Testing in a regulatory environment

Michael Sanders & Elisabeth Costa
UKCN, December 2016

Summary

These slides were presented by Michael Sanders and Elisabeth Costa of the Behavioural Insights Team (BIT) at the second workshop of the UKCN consumer remedies project held at the Competition and Markets Authority on 9 December 2016. The theme of this workshop was 'enhancing impact through customer testing'.

This presentation reviews the different customer testing techniques that the BIT uses to test proposed interventions.

The presentation begins with an overview of the main stages that form part of a randomised controlled trial, with a focus on the key concept on randomisation. The presentation then outlines some quasi-experimental methods available to researchers such as propensity score match and regression discontinuity design. Next, the presentation reviews the main pros and cons of conducting lab tests and introduces the BIT's new tool, Predictiv, which can be used to design lab trials.
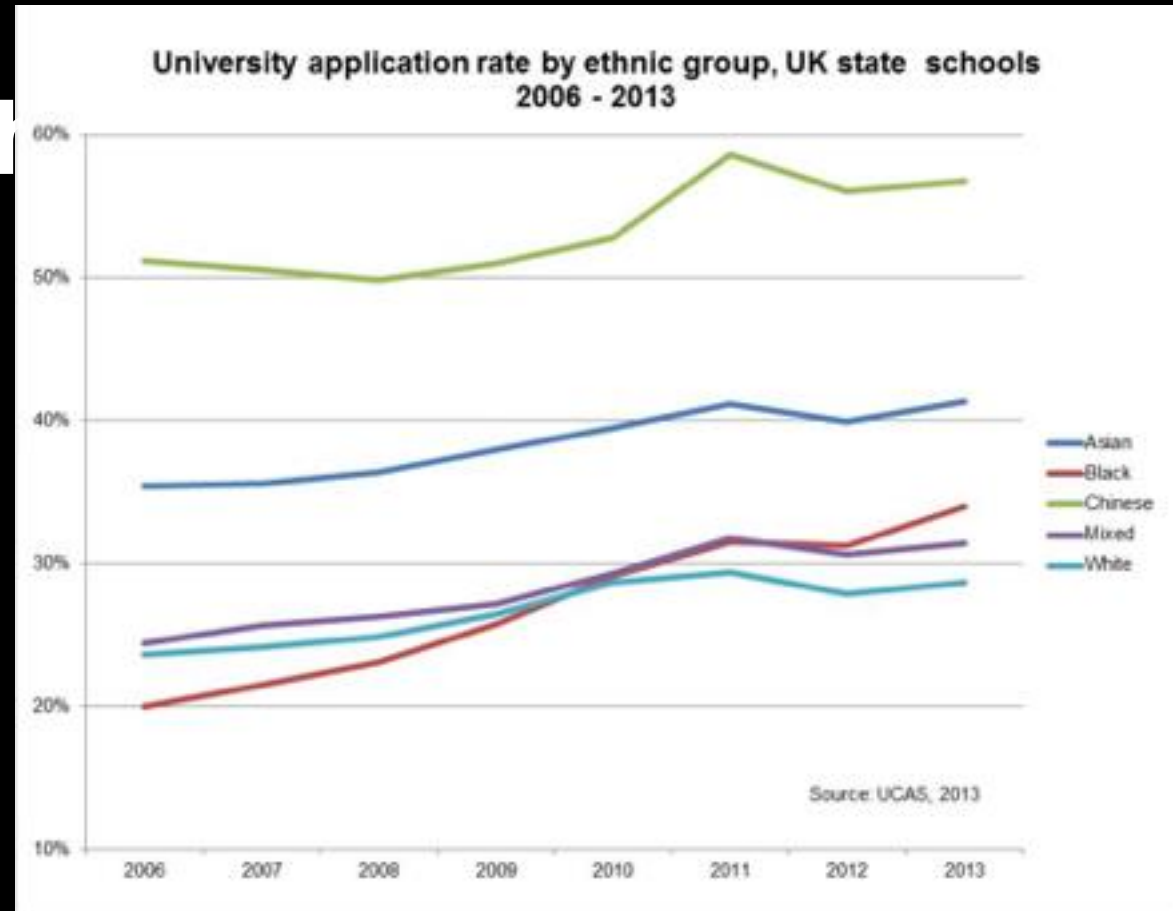
The presentation concludes with a set of lessons learned from previous experience of running trials.

Overview of today's session

An overview of consumer testing techniques
- Randomised control trials
- Quasi experimental designs
  - Matching
  - RDD
- Lab experiments with Predictiv

Let's start, with dubious distinction

University application rate by ethnic group, UK state schools 2006 - 2013

Asian
Black
Chinese
Mixed
White

Source: UCAS, 2013

What's the problem?

- Don't understand fees

- Don't understand ROI

- Don't understand scholarships

- Cultural barriers

- Just don't want to go

Parents don't earn that much? Your tuition fees can be halved.

You can also get over £3000 a year that you don't pay back.

