Quantitative Economics

Applied Microeconomics I - RCT's

Ian Crawford
(Material provided by Climent Quintana-Domeque)

Department of Economics

Nuffield College

Trinity Term, 2018

Applied Micro - Randomised Controlled Trails

- This week we will focus on an Applied Micro topic - the use of Randomised Controlled Trials
- In theory RCT's are dead simple.
- But in Reality^{TM} there are a number of issues which researchers need to consider.
- The aims are to introduce you to some of those.

Applied Micro - Randomised Controlled Trails

- Two important problems which we will focus on are
  1. compliance (non-random assignment)
  2. heterogeneity in treatment effects

Compliance

"plots of ground do not respond to anticipated treatments of fertilizer, nor can they excuse themselves from being treated."

Heckman, J.J. (1992) "Randomization and social policy evaluation"

Compliance

- Suppose that we can only randomise the offer of/assignment to treatment but we cannot mandate that the subjects actually comply.
- Let
  - the offer of treatment be denoted by: $Z \in \{0, 1\}$
  - the fact of treatment be denoted by: $D \in \{0, 1\}$
  - the outcome be denoted: $Y \in \mathbb{R}$
- Non-compliance means

$$Z \neq D$$

Compliance

- The problem is that even if assignment to treatment (Z) is random, treatment itself (D) might not be.
- People may choose "not to take the medicine" on the basis of their assessment of the potential outcomes. (They may know more about these than the experimenter).
- As a result selection bias is re-introduced and we cannot learn about the causal effect of the treatment by comparison across the $D = 0$ versus $D = 1$ groups.

Compliance

- In the case studied by Angrist (AER, 1990) of the Vietnam war the draft lottery determined *eligbility* for drafting - but you might still not get drafted for medical reasons or you could dodge the draft (e.g. head to Canada, or get yourself arrested for smoking weed outside a police station etc).
- In the case of medical trials assignment to the treatment/control group might be random but you might not take the medicine/access the treatment when you weren't supposed to or you might access it anyway.
- This makes D a choice variable and, if it is chosen on the basis of potential outcomes (and therefore correlated with it), it causes selection bias.

Compliance

- The causal path is extended from a world of perfect compliance

$$Z \equiv D \rightarrow Y$$

- ... to a world of endogenous compliance (which is to say, correlated with potential outcomes)

$$Z \rightarrow D \rightarrow Y$$

- To cope with this we extend the potential outcomes framework to also include potential treatments.

Compliance

- Suppose that we randomise the offer of/assignment to treatment
- Z_i is an assignment to treatment indicator

$$Z_i = \begin{cases} 0 & \text{if the individual is not assigned to treatment} \\ 1 & \text{if the individual is assigned to treatment} \end{cases}$$

- Whether the individual gets treated or not is a different matter

Compliance

- The individual has two potential treatments $\{D_i(0), D_i(1)\}$ which demand on assignment.
- D_i is the treatment indicator

$$D_i = \begin{cases} D_i(0) & \text{is the treatment status if } Z_i = 0 \\ D_i(1) & \text{is the treatment status if } Z_i = 1 \end{cases}$$

- Potential treatment is connected to assignment to treatment in the normal way:

$$D_i = Z_i D_i(1) + (1 - Z_i) D_i(0)$$

Compliance

- Potential outcome is defined by D_i and Z_i:

$$Y_i(d, z)$$

  - $Y_i(1,1)$ if treatment offered ($Z_i = 1$) and received ($D_i = 1$)
  - $Y_i(0,1)$ if treatment offered ($Z_i = 1$) but not rec'ed ($D_i = 0$)
  - $Y_i(1,0)$ if treatment not offered ($Z_i = 0$) but rec'ed ($D_i = 1$)
  - $Y_i(0,0)$ if treatment not offered ($Z_i = 0$) not rec'ed ($D_i = 0$)

