



Australian Government
The Treasury

Australian
Centre for
Evaluation

ace

Impact Evaluation: Bring together quantitative methods and program theory in mixed-method evaluations

Dr Vera Newman & Peter Bowers

Australian Centre for Evaluation (ACE), Commonwealth Treasury

Acknowledgement of Country



Introducing the Australian Centre for Evaluation

- **Our vision:** Evaluation evidence supports better policy and programs to improve the lives of Australians.
- **Our mission:** We help put evaluation evidence at the heart of policy design and decision-making.
 - One of the ways we fulfil this mission is to champion *high-quality* impact evaluation

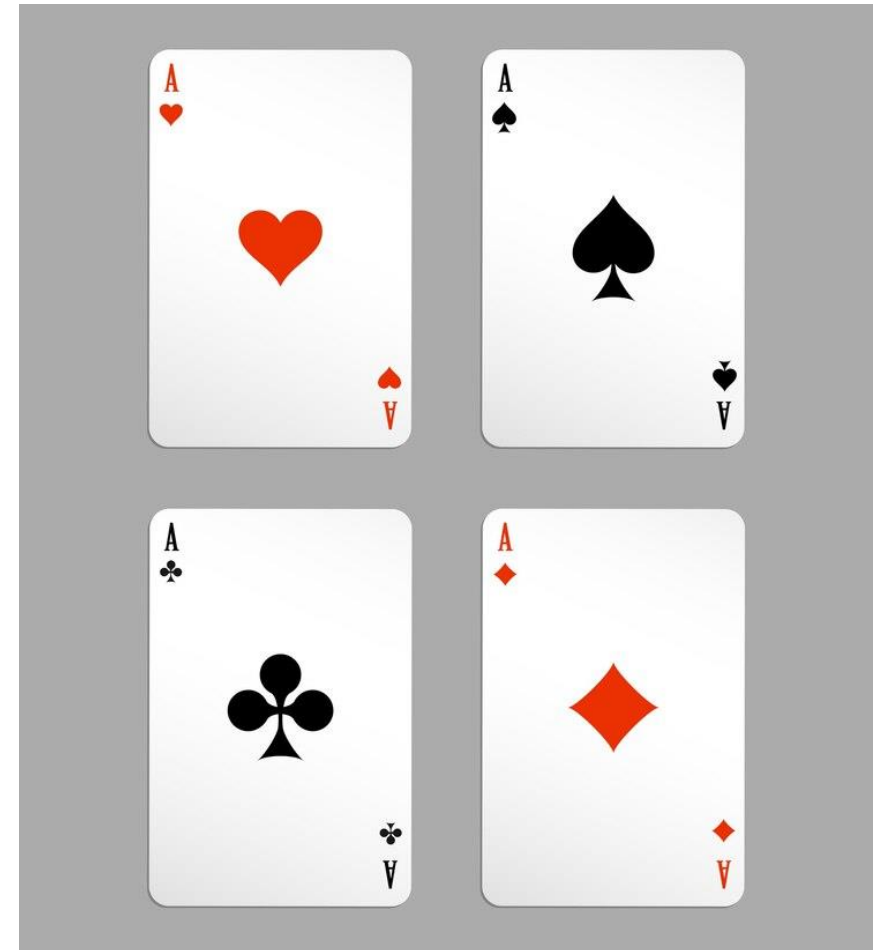


Image: Freepik.com

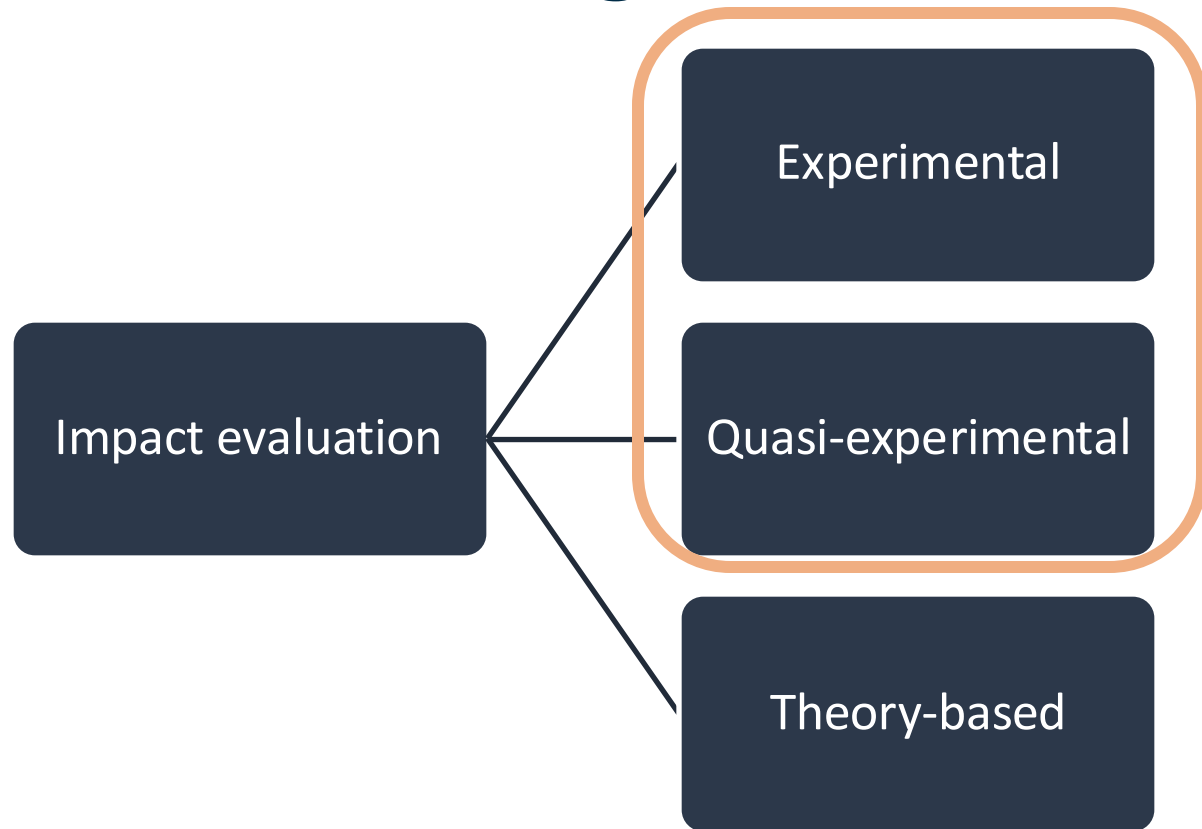
Presentation overview and key take-aways