You don't pay any tuition fees until you are earning more than £21,000 per year.
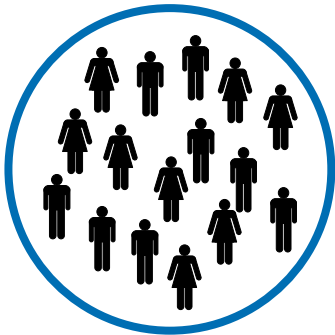
If you go to university, you're likely to earn £200,000 more over your lifetime.

going to university?

going to university?

going to university?
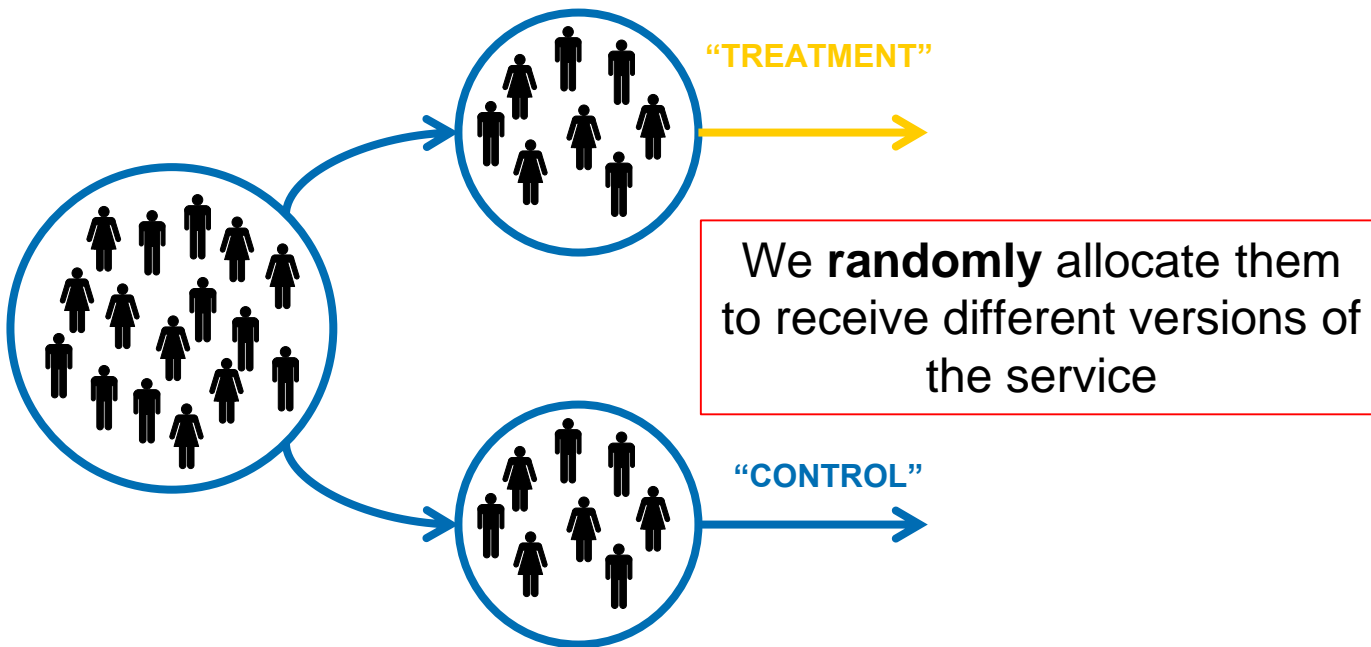
Are these going to work?

How can we tell if something really worked?

We test it with a <u>randomized control trial</u>!

We start with a group of people (or organisations) we want to try something new with

Random allocation

"TREATMENT"

"CONTROL"

