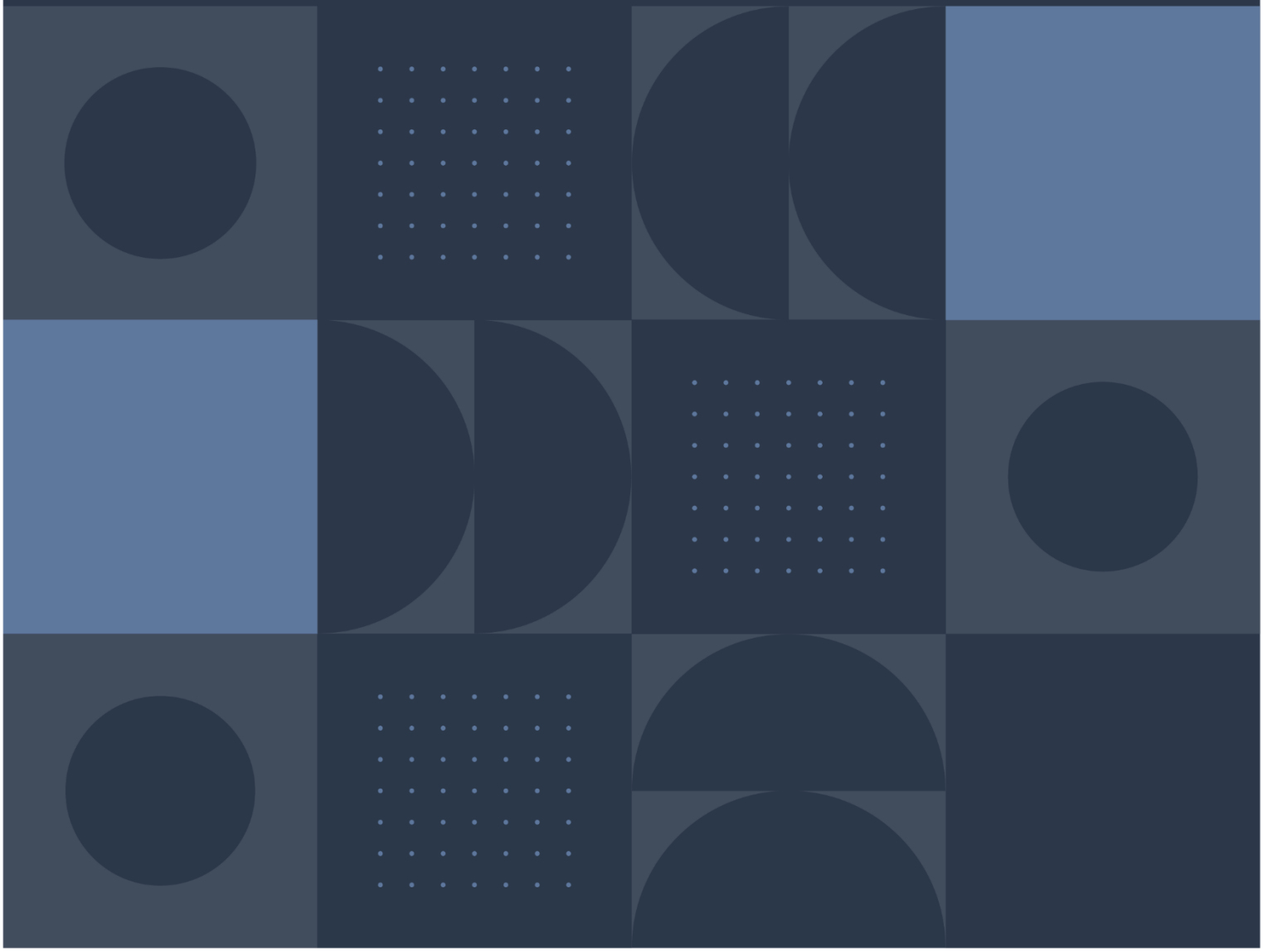
Australian Government
The Treasury

Australian Centre for Evaluation

Randomised trials in Australian public policy: a review

August 2025

Patrick Rehill, Ethan Slaven, Harry Greenwell, Peter Bowers, Scott Copley, Eleanor Williams



ISBN: 978-1-923278-12-7

The views expressed in this report are those of the authors and do not necessarily reflect those of the Australian Treasury or the Australian Government.

Revisions in this version

This report was originally published in March 2025. This version includes corrections to the summary of the nurse home visiting program, right@home, following correspondence with the evaluation's authors. Some typographical errors are corrected in the Appendix and, finally, this version includes a reference to the underlying database of randomised trials, which is available at: evaluation.treasury.gov.au/publications/randomised-trials-australian-public-policy-review

Contents

Executive summary

There is widespread recognition of the importance of evidence-informed policy (Kennedy 2024, Banks 2009). Policy evidence can take various forms. This report focuses on a particular method for generating such evidence: the randomised trial.

Randomised trials are a powerful tool for estimating the 'causal effects' of a program; that is, for separating correlation from causation. There is more to evaluation than just randomised trials, and often the most useful randomised trials are ones that are complemented by other research methods that, in combination, yield greater insights. Nonetheless, randomised trials are a valuable and under-utilised part of the evaluation toolkit.

This report starts by taking stock of the published randomised policy trials conducted in Australia. This is followed by a series of case studies, illustrating how randomised trials have informed policy decisions across various policy domains. The report then turns to the fundamental question of the ethics and cultural appropriateness of randomised trials. It concludes with a discussion of future directions for randomised policy trials in Australia.

The Australian Centre for Evaluation (ACE) has identified 369 published randomised policy trials going back to 1976 – mostly academic studies. The use of randomised trials is growing, with most of these studies (60%) having been published since 2017. Randomised trials are particularly popular in public health (even after excluding clinical trials), and they are also relatively common in education policy. Government behavioural insights units have started to run a substantial number of randomised trials in recent years, producing 30% of all published randomised trials.

The 4 case studies in this report are spread across different policy domains: early years child development, employment services, behavioural insights, and crime and justice. These case studies illustrate the range of different policy questions, randomised trial designs, results, and policy implications.

Randomised trials can be conducted in a way that meets high ethical standards. Furthermore, before governments introduce a policy or program to the whole population, there is an ethical imperative to seek the best available evidence about the effectiveness of the intervention (whenever this is uncertain). Nonetheless, it is always appropriate to subject randomised trials – like any human research – to ethics review, consistent with the National Statement on Ethical Conduct in Human Research. Likewise, evaluators should refer to the available guidance and experts to ensure that randomised trials are culturally appropriate.

While the number of randomised policy trials in Australia has grown in the past decade, they could still be used more often in Australian public policy. Further growth should occur in tandem with an expansion in the use of other evaluation methods and approaches. Beyond the volume of evaluation, there is also room for improvement on other dimensions: randomised trials must be ethical and culturally appropriate, well-designed, and used thoughtfully to inform policy.

The ACE aims to play a central role in improving the volume, quality and use of randomised trials, as part of its broader mission to improve the volume, quality and use of evaluation generally. It will do this in concert with others in the evaluation community—within government, and from academia and research, the not-for-profit sector, and the private sector.

The role of randomised trials in policy evaluation[1]

Why randomised trials matter

Randomised trials matter because – when they are feasible, ethical, and well-designed – they are a powerful tool for finding causal evidence on the effects of a policy or program. That is, for separating correlation from causation.

In a public policy setting, it can be tempting to assume some ideas are so obvious they do not require rigorous causal evidence. The history of randomised trials suggests such an assumption would be a mistake. Leigh (2018) uses as an example the following propositions:

- Asking teenage girls to care for a baby doll that's programmed to demand attention at all hours will discourage teen pregnancy.
- Juvenile delinquents can be 'scared straight' by spending a day in jail, and seeing how tough prison really is.

Both of these seemingly obvious statements were used to develop programs for, respectively, discouraging teen pregnancy, and discouraging criminal behaviour. And yet, in both cases, randomised trials produced convincing evidence that the programs were ineffective and produced suggestive evidence the programs may even have had the *opposite* effect of what was intended (Brinkman et al. 2016;[2] Petrosino et al. 2013; van der Put et al. 2024).

