

POLICY ON TRIAL

Randomised trials are the best tool we have for finding out if policies really work, writes **Ross Farrelly**

Policymakers aim to develop programs that will benefit citizens. They claim, implicitly or explicitly, to have certain knowledge of the causal relationship between the actions they plan to take and the outcomes they wish to achieve. This is emphasised when, as Prime Minister Kevin Rudd does so often, they express their wish to develop ‘evidence-based’ policy. It is well known in scientific circles that there is one gold standard technique for discovering such a causal relationship—the randomised trial. If policymakers want to be able to claim that their policies will work, they should subject them to randomised trials beforehand. Such trials present the policymaker who genuinely wants to know how to make a difference with a powerful and irrefutable tool to put theories to the test and to draw fact-based conclusions.

Randomised trials are, despite their name, the least random and most scientific method known for testing a hypothesis. They are the epitome of rational inquiry. In a randomised trial, the burden of proof is placed on the facts themselves, and ideology, beliefs, and vested interests are put to one side. In a randomised trial, the truth, as indicated by the data and as revealed by the experimental design, is laid bare for all to see and the facts are allowed to speak for themselves. Randomised trials, preferably double-blinded and placebo-controlled, have been the benchmark for scientific inquiry since R. A. Fisher’s ground-breaking work in the 1920s. Clinical trials are mandatory for every drug approved by the Therapeutic Goods Administration. In short, except for trivial and self-evident cases, the randomised trial is the one and only means of establishing a cause-and-effect relationship between one phenomenon and another.

What are randomised trials and what can they do?

A randomised trial starts with a hypothesis—a statement of fact that the trial puts to the test. For example, one randomised trial in Kenya tested the hypothesis that the provision of textbooks would raise students’ test scores.¹ (They didn’t.) Another in the Philippines tested whether or not regular visits from a bank representative would increase household savings.² (They did.) Stating a hypothesis can itself be problematic for policymakers, because it invites them to move from vague statements of intent to a specific measurable outcome they wish to achieve.

The second aspect of a randomised trial is to test the hypothesis by randomly selecting two groups of people. One group receives the treatment (the textbooks or the visit from the bank representative) and one group does not. The appeal of the randomised trial lies in the fact that the two groups are as alike as possible in every respect—geographical location, gender, age, socioeconomic status, education, and so on—except whether or not they receive the treatment. Thus, if a significant difference between the groups develops after the treatment has been applied, that difference can be attributed to the treatment and to the treatment alone. The researcher can conclude that the

Ross Farrelly works for a statistical software company. These are his personal views. He thanks Andrew Leigh for listening to and discussing some of the ideas contained in this article.

treatment *caused* the difference. This is a much stronger conclusion than discovering that the response and the treatment are merely correlated. A causal effect has been established, and therein lies the power of the randomised trial.

In the past, social experiments—such as the negative income tax experiment in the United States in the 1960s—have been conducted on a grand scale, with high ideals and enormous budgets. In contrast, the current trend is for randomised trials to address very specific questions and to be conducted on a small budget with minimal sample sizes. This makes the randomised trial a potent tool for economists working in developing countries.

The current trend is for randomised trials to address very specific questions and to be conducted on a small budget.

What randomised trials *can't* do

Randomised trials are not applicable in all situations. There are two main areas where randomised trials are not able to test a proposed social policy. The first is in trying to assess the effectiveness of very long-term policies. Claims that an intervention will increase the life expectancy of certain groups of people, benefit future generations, or affect global warming are not testable by randomised trial. Such trials are also inapplicable to policies that are not repeatable. The benefits or otherwise of going to war, holding the Olympics in a certain city, or signing an international treaty are not repeatable and therefore not testable by randomised trial. But this still leaves a vast array of policies that could easily be subjected to randomised testing.

Are there alternatives?