We **randomly** allocate them to receive different versions of the service

Easy and robust comparison

TREATMENT

CONTROL

We measure what happens
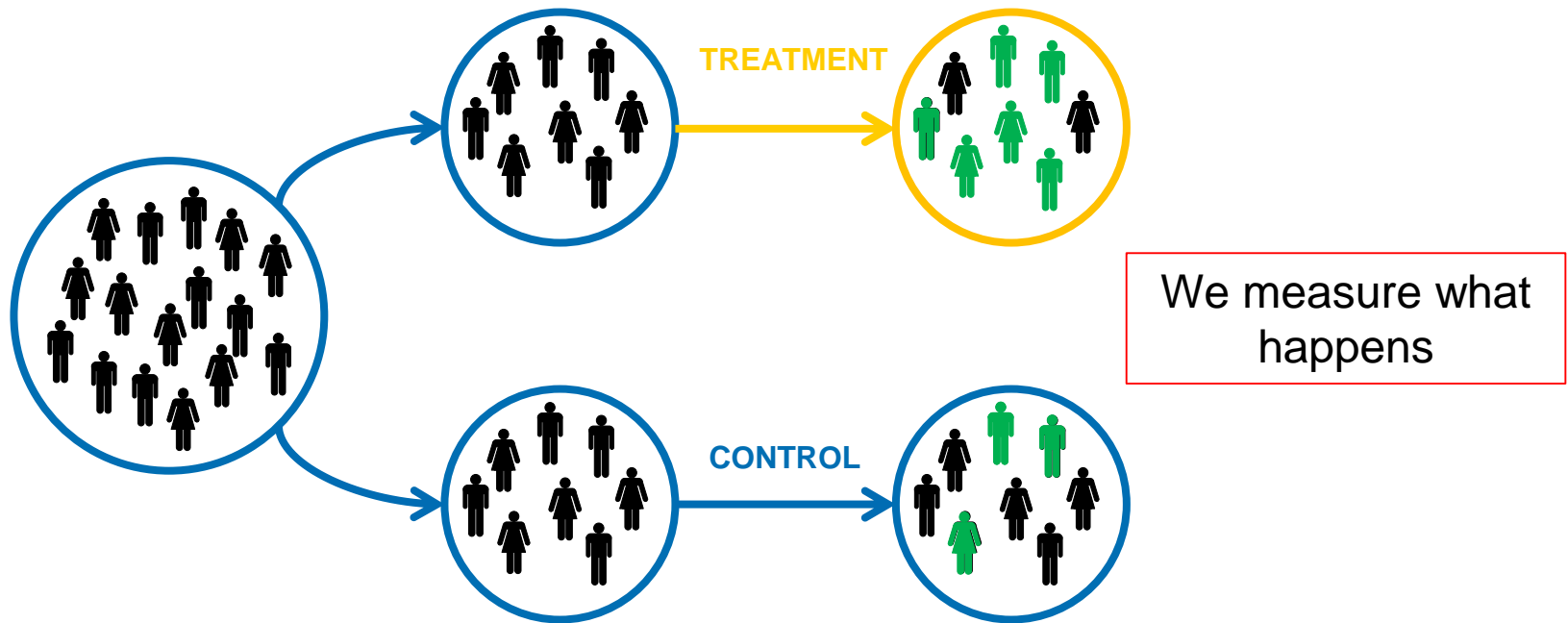
Can test multiple interventions simultaneously
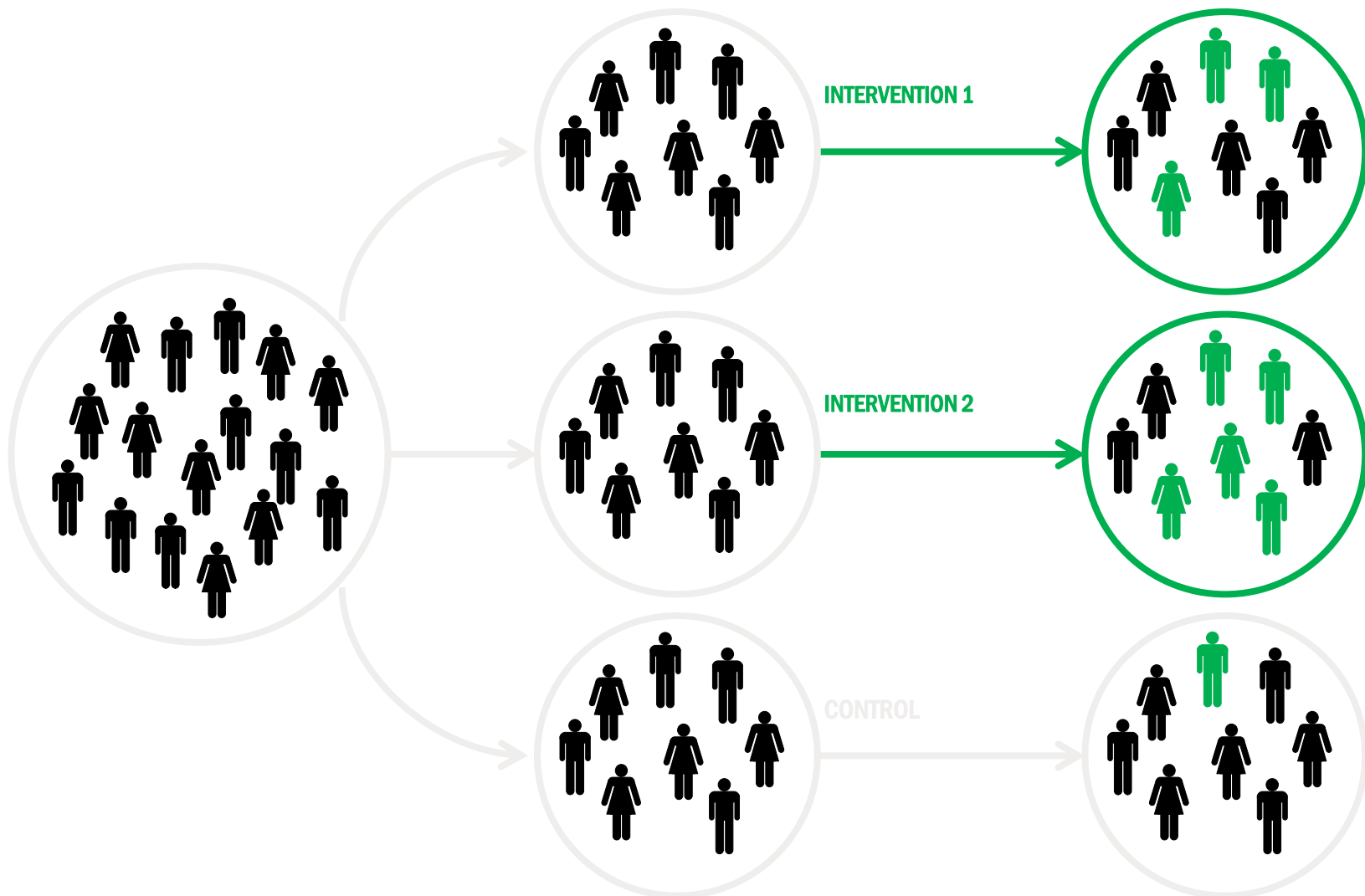
INTERVENTION 1

INTERVENTION 2

CONTROL

Parents don't earn that much? Your tuition fees can be halved.