Furthermore, in both cases, these evaluations overturned common wisdom. In the case of infant simulators, less rigorous evaluations had apparently shown that the simulators prevented pregnancy (Dinh et al. 2024:355-366). In the case of Scared Straight, there was apparent evidence of its effectiveness in early, less robust evaluations, and the program fit with popular notions about tough-on-crime strategies being the best way to reduce crime (Petrosino et al. 2013:3-4).

Rigorous causal evidence was only created in these cases because some decision-makers had the humility to admit they were not entirely sure what the effects of the programs were.

In short, public policy randomised trials matter because it is critical to establish whether public policies are achieving their intended outcomes, whether that is supporting the vulnerable, deterring crime, increasing employment, preventing biosecurity breaches, assisting small businesses, or something else entirely.

---

1 The term 'randomised trials' is used throughout this report, though these trials go by many other names including randomised controlled trials (RCTs), experimental evaluations, experimental trials or field experiments.

2 The Brinkman et al. (2016) study is probably the best single piece of causal evidence on this in the Australian context with a large sample size, and linked administrative data that allowed for measurement of outcomes for almost all participants. However, as the authors acknowledge, there was still potential bias in its results due to differences in participation rates: 45% in control, 58% in treatment (Brinkman et al. 2016:6-7; Brinkman et al. 2010:9). Thus, their apparent result, that the simulators backfired and *increased* teen pregnancies, should be treated as suggestive. But the evidence is sufficiently strong to support the authors' conclusion that the simulators were ineffective in *reducing* teen pregnancies.

How randomised trials work

Randomisation means that the people, schools, regions or businesses who receive a policy or program (the 'treatment group') and those who do not (the 'control group'), are assigned by chance. As a result, the treatment and control groups are likely to be similar in all relevant respects, whether that be age, income, education, motivation, resilience, and so on. Because the two groups are highly comparable, any difference in outcomes can be attributed to the program itself rather than other differences between the groups.

This allows the evaluator to overcome some of the usual obstacles to separating correlation from causation. For example, say the literacy skills of children who take part in a reading program are found to be better than those who do not take part, but randomisation was not used to assign children to each group. In that case, the children who did not take part in the program may be from families where the parents lacked the time to take them to the program and so were also not able to help with homework as often. This would mean any correlation between taking part in the program and reading skills would not be a causal relationship as it is mixed up with another causal factor—parental engagement.

There is more to evaluation than just randomised trials

The randomised trial is an important tool in the evaluation toolkit, but it is not the only one. There are many different ways to categorise evaluation approaches. A simple categorisation, provided in HM Treasury's 'Magenta Book' (HM Treasury 2020:14-15) identifies 3 broad types of evaluation, each of which is valuable for addressing different questions.

- Process evaluation (or implementation evaluation): Was the policy implemented as intended? Was it delivered efficiently (on time and on budget)? Was it well received by participants and providers/implementers?
- Impact evaluation (or outcome evaluation): What measurable short-term or long-term outcomes have occurred? Can changes in these outcomes be attributed to the policy? Did these changes vary for different groups? Why did these changes occur?
- Economic evaluation: Is this policy a good use of resources? Do the expected benefits outweigh the costs? Is the policy more cost effective than alternatives?

The randomised trial is one type of impact evaluation method. There are others. The field of 'causal inference' has developed various techniques, such as 'regression discontinuity design', that are used to address causal questions in epidemiology and public health, economics, education research, political science, criminology, and beyond. However, most experts in this field agree that, if it is feasible and ethical to run a well-designed randomised trial, that is the surest way to establish causal attribution.[3]

3 See, for example, Cunningham (2021, Section 4.1), Gertler et al (2016, Chapter 4) or Angrist and Pischke (2014, p1). Various other impact evaluation approaches consider other impact questions, instead of causal *attribution*, such as why or in what circumstances a program will work, or what contribution a program made. These approaches include realist evaluation, process tracing, and contribution analysis.

Mixed-method randomised trials are particularly insightful

Randomised trials are particularly insightful—and hence helpful for decision-makers—when they take a *mixed methods approach.* A mixed methods approach often means incorporating qualitative research, for example through interviews, focus group discussions or reviewing relevant literature. It may also mean incorporating other quantitative approaches that are not directly related to the randomised trial or the trial's primary outcomes, for example, through surveys of participants or analysing program data on the treatment group.

By incorporating these additional sources of information, a mixed methods approach will be able to go beyond *whether* the program is effective – in terms of average outcomes – to *why* it is or is not effective, for whom it is effective, and in what context it is effective. A mixed methods randomised trial will also be able to yield greater insights into how the program could be improved to start being effective if it is not already, or how it could be improved to be even more effective.

The mechanisms through which programs change outcomes are often complex in public policy. Most randomised trials in public policy can benefit from a mixed-methods approach as a way of unpicking some of this complexity.

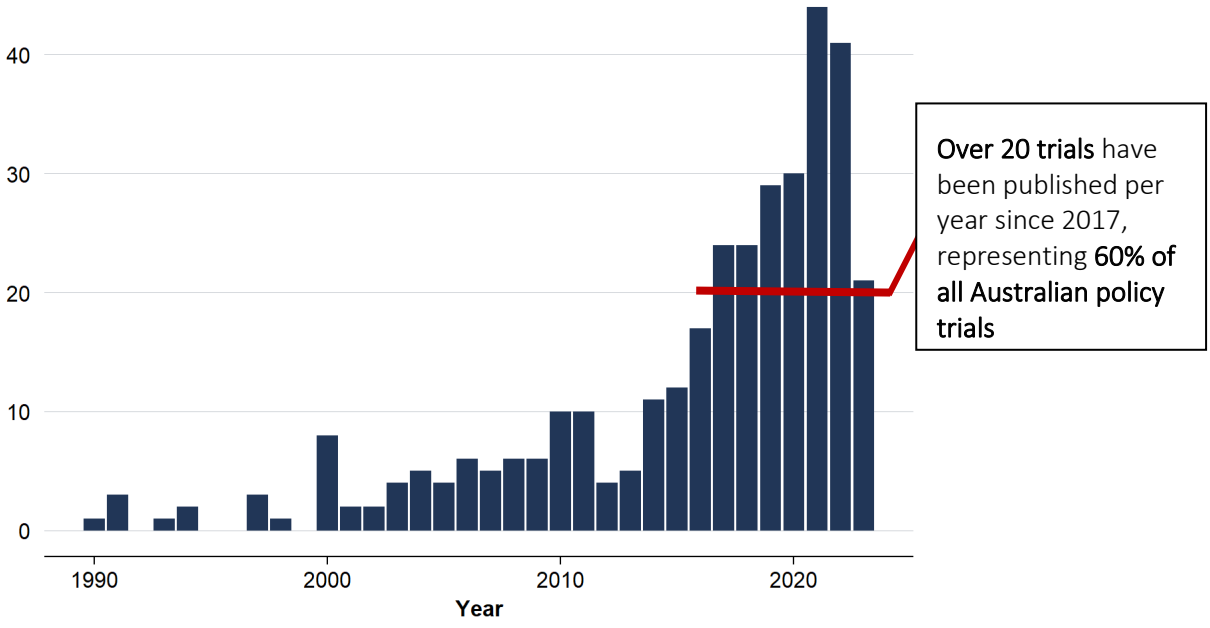
Randomised trials in Australian public policy

The Australian Centre for Evaluation (ACE) undertook a review of the literature to identify published randomised trials in Australian public policy. ACE then classified each study according to: the policy areas studied, whether the trial tested a policy or program for First Nations Australians, whether the trial was run by government (or had significant government involvement), whether the trial tested behavioural interventions. This section reports the results of this stocktake. See Appendix for methods and limitations.