Program theory and quantitative methods can be brought together in mixed-methods evaluations. In this presentation we will cover some ideas for *how* and *when* to do this:

- 1. Before an intervention:** Mixed-methods will help you to know if a particular method is feasible, and/or likely to generate biased estimates.
- 2. During the delivery of an intervention:** Mixed-methods can help to unpack experiences within the intervention and whether your evaluation is likely to suffer from confounds.
- 3. After an intervention:** Mixed-methods will help you to unpack the quantitative estimates of impact, and lend further insights into results.

What is impact evaluation, and why is a counterfactual important?

Some starting definitions and caveats



We use the term impact evaluation to refer to *causal relationships* between the intervention and an outcome, using a counterfactual

We will not be looking deeply into the theoretical background of:

- Program theory
- Multi- versus mixed-methods
- Sequential and parallel approaches

Our intent is to provide a broad overview of the kinds of ways we see mixed-methods contributing to quantitative impact evaluations

Why conduct an impact evaluation?

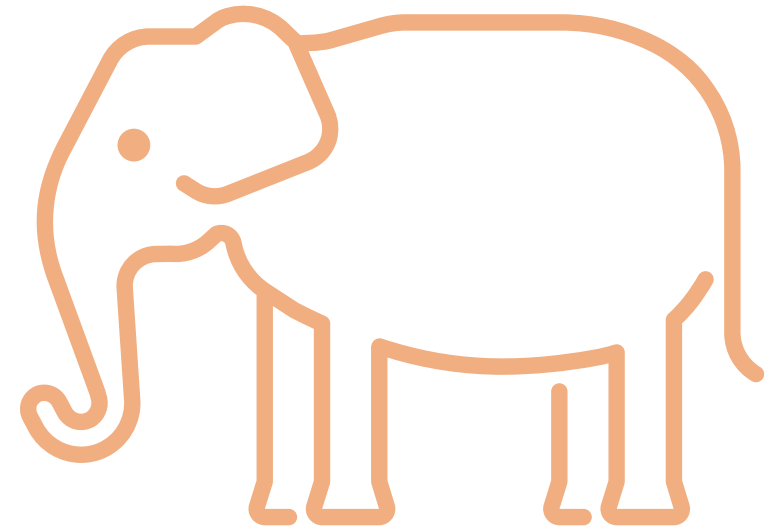
**To understand whether a
program works,
and how well it works.**

We want to be confident we can establish causality, not just correlation



Ethical research conduct

- We think it's most productive to have a conversation about ethics once we are all on the same page on the approaches in question.
- We will consider ethical research conduct later in the talk.
- We think that, where there is **genuine uncertainty** about the impact or relative merits of a program, it is ethical to pursue the rigorous generation of new knowledge – whether this is achieved through an RCT or another evaluation methodology. This aligns with the principle of equipoise.



Using observational data

Observational data *can* generate estimates of impact ...

- Using observational data, you can match your “treatment” group with a comparison group
- The comparison group would be matched with a group of similar individuals, based on *relevant characteristics*
 - Evaluate an activity affecting sleep by matching participants and non-participants on age, health and weight

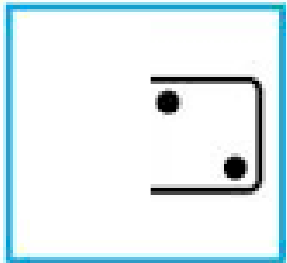
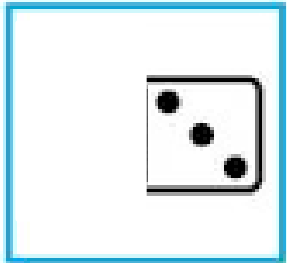
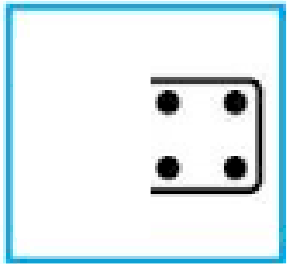
Observational data *can* generate estimates of impact ... If the right ingredients are present

- Using observational data, you can match your “treatment” group with a comparison group
- The comparison group would be matched with a group of similar individuals, based on *relevant characteristics*
 - Evaluate an activity affecting sleep by matching participants and non-participants on age, health and weight
- So what are the right ingredients?
 - Good theory to identify relevant characteristics
 - Data on most or all relevant characteristics
- **How do you know if you have these ingredients?**



A hypothetical example

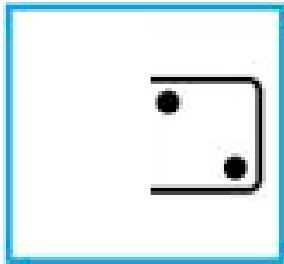
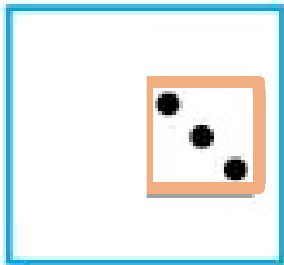
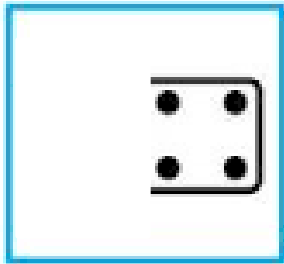
Treatment
group



