Fact sheet: Randomised controlled trials (RCTs)

Top TSIP evaluation tips

- The RCT is the gold standard for impact evaluations in social intervention research because it can confidently distinguish between the effect of your programme versus the effect of other factors on your outcome indicator(s)
- Randomisation ensures that your treatment and comparison groups are as comparable as possible
- However, RCTs are not be appropriate for all circumstances (e.g. if your programme has not been piloted yet)

Description

RCTs are the gold standard among experimental designs[1] in social intervention research. An RCT is defined as an evaluation of an intervention which is manipulated so that at least one randomly allocated sub-group receives the treatment and at least one does not.

Rationale: why randomise?

Through random assignment you largely remove the risk of allocation bias, allowing for two statistically identical groups to be created. If you then administer your programme to one group but not the other, any differences in outcome between these two groups after the programme can be confidently attributed to the programme (i.e. not to external factors) since the two groups only differ in whether or not they receive the treatment. This is because randomisation evenly distributes characteristics across groups (e.g. if you have 50 men and 50 women, and you randomise all of them to treatment and control groups, you will get close to 25 men and women in both groups, and all other observable and unobservable characteristics will also be evenly distributed)

Optimising the design

Since RCTs are the gold standard, it does not need to be improved if it is conducted properly. However, the following factors can be addressed to optimise the design:

- Sample size[2]: Ensure that a large enough sample size is available in order to be able to detect the expected effect
- Stratification[3]: If you have a small number of units for randomisation, you may want to stratify the units (according to important known confounders) to ensure that the two groups will be balanced on crucial characteristics. For example, if you suspect

[1] Learn more about randomised experiments on pp. 246-278 in Shadish 2002
[2] Learn more about sample sizes on pp. 127-138 in Togerson 2008, or on the fact sheet on quantitative methods
[3] Learn more about stratification on pp. 32-38 in Togerson 2008

that your programme has a different effect on boys compared to girls, and if you have only 8 pupils for randomisation, and four of them are male and four are female, you create separate allocation schedules for boys versus girls because this way you can ensure that you will randomly allocate equal percentages of each gender in both treatment and comparison groups than if you just randomised all 8 pupils together. This is however not necessary if you have large numbers of pupils (e.g. 200) because with this number of pupils, it's extremely likely that you will have equal percentages of each gender in both treatment and comparison groups.

- Attrition[4]: Ensure that all participants remain in the trial to avoid being left with only a subsample at post-test that could potentially differ systematically from the original participant group. This could then lead to misleading results.
- Blinding[5]: This denotes that the researcher is masked or 'blinded' to the identity of the group allocation of the participants when undertaking post-tests. For example, when measuring a child's problem behaviour by observation, the researcher should not know whether this child was in the treatment or comparison group. This prevents biased assessment. Sometimes participants are also blinded to the true nature of the experiment (double blinding) to simulate a placebo effect.
- Allocation concealment[6]: This prevents foreknowledge of allocation of an individual by the researcher, participant or practitioner. For example, you have generated a random allocation sequence for 6 participants (e.g. TTCTCC where T is treatment and C is control) and you have a pile of 6 cards with one participant name on each of them. If you spot that Mrs Smith (and you think she's just a lovely old lady who really deserves to be on the treatment group) is on the third card and you know in advance that this will mean that she will be put into the control group, you may cheat and swap her card with one that would be allocated to the treatment group. If random allocation is be undermined this could render the trial uninterpretable and/or produce misleading results.

Units of randomisation

You can either randomise individuals or clusters (i.e. groups of individuals, e.g. schools). Generally, it is better to randomise individuals because you need much smaller sample sizes in order to demonstrate a significant effect, but in some cases that is not possible. For example, if you want to test the effectiveness of school lunches on child growth, you will have to randomise schools rather than pupils because children are very likely to share their food (regardless if their friend is in the control group or not).

Conditions most conductive to random assignment

- When demand exceeds supply it may be fairest to allocated treatment on a random basis

[4] Learn more about attrition on pp. 51-54 and pp. 139-140 in Togerson 2008
[5] Learn more about blinding on pp. 71-72 in Togerson 2008
[6] Learn more about allocation concealment on pp. 45-50 in Togerson 2008

- When it is impossible to give treatment to everyone (e.g. staff constraints, resource constraints), this can be used to their advantage and the evaluation design can use a wait-list design, gradually phasing in the treatment to units.
- When it is unknown which treatment is the most effective out of several options
- When it is unknown which participants are the most in need for treatment
- When participants do not express any preference in treatments
- When lotteries are a socially accepted means of distributing resources

When random assignment is not feasible or desirable

- When quick answers are needed
- When there is already considerable evidence available on a treatment (e.g. experimental studies)
- When it's not possible to allocate people to variables (e.g. age)
- When a programme is in an early pilot stage (to avoid waste of resources)
- When you have very limited resources. However:
    - It is worth considering how much the answer to your research question is worth in terms of financial costs (e.g. good RCT can give better answer than 5 pre-posts, so it may actually be cheaper to do the RCT.)
    - Many quasi-experimental methods actually are similarly (or even more) expensive than an RCT since you have to collect additional data to be able to adjust for confounding variables in your analyses since your comparison group is not (statistically) identical to your treatment group at baseline. For example, if the outcome indicator data is collected by others and freely accessible to you (e.g. National Pupil Database), an RCT may actually not cost much at all since you don't have to do the data collection.

Books referenced in this fact sheet

Morgan, S. L., Winship, C. (2007), *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, Cambridge University Press, Cambridge

Shadish, W. R., Cook, T. D., Campbell, D. T. (2002), *Experimental and Quasi-experimental Designs for Generalized Causal Inference*, Wadsworth Cengage Learning, Belmont

Togerson, D. J., Togerson, C. J. (2008), *Designing Randomised Trials in Health, Education and the Social Sciences*, Palgrave Macmillan Basingstoke

Glossary

Attrition—when participants in either treatment or comparison groups leave the evaluation (e.g. they disappear or refuse to continue).

Allocation bias—when the intervention is differentially assigned to the population, which leads to selection bias.

Bias—A term denoting that a known or unknown variable is or may be responsible for an observed effect other than the intervention.

Confounders— A variable associated with cause and outcome; can mask a true relationship between another variable and outcome. E.g. if you are trying to measure the impact of an afterschool maths club on children's numeracy in London, and your comparison group consists of schools in a less affluent area of London, the general quality of the schools' teaching might be a confounding variable – i.e. if numeracy scores improve less among the comparison group, that might be due to a poorer quality of teaching rather than the lack of the afterschool maths club.

Counterfactual—what change in outcome indicators would have been seen in the absence of the intervention. It is not possible to observe the counterfactual (since it does not actually exist), but it can be simulated approximately with a comparison group.

Experimental designs—defined as a design that involves deliberatively introducing an intervention to observe its effects against a (randomly assigned) comparison group.

Independent variable—the things whose impact is being measured (in our case this is generally the intervention).

Natural experiment—evaluation of an intervention not manipulated by the researcher, meaning that cases cannot be randomly allocated to control and treatment groups.

Quasi-experimental study—a design that involves deliberatively introducing an intervention to observe its effects against a non-randomised counterfactual.

Selection bias—a bias that happens when choosing to put participants into a treatment group and comparison group for a specific reason (e.g. because they failed in their application to be on the programme) rather than through randomisation. This means that the treatment and control groups differ already *before* they receive the intervention.

Stratification (or blocked randomisation)—This is a process whereby randomisation is restricted (e.g., by blocking) such that any important known confounders are balanced between the groups.