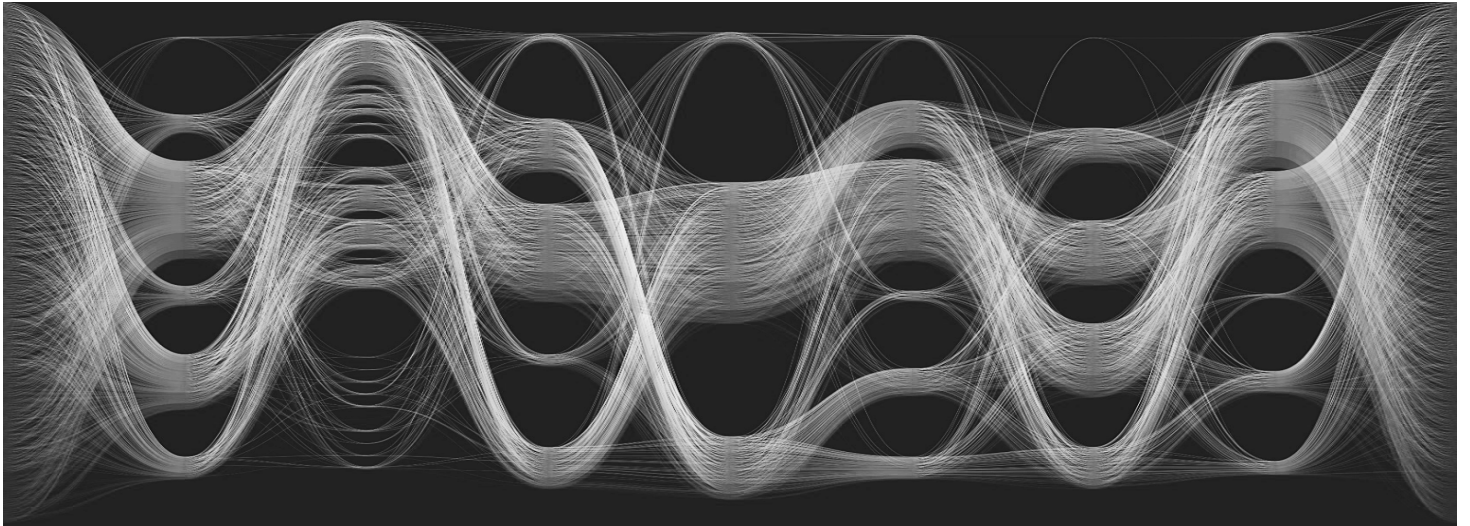




Australian  
National  
University



# Running Randomised Trials and the ANU Centre for Social Research & Methods

ANU Centre for  
Social Research  
& Methods

## 1. Introduction

Randomised controlled trials (RCT's) are an increasingly used method for evaluating public policy across developed and, in particular, developing countries. RCT's are reliable and effective and in a resource – limited environment it is essential to make sure funding is being spent on policies and programmes that work.

RCT's are a useful addition to a suite of policy and programme evaluation tools but many people are uncertain about how to begin using RCT's in their policy evaluations. Simply, RCT's randomise who receives an intervention and measures the differences between those who received the programme/service (treatment) and those who did not (control). A good RCT will then be able to tell us whether there was a causal link between the intervention and the outcome.

This guide outlines the key steps in designing and running a good RCT in a public policy environment, and introduces key staff associated with the ANU Centre for Social Research and Methods who have experience designing, running and analysing RCTs.

## 2. Preparation - It's better to do the right thing slowly than the wrong thing quickly.

The planning and preparation stage is critical for an RCT. Reaching the end of a project only to discover during analysis that you have missed a key variable or incorrectly measured a relevant indicator will lead to significant wasted resources. It is also important to remember that RCTs cannot be run after an intervention has begun, so it will need to be a part of early stage policy design.

The first step is developing your understanding of the policy context. This may include running a needs assessment, developing a logic model or programme theory. These steps will help to articulate any assumptions or risks that might prevent the programme or policy from being successful. The development of a literature review, including exploring the findings from other relevant evaluations will also help support your trial by highlighting the lessons learned by others in the field.