369

369 policy randomised trials have been published in Australia, mostly over the last 35 years (excluding medical clinical trials, lab experiments and survey experiments)[4]

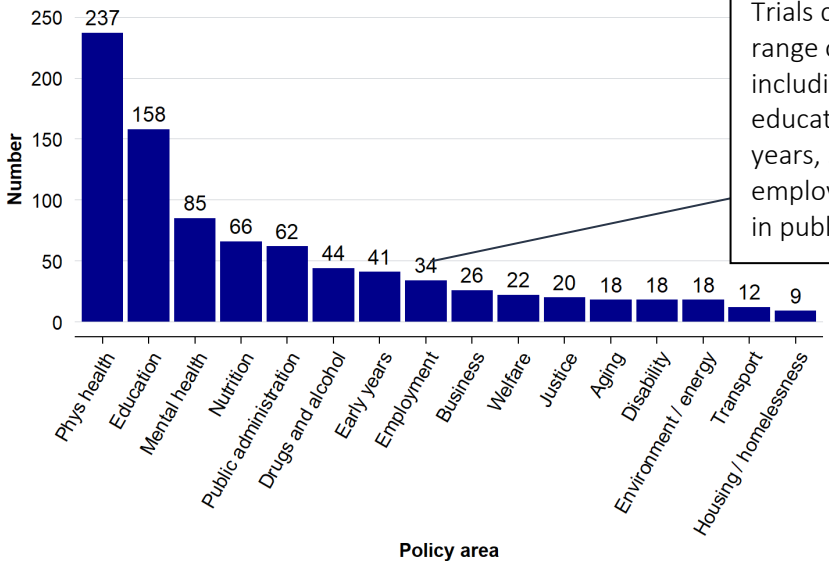
Figure 1: Australian policy randomised trials by year, 1990–2023



24 randomised trials on programs/ policies for First Nations Australians

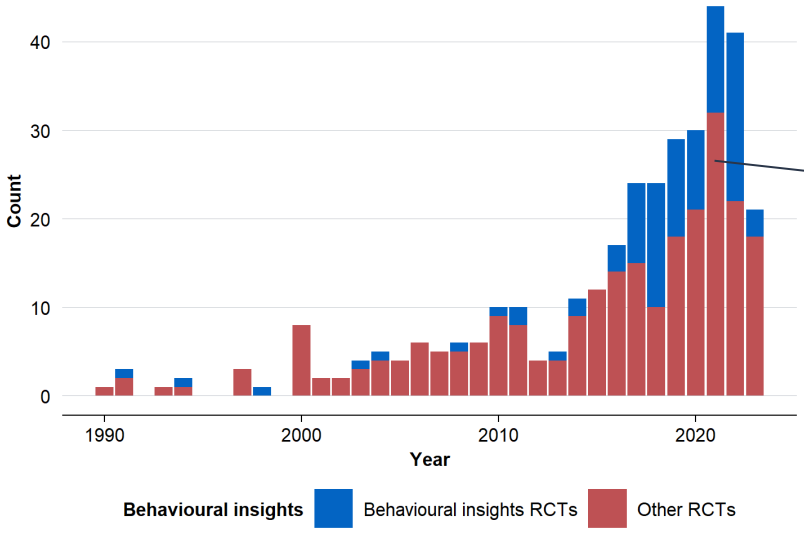
4 This is a conservative estimate of the number of randomised trials based on those identified through a search of journal databases and the grey literature. See Appendix for details of the search method and limitations.

Figure 2: Australian policy randomised trials by policy area[5]



Trials covered a wide range of policy domains including 158 in education, 41 in early years, 34 in employment and many in public health policy

Figure 3: Australian behavioural insights randomised trials



Behavioural insights trials were a driver of growth in trials over the past decade, although not the only driver

5 The policy area categories are not mutually exclusive, and many trials' studies covered multiple policy domains.

Case studies of randomised policy trials in Australia

Improving the quality of evaluation through randomised trials allows government agencies to better serve citizens, save money and make informed investment decisions. The 4 case studies outlined below illustrate the way that randomised trials have informed Australian policymaking across the fields of:

- Early years development
- Employment services
- Behavioural insights
- Crime and justice

These case studies were selected where there was a clear link between the trial results and subsequent policy decisions. They were also selected to reflect a variety of policy domains. Like most research, these trials have some limitations, which are noted where relevant. However, in each case, the results remain sufficiently robust to justifiably inform policy.

Early years child development

There have been at least 41 Australian randomised trials of early years development programs. This relatively large number could be partly explained by the overlap between the public health and medical research communities. Another contributing factor could be some of the influential American randomised trials that found dramatic positive effects such as the Perry Preschool Study and the Abecedarian Early Intervention Project (Conti et al. 2016).

Case study: Nurse home visiting program

Intervention: right@home is a sustained nurse home visiting program for pregnant mothers facing adversity, designed to improve child and family wellbeing. The program consisted of 25 home visits by a specially-trained Child and Family Health nurse with support from a social care practitioner during the mother's pregnancy and the first 2 years of childhood.

Source: Murdoch Children's Research

Evaluation question: Did right@home improve child and family wellbeing?

Trial: The program recruited 722 pregnant women experiencing adversity across Victoria and Tasmania and randomised whether they received home visits from right@home, or usual child and family health services. (Usual care involved a nurse home visit for the first appointment, and later appointments at a local clinic.) The study followed families until children turned 6 years (Goldfeld et al. 2017).

Results: The trial found improvements in multiple domains. At 2 years, the program reduced hostile parenting, improved warm parenting and parental involvement, home safety, bedtime routines, and variety in children's experiences. Improvements in women's mental health and wellbeing, self-efficacy, parenting and family relationships were evident from 3-5 years. At 5-6 years, emerging benefits for children's behaviour, executive function and early academic skills were evident (Price et al. 2024; Goldfeld et al. 2022).[6] This illustrates the value of long-term follow-up.

An initial cost-benefit analysis at 3 years showed right@home cost an additional $7,700 per woman and was not yet cost-effective (Bohingamu Mudiyanselage et al 2021:1). This is consistent with international cost-effectiveness evaluations, which show that similar nurse home visiting programs recoup costs over longer time periods (WSIPP 2023).

Policy impact: The right@home program was subsequently adopted in Queensland and the Northern Territory. The program also bolstered the evidence base on the value of early, relational, integrated support for vulnerable children and families. As a well-implemented program with a robust evaluation, the evidence from right@home continues to be used to inform policy around child and family wellbeing (Molloy et al. 2019, DFFH 2022).

6 While some of these later results were not statistically significant, and the 5-year results were subject to greater loss to follow-up, the evidence was nonetheless suggestive of an effect on these outcomes. Furthermore, the evaluation followed best practice in transparently reporting all effect sizes and confidence intervals, and used statistical techniques to address missing outcome data.

Employment services

Randomised trials have long been important in testing the effectiveness of employment programs (Ashenfelter 1987). This includes a range of robust evidence on whether employment training programs have improved employment outcomes (WWCLEG 2016). In Australia, there have been at least 34 randomised trials on employment-related programs.

Case study: Online Employment Services

Intervention: Unemployed job seekers are required to participate in employment services as a condition of receiving the JobSeeker Payment. Until mid-2020, these employment services were only delivered by government-contracted providers, typically through face-to-face meetings. However, this approach was expensive, often inconvenient for job seekers, and did not take advantage of advancements in digital technology. This trial evaluated the impact of switching from provider services to *online* employment services for certain job seekers who were considered able to self-manage their job search (DESE 2021).

Source: Workforce Australia

Evaluation question: Do job seekers achieve similar or better outcomes with online employment services?

Trial: The trial randomised 17,810 job seekers into control and treatment groups.[7] It ran from July 2018 until mid-April 2020 when online employment services were rolled out as the default for all job-ready job seekers (partly due to the COVID-19 pandemic).

Results: Participants who received online employment services had similar or better outcomes across several measures of interest. The two groups were equally as likely to have left income support after six months. Similarly, 91% of the treatment group had not re-entered income support within 6 months of leaving compared to 89% in the control group (DESE 2021:54). While these effects may not seem large, the aim of online services was to just to achieve similar outcomes to provider services, since the former are much cheaper to provide.

Complementary research revealed that job seekers generally found the online services to be easy-to-use and were satisfied with their quality. The evaluation also identified some job seekers who had difficulties using the online services due to poor technology skills or poor internet access. These job seekers would benefit from continued support from a job services provider.