Compliance

- The observed treatment status D_i can be written as

$$\begin{aligned} D_i &= Z_i D_i(1) + (1 - Z_i) D_i(0) \\ &= D_i(0) + [D_i(1) - D_i(0)] Z_i \end{aligned}$$

- The causal effect of Z_i on D_i for individual i is

$$D_i(1) - D_i(0)$$

Compliance

- The observed outcome Y_i can be written as

$$\begin{aligned} Y_i &= Z_i Y_i \left(D_i(1), 1\right) + \left(1 - Z_i\right) Y_i(D_i(0), 0) \\ &= Y_i(D_i(0), 0) + \left[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)\right] Z_i \end{aligned}$$

- The causal effect of Z_i on Y_i for individual i is

$$Y_i(D_i(1), 1) - Y_i(D_i(0), 0)$$

Compliance

- If the treatment is randomly offered:

$$Z_i \perp\!\!\!\perp \{D_i(0), D_i(1), Y_i(1,1), Y_i(0,1), Y_i(1,0), Y_i(0,0)\}$$

- Two causal effects can be recovered:
  1. The causal effect of Z_i on D_i: the causal effect of being assigned to (offered) treatment on actual treatment status
  2. The causal effect of Z_i on Y_i: the causal effect of being assigned to (offered) treatment on the outcome of interest

Compliance

Causal effect of assignment to treatment on treatment status

- Since Z_i is randomly assigned, this can be obtained by comparing conditional means of treatment status by assignment to treatment:

$$\begin{aligned} &\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0] \\ &= \mathbb{E}[D_i(1)|Z_i = 1] - \mathbb{E}[D_i(0)|Z_i = 0] \\ &= \mathbb{E}[D_i(1) - D_i(0)] \end{aligned}$$

- This can also be obtained by regressing D_i on Z_i:

$$D_i = \alpha_0 + \beta_0 Z_i + e_i$$

where

$$\beta_0 = \frac{Cov(D_i, Z_i)}{Var(Z_i)}$$

Compliance

Causal effect of assignment to treatment on the outcome of interest

- Since Z_i is randomly assigned, this can be obtained by comparing conditional means of the outcome by assignment to treatment:

$$\begin{aligned} &\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0] \\ &= \mathbb{E}[Y_i(D_i(1), 1)|Z_i = 1] - \mathbb{E}[Y_i(D_i(0), 0)|Z_i = 0] \\ &= \mathbb{E}[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)] \end{aligned}$$

- This can also be obtained by regressing Y_i on Z_i:

$$Y_i = \alpha_1 + \beta_1 Z_i + u_i$$

where

$$\beta_1 = \frac{Cov(Y_i, Z_i)}{Var(Z_i)}$$

- This is what we call **ITT** (**intent-to-treat**) effect

Compliance

Example: Randomly offering the opportunity to participate in a free health insurance scheme

$$Z_i = \begin{cases} 0 & \text{if the individual is not offered such an opportunity} \\ 1 & \text{if the individual is offered such an opportunity} \end{cases}$$

Randomised controlled trials in the real world

- With imperfect compliance, we can learn about the **intent-to-treat effect** (**ITT**)
- Comparison of average outcomes across groups that are randomly assigned to treatment,without consideration as to whether the subjects actually take up the treatment or not
- Why? Because of RA of treatment
- **ITT** is the average causal effect of a program/policy that is introduced to a group of individuals, though subjects retain discretion as to whether or not they actually participate
- **ITT** is not the average causal effect of job training on employment
- Can we learn about (real) causal effects?