You can also get over £3000 a year that you don't pay back.
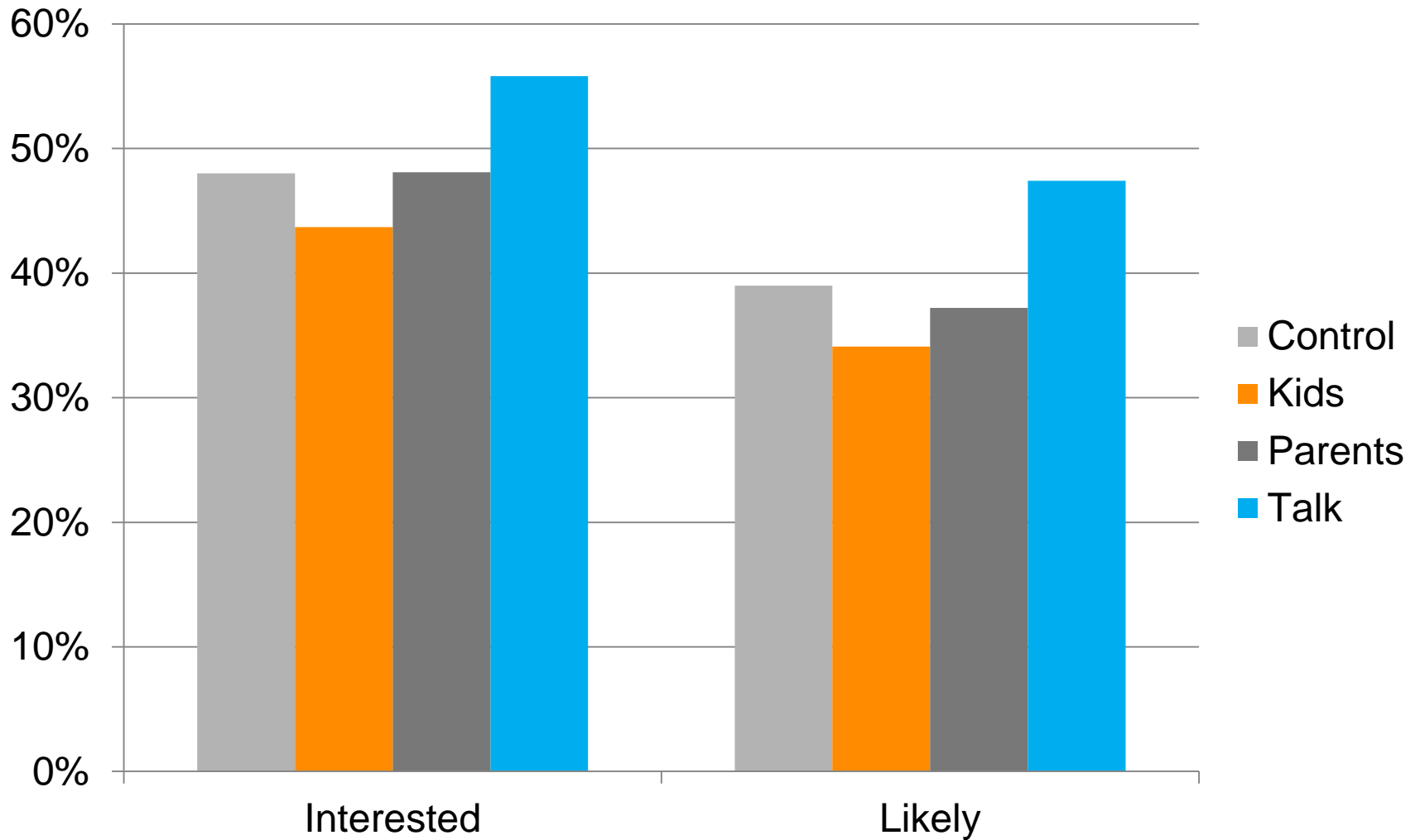
You don't pay any tuition fees until you are earning more than £21,000 per year.

If you go to university, you're likely to earn £200,000 more over your lifetime.

going to university?