Policy impact: During the COVID-19 pandemic, the government needed new options to deal with the influx of unemployed workers and lockdown policies that made face-to-face meetings impossible. This trial provided robust evidence that the switch to online services—for job seekers who were considered able to self-manage their job search—was a worthwhile decision. Furthermore, it provided evidence that the new approach could be used as a cost-saving measure even after the pandemic, as well as providing insights on how best to target the program and support job seekers who had difficulties with the online services.

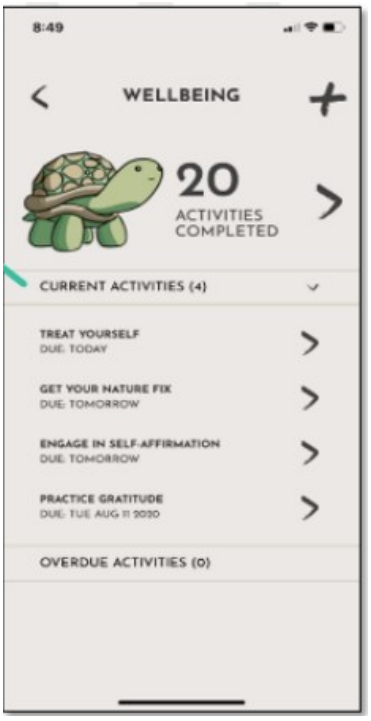
7 A small number of participants (5.5%) opted out of the treatment (online services) after randomisation which may have introduced some bias in the estimate.

Behavioural insights

One of the drivers of the large increase in randomised trials over the last few decades is the widespread use of 'behavioural insights' or, in other words, the application of behavioural science to public policy. In behavioural insights, randomised trials are the norm. This is largely due to experimental tradition in psychology, and the relative ease with which simpler behavioural interventions can be randomly allocated. In recent years, government behavioural insights units around the world have run many policy-relevant randomised trials (OECD 2017). This report identified 116 behavioural insights-related studies conducted in Australia, across government and academia.

Case study: The Grok app for university student retention

Intervention: University student drop-out is linked to several negative outcomes, like worse financial prospects (due to lost time in the labour market, lower earnings due to not having obtained a qualification, and debt burden from incomplete courses (Norton and Cherastidtham 2018) and worse mental health outcomes due to feelings of failure (Trusty et al. 2025). Therefore, policymakers and universities are interested in ways to strengthen students' resilience, peer connections, and academic success, especially for students who only have a tenuous connection to university. The Behavioural Economics Team of the Australian Government (BETA) designed an app, Grok, to encourage students to connect with peers, challenge negative thoughts about setbacks, and provide tips on study and wellbeing (BETA 2021).



Source: BETA

Evaluation question: Can a semester-long app (the Grok app) increase academic achievement, course completion rate, subjective wellbeing and sense of student identity?

Trial: The app was rolled out to 4,463 students across two universities – Western Sydney University and University of Newcastle – in the first semester of 2020 (the start of the pandemic). The treatment group receiving the full Grok app and the control group only receiving standard information they would normally receive from their university.

Results: The app had a high rate of initial adoption and got positive feedback from users. However, there was no meaningful difference between the treatment group (who received the app) and the control group in terms of unit completions or grades. One likely explanation for the app's failure is that it proved challenging to get students to continue to engage with the app after downloading it. It is also possible that pandemic-related lockdowns reduced the app's appeal or relevance.

Policy impact: Grok illustrates the value of a 'null finding'. The lack of a statistically significant result was not because of a lack of data: there was a substantial sample size and yet the treatment and control groups had similar outcomes. This is evidence that there was little or no effect. The positive feedback from students who used the app was deceptive: even though they enjoyed using it, it was not enough to shift the outcomes of policy interest. While the app could have been tested again, outside of pandemic conditions, the evidence was judged strong enough to discontinue its use, and shift resources to finding other ways to try to improve students' outcomes.

Crime and justice

Randomised trials have a long history in criminal justice. Due to the ability to impose treatments through courts or prisons it has been relatively easy to randomise the application of treatments (Farrington 2003). With careful design it is possible to run ethical studies that can generate robust evidence about a high-stakes policy area without exploiting participants, although care must be taken given the power the state has over those in the criminal justice system (Smyth et al. 2024, Chapter 19). Importantly, where a change of practice is expected to produce positive policy outcomes, but the effectiveness has not been tested, the is a strong ethical case to test that practice before rolling out to the whole system.

Trials have helped shape best practice in a range of areas including policing, sentencing, drug and alcohol treatment and community programs (Lum and Mazerolle 2014). The Australian Centre for Evaluation's stocktake found 20 Australian crime and justice evaluations used randomised trials.

Case study: The Canberra restorative justice experiments

Intervention: The justice system has long faced the challenge of how to reduce the risk of repeat offending. Starting in the mid-1990s, a series of 4 groundbreaking randomised trials in Canberra tested the impact of 'restorative justice' as a diversion from the standard court system (Strang 2011). Restorative justice processes involve bringing together an offender in conversation with the person they have harmed.

Evaluation question: Is restorative justice effective at creating a feeling of fairness for victims and offenders, and reducing re-offending rates?

Trial: Randomised trials tested the impact of restorative justice for 4 different offence categories: drink driving (900 offenders), juvenile property offending (249 offenders), juvenile shoplifting (143 offenders), and youth violence (121 offenders). In addition to measuring re-offending rates, the research team had an observer at almost all court and conference events, and conducted interviews with offenders in the first and second year after treatment.

Source: ACT Courts

Results: Victims and offenders who went through the restorative justice conferences had higher perceptions of procedural fairness than those who went to court across all 4 trials. Recidivism rates were substantially lower among those who were diverted into restorative justice conferences than the usual court process for *violent juvenile* offenders but not for other types of offences (Sherman et al. 2000:12-15). This suggests the effect of restorative justice might be dependent on the type of crime.

Policy impact: The RISE trials led to a series of 8 larger-scale trials in the United Kingdom, which studied the effect of restorative justice in addition (not as an alternative) to regular court proceedings. They showed that restorative justice caused a 27% reduction in offending, on average. Unlike Australia, they found restorative justice was more effective for adults than juveniles and more effective for serious crime than less serious crimes (Strang 2017:493-494). This body of evidence, from the RISE trial, UK trials and others around the world, led to adoption of restorative justice conferences in the ACT and across most other Australian jurisdictions, albeit predominantly for juvenile offenders (Larsen 2014:5-21).

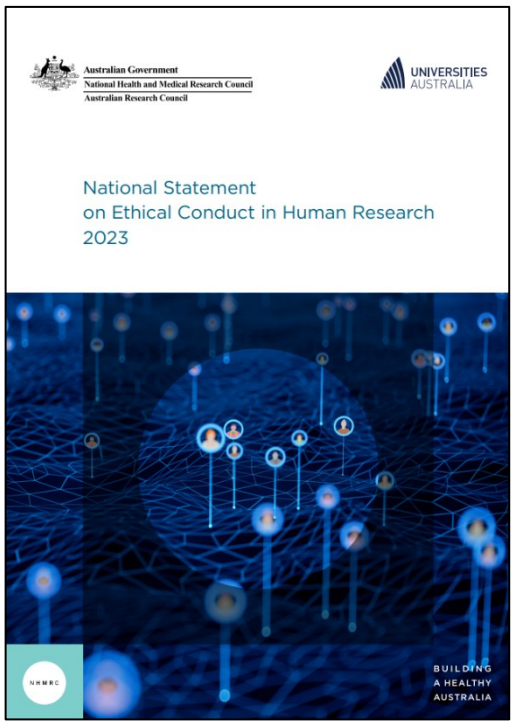
The ethics of randomised trials

One key principle of the Commonwealth Evaluation Policy is that evaluations need to be 'robust, ethical and culturally appropriate'. Randomised trials can be conducted in a way that meets high ethical standards. Furthermore, there is an ethical imperative for governments to seek convincing evidence about the effectiveness of a policy or program (whenever this is uncertain) before introducing interventions to the whole population. A well-designed randomised trial can be a highly ethical form of policy evaluation precisely because it provides such robust evidence to inform decision-making.

