficiaries and implementers have may be better investigated using qualitative or survey interviews.

Some bibliographical references to go further

De Neve, Jan-Emmanuel. and Imbert, Clement. and Spinnewijn, Johannes. and Tsankova, Teodora. and Luts, Maarten. 2019. "How to Improve Tax Compliance? Evidence from Population-wide Experiments in Belgium." Working paper.

White, Howard. and Sabarwal, Shagun. and De Hoop, Thomas. 2014. Randomised Controlled Trials, Notes méthodologiques, Évaluation d'impact no 7, Unicef. <https://www.unicef-irc.org/publications/pdf/MB7FR.pdf>

Gibson, Michael. and Sautmann, Anja. Last updated April 2021. Introduction to randomized evaluations, Abdul Latif Jameel Poverty Action Lab. <https://www.povertyactionlab.org/resource/introduction-randomized-evaluations>

Gertler, Paul. and Martinez, Sebastian. and Premand, Patrick. and Rawlings, Laura. and Vermeersch, Christel. 2016. Impact Evaluation in Practice, second ed. World Bank Group <https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464807794.pdf?sequence=2&isAllowed=y> (Chapters 3 and 4 of this manual)