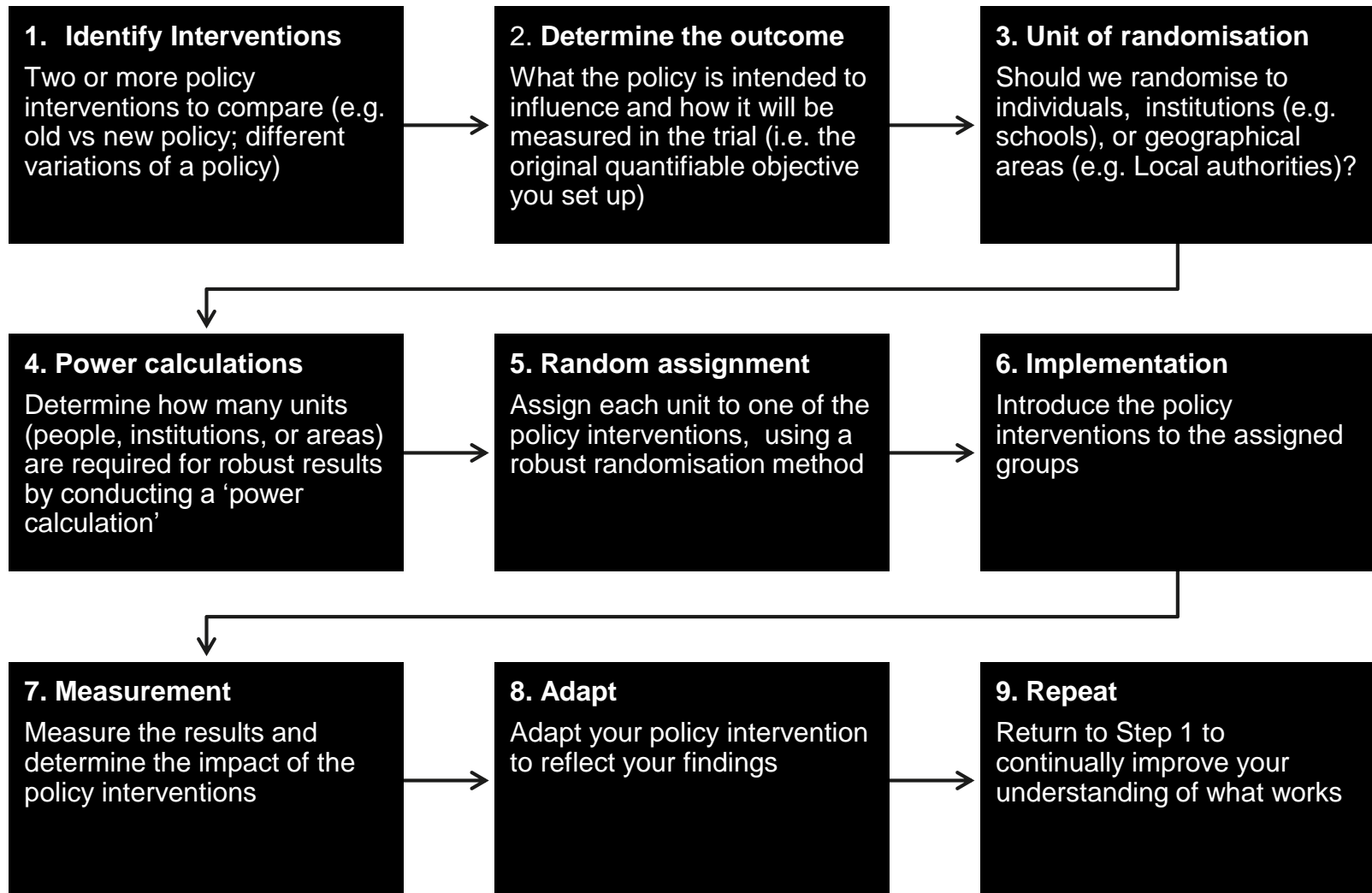*going to university?*

going to university?

Results

Why conduct (causal) research?

1. Because *not* testing is ill-advised

2. Because intuitions sometimes fail us (…even the experts)
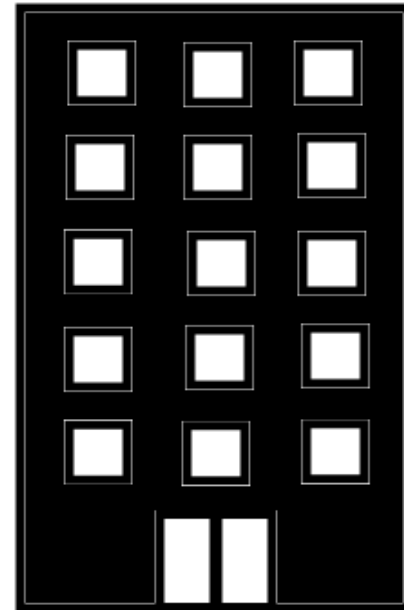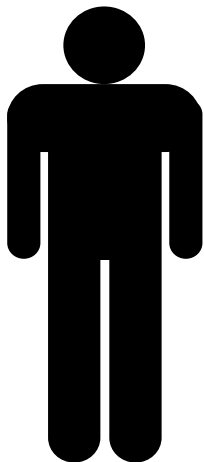
Stages of a robust trial

1. Identify Interventions
Two or more policy interventions to compare (e.g. old vs new policy; different variations of a policy)

2. Determine the outcome
What the policy is intended to influence and how it will be measured in the trial (i.e. the original quantifiable objective you set up)

3. Unit of randomisation
Should we randomise to individuals, institutions (e.g. schools), or geographical areas (e.g. Local authorities)?

4. Power calculations
Determine how many units (people, institutions, or areas) are required for robust results by conducting a 'power calculation'

5. Random assignment
Assign each unit to one of the policy interventions, using a robust randomisation method

6. Implementation
Introduce the policy interventions to the assigned groups

7. Measurement
Measure the results and determine the impact of the policy interventions

8. Adapt
Adapt your policy intervention to reflect your findings

9. Repeat
Return to Step 1 to continually improve your understanding of what works

When we randomly allocate people to receive different versions of a service, it...

A. Allows us to make causal statements about the impact of the service

B. Creates a control group that provides information on what would've happened in the absence of the program

C. Creates a treatment group and control group that are similar to each other on observable and unobservable characteristics