Randomised controlled trials in the real world

- In the imperfect compliance world the population gets partitioned into 4 groups:
  - Compliers: individuals who are treated (not treated) if they are (are not) offered the treatment
  - Always-takers: individuals who are treated no matter what
  - Never-takers: individuals who are not treated no matter what
  - Defiers: individuals who are **treated** (*not treated*) if they are **not offered** (*offered*) the treatment

Randomised controlled trials in the real world

In terms of potential outcomes:

- Compliers: $D_i(1) = 1,\ D_i(0) = 0$
- Always-takers: $D_i(1) = 1, D_i(0) = 1$
- Never-takers: $D_i(1) = 0, D_i(0) = 0$
- Defiers: $D_i(1) = 0,\ D_i(0) = 1$

Randomised controlled trials in the real world

- Only for the *Compliers* does the random assignment of the instrument translate into the random assignment of the treatment (they do exactly as they were supposed to).That is

$$Z_i = D_i \text{ [for Compliers]}$$

- That means that for *Compliers* the treatment is randomly assigned. Therefore there is no selection bias for this subpopulation.
- The problem is that you cannot easily identify *Compliers*.

Randomised controlled trials in the real world

		$D_i =$ 0	1
$Z_i =$	0	Never Taker/Complier	Always Taker
	1	Never Taker	Always Taker/Complier

- The group with $\{D_i = 1, Z_i = 1\}$ mixes *Compliers* with *Always-Takers*
- The group with $\{D_i = 0, Z_i = 0\}$ mixes *Compliers* with *Never-Takers*

Randomised controlled trials in the real world

- But there is a way to estimate the average causal effect for the complier sub-population: *the sub-population of subjects who would switch treatment status if the assigned treatment were changed*.
- This is known as the Local Average Treatment Effect (LATE).
- The sense in which it is "local" is that it only applies to this group

The Local Average Treatment Effect

$$LATE \equiv \mathbb{E}\left[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)\right]$$

The LATE Theorem

The LATE Theorem

Under some assumptions,

$$LATE = \frac{ITT}{Proportion\ of\ \ Compliers}$$

$$\mathbb{E}\left[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)\right] = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]}$$

The LATE Theorem - Assumptions

1. Independence assumption:

$$[\{Y_i(d,z); \forall d, z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$$

2. Relevance condition assumption: $Z_i \to D_i$

$$\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0] \neq 0$$

3. Exclusion restriction assumption: $Z_i \to D_i \to Y_i$

$$Y_i(d,0) = Y_i(d,1) = Y_i(d) \;\; \textit{for} \;\; d = 0,1$$

4. Monotonicity assumption:

$$D_i(1) - D_i(0) \geq 0 \;\; \text{or vice versa} \;\forall \;\; i$$

The LATE Theorem

Exclusion restriction?

- This says

$$Y_i(d, 0) = Y_i(d, 1) = Y_i(d) \quad \textit{for} \quad d = 0, 1$$

- This means that the potential outcome only directly depend upon/vary by treatment (D_i), not assignment to treatment (Z_i).
- This dimension-reduction reduces the number of potential outcomes from 4 to just 2 again.

The LATE Theorem

Monotonicity?

- Monotonicity rules out the presence of **defiers**: individuals who "do the opposite of what they are told".
- Recall:
  - Compliers: $D_i(1) = 1$, $D_i(0) = 0$ so $[D_i(1) - D_i(0) > 0]$
  - Always-takers: $D_i(1) = 1, D_i(0) = 1$ so $[D_i(1) - D_i(0) = 0]$
  - Never-takers: $D_i(1) = 0, D_i(0) = 0$ so $[D_i(1) - D_i(0) = 0]$
  - Defiers: $D_i(1) = 0$, $D_i(0) = 1$ so $[D_i(1) - D_i(0) < 0]$

The LATE Theorem

- Causal effects for:
- Compliers: $Y_i(1,1) - Y_i(0,0) = Y_i(1) - Y_i(0)$ [exclusion]
- Always-takers: $= 0$ [exclusion]
- Never-takers: $= 0$ [exclusion]
- Defiers: $\nexists$ [monotonicity]

The LATE Theorem

The Late Theorem:

Under assumptions 1 to 4

$$\frac{\mathbb{E}[Y_i|Z_i=1]-\mathbb{E}[Y_i|Z_i=0]}{\mathbb{E}[D_i|Z_i=1]-\mathbb{E}[D_i|Z_i=0]} = \mathbb{E}\left[Y_i(1)-Y_i(0)|D_i(1)>D_i(0)\right]$$

The LATE Theorem

The Late Theorem (Proof):

Numerator (first term)

$$\mathbb{E}[Y_i|Z_1 = 1] \underbrace{=}_{(exclusion)} \mathbb{E}[Y_i(0) + (Y_i(1) - Y_i(0))D_i(1)|Z_i = 1]$$

$$\underbrace{=}_{(independence)} \mathbb{E}[Y_i(0) + (Y_i(1) - Y_i(0))D_i(1)]$$

Numerator (second term) similarly:

$$\mathbb{E}[Y_i|Z_1 = 0] \underbrace{=}_{(exclusion)} \mathbb{E}[Y_i(0) + (Y_i(1) - Y_i(0))D_i(0)|Z_i = 0]$$

$$\underbrace{=}_{(independence)} \mathbb{E}[Y_i(0) + (Y_i(1) - Y_i(0))D_i(0)]$$

The LATE Theorem

The Late Theorem (Proof):

Numerator (difference):

$$\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0] =$$

$$\mathbb{E}[Y_i(0) + (Y_i(1) - Y_i(0))D_i(1)] - \mathbb{E}[Y_i(0) + (Y_i(1) - Y_i(0))D_i(0)]$$

$$\underbrace{=}_{\textit{linearity } \mathbb{E}} \mathbb{E}\left[(Y_i(1) - Y_i(0))\left(D_i(1) - D_i(0)\right)\right]$$

The LATE Theorem

The Late Theorem (Proof):

Monotonicity: $D_i(1) - D_i(0) = 1$ or $D_i(1) - D_i(0) = 0 \ \forall \ i$

$$\mathbb{E}\left[\left(Y_i(1) - Y_i(0)\right)\left(D_i(1) - D_i(0)\right)\right]$$

$$= \mathbb{E}\left[\left(Y_i(1) - Y_i(0)\right)|D_i(1) > D_i(0)\right] P\left[D_i(1) > D_i(0)\right]$$

The LATE Theorem

The Late Theorem (Proof):

Hence the numerator is

$$\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]$$

$$= \mathbb{E}\left[\left(Y_i(1) - Y_i(0)\right)|D_i(1) > D_i(0)\right] P\left[D_i(1) > D_i(0)\right]$$

The LATE Theorem

The Late Theorem (Proof):

Denominator:

$$\mathbb{E}[D_i|Z_i = 1] = \mathbb{E}[D_i(1)|Z_i = 1] \underbrace{=}_{(independence)} \mathbb{E}[D_i(1)]$$

$$\mathbb{E}[D_i|Z_i = 0] = \mathbb{E}[D_i(0)|Z_i = 0] \underbrace{=}_{(independence)} \mathbb{E}[D_i(0)]$$

Hence

$$\mathbb{E}\left[D_i|Z_i = 1\right] - \mathbb{E}\left[D_i|Z_i = 0\right] = \mathbb{E}\left[D_i(1) - D_i(0)\right]$$

$$\mathbb{E}\left[D_i|Z_i = 1\right] - \mathbb{E}\left[D_i|Z_i = 0\right] = P\left[D_i(1) > D_i(0)\right]$$

The LATE Theorem

The Late Theorem (Proof):

Finally

$$\frac{\mathbb{E}[Y_i|Z_i=1]-\mathbb{E}[Y_i|Z_i=0]}{\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right]}$$

$$=\frac{\mathbb{E}\left[\left(Y_i(1)-Y_i(0)\right)|D_i(1)>D_i(0)\right]P\left[D_i(1)>D_i(0)\right]}{P\left[D_i(1)>D_i(0)\right]}$$

$$=\mathbb{E}\left[\left(Y_i(1)-Y_i(0)\right)|D_i(1)>D_i(0)\right]$$

Numerical example: Siddique (2014)