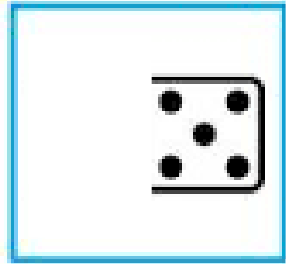
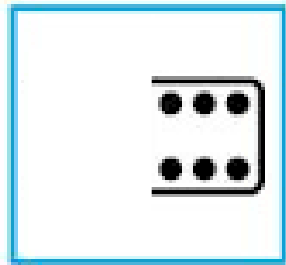
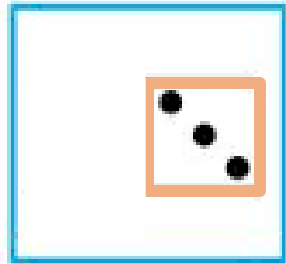
- School-based tutoring program to teach maths to Year 5 children
- Kids or their parents can choose to go into the program, and receive tutoring if they sign-up
- Outcomes are measured via the NAPLAN

A hypothetical example

Treatment group



Comparison group



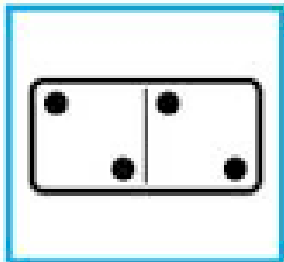
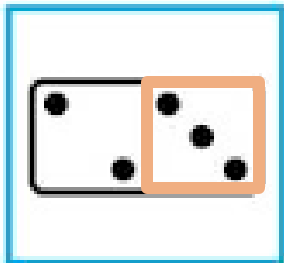
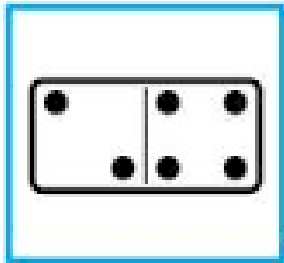
- School-based tutoring program to teach maths to Year 5 children
- Kids or their parents can choose to go into the program, and receive tutoring if they sign-up
- Outcomes are measured via the NAPLAN

- We can use administrative data to find children who have *similar observable characteristics* to those in the treatment group.
- We select some of those children and include them in a comparison group, based on their similarity on observable characteristics.

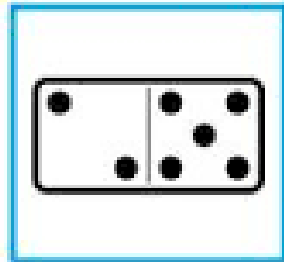
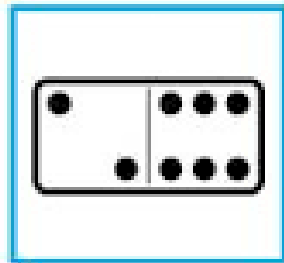
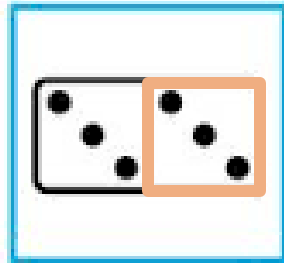
A hypothetical example