Some claim that there are attractive alternatives to randomised testing, the main candidates being observational studies, pilot programs, and surveys. A pilot program in which the intended intervention or treatment is applied to a small sub-population to test its efficacy has one major drawback. Having

applied the treatment and seen an improvement in the desired outcome, researchers usually go on to assert that the treatment caused the change in response. In doing so, however, researchers imply that they know how the targeted population would have fared in the absence of the treatment. In fact, there is no way of knowing this, and therefore pilot programs are not able to establish a causal relationship between the treatment and the effect.

Observational studies are also proposed as valid alternatives to randomised trials. Yet, not only are observational studies unable to establish a correlation between two phenomena, they are also subject to bias. With an observational study, there is scope for researchers to look for, discover, and report findings that fit with their preconceived views. They may choose to overlook or not report findings that do not agree with their previous publications, and they may choose to include certain covariates in their regression analyses that corroborate the conclusion they wish to find. I am not commenting on the prevalence of such biased researching methods, but merely indicating that observational studies contain within them scope for such bias.

In contrast, randomised trials, if rigorously conducted, are not open to such abuse. In a well-conducted randomised trial, the hypothesis should be stated and publicised beforehand. A finding of no effect is important information because it establishes the absence of a causal link, so results tend to be published whether or not the treatment proves to have a statistically significant effect.

A survey is also a poor alternative to a randomised trial. Surveys are notoriously unreliable at predicting the outcome of planned interventions. Asking people how they think they would react if a certain change were to be made in some aspect of social policy is one thing. It's quite another to intervene and observe how people actually react. Life is full of unexpected consequences, and the only reliable way to discover the true reactions to a social intervention is to trial the intervention first. Surveys are also subject to selection bias. Only the views of those who choose to respond to the survey are recorded and analysed, but these people do not always comprise a sample representative of the entire target population.

Are there limits to randomised trials?

Randomised trials are an effective means to answer microeconomic questions. They will tell you about the efficacy of a single planned intervention in a particular setting. They will not tell you much about macroeconomic strategies, and nor will they be able to predict the interactive effect of a large number of policies. Randomised trials are not the one and only sound way to develop good policy. But they should be viewed as a very powerful piece of the policymaker's toolkit. Having said that, the scope of randomised trials can be very wide. If the experiment is well designed, the outcome of the trial will answer the question you are seeking to address. The results of randomised trials have been criticised as too narrow and not easily generalised.³ If randomised trials are promoted as the silver bullet for poverty alleviation, this is a fair criticism. If they are viewed as an additional weapon in the economist's arsenal, it does not hold water. For developed countries like Australia, where macroeconomic questions such as those about long-term growth and interest rates are well addressed by other means, randomised trials to examine microeconomic issues have particular relevance.

How have randomised trials been used elsewhere?

One of the most outstanding examples of randomised trials in social reform is the *Progresa* program in Mexico (later known as *Oportunidades*). The aim of the program was to close the gap between rich and poor in Mexico in terms of nutrition and education. The program was planned as a randomised trial from the outset because the incumbent president knew that without hard evidence the program would not survive a change of government.⁴ A secondary consideration that led to *Progresa* being implemented as a randomised trial was that budgetary constraints meant the program could not be delivered to all families that might have benefited from it.⁵ What could have been seen as a deficiency was turned into a positive attribute through randomisation.

Funding was made available to poor rural families for education and improved nutrition. However, that funding was conditional on attendance at both school and a government-funded infant

health clinic. Independent consultants from the International Food Policy Research Institute were engaged to evaluate the trials and compare the families that were offered the incentives with those that were not. The results have been encouraging, and the program has been expanded into urban areas and extended to target youth up to the age of twenty-two.⁶

This program has a number of striking features. First, it worked! The families that received the conditional funding benefited significantly from the intervention. This might sound obvious, but there has been plenty of funding given to programs that have not made people better off. Second, we know they benefited because of the program. The improved outcomes cannot be attributed to another cause because the control group, who were like the treatment group in every other respect, did not benefit. Third, the evidence was so overwhelming that the program survived a change of government. Objective evidence proved to be more persuasive than ideology.

