

Types and strength of evidence: assessing sources, and producing your own evidence

How (and why) to do an RCT (intermediate/advanced)

May 2023

Overview of session

1

RCTs

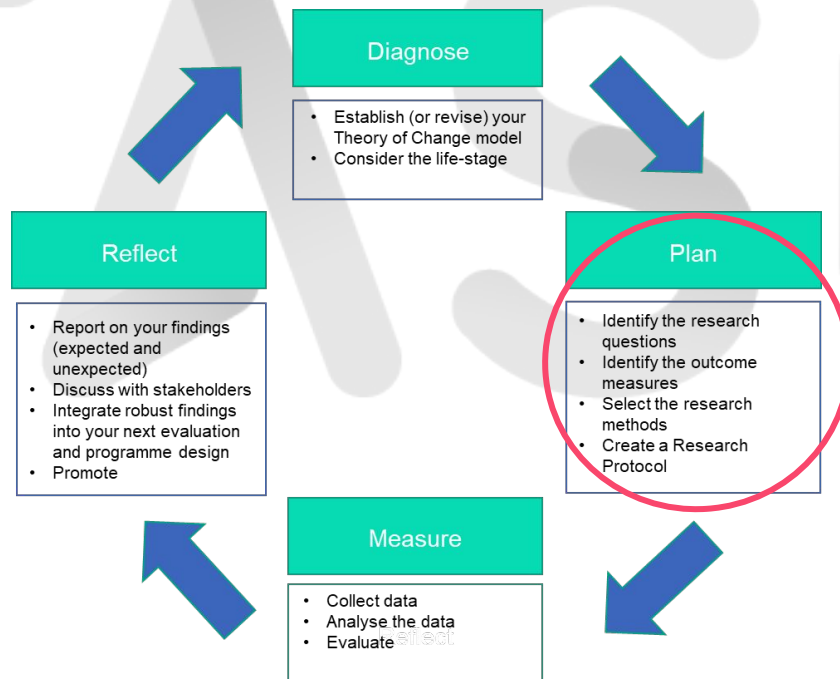
- Theory and background
- Examples

2

Assessing evidence

NO MATHS!