Treatment group

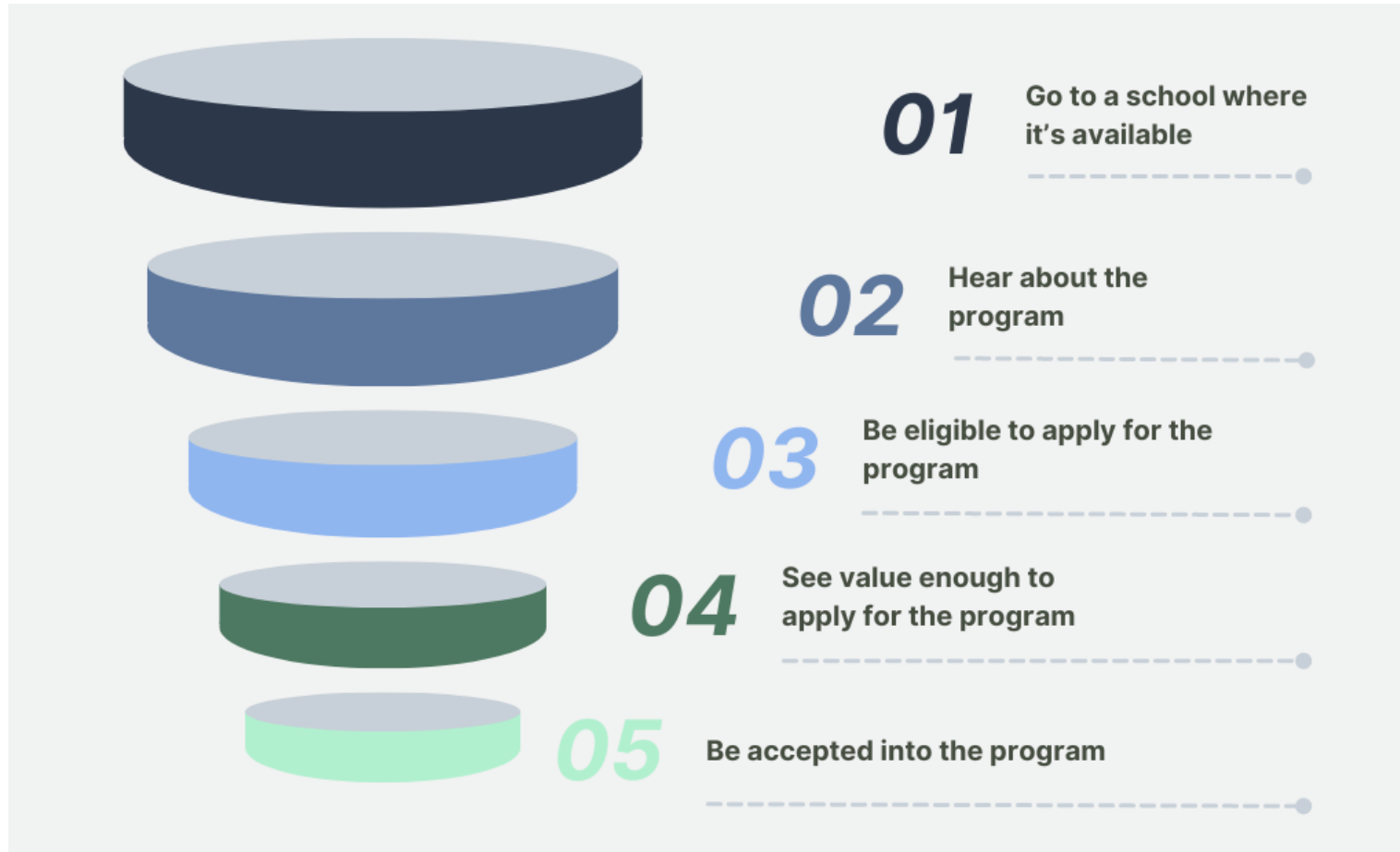


Comparison group



- There might be *unobservable* characteristics that differ between the two groups.
- You can use mixed-methods to understand these unobservable characteristics:
 1. Program implementation
 2. How different cohorts might end up in the program
 3. Whether these factors matter for outcomes

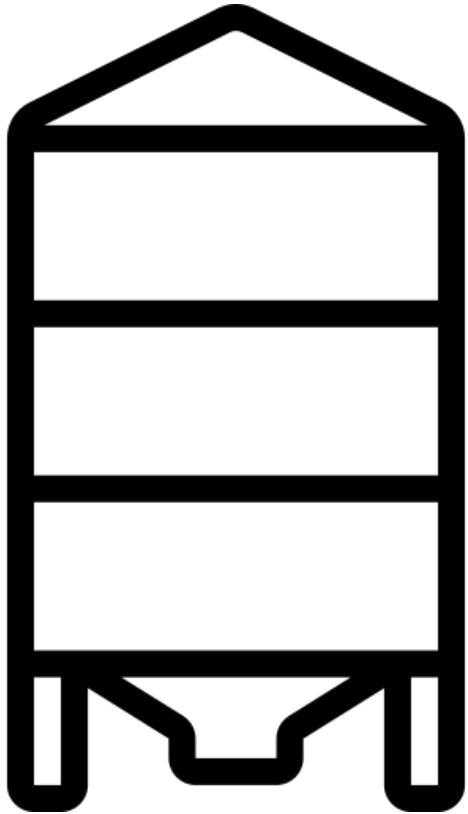
To access this program a child needs to



- Mixed-methods can unpack these factors
- *And whether these factors are important for outcomes*



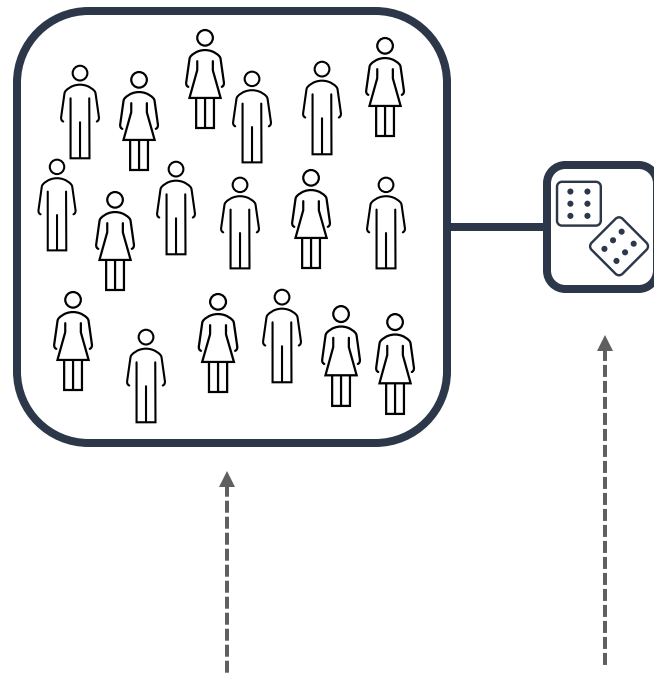
Control on observables approaches don't generate estimates of impact in a silo



Incorporating program theory and using a mixed-methods approach **before** an intervention (during evaluation design) can strengthen evaluations, and help to know whether you *can* or *cannot* generate credible causal estimates