The next step in the process will include considering what indicators you hope to use to measure change. What would success look like? Specifying these key evaluation questions early in the process is critical to building your sample, your units and method of randomisation and the contextual variables.

Your indicators should follow the SMART test:

**Specific** – your evaluation measures the impact on a specific population, so keep indicators specific as well. Your findings can be extrapolated, but broad indicators make change harder to measure.

**Measurable** – measuring some indicators can be difficult, but many proxies are available? Spend time researching how you can operationalise your variables to find better measures for your indicators.

**Achievable** – Use indicators that can measure change as well as the total elimination of an issue. Indicators should be realistic, not aspirational.

**Relevant** – make sure your indicators are defensibly linked to your desired outcomes

**Timed** – can your indicators reasonably be met within your data collection timeframe?

## 3. Design - Bad design is smoke, while good design is a mirror.

Once you are confident in what your policy will look like, and who it will be applied to, you can begin to design your trial.

The first step is *choosing the right population*, selecting an appropriate and representative sample of the population from which to recruit. You will need to select a sample size which has sufficient statistical power to be able to illustrate statistical significance in the results.

**Statistical significance** refers to the probability that the results we observe are not purely based on chance. Conventions in the literature state that significance levels above 90% - preferably at 95% - are sufficient. This means that, either 5% or 10% of the time, the results we observe are by chance.

**Statistical power**, instead, refers to the probability of detecting an impact when there is one. The inverse, then, is how likely are we to miss impact when it occurs (thus generating a “false negative”)? A number of factors determine statistical power: the sample size, the minimum detectable effect size (i.e. how sensitive must the test be), the outcome variable’s underlying variance, the proportion that are in treatment and control, and – if it is a cluster RCT – the intra-cluster correlation. Convention allows 80% to be a sufficient level of power.

The next step is choosing the *sampling strategy*. Exclusion and/or inclusion criteria should be developed. You will also need to decide if you need to stratify or use sub-groups to strengthen the representativeness of your sample. Your *method of randomisation* will also need to be considered at an early stage. Key approaches include<sup>1</sup>;

Design	Most useful when	Advantages	Disadvantages
Basic Lottery	<ul style="list-style-type: none"> <li>– Program oversubscribed</li> <li>– OK for some to get nothing</li> </ul>	<ul style="list-style-type: none"> <li>– Familiar</li> <li>– Easy to understand</li> <li>– Easy to implement</li> <li>– Can be implemented in public</li> </ul>	<ul style="list-style-type: none"> <li>– Control group may not cooperate</li> <li>– Differential attrition</li> </ul>
Phase in	<ul style="list-style-type: none"> <li>– Expanding over time Everyone must receive treatment eventually</li> </ul>	<ul style="list-style-type: none"> <li>– Easy to understand</li> <li>– Constraint easy to explain</li> <li>– Control comply as expect to benefit later</li> </ul>	<ul style="list-style-type: none"> <li>– Anticipation of treatment may impact short run behaviour</li> <li>– Difficult to measure long term impact</li> </ul>
Rotation	<ul style="list-style-type: none"> <li>– Everyone must get something at some point, not enough resources a year for all</li> </ul>	<ul style="list-style-type: none"> <li>– More data points than phase in</li> </ul>	<ul style="list-style-type: none"> <li>– Difficult to measure long term</li> </ul>
Encouragement	<ul style="list-style-type: none"> <li>– Program has to be open to all comers</li> <li>– When take up in general is low but can be impacted with incentive easily</li> </ul>	<ul style="list-style-type: none"> <li>– Can randomize at individual level even when program isn't</li> </ul>	<ul style="list-style-type: none"> <li>– Measures impact of those who respond to the incentive</li> <li>– Need big enough inducement to get change in take up</li> <li>– Encouragement may have direct effect</li> </ul>

Another important issue to take into consideration is how you will recruit your participants. Each method will have its own risks and biases. For example, the use of volunteers can lead to bias as there are some people who are more likely to self-select into an intervention than others, usually people who are already interested in the intervention. This may skew your results to imply greater take up than will be replicated when the programme is rolled out more widely.