The evaluation cycle



Correlation

Correlation means there is a relationship between variables

Variable **A** and Variable **B** are related somehow



Causation

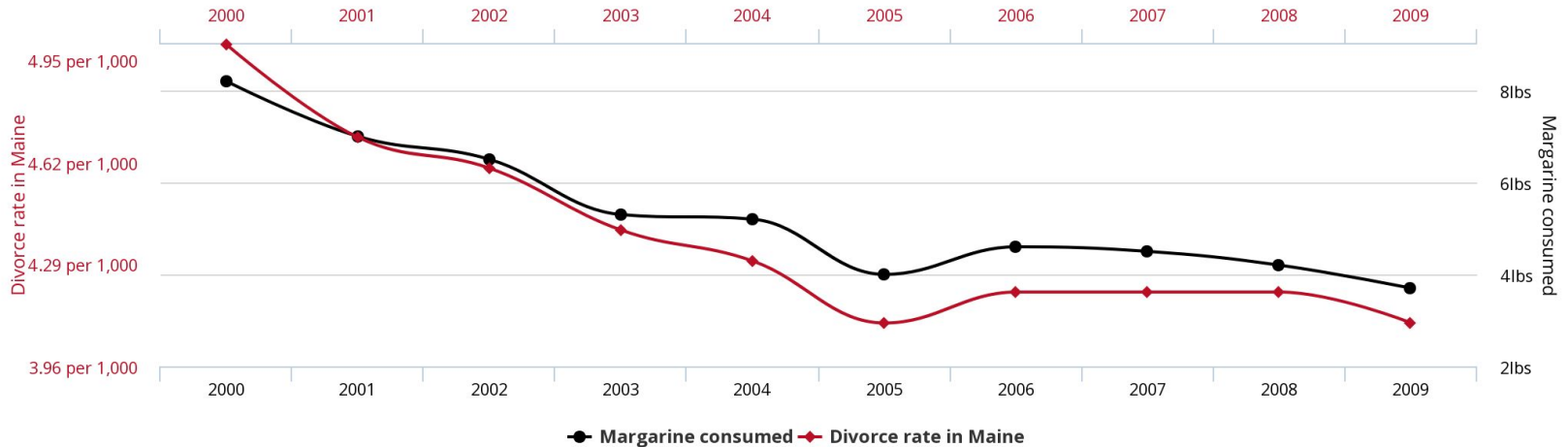
Causation means A causes B

Aim of impact evaluation is to demonstrate causation



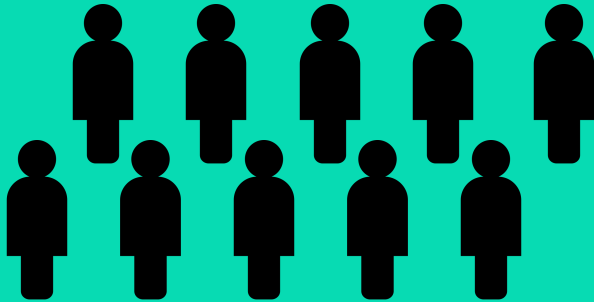
Correlation is not causation

Divorce rate in Maine
correlates with
Per capita consumption of margarine

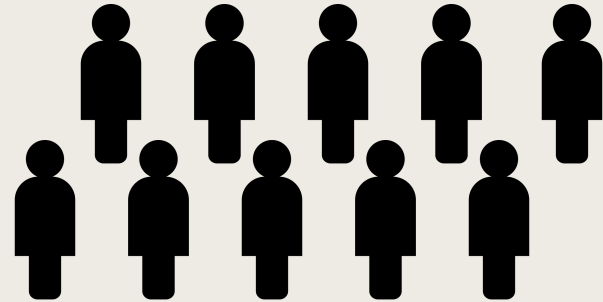


Example: WP programme

Apply for WP programme

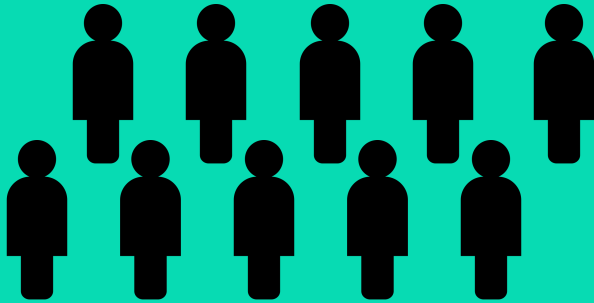


Do not apply



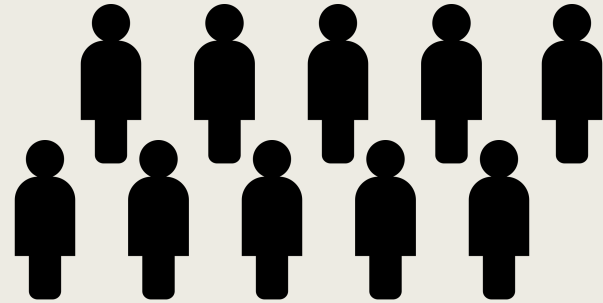
Example: WP programme

Apply for WP programme



80% enrol in HE

Do not apply



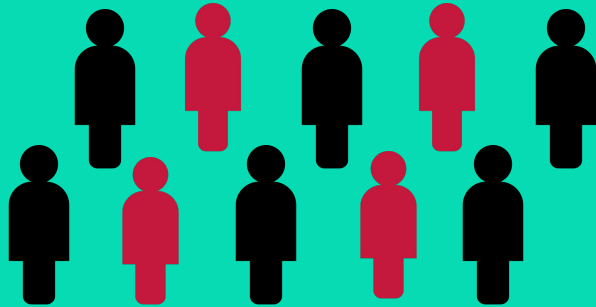
50% enrol in HE

Differences between groups?

- **Demographic differences?** (e.g. gender, prior attainment, location)
- **Other differences?** (e.g. family support, individual motivation, other barriers)?