Randomised controlled trials

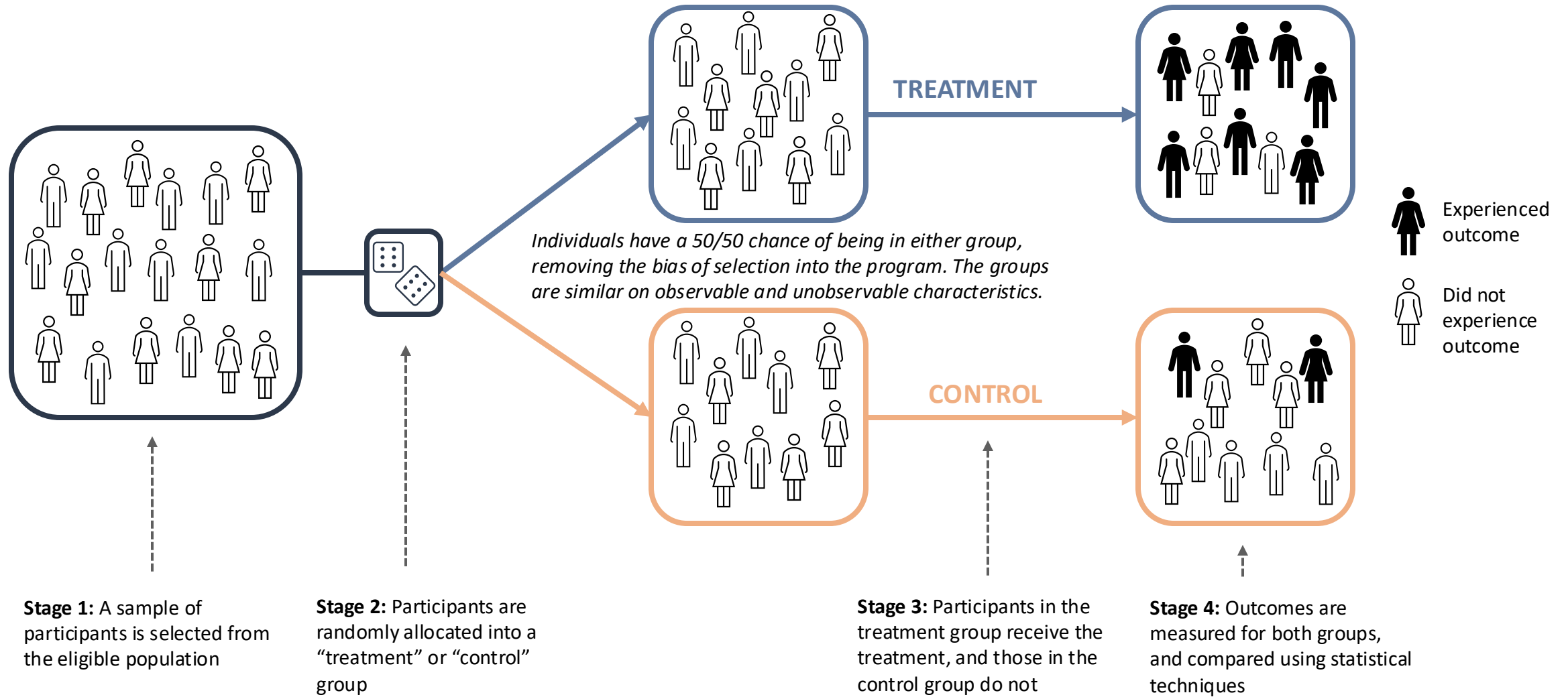
What is the intuition of an RCT?



Stage 1: A sample of participants is selected from the eligible population

Stage 2: Participants are randomly allocated into a “treatment” or “control” group

What is the intuition of an RCT?

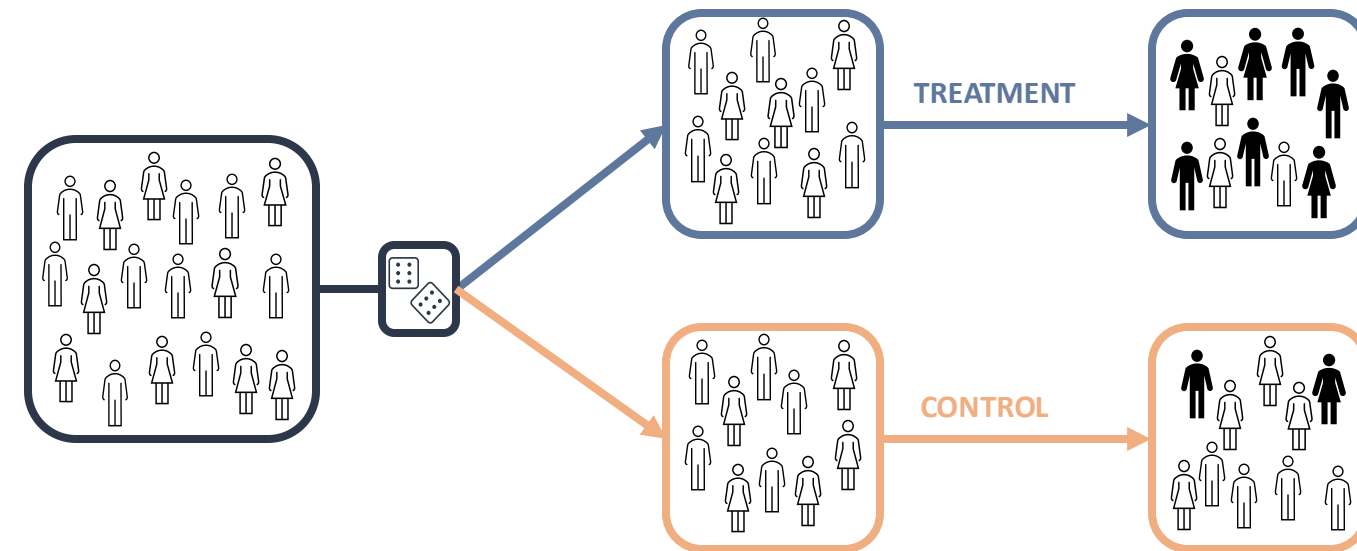




The reality of an RCT is much more complex!

Photo by [Ivan N](#) on [Unsplash](#)

How to incorporate program theory into an RCT?



Most discussion of RCTs ends here.

But what about what happens after you compare the quantitative outcomes, in the difference between the treatment and control group?



How to incorporate program theory into an RCT?

“Why are we seeing the differences we are, and how did they occur?”