This section discusses the ethical considerations specific to randomised trials. It also summarises the available guidance to help ensure that trials are conducted in a culturally appropriate manner.

The National Statement provides a framework for research ethics

In Australia, the National Statement on Ethical Conduct in Human Research ('National Statement') provides an ethical framework for research involving humans, including randomised trials (NHMRC 2023; Smyth et al. 2024).[8] The National Statement requires that human research is subject to an ethics risk assessment. If that assessment determines that there is a risk of harm to participants as part of the research project, a Human Research Ethics Committee (HREC) must assess whether a research proposal is consistent with the National Statement. Ethics committees are required to have a range of experience and expertise, and so are well-placed to assess the ethics of proposed randomised trials.



While there is no legal or policy requirement for government agencies to abide by the National Statement, the ACE submits all its randomised trials to an ethics risk assessment and, where appropriate, to ethics review. ACE encourages other government evaluations involving human research to do the same. Also, all major Australian universities, most academic journals, and any projects funded by the Australian Research Council or National Health and Medical Research Council are required to comply with the National Statement.

The National Statement includes four overarching values that inform an assessment of research ethics:

8 In addition to the National Statement, there are several other touchstone documents for Australian researchers or evaluators in relation to ethical and culturally appropriate research and evaluation. These are: the Ethical guidelines for research with Aboriginal and Torres Strait Islander peoples; the AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research; the Australian Code for the Responsible Conduct of Research; Ethical considerations in quality assurance and evaluation activities; and the Australian Evaluation Society's Guidelines for the Ethical Conduct of Evaluations and First Nations Cultural Safety Framework.

- Research merit and integrity – Researchers should have integrity, capability and a good research design. Poor quality research wastes resources and involves risks with no benefit.
- Justice – Research should treat individuals and groups involved in the research fairly.
- Beneficence – The likely benefit of the research must justify any risks of harm or discomfort to participants. The likely benefit may be to the participants, to the wider community, or to both.
- Respect – Research should respect the autonomy and dignity of participants. Informed consent should be sought unless specific conditions are satisfied that justify not doing so.

Is it ethical to withhold a policy or program from some people?

The principle of 'equipoise'

The fundamental concern raised by randomisation is whether it is ethical to withhold a policy or program from some people who would otherwise be eligible for it. The answer to this question must be assessed on a case-by-case basis. It is clearly incorrect to claim that all possible randomised trials are ethical. But it is equally mistaken to make general claims that randomised trials are, by their nature, unethical.

The foundation for establishing that randomisation is ethical, in a particular instance, is the principle of equipoise. That is, there is genuine uncertainty about the effects of a program or policy. In this case, it is not clear that people who are randomised to be in the control group are missing out because they do not receive the program. In fact, if a program is not effective, there may be saved time wasted and inconvenience from *not* being included.

Furthermore, if there is uncertainty about the effects of a policy or program, that makes it more vulnerable to a policy change, for example through a loss of funding, or a legislative change. The ethics of randomisation must also take account of the potential long-term benefits of putting future policy decision-making on a surer footing.

It can be tempting to point to expert opinion, prior evidence, or theory to argue the principle of equipoise does not apply. However, equipoise does not mean there is *complete* uncertainty about a policy's effects. Instead, it means, there is not yet certainty. And the history of randomised trials (and other impact evaluations) has often shown that seemingly common-sense interventions had neutral or negative effects, while interventions that were believed to only have a moderate effect were found to be highly effective (Leigh 2018:151; Jamison 2019:14-17).

Opt-in consent is the default

In the National Statement, there is a presumption that participants will be provided with appropriate information about the evaluation research before their consent to participate is sought. The information typically includes an explanation that they may or may not receive the policy or program that is being evaluated.

There are some circumstances where opt-in consent is not feasible, and these cases are covered in Chapter 2.3 of the National Statement. However, where potential participants are informed that they may be part of a randomised trial, and are given the opportunity to not participate, the ethical issues are substantially reduced.

The implementation approach can help mitigate concerns about randomisation

There are several ways that a randomised trial can be designed to minimise ethical concerns that participants will miss out on the *hoped-for* benefits of a policy or program. First, the treatment can be rolled out to all the neediest people, then randomised for the rest of the population. Second, instead of a 50/50 split, the assignment ratio can be skewed towards treatment (for example an 80/20 split between treatment and control). Third, the control group can be provided with other services such as the current best standard of care. An example of this was the right@home study profiled earlier where children in the control group were not denied care, just given the current standard of care.

There may be practical reasons to use randomisation

If there is limited funding for a program or it will take time to roll out, then some potentially eligible participants will for practical reasons miss out on the program initially if not entirely. In other words, the principal ethical concern above – that some people may miss out *due to randomisation* – does not apply. In such cases a randomised trial that uses an *over-subscription* design may be appropriate. This is where the limited spaces available for the program are allocated randomly. This can sometimes be fairer than other approaches to determining eligibility that can be more arbitrary or can rely on individuals' subjective judgement. Leigh (2018:36) describes how for a randomised trial on homelessness, some homeless participants came to recognise the fairness of randomisation. Many people dealing with homelessness are used to missing out on programs because such programs usually have limited resources, and staff and not-for-profits can be tempted to select for those who do not have complex support needs or who have not been difficult to work with in the past. Some homeless people recognised the fairness of being given an equal chance of receiving the program as everyone else, regardless of the barriers they faced.

It may also be possible to exclude some participants temporarily rather than entirely. For example, a randomised trial with a *waitlist control design* involves randomising who comes off a waitlist first, with all participants eventually gaining access to the program. In some cases, there may be a waitlist because the program takes time to rollout or because resourcing has been phased over multiple years. Alternatively, the waitlist could be created purely to run a randomised trial. Similarly, a *stepped wedge design* typically involves rolling out a program to different geographic areas at different times. If some regions are randomised to receive a program 12 months before other regions, then a randomised trial has occurred.

Another practical way that randomisation can be less controversial is to use an *encouragement design*. In this case, all individuals are allowed to enter a program, but some are randomised to receive direct advertising informing them of their eligibility, while others do not receive the advertising. The advantage of this approach is it prevents disappointment for people in the control group (no advertising) who hear about the program and would like to participate because they are still able to do so. The disadvantage is that it typically requires a larger sample size than is the case with other randomised trials.

It helps to bring stakeholders on the journey

In some cases, there may be a clear rationale for using randomisation however there might still be concern that some stakeholders will react negatively to the idea. This requires clearly explaining the rationale for using randomisation to stakeholders, and listening to their concerns.

Recent evidence suggests the public has a surprisingly positive view of randomised trials when it is explained to them. In a survey of the Australian general public, Biddle et al. (2023) found that after having the concept of randomised trials explained to them, 83% of people supported or strongly supported their use. Further, Ames and Wilson (2016) found 73% of Australian members of parliament supported the use of randomised trials.

Culturally appropriate evaluation

Cultural appropriateness is an important standard for all evaluations, and especially for evaluations involving Aboriginal and Torres Strait Islander people. However, a recent review of 17 health-related randomised trials involving Indigenous Australians concluded that most rated poorly against the Aboriginal and Torres Strait Islander Quality Appraisal Tool (Esgin et al. 2023). This suggests there may also be room for improvement for policy-related randomised trials.

There are several important sources of guidance on conducting ethical and culturally appropriate evaluations with Aboriginal and Torres Strait Islander peoples:[9]

- Ethical guidelines for research with Aboriginal and Torres Strait Islander peoples (NHMRC)
- The AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research
- The Australian Evaluation Society's First Nations Cultural Safety Framework

In addition, the Indigenous Evaluation Strategy, developed by the Productivity Commission, provides a whole-of-government framework to use when selecting, planning, conducting and using evaluations of policies and programs affecting Aboriginal and Torres Strait Islander people. The Strategy centres Aboriginal and Torres Strait Islander people, perspectives, priorities and knowledges. It sets out what good practice looks like, and what agencies should consider when undertaking evaluations.

These resources are as relevant for randomised trials as they are for other evaluations.