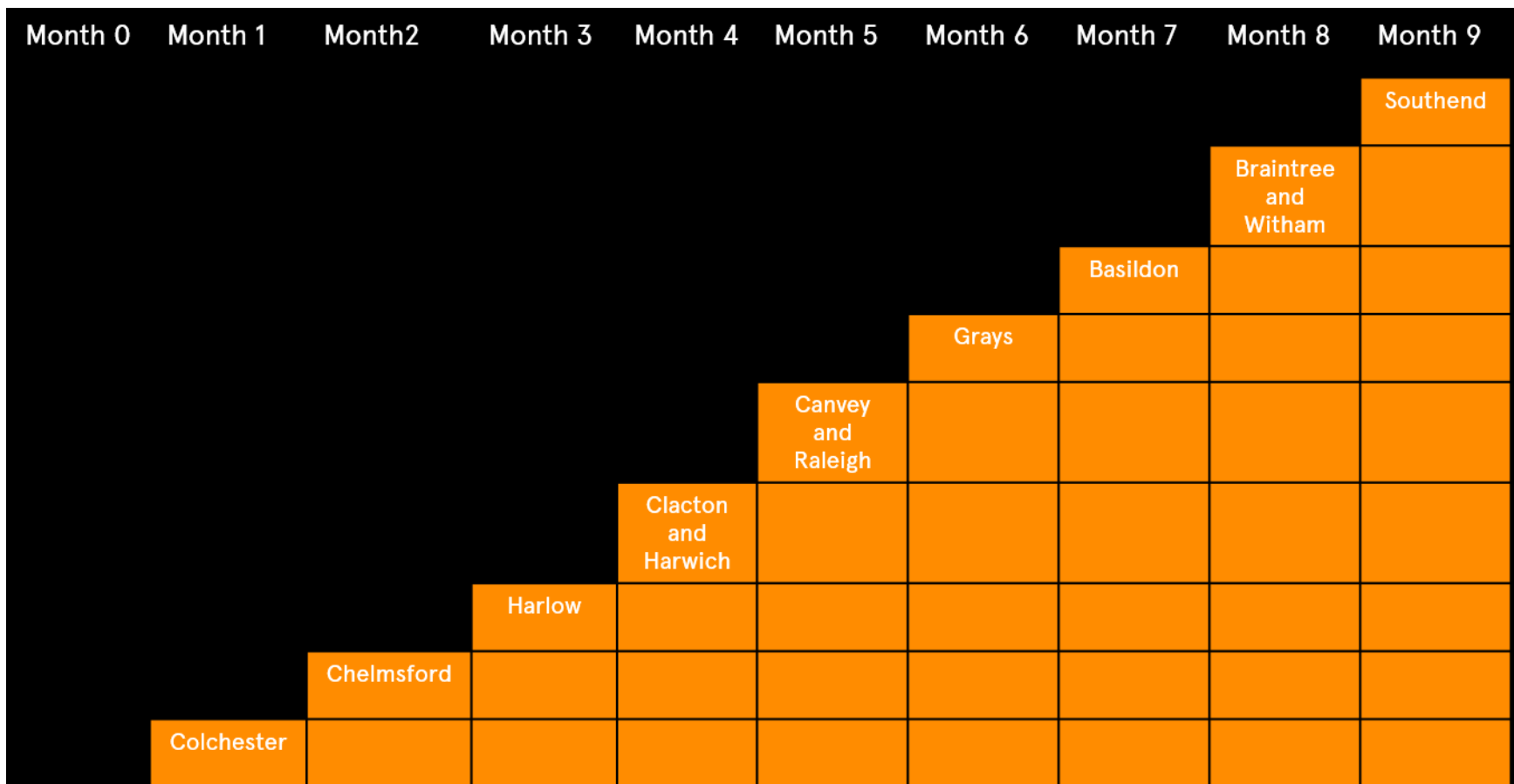**D.** A and C

E. All of the above

Units of randomisation

- Choosing a unit of randomization
    - Individual
    - Cluster (e.g., household or neighbourhood)
- When the benefit of the program spillovers to the control group.

Step wedge & phased roll outs

When you can't go live all at once!

Month 0	Month 1	Month2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9
									Southend
								Braintree and Witham	
							Basildon		
						Grays			
					Canvey and Raleigh				
				Clacton and Harwich					
			Harlow						
		Chelmsford							
Colchester									

Quasi experimental: propensity score matching

- Used in non-experimental settings.

- We estimate the likelihood of people taking up treatment, given their characteristics

- We match people with the same probabilities of taking up treatment (not necessarily the same characteristics

- Compare the matched groups

- As long as selection is only driven by things we observe, we get a estimate of the effect.

Propensity Score Matching
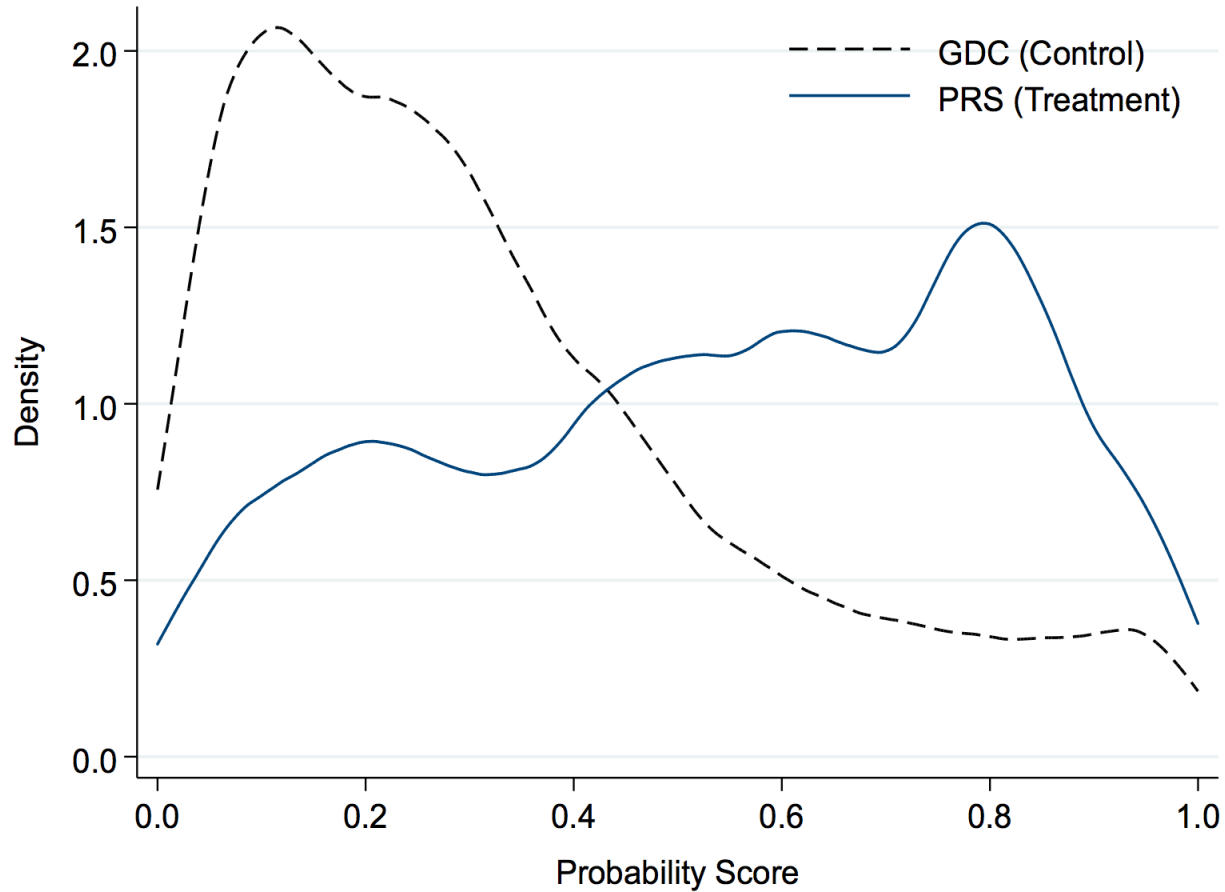
Quasi experimental: Regression Discontinuity Design (RDD)

- Many policies are applied with an arbitrary cut-off in eligibility

- Above some level of a "forcing variable" everyone is treated, below it nobody is treated.