Randomised trials ... should be viewed as a very powerful piece of the policymaker's toolkit.

Progresa is just one example where randomised trials have been used to test social policy. Randomised trials have been used to test policies as diverse as the effectiveness of driver education programs,⁷ the effect of class size,⁸ and the performance of phonics versus whole-language reading tuition.⁹

Randomised trials are becoming increasingly well-established in social policy assessments. There is now a think tank solely dedicated to such trials, the Abdul Latif Jameel Poverty Action Lab (J-PAL) at MIT.¹⁰ J-PAL has run randomised trials to test many social programs, mostly in developing countries. Issues examined include the effect of remedial education programs on school quality and test scores; the effect of microcredit in Hyderabad slums; and a comparison of electronic surveillance, documented teacher attendance, and incentive pay as a means to improve student performance.

For-profit microcredit institutions are also turning to randomised trials to test the best ways to

serve their markets. The Centre for Micro Finance in India has coordinated a number of random trials on microcredit financial projects. Projects include a trial of smokeless cooking stoves as an alternative to traditional cooking methods that lead to serious respiratory infections in many young children, a trial comparing the difference between weekly and monthly repayment schedules on loan default rates, and a trial measuring the impact of micro-health-insurance products on clients and their families.¹¹

The time is ripe for randomised trials in Australian politics. Kevin Rudd speaks often about his preference for ‘evidence-based’ policy.

How have randomised trials been used in Australia?

Despite some recent interest, randomised trials are yet to be used extensively to test policy in Australia.¹² However, policymakers here have experimented with randomised trials a number of times. Between 1999 and 2001, the Department of Family and Community Services conducted two randomised trials on the Job Network, examining the effect of interviews and follow-up contact from professional staff on workforce participation by the long-term unemployed.¹³ They found that the intervention led to a reduction in the number of hours worked but an increase in the number of hours spent in studying or training.

In 2002, the effectiveness of the Drug Court of New South Wales in reducing recidivism was tested in a randomised trial where 514 offenders who met certain criteria were randomly assigned to either the standard court system or the Drug Court, which took them through a detoxification program.¹⁴ The trial showed that not only did the Drug Court reduce recidivism, it was also more cost-effective when measured in cost per offence averted.

Is there scope for further randomised trials in Australia?

In theory, the time is ripe for randomised trials in Australian politics. Kevin Rudd speaks often about his preference for ‘evidence-based’ policy.¹⁵ A raft of new policies is being introduced by his

enthusiastic, newly elected government. The government’s responses to the 2020 summit are to be built on ‘a strong evidence base.’¹⁶ Many of these are candidates for testing by randomised trials. Let’s examine two proposed policies that lend themselves to objective testing.

Behind the introduction of the national welfare card lies the following hypothesis: that making welfare payments available to delinquent parents through a national welfare card will benefit the children of these parents. Some agree with this policy, while others doubt it will work.¹⁷ The hypothesis would need to be more clearly defined before randomised trials could test it, and the exact benefit that was supposed to accrue to the children would need to be specified. Once this had been done, there would be no reason why the hypothesis could not be tested. As child-protection authorities identified delinquent parents, each family could be randomly assigned either to a control group with no curb on their welfare spending, or to a treatment group that received welfare payments through the card. The hypothesised good that was supposed to accrue to children could be measured before and after the trial, and the efficacy or otherwise of the welfare card could be determined.

A similar analysis could be applied to the provision of high-speed internet access to schools, another initiative of the Rudd government.¹⁸ The hypothesis behind the initiative is that it will benefit students; that is, it will improve their grades. By randomly assigning high-speed internet access to one group of schools and leaving another group as it is, we could discover if such technology made any difference to student achievement.