9 Links to these and other resources can be found at the 'Indigenous evaluation' page on the ACE's web site: https://evaluation.treasury.gov.au/about/indigenous-evaluation.

Future directions for randomised trials in Australian public policy

The number of randomised trials in Australia has grown in the past decade, relative to the preceding two. There is, however, substantial room for further growth. It is still true that few government policies or programs in Australia are evaluated using randomised trials (CEDA 2023).

For the number of randomised trials in Australian public policy to grow, the policy community will need a certain humility to recognise that there is often still uncertainty about the effectiveness of new and existing programs and policies. Wherever this uncertainty exists, there is likely to be value in exploring whether a randomised trial is feasible.

Further growth in randomised trials should occur alongside growth in the use of other evaluation methods and approaches. This should be guided by what questions the evaluations seek to answer, and what evaluation methods are feasible. This is consistent with the principle of fit-for-purpose evaluation, as stipulated in the Commonwealth Evaluation Policy.

It is not enough, however, for just the number of randomised trials to increase. It is also important that these trials be ethical and culturally appropriate (as discussed above), well-designed, and used thoughtfully to inform policy.

The ACE and other organisations in government, academia and research, the not-for-profit sector, and private sector all have roles to play in achieving this vision.

Well-designed randomised trials

Randomisation is a powerful tool for creating treatment and control groups that are truly comparable. However, randomisation alone is not enough. Randomised trials will not be able to make credible estimates of causal effects if they suffer from poor design: insufficient sample size, or failure to address possible sources of bias that can re-emerge after randomisation. And randomised trials will often provide fewer insights if they are not designed as part of a mixed-methods evaluation. Similarly, the causal estimates from randomised trials will be less credible if they suffer from errors, omissions or questionable research practices in the statistical analysis and reporting.

This section elaborates on 6 good practices the community of randomised trial practitioners should adopt, at a minimum.

Trial design

- Use a sufficient sample size: The minimum sample size needed for a particular trial varies widely and depends on several factors, including how transformative the program is (programs with larger effects require smaller sample sizes). If a randomised trial is conducted in a sample that is too small, the trial will not be able to draw meaningful conclusions, which may mean the trial was not a good use of time and resources. It is therefore critical to calculate the minimum sample size needed for the trial beforehand. This is called a ‘power analysis’.
- Identify and mitigate the risk of bias: While randomised trials are very effective at minimising the risk of bias, it is still possible for results to be biased if the trial is designed or implemented poorly. For example, if people in the control group also receive the benefits of the treatment (either by mistake or intentionally), this will introduce bias into impact estimates. One tool for assessing and addressing the risk of bias is the *Risk of Bias 2 (RoB 2)* tool (Sterne et al. 2019).

- Adopt a mixed-methods approach where feasible and appropriate: The most common use of mixed methods in randomised trials is to collect additional data about the treatment group through surveys, interviews or program data. This can shed light on why a policy or program worked (or not), and for whom it was most likely to be effective. Mixed methods can also be used to inform the design of the randomised trial, for example, to guide the choice of outcome variables or the approach to data collection.

Trial analysis, review and reporting

- Prepare and pre-register a pre-analysis plan: Pre-analysis plans (or 'trial protocols') are documents that lay out exactly how the data from the trial will be analysed. Pre-registering involves publishing a pre-analysis plan online, for example on the American Economic Association's RCT Registry or the Open Science Framework. By doing this, researchers and evaluators force themselves to avoid 'p-hacking', which is where a researcher tries multiple different approaches to analysing data to find the results they want (e.g., a statistically significant result). The ACE provides a pre-analysis plan template on its website.[10]
- Invite peer review: Expert peer review is an important method for ensuring the quality of the design, analysis and interpretation of impact evaluations, including randomised trials. In public policy, there is considerable merit in emulating the 'Registered Reports' process. This is a 2-stage model of peer review. The first stage occurs before data collection or analysis, and assesses the quality of the study design (emulating the pre-analysis plan described above). The second stage reviews the analysis against the original plan, ensures all planned analyses are reported, and assesses whether any deviations from the original plan were justified (Chambers and Tzavella 2022).
- Provide reproducibility and replicability packages where possible: Randomised triallists should, where possible, share the data and code, so that others can check that their results are reproducible. Privacy or ethical restrictions on data sharing will sometimes mean this is not possible however, in this case, researchers could consider creating synthetic datasets that mimic the properties of their actual data. Furthermore, triallists should provide enough details of the intervention to allow others to replicate the evaluation in different populations or contexts.

Using evidence from randomised trials thoughtfully

Like any evidence, that from randomised trials can be used thoughtfully or carelessly. Well-designed randomised trials provide convincing estimates of the *average* impact of a program or policy. However, the average impact may hide variation in impacts for different groups or individuals. Sometimes trials can estimate the different effects for different groups. For example, a randomised trial of the Early Head Start program in the United States examined the program's impacts for 27 sub-groups (Love et al. 2004). However, sub-group analysis is often constrained by sample size. Otherwise, as mentioned above, qualitative and survey data on the experiences of the treatment group can often complement trial results to give an indication of whether there was variability in impacts.

10 Available at: https://evaluation.treasury.gov.au/toolkit/preregistration-and-pre-analysis-plans.

Randomised policy trials often test a policy in the relevant policy setting, with a sample of the eligible population. Provided there is no dramatic society-wide change (such as a pandemic or economic crisis) during the trial, it is reasonable to expect that the same policy with a similar population will have similar impacts in future. However, generalising the results to different populations or contexts should be done with caution, and likewise if there are substantive changes to the design or implementation of the policy or program. For a discussion of how to thoughtfully apply evidence from randomised trials in such circumstances, see 'The Generalizability Puzzle' (Bates and Glennerster 2017).

Government-academia collaboration

Collaboration between government and academia promises better, more impactful randomised trials. Researchers have deep expertise in running and analysing randomised trials, while public servants can identify policy questions, help provide access to large administrative datasets, and apply trial findings to policy advice. Around the world there are many examples of organisations working on combining the rigour of academic research and the practicality of government to allow for robust evaluations that can have a substantial policy impact. For example, in the UK, the Evaluation Trial and Advice Panel, and What Works Network, and in the US, the not-for-profit Research4Impact, and Penn State's Research-to-Policy Collaboration.

The ecosystem for this kind of work in Australia is less mature but there are some similar initiatives. In 2024, the ACE joined with the Australian Education Research Organisation and other government and research agencies to establish an Impact Evaluation Practitioners Network, which aims to be an interface between government and academia. This network is open to evaluators and policy-makers across government, not-for-profits, consulting, and academia. To join the Impact Evaluation Practitioners Network, please email evaluation@treasury.gov.au.

The role of the Australian Centre for Evaluation

The Australian Centre for Evaluation (ACE) was established in mid-2023, with a broad mission to help put evaluation evidence at the heart of policy design and decision making. The ACE is working to improve the volume, quality and use of evaluation evidence across the Australian government to support better policy and programs that improve the lives of Australians.

The ACE's strategy for delivering on this mission is organised around four interconnected work streams. One of these work streams is that the ACE champions high-quality impact evaluations, such as well-designed randomised trials.[11] The ACE does this by delivering training, supporting evaluation delivery, sharing resources, and building networks amongst practitioners.

- Impact evaluation training – The ACE had delivered advanced training courses on impact evaluation to over 250 public servants at the time of publication.

11 The other 3 work streams are: evaluation leadership and promotion; evaluation planning and use; and evaluation capability building. More information on the ACE and its service offer can be found on the ACE's website (https://evaluation.treasury.gov.au) and its strategy (https://evaluation.treasury.gov.au/publications/australian-centre-evaluation-strategy).

- Supporting delivery of impact evaluations – the ACE is currently contributing to the delivery of various impact evaluations, including several randomised trials testing the impact of different features of online employment services on employment outcomes.[12]
- Impact evaluation resources – The ACE has published a pre-analysis plan template, and a guide to government administrative data for evaluation.
- Networks – As noted above, the ACE has collaborated with others to establish the Impact Evaluation Practitioners Network.