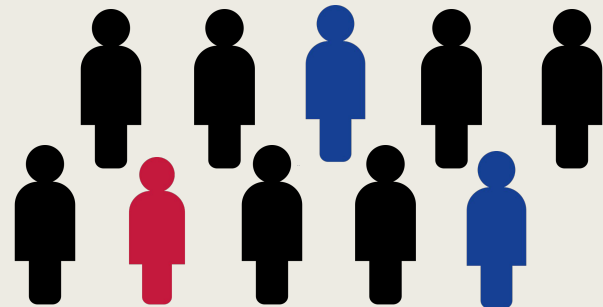
Selection bias

Apply for WP programme



80% enrol in HE

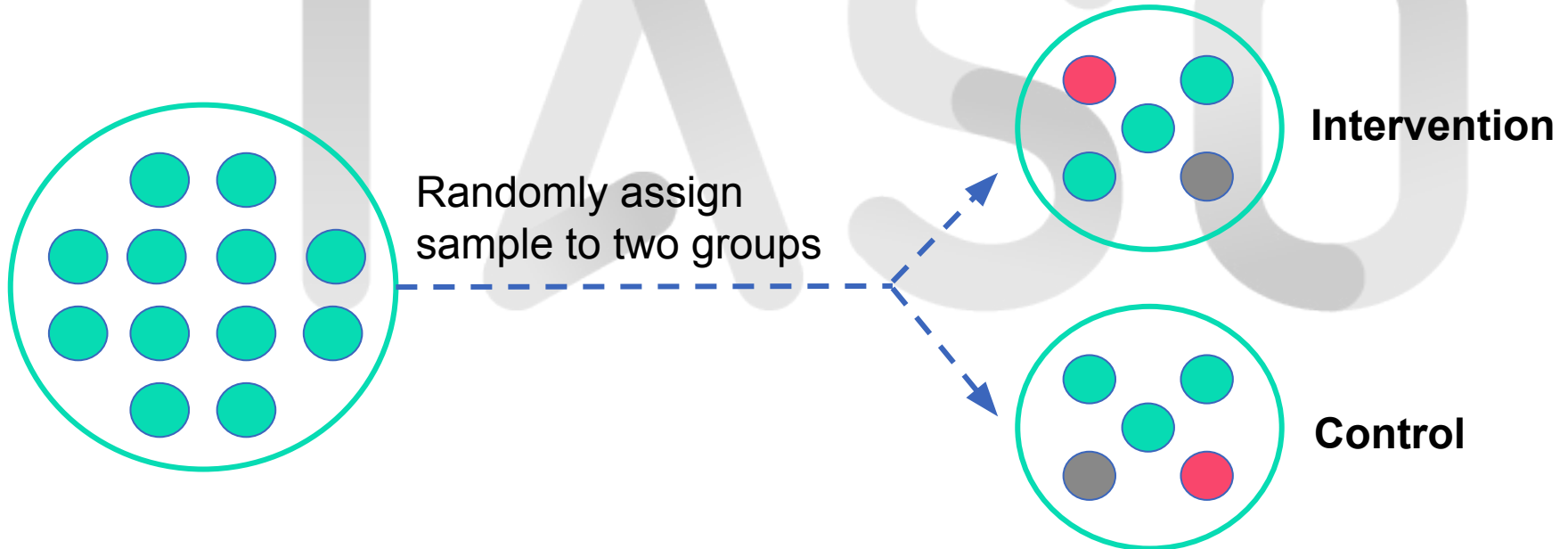
Do not apply



50% enrol in HE

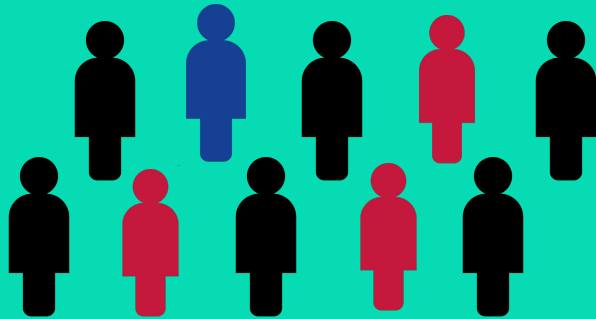
Why randomise?

Control for **individual differences** between participants by using **random assignment**

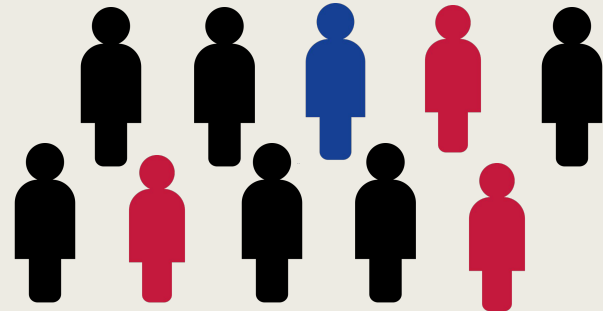


Randomisation removes selection bias

Apply and get WP programme



Apply and don't get WP programme



Why randomise?

When applied properly, we can **estimate the causal effect** of an intervention on the measured outcomes



What are RCTs good for?

**What
works?**



**Why it
works?**



**How it
works?**



OfS standards of evidence

Type of evidence	Description	Evidence	Claims you can make
Type 1 - narrative	The impact evaluation provides a narrative or a coherent theory of change to motivate its selection of activities in the context of a coherent strategy.	Evidence of impact elsewhere and/or in the research literature on access and participation activity effectiveness or from your existing evaluation results.	We have a coherent explanation of what we do and why our claims are research-based.

OfS standards of evidence

Type of evidence	Description	Evidence	Claims you can make
Type 2 – empirical enquiry	The impact evaluation collects data on impact and reports evidence that those receiving an intervention have better outcomes, though does not establish any direct causal effect.	Quantitative and/or qualitative evidence of a pre/post intervention change or a difference compared to what might otherwise have happened.	We can demonstrate that our interventions are associated with beneficial results.

OfS standards of evidence

Type of evidence	Description	Evidence	Claims you can make
Type 3 – causality	The impact evaluation methodology provides evidence of a causal effect of an intervention.	Quantitative and/or qualitative evidence of a pre/post treatment change on participants relative to an <u>appropriate control or comparison</u> group who did not take part in the intervention.	We believe our intervention causes improvement and can demonstrate the difference using a control or comparison group.

Considerations

Strengths

- A Type 3 method under the OfS Standards of Evidence
- Ideal tool for causal inference
- Helpful in determining whether an intervention / public policy works
- Typically have high internal validity

Limitations

- Can be time consuming and expensive
- Hawthorne effect and other internal validity concerns
- Sometimes difficult to randomize participants
- Can require large sample
- External validity considerations

Ethics

- **If there is substantial, consistent, high-quality evidence that something is effective, it shouldn't be withheld from anyone who might benefit**
- But it is surprising how often this is not the case
- We also should consider whether the benefit justifies the cost (including costs to participants such as opportunity costs)



Key steps

1. Identify your intervention(s)
2. Define the outcome
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

Test, Learn, Adapt:

Developing Public Policy with
Randomised Controlled Trials

Laura Haynes

Owain Service

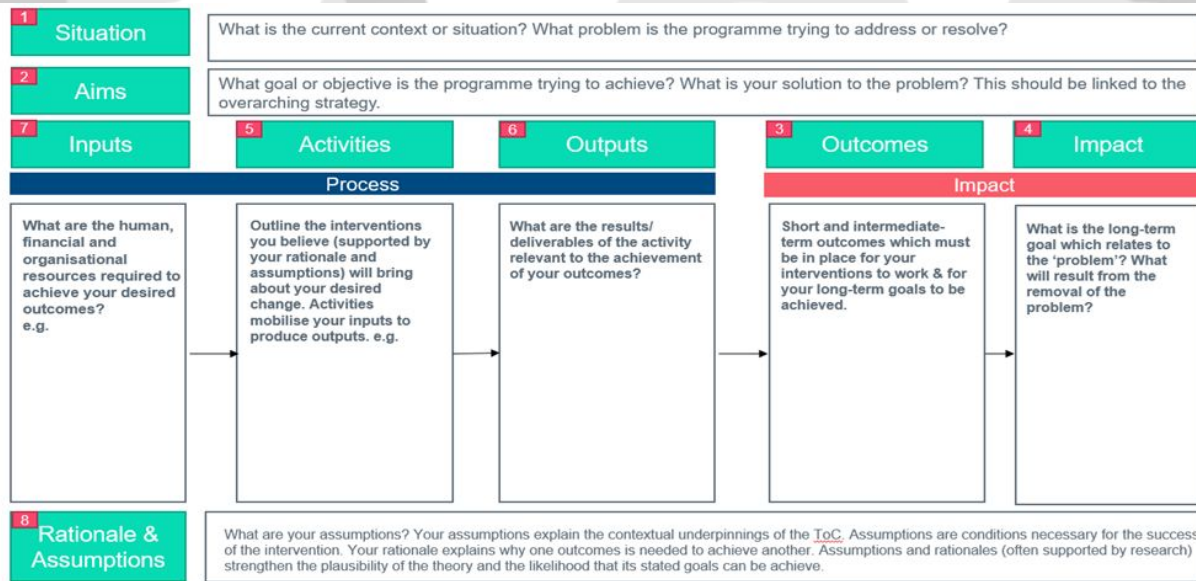
Ben Goldacre

David Torgerson

Key steps

1. Identify your intervention(s)
2. Define the outcome
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

Intervention mapping in ToCs



What is your research question?

Primary Research Question

Causal impact of your evaluation

Did [scheme] increase [main outcome] among [group]?

Did Summer School attendance improve enrolment rates among participants?

Secondary Research Question

Focus on specific **groups** or **intermediate outcomes**

Did [scheme] increase [main outcome/ secondary outcome] among [group/subgroup]?

Did Summer School attendance improve enrolment rates among estranged students?

Activity (3 minutes)

- Consider a specific intervention you run at your institution
- Discuss with your neighbour your primary research questions
- Consider:
 - Who will use the findings and how?
 - What do stakeholders need to learn from the evaluation?
 - What questions will you be able to answer and when?

Key steps

1. Identify your intervention(s)
2. **Define the outcome**
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

Outcome measures

- **“I’ll know [outcome reached] when I see [indicator]”**
- Data need to be available at the level of your outcome - normally at the individual level
- Examples:
 - Core impact (e.g. A level attainment, university acceptances, continuation).
 - Interim or proxy outcome (e.g. GCSE selections, sign-ups to events)
 - Validated scales (e.g. from academic research, externally-administered tests),
 - Self-report objective (e.g. actual knowledge),
 - Self-report subjective (e.g. perceived knowledge)

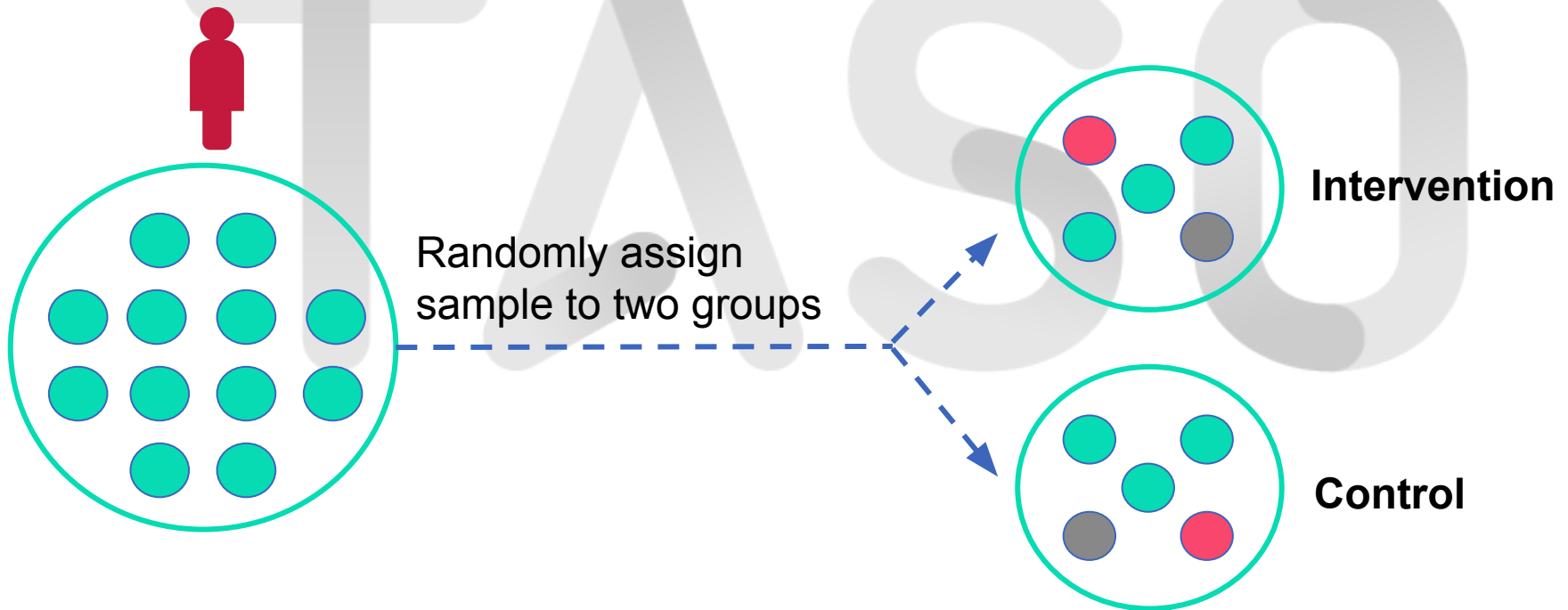
ACTIVITY: Outcomes (2m)

- Identify an appropriate outcome indicator for the intervention you want to evaluate.