- *Or*

- Above some level of a "forcing variable" more people are treated, below it fewer people are treated (fuzzy RDD)
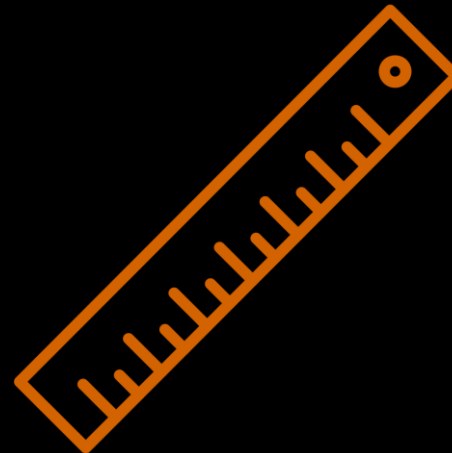
Likelihood of making a bequest

No prompt to donate	Prompt	Social + Emotional Prompt
5.9%	11.3%	17.3%
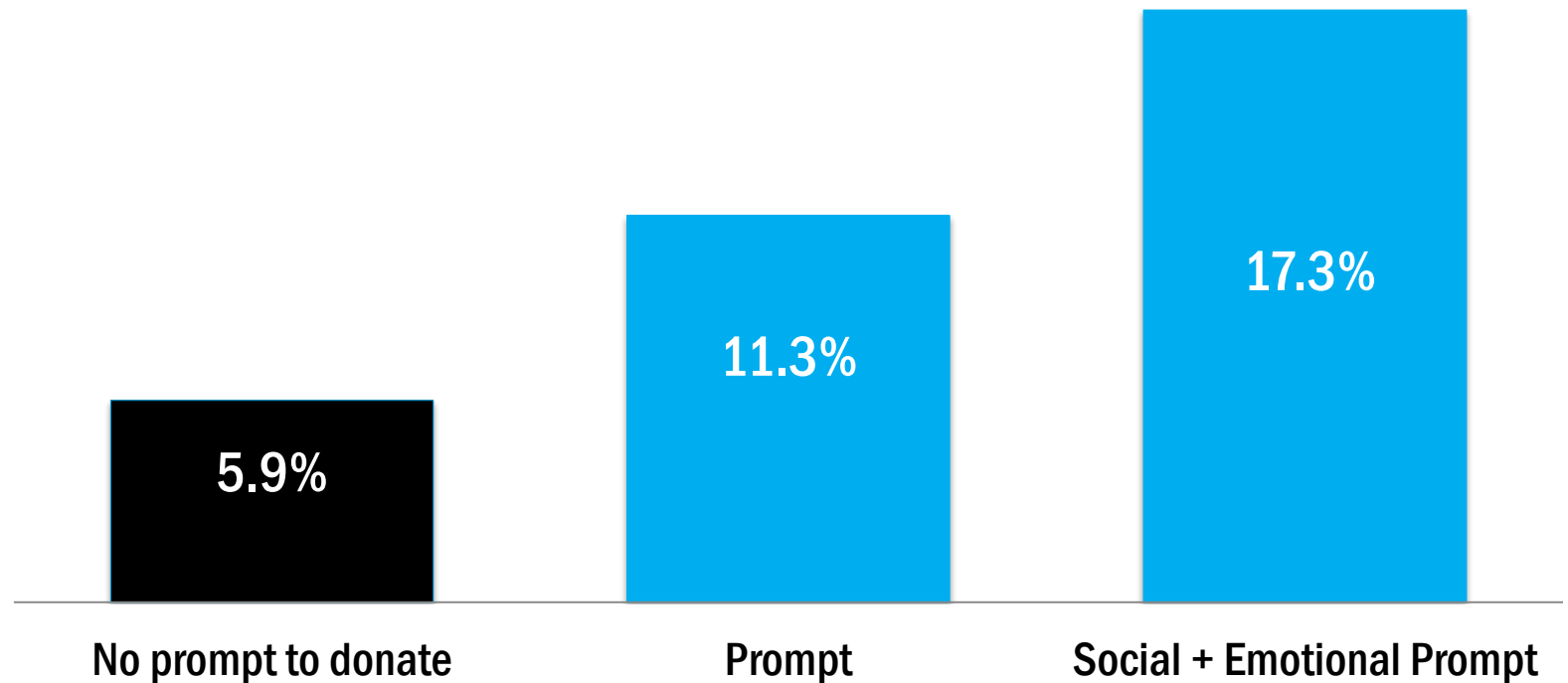
Quasi experimental: regression discontinuity design
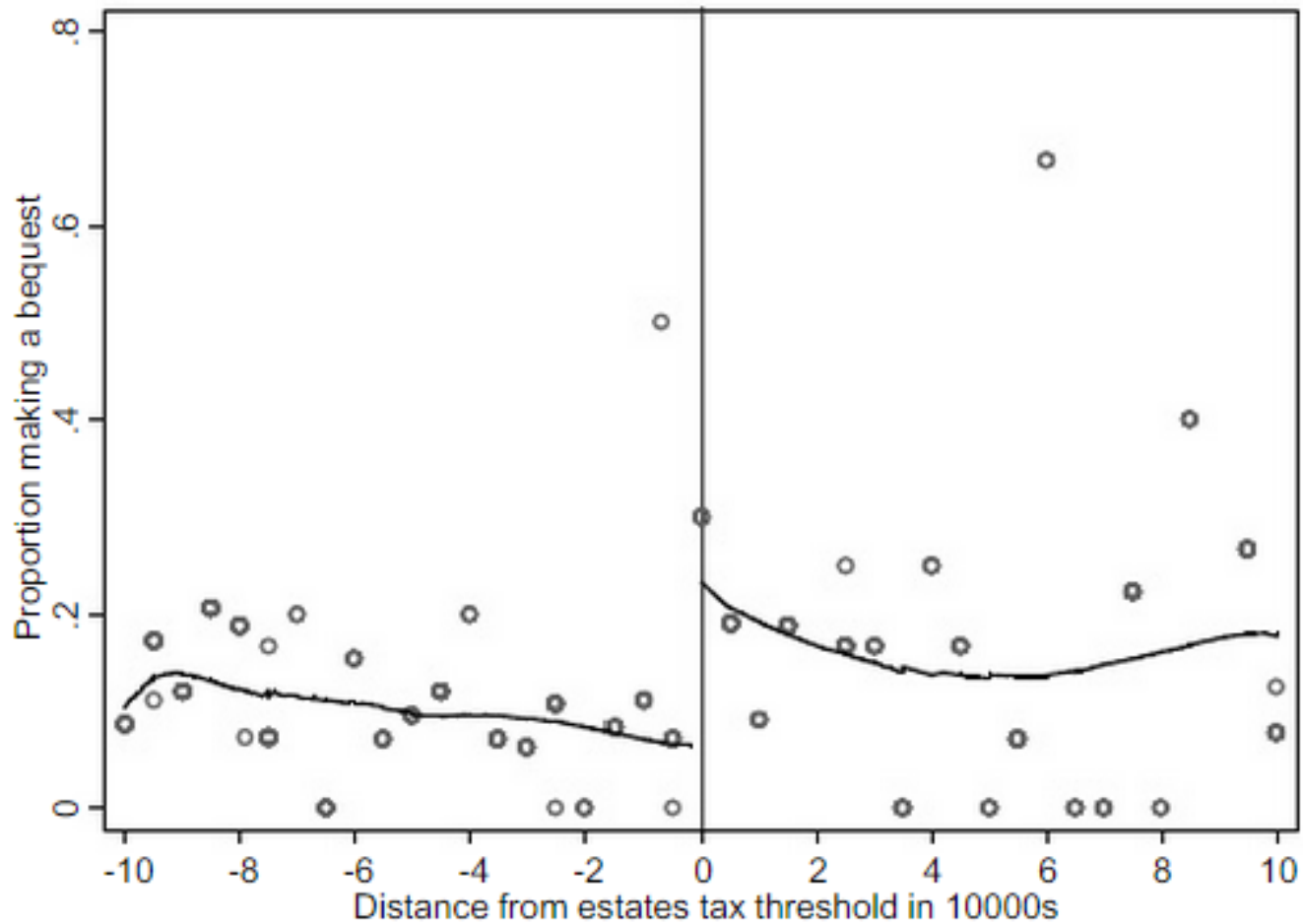
Quick introduction to laboratory experiments