12 For further details about these trials, see 'Service improvement trials for Workforce Australia Online', Joint Media Release from the Minister for Employment and Workplace Relations, and the Assistant Minister for Employment, 14 December 2023, https://ministers.dewr.gov.au/burke/service-improvement-trials-workforce-australia-online.

Appendix: method and limitations of literature search

The aim of this report is to provide a comprehensive picture of policy-related randomised trials in Australia. This Appendix details how studies were collected for this stocktake, how they were categorised (for example, by policy area), and finally notes the limitations of this approach.

Method – finding relevant papers

There were 3 criteria for a paper to be included in this stocktake. The randomised trial needed to be:

- conducted in Australia
- a field experiment (as opposed to a lab or survey experiment)
- a policy evaluation (as opposed to a clinical trial).

The starting point for our literature search were the studies identified in Ames and Wilson's (2016) review of randomised trial evaluations in Australia. From there, the literature search consisted of two parts: an effort to collect academic papers, and another to collect studies defined in the 'grey literature' of reports and similar documents, usually from governments or non-government organisations.[13]

To search for academic papers, the authors undertook a search of the Scopus database searching for any papers that have been published by Australian-based authors. The search used the terms 'RCT', 'randomised controlled trial', 'randomised evaluation', and 'field experiment'. All papers with the Scopus subject area 'Medicine' were excluded, since they were not considered to be policy evaluations.

The search for grey literature was more difficult because there is no authoritative repository for such publications. Those services that do exist are generally not comprehensive, so the process of trying to find these reports was less well-defined. The authors searched the Analysis and Policy Observatory with the same terms used for the Scopus search. This yielded 14 articles. The search also included a review of publications from several organisations that routinely conduct and publish randomised policy trials: the Behavioural Economics Team of the Australian Government (BETA), the Behavioural Insights Team (BIT-Australia), and the NSW Behavioural Insights Unit.

The search process described above produced an initial 'long list' of papers. This list was then reviewed against the 3 inclusion criteria set out above, using a custom large language model tool, as a 'first pass', followed by an assessment by the authors as needed. The following paragraphs describe some of the details of the inclusion/exclusion assessments.

First, was the study a policy evaluation? For health-related trials, there was no clear dividing line between what was a 'policy evaluation' and what was not. Instead, the authors used a rule-of-thumb based on who the intended target audience was. Where a trial was primarily aimed at informing government policy, it was considered a policy evaluation, while trials aimed at informing other kinds of decision-makers—like medical practitioners or the Therapeutic Goods Administration—were not considered policy evaluations. For example, a study on the effects of a certain program of physical

13 'Grey literature' is used broadly to refers to published studies or reports that do not appear in academic journals.

activity as part of a cardiac rehab would be excluded, but a study on the effects of a similar program of physical activity in schools would be included.

Second, was the study a field experiment? Several papers identified in the search referred to a randomised trial or field experiment because they were: the trial protocol, a meta-analysis, a cost benefit analysis that used data from randomised trials, or a quasi-experimental study. These were also excluded.

Third, was the study conducted in Australia? This step meant that, for example, randomised trials were excluded if they were undertaken by Australian evaluators or researchers but related to overseas programs overseas, including Australian aid programs.

This process led to a final list of 369 papers, and excluded most of the papers initially identified in the search. This was mainly because many experiments in the long list assessed the effects of specific medical interventions or organisational changes in health care systems (and hence were not classified as 'policy evaluations').

Method – categorising papers

This analysis involved categorising papers in various ways:

- by policy area
- whether a paper tested a behavioural insights intervention
- whether the paper tested a policy or program that was primarily intended for Indigenous Australians
- whether the randomised trial had 'significant government involvement.

In each case, papers were categorised based on a review of the abstract. In a handful of ambiguous cases where a full paper was available, one of the authors reviewed the paper in full. For the 'policy area' categories and the assessment about whether a study involved behavioural insights, the categorisation process followed a 2-step process. A custom large language model tool made a first pass at categorisation for each paper using key information (title, abstract, tags). Next, these categorisations were then reviewed and changed if needed by a human researcher.

Policy area was coded into one of the 18 areas, as shown in Figure 1 above. One paper could be tagged with multiple policy areas, and some judgement (and hence potential error) was involved in deciding which policy areas applied. The policy areas do not come from a standard classification: they were chosen by the authors simply to convey the breadth of policy domains.

Whether a paper tested a behavioural insights intervention also required some judgement. Often, behavioural interventions are relatively cheap or unobtrusive treatments that seek to change behaviour. However, some interventions that were more complex or expensive, but which still sought to test the application of behavioural science, were included. (For example, the Grok app described in the case studies above was not a simple, cheap nudge intervention but its design was guided by evidence from behavioural science). Finally, this assessment was also informed by who conducted the trial so that, for example, papers from government behavioural insights units were classified as behavioural insights studies.

Whether a paper was classified as having a specific focus on Indigenous Australians was determined by whether the policy or program was targeted at Indigenous Australians. This excluded many other evaluations that might have disproportionately affected or assisted Indigenous Australians.

Whether a study had **significant government involvement** depended on meeting at least one of three criteria: whether it was funded by government (excluding research grants), whether government delivered the evaluation, or whether the evaluation focused on a significant government program. This was determined from the abstract.

Limitations

A limitation of this review is that it is unlikely to have achieved perfect coverage across all published randomised trials in Australia. In particular, searching the grey literature for randomised policy trials made the task of getting full coverage substantially harder. Furthermore, as described above, applying the inclusion and exclusion criteria required judgement, especially the distinction between medical and policy trials. Similarly, the categorisation of papers also involved judgement. While every effort was made to apply an accurate and consistent approach to these judgements, no doubt some may be mistaken. This review then represents a good faith effort to get coverage as good as possible.

The full list of randomised policy trials identified is publicly available at: evaluation.treasury.gov.au/publications/randomised-trials-australian-public-policy-review.

References

Ames P and Wilson J (2016) *Unleashing the potential of randomised controlled trials in Australian governments*, Mossavar-Rahmani Center for Business and Government, Harvard Kennedy School, Cambridge MA.

Angrist J and Pischke JS (2014) *Mastering 'Metrics: The Path from Cause to Effect*, Princeton University Press, Princeton NJ.

Ashenfelter O (1987) 'The case for evaluating training programs with randomized trials', *Economics of Education Review*, 6(4):333–338, doi: 10.1016/0272-7757(87)90016-1.

Banks G (2009) *Evidence-based policy making: What is it? How do we get it?* Productivity Commission.

Bates M and R Glennerster (2017) 'The Generalizability Puzzle', *Stanford Social Innovation Review*, Summer 2017.

BETA (Behavioural Economics Team of the Australian Government) (2021) *Gather. Grow. Graduate: Using behavioural insights and technology to help students graduate from university*, Department of Prime Minister and Cabinet, Australian Government.

Biddle N, Gray M and Hiscox M (2022) *Public support for randomised controlled trials and nudge interventions in Australian social policy*, Centre for Social Research and Methods, Australian National University.

Brinkman S, Johnson S, Codde J, Hart M, Straton J, Mittinty M and Silburn S (2016) 'Efficacy of infant simulator programmes to prevent teenage pregnancy: a school-based cluster randomised controlled trial in Western Australia', *The Lancet,* 388(10057):2264-2271.

Brinkman S, Johnson S, Lawrence D, Codde J, Hart M, Straton J and Silburn S (2010) 'Study protocol for the evaluation of an Infant Simulator based program delivered in schools: a pragmatic cluster randomised controlled trial', *Trials, 11*(100):2-11, doi:10.1186/1745-6215-11-100.

Bohingamu Mudiyanselage S, Price A, Mensah F, Bryson HE, Perlen S, Orsini F, Hiscock H, Dakin P, Harris D, Noble K, Bruce T, Kemp L, Goldfeld S and Gold L (2021) 'Economic evaluation of an Australian nurse home visiting programme: A randomised trial at 3 years', *BMJ Open*, 11(12): e052156, doi: 10.1136/bmjopen-2021-052156.

Boreland F, Lesjak M and Lyle D (2009) 'Evaluation of home lead remediation in an Australian mining community', *Sci Total Environ*, 408(2):202-208, doi: 10.1016/j.scitotenv.2009.10.013.