Such a proposal would no doubt raise objections. On what grounds could the government possibly deny schools access to high-speed internet? Wouldn’t that be inequitable? This assumes that high-speed internet access is beneficial to students, the very question the trial is designed to test. Temporarily denying a group of people a service that may or may not benefit them is a reasonable price to pay to discover if it is actually beneficial.

Clearly, such randomised trials would be one of the most effective possible uses of public funds. Instead of rolling out untested programs

that have not been proved to deliver benefits, but which draw heavily on taxpayer dollars, the government would be judiciously screening proposed new programs before they were introduced on a wider scale.

Why are randomised trials not being used?

In the financial year 2005–06, the federal government spent an estimated \$90.2 billion of taxpayers' hard-earned cash on programs that purported to benefit Australians.¹⁹ None of these were tested by randomised trial.

A number of factors make randomised trials unattractive to politicians.

Making a real difference is hard, and many randomised trials often show that the intervention made no difference. There are two ways of looking at this. One is to celebrate that the intervention is now known to be ineffectual, and that it can be discarded as a possible solution to the problem. One could also acknowledge that without the trial, large amounts of public funds could have been wasted on a 'solution' that was no solution at all. Alternatively, one could take the view that the experiment was a 'failure,' that the researcher's hypothesis was 'wrong,' and that funds that could have been better used elsewhere had been squandered on a frivolous investigation that bore no fruit. The former interpretation of the outcome is based on knowledge of the scientific method. Unfortunately, the media loves bad news and often favours the latter interpretation.

But suppose the randomised trial shows that the intervention significantly benefits the participants. Suppose the national welfare card really does benefit children, or that broadband internet access really does improve student grades. Surely, that would be a coup for the government. Not necessarily so. They may find themselves open to accusations of withholding a beneficial treatment from the control group. In retrospect, this would be true, but at the time the randomised trial was conducted, it wouldn't have been known whether or not the treatment was beneficial. But such subtleties are often lost on the popular press, and consequently it is understandable that politicians do not see randomised trials in a favourable light.

These are not the only reasons randomised trials are unpopular with politicians. Our

elected representatives like to be seen as decisive, energetic, and positive—especially when there is a crisis. They like to be seen doing something about problems and demonstrating leadership where others will not. Randomised trials require an investment of time and money, show no immediate results, and are based on the premise that no one actually knows what will work. That they may lead to certain knowledge about real solutions is often not enough to recommend them to many politicians.

Since governments spend far more on implementing social policy than any other body in Australia, it would be preferable if they were the primary champions of randomised trials. But because of the political and ideological factors mentioned above, this is unlikely to happen in the short term. It is more likely that NGOs or charities would be open to possibility of testing their interventions through randomised trials. NGOs are less subject to popular opinion, and are under no obligation to be seen benefiting the entire population. Therefore, small-scale randomised trials may fit within their charters. They may also find randomised trials an attractive means of providing hard evidence for the efficacy of their programs, which could attract additional funding.

Randomised trials are the best way of establishing a causal relationship between one phenomenon and another.

Conclusion

There is little doubt that randomised trials are the best way of establishing a causal relationship between one phenomenon and another. Because of their inherent sophistication, there are serious challenges that need to be overcome before an elected body in Australia will take up randomised trials to test the efficacy of proposed social policy. However, if elected officials really want to make a difference, and not just be *seen* to be making a difference, this is exactly what they need to do.