5

Why use the lab?

Controlled variation

Control and systematically vary possible confounding factors to study a phenomenon rigorously and understand <u>why</u> something works

Allows researchers to study difficult interventions, such as testing the effect of sanctions or increases in wage on worker productivity.

Usually quicker and cheaper than a large scale field study

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, *312* (5770), 108-111

Lack of realism

"The casual reader may mistakenly interpret arguments about "realism" as an effective critique against the lab, potentially discouraging lab experimentation, and slowing down the production of knowledge in economics and other social sciences. The issue of "realism", however, is not a distinctive feature of lab vs. field data. The real issue is determining the **best way to isolate the causal effect of interest**."

Armin Falk & James Heckman, 2009. Lab experiments as a major source of knowledge in the social sciences. *Science*, 326

25

Introducing Predictiv

Predictiv is...

- A **website**, built on the foundations of an existing market research website
- With **experimental features & templates** that we are designing and validating
- Currently **back end only**, i.e. you can't yet build and run experiments without us manually creating it

- Able to access **4m UK participants** for very short tests from content partners
- **100k+ incentivised UK participants** from panels
- Can bring on other markets (European, US)
- Manages **incentivisation** – fixed and variable
- Filtering & screening

- **End-to-end**: right now, it manages test creation + recruitment + incentives.
- Next year it will also enable users to select a pre-designed experiment, run it and generate a basic report
- **In 2017: Predictiv becomes self-service & for use outside BIT projects**

Predictiv in action

Comprehension	**Recall**	**Consideration set**	**Social norms**	**Hassle costs**
Do people understand the message?	Do people remember information at a later date?	What information do people consider when making a decision and do they use this to make good decisions?	What do people believe others will do?	How complicated do people expect a process to be?
December 2016	March 2017	January 2017	January 2017	December 2016

Bespoke experiments

- Financial decision-making tests (both simple and complex) developed through MAS — Q1/Q2 2017
- Multiplayer: participants are connected together. Useful for testing trust, altruism, guilt/shame etc. — Q2 2017

Integration of lab and field experiments

THE
BEHAVIOURAL
INSIGHTS TEAM.

Lab

Support 'Explore' phase

Test different interventions when a field RCT is not an option

Integration

Validating material tested on Predictiv in field RCTs

Use Predictiv tests to directly inform intervention design

Andreoni, Callen, Jaffar, Khan & Sprenger (2016). Tailoring Intertemporal Incentives: An Application to Polio Vaccination Drives in Pakistan. *Work in progress.*

Field

Complement field RCT outcomes:
- Institutional trust
- Change in managerial attitudes
- Social norms

Lessons learned from running trials

- Good trial design takes time and expertise

- Data collection can be challenging

- If sample size is no problem, run as many arms as you can

- Value in sharing results to enable continuous improvement

THE
BEHAVIOURAL
INSIGHTS TEAM.

Thank you

www.behaviouralinsights.co.uk