Interviews

Surveys

Focus groups

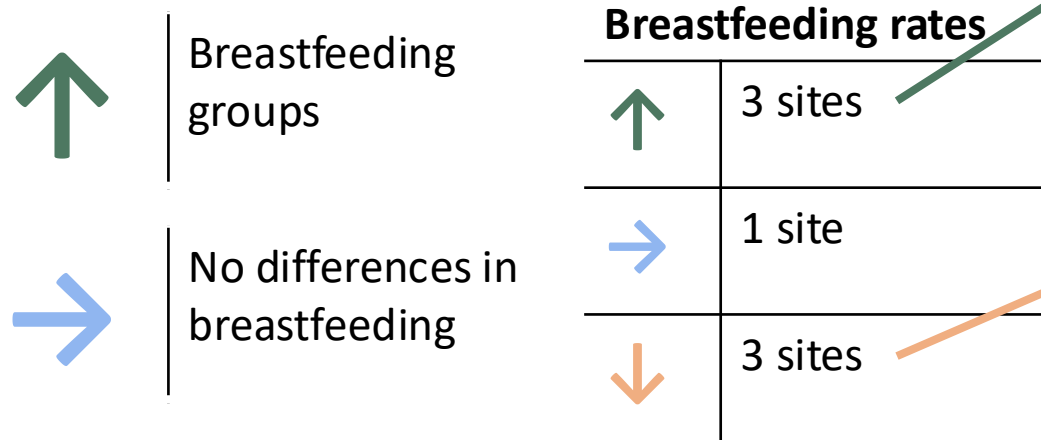
etc ...

Help us to understand the results better, leading to further nuance in the findings

Where have we seen this done in the field?

- Scotland has a particularly low rate of breastfeeding.
- A clustered RCT in Scotland evaluated a policy to provide breastfeeding groups for pregnant/breastfeeding women.
- Intended primary outcome: **increase breastfeeding rates at 6-8 weeks.**

Ethnographic in-depth interviews, focus groups, observations and survey data helped unpack the theory of how organisations would implement the policy:



Evidence of a strong focus of the policy delivery, including strong leadership, multi-disciplinary partnership approach, and reflective action cycles

A focus on the following issues, versus delivering the policy itself:

- Site issues: Relative poverty, geographical barriers
- Personnel issues: Staff shortages and high workloads
- Organisational change

Hoddinott et al., 2010

Ethical research conduct

RCTs are being conducted across Australia

We are aware of...

80+

Non-medical RCTs in Australia

Thousands

of non-medical RCTs internationally

Some examples...

Randomised trial	Organisation
Victorian Healthy Homes Trial (2018 – 2021)	Sustainability Victoria
Job readiness training for job seekers (2019)	Department of Education, Skills and Employment
Online Employment Services Trial (2018 – 2019)	Department of Education, Skills and Employment
Strengthening student's resilience (2021)	Department of Social Services
Nudge vs Superbugs (2018)	Department of Health and Aged Care



Is it ethical to exclude some people from a program?

1 First check if you will already be forced to exclude some people

- Limited funding may mean some people are already excluded so an **over-subscription design** might work (i.e., randomise who is eligible)
- Can be fairer than picking favourites

2 Then check if you actually need to exclude people entirely

Delay participation rather than exclude with:

- **Waitlist control design**
- **Stepped wedge design**

3 Then think about ethics

An RCT might be ethical if at least one of the following are true:

- **Genuine uncertainty (i.e., ‘ equipoise ’)** about the *exact* effects of the program (surprisingly common!)
- **Informed, opt-in consent** gives the participants the right to decide
- **People in the control group** still get something
- **There are long-term benefits** of knowing the truth
- **There is significant harm to a less rigorous evaluation**

Who was Frances Kelsey?



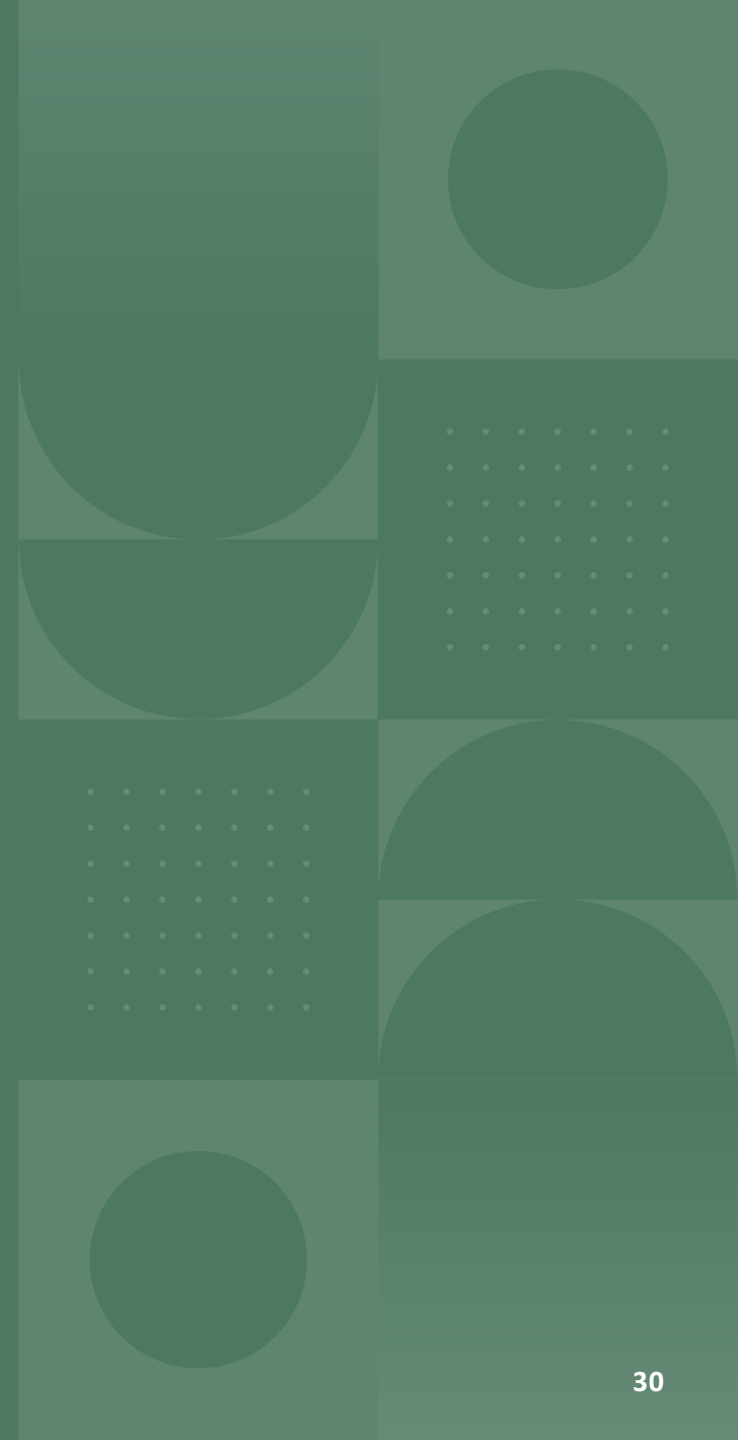
Who was Frances Kelsey?