Key steps

1. Identify your intervention(s)
2. Define the outcome
3. **Decide on randomisation unit**
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

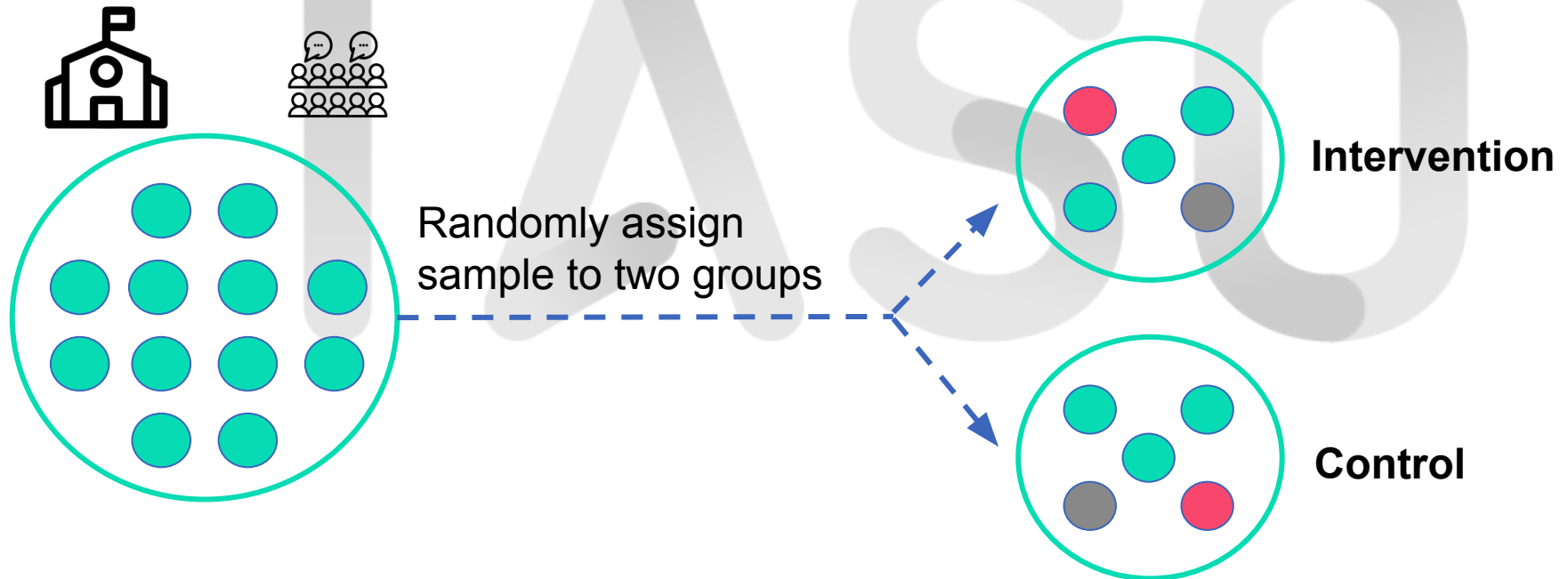
Randomisation unit: people



Randomisation unit

- Where programmes are oversubscribed, or there is only funding for a certain number of people to undertake a certain activity, this offers a great opportunity to randomise at the level of individuals who apply
 - **Intervention:** people who get a spot
 - **Control:** those who don't
- Sometimes, your control group might get a spot later (e.g. next term)
- If this is the case you could conduct an RCT using short-term outcome measures, but not longer term outcome measures

Randomisation unit: clusters (e.g. schools/classes)



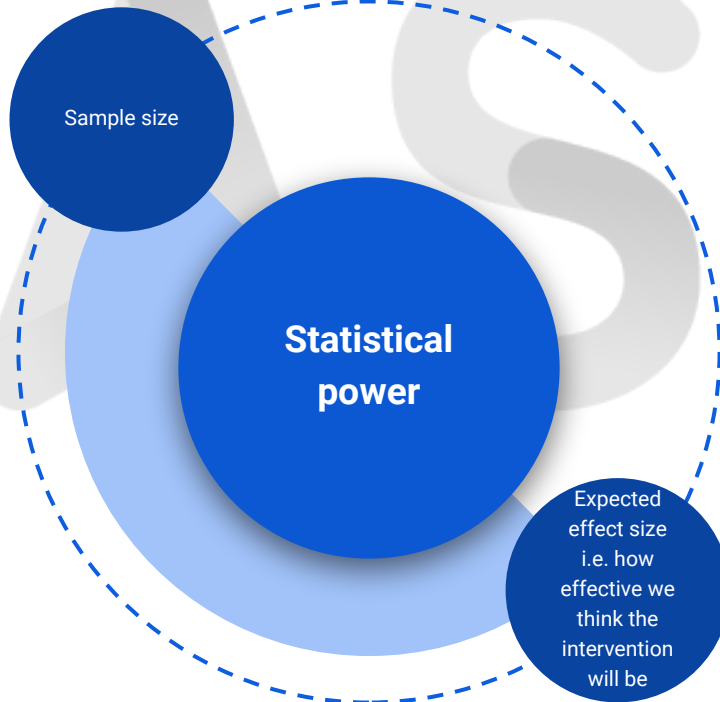
Randomisation unit: clusters (e.g.schools/classes)

- Why cluster?
 - Practical reasons
 - The intervention might need to be delivered to groups
 - Avoid 'spillover'
 - E.g. students might share education resources with their friends
- **But you'll need a bigger sample**