Endnotes

- ¹ Paul Glewwe, Michael Kremer, and Sylvie Moulin, *Many Children Left Behind? Textbooks and Test Scores in Kenya*, Poverty Action Lab Paper 44 (Cambridge, MA: Poverty Action Lab, 2007), www.povertyactionlab.com/papers/Textbooks%20and%20Test%20Scores%20Kenya.pdf.
- ² Nava Ashraf, Dean Karlan, and Wesley Yin, 'Deposit Collectors,' *Advances in Economic Analysis & Policy*, 6:2, (2006), ipa.phpwebhosting.com/images_ipa/DepositCollectors.AshrafEtAl.2006_1.pdf
- ³ The Economist, 'Control Freaks,' *The Economist* (12 June 2008), www.economist.com/finance/displaystory.cfm?story_id=11535592.
- ⁴ Ian Ayres, *Super Crunchers: How Anything Can Be Predicted* (London: John Murray, 2007), 76.
- ⁵ Esther Dufloy, Rachel Glennerster and Michael Kremer, 'Using Randomization in Development Economics Research: A Toolkit' (2006), www.povertyactionlab.com/papers/Using%20Randomization%20in%20Development%20Economics.pdf, 20.
- ⁶ 'Mexico's Oportunidades Program,' info.worldbank.org/etools/reducingpoverty/docs/newpdfs/case-summ-Mexico-Oportunidades.pdf.
- ⁷ The Economist, 'Try it and See,' *The Economist* (3 February 2002), 97–98.
- ⁸ As above.
- ⁹ As above.
- ¹⁰ Abdul Latif Jameel Poverty Action Lab, 'Poverty Action Lab,' www.povertyactionlab.com.
- ¹¹ Institute for Financial Management and Research Centre for Micro Finance, *Centre for Micro Finance: An Institute for Financial Management and Research*, brochure, ifmr.ac.in/cmfc/CMF_Brochure.zip.
- ¹² A conference held 19–20 June 2008 at the Australian National University, 'New Techniques in Development Economics,' included a session entitled 'The Economics, Ethics and Politics of Randomised Policy Trials.' See 'New Techniques in Development Economics,' econrsss.anu.edu.au/developmenteconconf.htm.
- ¹³ Garry Barrett and Deborah Cobb-Clark, 'The Labour Market Plans of Parenting Payment Recipients: Information from a Randomised Social Experiment,' *Australian Journal of Labour Economics* 4:3 (2001), 192–205; Robert Breunig, Deborah Cobb-Clark, Yvonne Dunlop, and Marion Terill (2003), 'Assisting the Long-term Unemployed: Results from a Randomised Trial,' *Economic Record* 79:244 (2003), 84–102.
- ¹⁴ Karen Freeman, 'Evaluating Australia's First Drug Court: Research Challenges' (paper presented to 'Evaluation in Crime and Justice: Trends and Methods,' a conference convened by the Australian Institute of Criminology in conjunction with the Australian Bureau of Statistics, Canberra, 24–25 March 2008), www.aic.gov.au/conferences/evaluation/freeman.pdf.
- ¹⁵ As of 20 May 2008, the phrase 'evidence based' occurred at least six times on the official prime ministerial website. The number of mentions is increasing: by 9 August 2008, 'evidence based' appeared at least eleven times. See results of Google Search query ' "evidence based" site:pm.gov.au,' www.google.com/search?sourceid=navclient&ie=UTF-8&rlz=1T4SKPB_enAU217AU217&q=%22evidence+based%22+site:pm%2egov%2eau
- ¹⁶ Australian Government, *Australia 2020 Summit: Initial Summit Report* (Canberra: Australian Government, 2008), www.australia2020.gov.au/docs/2020_Summit_initial_report.doc, 38.
- ¹⁷ Patricia Karvelas, 'Welfare Curbs on Parents,' *The Australian* (9 May 2008), www.theaustralian.news.com.au/story/0,25197,23668892-601,00.html.
- ¹⁸ Australian Government, *First 100 Days: Achievements of the Rudd Government* (Canberra: Australian Government, 2008), www.pm.gov.au/docs/first_100_days.doc, 5.
- ¹⁹ Jenny Hargreaves, 'Welfare Services Resources: Financial and Human' (6 December 2007), www.aihw.gov.au/eventsdiary/aw07/presentations/jenny_hargreaves_welfare_services_resources.pdf, 5.