“[Kelsey] prevented ... the birth of hundreds or indeed thousands of armless and legless children.”

- The Washington Post

Regression discontinuity designs



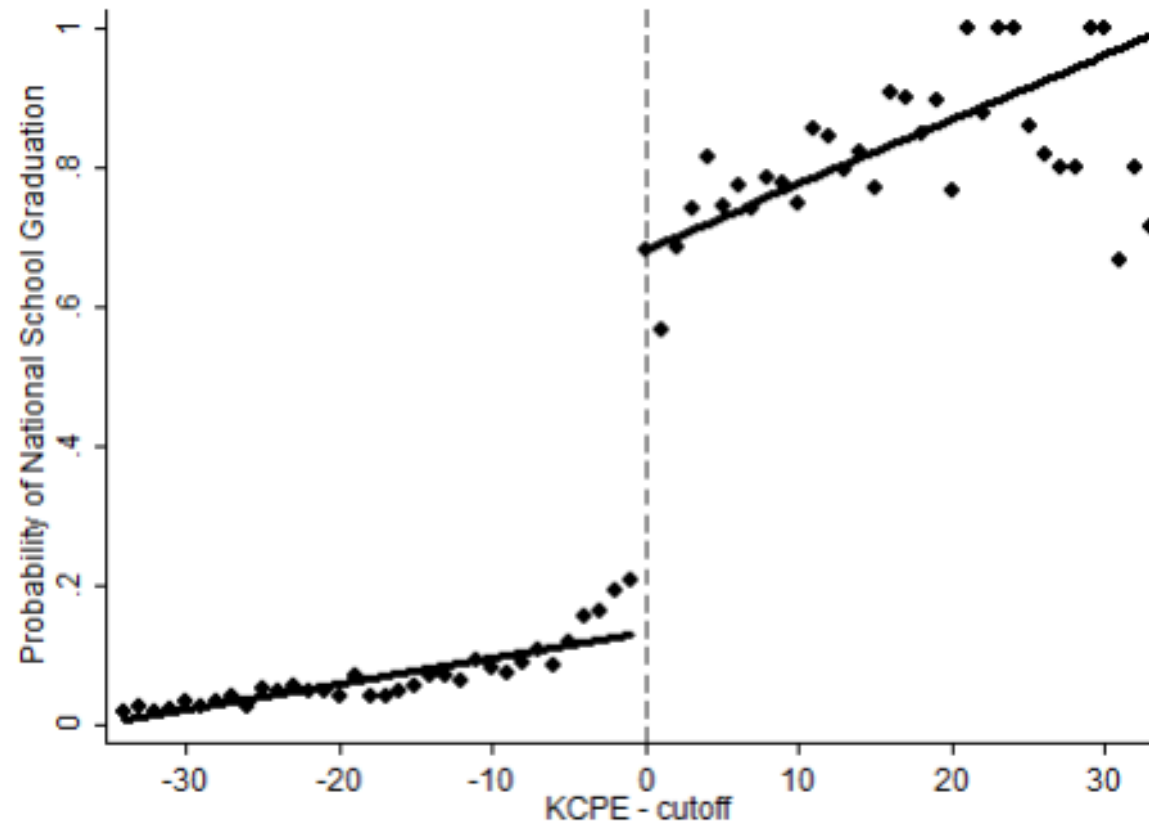
Regression discontinuity design (RDD) uses hard cut-offs on eligibility

For example:

- Minimum test mark to attend a selective school
- People who were employed before 1 March 2020 eligible for JobKeeper
- People on one side of a street in a certain school district so allowed to attend sought after school

Students above the cut-off attended an elite selective school, those below did not