Key steps

1. Identify your intervention(s)
2. Define the outcome
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

Sample size



**OFTEN
CONSTRAINED BY
SAMPLE! See
further reading**

Sample size

- In an ideal world, >1000 (across treatment and control)
- Can reduce to a couple of hundred with good baseline data
- Also depends on:
 - The outcome (how hard it is to influence)
 - How effective you think the intervention will be
 - If your intervention is light-touch (e.g. a text) you need a large sample
 - If it is very substantial (e.g. a huge bursary) you would expect (or want) a large effect so the sample size required would be smaller
- A 50:50 split between intervention/control is the most conservative approach for power

Key steps

1. Identify your intervention(s)
2. Define the outcome
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
- 5. Randomly assign treatments**
- 6. Roll out your intervention**
- 7. Measure your results and determine impact**
- 8. Reflect and adapt your intervention**
- 9. Repeat...**



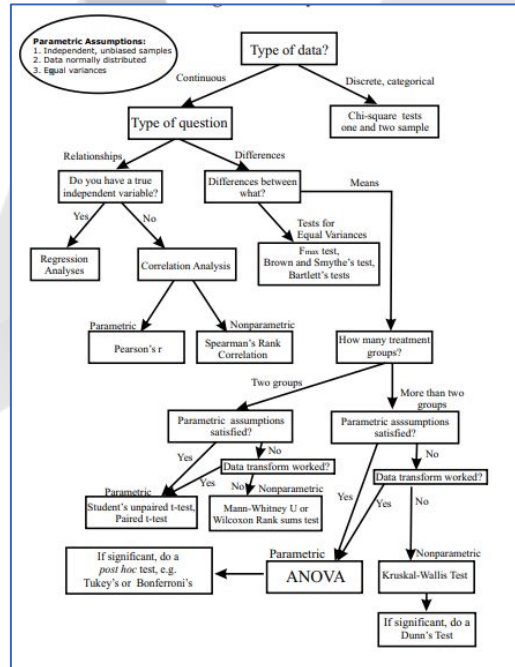
Balance checks



Balance checks

Variable	Treatment	Control	P-value
Female	0.67	0.7	0.51
White	0.3	0.31	0.81
IMD Q	2.46	2.31	0.34
POLAR Q	3.65	3.48	0.59
FSM	0.32	0.33	0.29
Family history of HE	0.32	0.32	0.87
Number of top grades	7.82	7.3	0.99

Analysis - stats required



Example 1: Aston Pathways to STEM

TASU

Key steps

1. Identify your intervention(s)
2. Define the outcome
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

Intervention

- The Pathway to Science, Technology, Engineering and Maths (STEM) programme aims to assist students who are considering a career in STEM in their journey to studying STEM subjects in HE.
 - 12-month programme
 - Induction session
 - Careers advice sessions
 - UCAS personal statement day
 - Summer school
 - Graduation and transition event.



Intervention versus control

Intervention: standard programme

Control: flexible programme

**Delivered with fewer events
and sometimes virtually**

Key steps

1. Identify your intervention(s)
2. **Define the outcome**
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

Outcome

Primary outcome measures:

- whether students enrol in a STEM-related course
- whether students enrol at HE

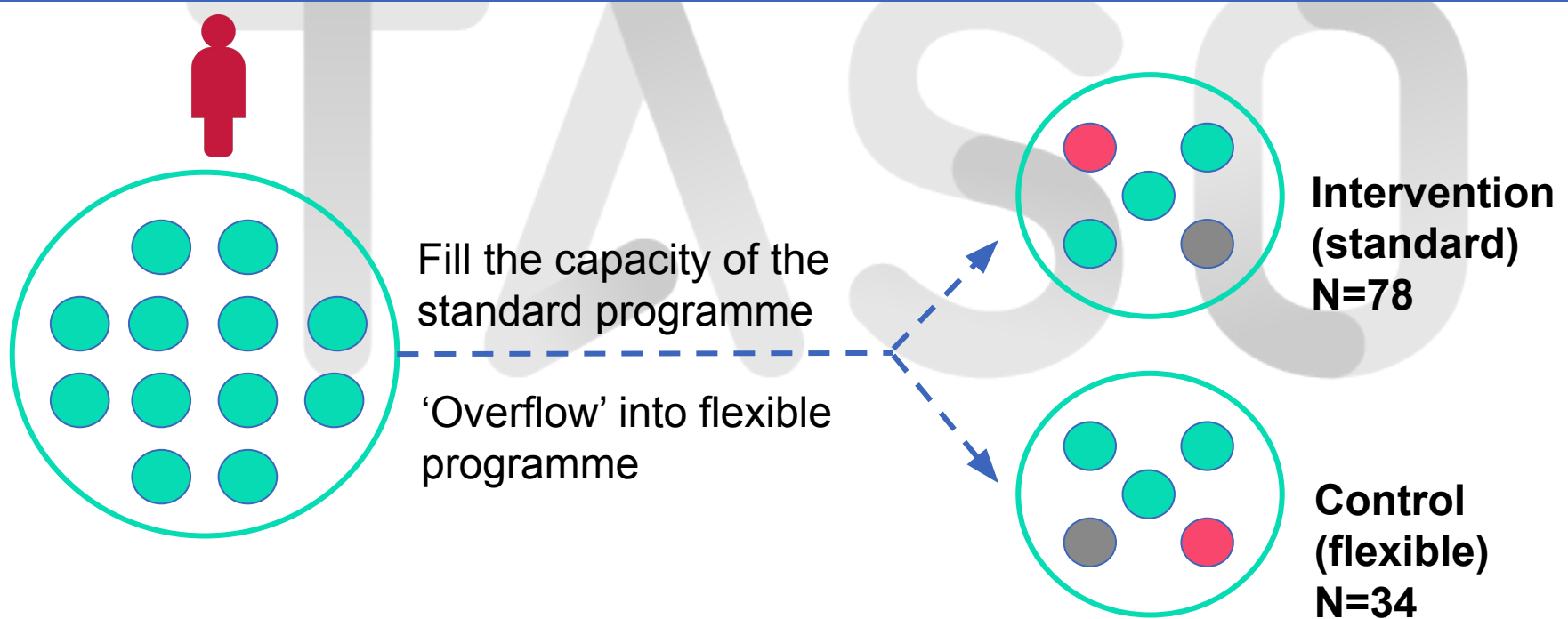


Can be collected at student-level via HEAT...but a time delay!

Key steps

1. Identify your intervention(s)
2. Define the outcome
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

Randomisation unit: applicants



Randomisation unit: applicants

- Sample constrained by number of eligible applicants
- RCT run as pilot to practice using method and learn what was viable
- Opportunity to re-run and combine data across multiple years

Key steps

1. Identify your intervention(s)
2. Define the outcome
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
- 6. Roll out your intervention**
- 7. Measure your results and determine impact**
- 8. Reflect and adapt your intervention**
- 9. Repeat...**

Early findings - caveats apply

- Using interim data, there is not evidence that the standard programme improves the likelihood of students attending HE over those on the less resource intensive flexible programme...that is not necessarily a bad thing...
- **Full data is needed to properly analyse the results of the RCT and this will flow through in future years.**

Example 2: Bursary information RCT

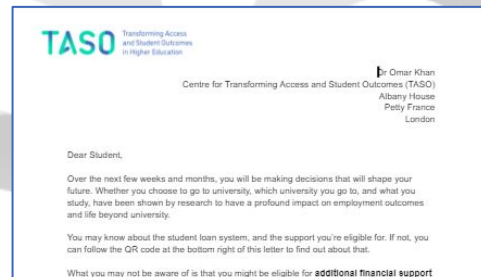


Key steps

1. Identify your intervention(s)
2. Define the outcome
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

Intervention

- Brown envelopes
 - Letter for student
 - Booklet
 - Information about bursaries which are available at university



You may know about the student loan system, and the support you're eligible for. If not, you can follow the QR code at the bottom right of this letter to find out about that.

What you may not be aware of is that you might be eligible for **additional financial support** from the university you go to. Many universities offer additional support, in the form of bursaries and scholarships. For example:

- Bath University offers bursaries of £3,000 if your household income is less than £25,000
- Imperial College London offers bursaries of £5,000 if your household income is less than £16,000

Intervention versus control

Intervention: get booklet and letter*

Sent to school and (we hope) distributed to students - pragmatic choice

Control: don't get booklet and letter

Key steps

1. Identify your intervention(s)
2. **Define the outcome**
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

Outcome

Primary outcome measure:

- whether students applies to HE

Will be accessed via UCAS Outreach Evaluator



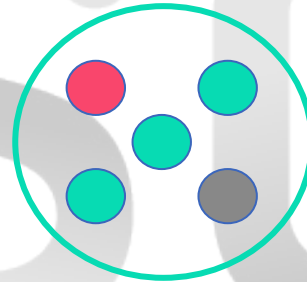
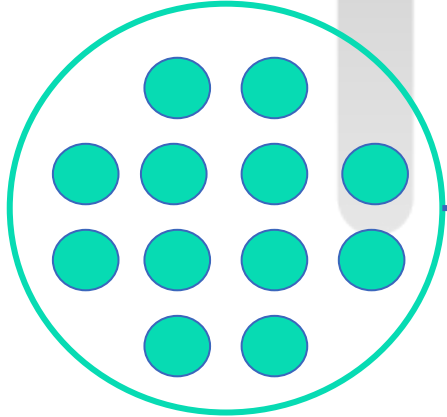
Key steps

1. Identify your intervention(s)
2. Define the outcome
3. Decide on randomisation unit
4. Decide how big your sample needs to be (or can be)
5. Randomly assign treatments
6. Roll out your intervention
7. Measure your results and determine impact
8. Reflect and adapt your intervention
9. Repeat...

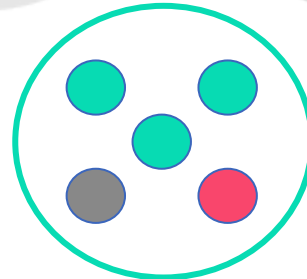
Randomisation unit: schools



570 schools/colleges with >1 A-level student



Intervention
N= 380



Control
N=190

Randomisation unit: schools

- Large number of units
- Average size of a cluster estimated to be 37 students
- Affects overall 'statistical power' but large sample size balances this out
- Data only available at level of the school via UCAS - will be 'expanded' into a pseudo-individualised dataset
- Would always be better to have individual data - consent and tracking are challenging

Activity (5 minutes)

Thinking about the intervention you want to evaluate:

- Decide on randomisation unit and design

Consider:

- Will you be able to get the data you need?