You may decide to undergo a baseline survey of your population. Running a baseline survey is not necessary for measuring the impact of a well-designed policy or programme. However, baseline surveys allow you to validate your treatment and control samples are equivalent in terms of indicators and relevant contextual variables. Baseline results also allow us to measure heterogeneous effects (i.e. subgroup analysis) when the groups are defined by variables that could change over time. Finally, if sample sizes are

<sup>1</sup> Table taken from Abdul Latif Jameel Poverty Action Lab Executive Training Course material found at <http://ocw.mit.edu/resources/res-14-001-abdul-latif-jameel-poverty-action-lab-executive-training-evaluating-social-programs-2009-spring-2009/>

reasonably small, baseline surveys increase the precision of the estimates by allowing one to control for observed heterogeneity. Depending on the size and cost of the RCT, or the intervention to be tested, a pilot study can be a good opportunity to refine methods and make sure the approach is a good fit for purpose.

It can be helpful to collect qualitative data as well as administrative or quantitative data. This will help support your findings. A simple exit survey may be useful.

#### **4. Pre-analysis – Maximising validity by locking yourself in.**

At this stage, it is also worth considering a pre-analysis plan that can then be registered with the American Economic Association (AEA)—<https://www.socialscienceregistry.org/> or similar registries. Pre-analysis plans represent ‘best practice in running trials as they ensure all details of the trial have been thought through before collection of data and significantly increase the robustness with which findings are viewed. According to the World Bank’s Development Impact discussion, “A pre-analysis plan is a step-by-step plan setting out how a researcher will analyze data which is written in advance of them seeing this data (and ideally before collecting it in cases where the researcher is collecting the data).” A Pre-analysis plan generally includes most of the following: A description of the sample to be used in the study; Key data sources; Hypotheses to be tested throughout the causal chain; How variables will be constructed; The treatment effect equation to be estimated; The plan for how to deal with multiple outcomes and multiple hypothesis testing; and The model to be tested.

#### **5. Conducting your trial – Now is not the time to relax**

The key in successfully running a trial is to ensure that the intervention and measurement of outcomes are done in the way they were intended. This will require the development of standard operating procedures for those involved in the trial, including quality control procedures and issues management. You will also need to ensure that there is a way to monitor the implementation and compliance throughout. This may be through recordkeeping or having a member of the team involved in oversight. There is also a risk of problems of human error that will need to be addressed through training or instruction. It is also a good idea to have a contingency plan for technical problems. You will need to consider how will data be collected and entered. Will you need to purchase new software or provide training? Will the software you already have be sufficient for your needs?

#### **6. Analysis – Good design trumps good econometrics**

Analysing the data you have collected can be relatively easy if it follows a simple test and control group design. However, more complex samples may require more detailed econometric analysis. Potential methods could include difference-in-difference, instrumental variables, matching or non-linear approaches. If you are using more than one outcome variable, then you may need to adjust standard errors to avoid ‘fishing’ for results. Importantly though the timing and method of outcome assessment will have been decided during the pre-analysis stage based on how long the intervention is likely to need to work.

Look back on your logic model and analysis plan and revisit the early assumptions to see whether the indicator results are what you expected. If results are in line with your earlier expectations then determining effectiveness should be relatively easy. If any indicators show a different result, however, you will need to investigate the result to understand how it impacts on your conclusions.

Some questions you may need to ask during this stage include;

- Is your result supported by a range of indicators or are you focusing in on one in particular? Don’t ignore or explain away small, possibly negative, findings changes just because another indicator returned a result you like. Explore the results as a whole.
- Could any of your findings be the result of an unintended consequence, spill over or bias? These risks must be explored, even when you have taken steps to alleviate their impact.

- Is further secondary analysis necessary? You may need to do further analysis to understand the reason for a particular result. Consider issues such as a flaw in the data that was not previously identified or a mistaken assumption in the original thinking.

## **7. Conclusions – Figuring out what it all means**

There is always a risk that you will find the intervention had no impact, or perhaps even a negative impact. Remember that this is a good result. You have discovered an issue that can, hopefully, now be fixed. Positive findings should also be contextualised to make sure that the internal validity of a randomised trial is supported by external validity and relevance to the policy process. Sharing your results will require you to distil your results into a narrative that makes the policy implications clear to policy makers. Avoid technical jargon and focus on the lessons learned through the process. You may also like to include a message on the value of the randomised trial to the policy development process to help build a culture of experimentation in your Department.