Figure 2: Probability of National School Graduation

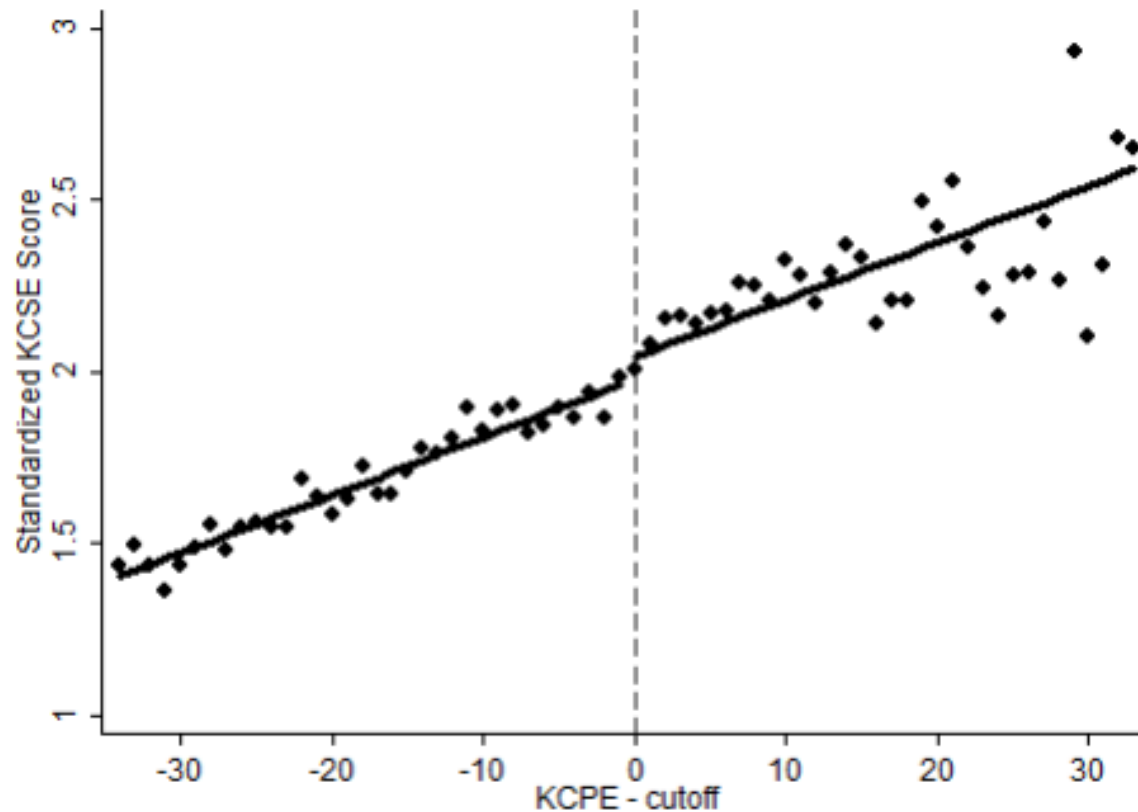


Source: Adrienne M. Lucas and Isaac Mbiti, "Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya", *American Economic Journal: Applied Economics* July 2014 6(3): 234-263



Does attending a selective school have a causal effect on performance?

Figure 4: Student Achievement



Results:

- No statistically significant effect from attending a selective school
- There may be a very small effect – but if so, it would be <0.12 standard deviations

As always, a mixed methods approach is key

Qual used before intervention

- If the research question matters
- If the assumptions of the method are satisfied – *e.g. Requires understanding Ministry of Education processes*

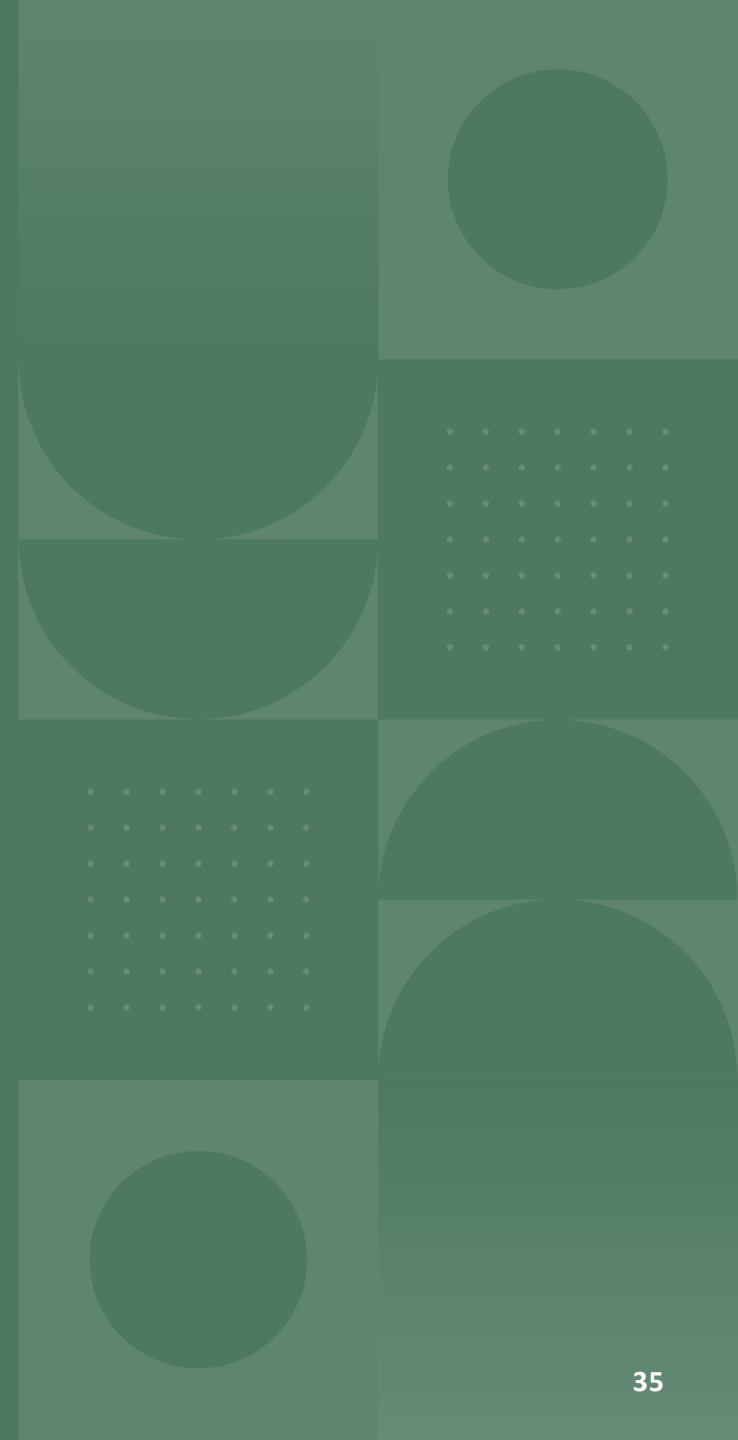


Qual used after intervention

- ‘Why’ selective schools did not have a significant effect
- The policy implications of the results



Conclusion



In summary

A mixed methods approach improves both the usefulness and accuracy of impact evaluations

This is true for all impact evaluation approaches – including those discussed today:

- Randomised trials
- Regression discontinuity
- Matching





Australian Government
The Treasury

Australian
Centre for
Evaluation

ace

Thank you

peter.bowers@treasury.gov.au

vera.newman@treasury.gov.au

evaluation.treasury.gov.au