CEDA (Committee for Economic Development of Australia) (2023) 'Disrupting disadvantage 3: Finding what works', Committee for Economic Development of Australia.

Chambers C and Tzavella L (2022) 'The past, present and future of Registered Reports' *Nature Human Behaviour* 6:29–42, doi: 10.1038/s41562-021-01193-7.

Conti G, Heckman J and Pinto R (2016) 'The effects of two influential early childhood interventions on health and healthy behaviour', *The Economic Journal*, 126(596): F28–F65, doi: 10.1111/ecoj.12420.

Cunningham, S (2021) *Causal Inference: The Mixtape*, Yale University Press, New Haven Connecticut.

DESE (Department of Education, Skills and Employment) (2021) *Online employment services trial evaluation report*, Department of Education, Skills and Employment, Australian Government.

DFFH (Department of Families, Fairness and Housing) (2022) *Right@Home Menu of Evidence*. Department of Families, Fairness and Housing, Victorian Government.

Dinh T, Lees D, Van Dam P and McGarry D (2024) 'Use of infant simulators as an aid in pregnancy and parenting educational interventions for school-aged students: a scoping review', *Health Educ Res.* 39(4):351-374, doi: 10.1093/her/cyae012.

Esgin T, Macniven R, Crouch A and Martiniuk A (2023) 'At the cultural interface: A systematic review of study characteristics and cultural integrity from twenty years of randomised controlled trials with Indigenous participants', *Dialogues in Health*, doi: 10.1016/j.dialog.2023.100097.

Farrington D (2003) 'A short history of randomized experiments in criminology: A meager feast', *Evaluation Review*, 27(3): 218–227, doi: 10.1177/0193841X03027003002.

Gertler P, Martinez S, Premand P, Rawlings L and Vermeersch C (2016) *Impact Evaluation in Practice, Second Edition*, Inter-American Development Bank and World Bank.

Goldfeld S, Bryson H, Mensah F, Price A, Gold L, Orsini F, et al. (2022) 'Nurse home visiting to improve child and maternal outcomes: 5-year follow-up of an Australian randomised controlled trial', *PLoS ONE* 17(11), doi: 10.1371/journal.pone.0277773.

Goldfeld S, Price A, Bryson H, Bruce T, Mensah F, Orsini F, Gold L, Hiscock H, Smith C, Bishop L, Jackson D and Kemp L (2017) 'right@home: A randomised controlled trial of sustained nurse home visiting from pregnancy to child age 2 years, versus usual care, to improve parent care, parent responsivity and the home learning environment at 2 years', *BMJ Open*, 7(3), doi: 10.1136/bmjopen-2016-013307.

HM Treasury (2020) *Magenta Book*, HM Treasury, United Kingdom Government.

Jamison J (2019) 'The Entry of Randomized Assignment into the Social Sciences' *Journal of Causal Inference* 7(1), doi: 10.1515/jci-2017-0025.

Kennedy S (2024) *Evidence-informed policy making* (speech), University of Adelaide, South Australian Centre for Economic Studies Corporate Members Luncheon.

Larsen J (2014) *Restorative justice in the Australian criminal justice system,* Australian Institute of Criminology.

Leigh A (2018) *Randomistas: How radical researchers changed our world*, La Trobe University Press and Black Inc, Melbourne.

Love JM, Kisker EE, Ross C, Schochet PZ, Brooks-Gunn J, Paulsell D, Boller K, Constantine J, Vogel C, Fuligni AS and Brady-Smith C (2004) *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start, Volume I: Final Technical Report*, Mathematica Policy Research.

Lum C and Mazerolle L (2014) 'History of randomized controlled experiments in criminal justice', in G Bruinsma and D Weisburd (eds), *Encyclopedia of Criminology and Criminal Justice*, 2227–2239, Springer, New York NY.

Molloy C, Beatson R, Harrop C, Perini N and Goldfeld S (2019) *Sustained nurse home visiting: An evidence-based review of indicators to assess quality, quantity, and participation*, Restacking the Odds Collaboration, Murdoch Children's Research Institute, Social Ventures Australia, Bain & Company.

NHRMC (National Health and Medical Research Council) (2023) *National statement on ethical conduct in human research 2023*, National Health and Medical Research Council.

NSW DPE (New South Wales Department of Planning and Environment) (2023) *NSW Land and Housing Corporation Environmental Sustainability Strategy 2024–2026*, Department of Planning and Environment, New South Wales Government.

Norton A and Cherastidtham I (2018) *Dropping out – the benefits and costs of trying university*, The Grattan Institute.

Olive L, Telford R, Westrupp E and Telford R (2024) 'Physical activity intervention improves executive function and language development during early childhood: The active early learning cluster randomized controlled trial', *Child Development*, 95(2): 544–558, doi: 10.1111/cdev.14014.

OECD (Organisation for Economic Co-operation and Development) (ed) (2017) *Behavioural insights and public policy: Lessons from around the world*, OECD.

Petrosino A, Turpin-Petrosino C, Hollis-Peel ME and Lavenberg JG (2013) ''Scared Straight' and other juvenile awareness programs for preventing juvenile delinquency'. *Cochrane Database System Rev*, doi: 10.1002/14651858.CD002796.pub2.

Price A, Zhang J, Goldfeld S, Carlin J and Mensah F (2024) 'Evaluating the "right@home" randomised trial of nurse home visiting using linked administrative data at school transition', *International Journal of Population Data Science* 9(2), doi: 10.23889/ijpds.v9i2.2400.

Sherman L, Strang H and Woods D (2000) *Recidivism patterns in the Canberra Reintegrative Shaming Experiments (RISE)*, Centre for Restorative Justice, Australian National University.

Smyth B, Martin M and Downing M (eds) (2025) *The Routledge handbook of human research ethics and integrity in Australia*. Routledge, doi: 10.4324/9781003319733.

Sterne J, Savović J, Page M, Elbers R, Blencowe N, Boutron I, Cates C, Cheng H, Corbett M, Eldridge S, Emberson J, Hernán M, Hopewell S, Hróbjartsson A, Junqueira D, Jüni P, Kirkham J, Lasserson T, Li T et al. (2019) 'RoB 2: a revised tool for assessing risk of bias in randomised trials'. *BMJ* l4898 doi: 10.1136/bmj.l4898.

Strang H (2017) 'Experiments in restorative justice' in *Experiments in Restorative Justice* Drahos P (ed), ANU Press, Canberra.

Strang H, Sherman L, Woods D and Barnes G (2011) Experiments in restorative policing: Final report – Canberra Reintegrative Shaming Experiments (RISE), Australian National University.

Trusty W, Scofield B, Christensen A, White T, Murphy Y, Janis R, Tan H, Hernandez N and Hochstedt K (2025) 'Psychological Symptoms and Academic Dropout in Higher Education: A Six-Year Cohort Study', *Journal of College Student Mental Health*, doi: 10.1080/28367138.2024.2444883

van der Put C, Boekhout van Solinge N, Stams G, Hoeve M and Assink M (2021) 'Effects of Awareness Programs on Juvenile Delinquency: A Three-Level Meta-Analysis', *International Journal of Offender Therapy and Comparative Criminology*, 65(1):68-91, DOI: 10.1177/0306624X20909239.

VCOSS (Victorian Council of Social Service) (2023) *Building for the Future: VCOSS submission to Infrastructure Victoria on updating Victoria's 30 year infrastructure strategy,* Victorian Council of Social Service.

WSIPP (Washington State Institute for Public Policy) (2023) *Nurse Family Partnership Public Health & Prevention: Home- or Family-based*, WSIPP

Watchorn D, Cowan M, Driscoll D, Nimmo D, Ashman K, Garkaklis M, Wilson B and Doherty T (2022) 'Artificial habitat structures for animal conservation: design and implementation, risks and opportunities', *Frontiers in Ecology and the Environment*, 20(5):301-309, doi: 10.1002/fee.2470.

WWCLEG (What Works Centre for Local Economic Growth) (2016) *Evidence review: Employment training—Updated June 2016*, What Works Centre for Local Economic Growth.