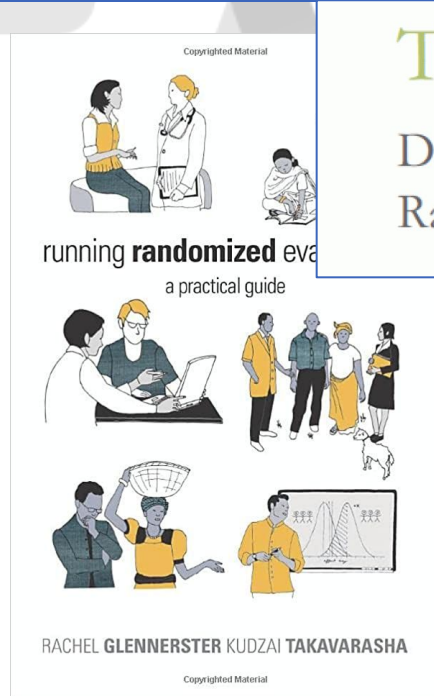
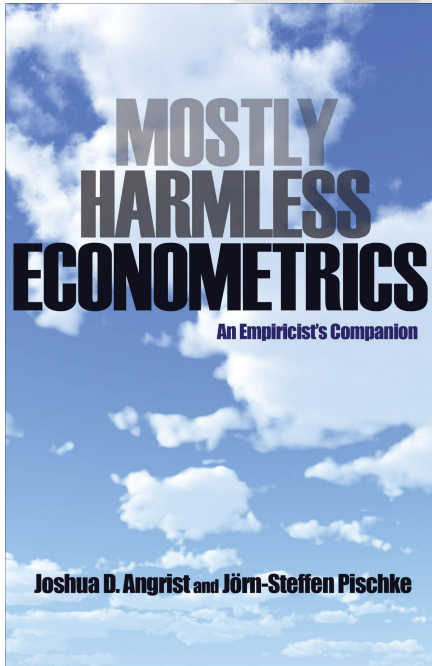
Questions

Further TASO reading/resources

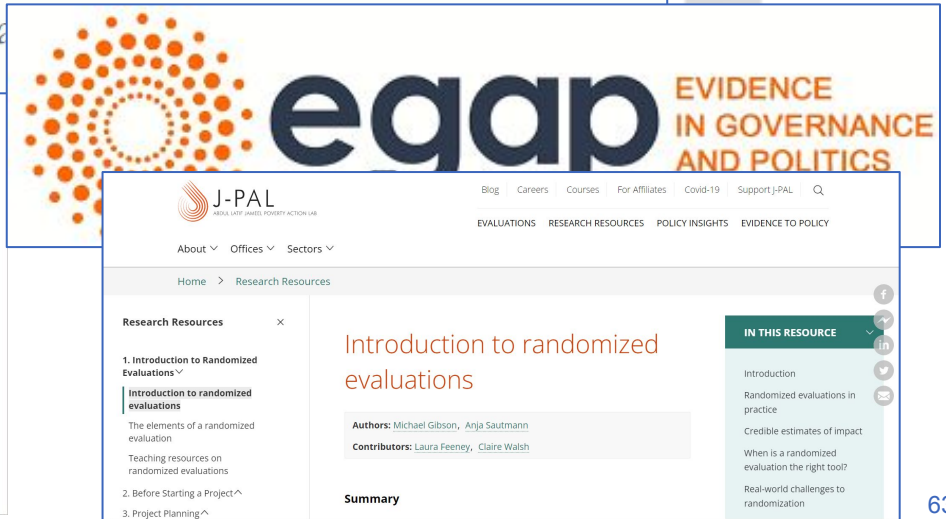
- TASO webinar on randomisation
 - More detail on different randomisation mechanisms
- Ethics case study on RCT
- **Protocols on our website - provide step by step guides on our RCTs and templates which you can use!**
 - Summer schools RCT
 - Further multi-intervention outreach RCTs
 - Learner analytics RCT

The screenshot shows the TASO website page for 'Randomised Controlled Trials (RCTs) - theory, methods and practice'. The page features a navigation bar with 'RESEARCH', 'EVIDENCE & EVALUATION', 'GET INVOLVED', 'NEWS & BLOG', and 'ABOUT'. The main content area includes a breadcrumb trail: 'Home > Evidence & Evaluation > Evaluation guidance resources > Randomised Controlled Trials (RCTs) - theory, methods and practice'. The title 'Randomised Controlled Trials (RCTs) - theory, methods and practice' is prominently displayed. Below the title, there is a 'Share content' section with social media icons. The main text describes RCTs as commonly used in research but challenging to implement, and mentions a webinar recording. A list of sections is provided: 'RCTs - theory and background', 'Summer Schools project - a practical example', 'Early findings', and 'Q&A'. A video player thumbnail for 'Unlocking Effective Evaluation: Randomised Controlled Trials (RCTs) - Theory, methods and practice' is shown, featuring speakers Rain Sherrick and Sarah Chappell. A right-hand sidebar contains a 'EVIDENCE & EVALUATION' menu with items like 'Evidence toolkit', 'Evidence ratings', 'Our approach to evaluation', 'Evaluation guidance resources', 'Correlation versus causation', 'Randomised Controlled Trials (RCTs) - theory, methods and practice', 'Introduction to quasi-experimental designs', 'Impact evaluation with small cohorts', 'Theory of Change resources', 'Pre and post-test design', 'Survey design and validation', 'Question time - addressing common challenges in evaluation', 'Research ethics guidance', and 'Resource hub'.

Further reading/resources



Test, Learn, Adapt:
Developing Public Policy with
Randomized Evaluations



Takeaways

- RCTs are a strong and viable impact evaluation method
- Start with the key steps outlined here
- Speak to TASO if you want to undertake one

Return to main room

TASU