## **8. The ANU Centre for Social Research and Methods**

There is a wealth of experience in designing and running trials within the ANU Centre for Social Research and Methods, with a particular focus on complex social and economic policy issues. Below are brief bios of key staff from within the centre who could work on trials.

### Prof. Matthew Gray, Director

Matthew is Professor of Public Policy at The Australian National University. Previous appointments include Director of the Centre for Aboriginal Economic Policy Research and Deputy Director of the Australian Institute of Family Studies. He has published research on a wide range of social and economic policy issues and has undertaken major evaluations of government policies and programs including the family law system, income management, service delivery models and place-based interventions. From 2005 to 2010 he was responsible for the Longitudinal Study of Australian Children.

### Associate Professor Nicholas Biddle, Deputy Director

Nicholas has extensive experience in the design and analysis of survey data on economic and social policy issues. He has been involved in the development, implementation and reporting of a range of policy and program evaluations with a particular emphasis on evaluations that utilise both qualitative and quantitative information. He has a Bachelor of Economics (Honours), a Masters of Education, and a PhD in Public Policy. He has published in national and international journals with a particular focus on education and labour market outcomes. He previously held a Senior Research Officer and Assistant Director position in the Methodology Division of the Australian Bureau of Statistics working on the design and analysis of a range of economic and social collections.

### Associate Professor Ben Edwards, Senior Fellow

Associate Professor Ben Edwards is a Senior Fellow at the ANU Centre for Social Research and Methods where he is focused on policy relevant research on child and youth development and advising and supporting longitudinal studies (the Longitudinal Study of Australian Youth, Ten to Men, Home Interaction Program for Parents and Youngsters (HIPPY) and a new longitudinal of children in the Philippines). Internationally, he advised the Organisation for Economic Cooperation and Development (OECD) on the measurement of non-cognitive skills in longitudinal studies. Previously as Executive Manager of Longitudinal Studies at the Australian Institute of Family Studies he had leadership role in the development of Growing Up in Australia: The Longitudinal Study of Australian Children, Australia's national longitudinal study of children of over 10,000 children; The Australian Temperament Project, a birth cohort study of Victorian children born in 1983; and Building a New Life in Australia: The Longitudinal Study of Humanitarian Migrants, and in the development of a Commonwealth data linkage integration unit.

#### Dr Naomi Priest, Fellow

Naomi's broad research interest is to integrate social and epidemiologic methods to examine and address racial-ethnic inequalities in child and youth health and development across populations and place. This includes social epidemiology and qualitative research to understand racial-ethnic differences in child health and development and explanations for observed differences, particularly the patterns, mechanisms and prospective influence of adverse early life exposures and stressors, including racial discrimination. She is also interested in development of racial/ethnic attitudes, bias, stereotypes and prejudice and ethnic-racial socialisation processes and outcomes among children from stigmatised and non-stigmatised groups. A third area of her research is focused on developing, implementing and evaluating initiatives to counter discrimination and promote diversity and inclusion. Naomi has a PhD in population health (Melbourne) where she conducted a qualitative participatory study exploring Aboriginal perspectives of child health and wellbeing in an urban area. She also has a B App Sc honours in Occupational Therapy (University of South Australia). In 2014-15 she was a Visiting Scientist at Harvard T.H. Chan School of Public Health.

#### Rob Bray, Research Fellow

Rob Bray joined the ANU in 2010 after a long career in the Australian Public Service working in a diverse range of policy areas including employment, health, housing, regional development and social security. He was awarded the Public Medal in 2010 for his work as a policy analyst and researcher. He has a BA from Adelaide University. He has undertaken extensive consultancy work. This includes projects for the New Zealand Government and the Departments of Social Services, Human Services and Employment. Recent projects include the evaluation of Income Management in the Northern Territory and reviewing the Evaluation Plan for 'jobactive'. He has worked extensively on the interaction of the tax and transfer systems, both in the public service as part of the Pension Review and Henry Review, and at the ANU.