- Random sample of 1,000 unemployed workers
- 500 are RA to take a job training (JT) program: $Z = 1$
- 500 are RA not to (control group): $Z = 0$
- Among those RA to take a JT program ($Z = 1$), 20% refuse to do so ($D = 0$):

$$P(D = 1|Z = 1) = 0.8 \quad \text{and} \quad P(D = 0|Z = 1) = 0.2$$

- Among those RA not to ($Z = 0$), 20% refuse to do so ($D = 1$)

$$P(D = 0|Z = 0) = 0.8 \quad \text{and} \quad P(D = 1|Z = 0) = 0.2$$

Numerical example: Siddique (2014)

- **What is the causal effect of treatment assignment on treatment delivery?**

$$\begin{aligned}\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0] &= \mathbb{E}[D(1)|Z=1] - \mathbb{E}[D(0)|Z=0] \\ &= \mathbb{E}[D(1)] - \mathbb{E}[D(0)] = \mathbb{E}[D(1) - D(0)]\end{aligned}$$

- Sample counterpart:

$$\frac{1}{n_1} \sum_{i \in Z_i = 1} D_i - \frac{1}{n_0} \sum_{i \in Z_i = 0} D_i = 0.8 - 0.2 = 0.6$$

- Individuals offered participation in the JT program are 60 pp (percentage points) more likely to participate in the JT program

Numerical example: Siddique (2014)

Average outcome: Employment rate	No Job Training ($D = 0$)	Job Training ($D = 1$)
Assigned to control ($Z = 0$)	$\frac{140}{400} = 0.35$	$\frac{40}{100} = 0.4$
Assigned to treatment ($Z = 1$)	$\frac{40}{100} = 0.4$	$\frac{180}{400} = 0.45$

Numerical example: Siddique (2014)

- **What is the causal effect of treatment assignment on the outcome of interest?**

$$\begin{aligned}\mathbb{E}[Y|Z=1]-\mathbb{E}[Y|Z=0] &= \mathbb{E}[Y(D(1),1)|Z=1]-\mathbb{E}[Y(D(0),0)|Z=0] \\ &= \mathbb{E}[Y(D(1),1)]-\mathbb{E}[Y(D(0),0)] = \mathbb{E}[Y(D(1),1)-Y(D(0),0)]\end{aligned}$$

- Sample counterpart:

$$\frac{1}{n_1}\sum_{i\in Z_i=1} Y_i - \frac{1}{n_0}\sum_{i\in Z_i=0} Y_i = \left[\frac{40+180}{100+400}\right] - \left[\frac{140+40}{400+100}\right] = \underbrace{0.44-0.36}_{0.08}$$

  - Individuals offered participation in the JT program are 8 pp more likely to find a job
  - This is the **ITT**
  - Average causal effect on employment rates of making unemployed workers eligible for the JT program with discretion as whether or not the worker chooses to take up the program.

Numerical example: Siddique (2014)

- **What is the causal effect of treatment on those whose treatment status depends on being offered the treatment or not?**

$$LATE = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]}$$

  - Sample counterpart?

$$\frac{0.08}{0.6} = 0.13$$

  - Individuals who participate in the JT program because they were offered the program and would not have participated otherwise are 13 pp more likely to find a job
  - This is the **LATE**
  - Average causal effect of the JT program on employment rates for the sub-population of unemployed workers who would switch to taking or not taking the program if assigned to do so by the randomisation mechanism.

Better LATE than nothing?

- The population of **treated** individuals is composed of two types: **always-takers** (always treated) and **compliers** (treated only when offered treatment)
- The ATT is a weighted average of two average causal effects:
  - Average causal effect of the treatment on always-takers (ATAT)
  - Average causal effect of the treatment on compliers (LATE)

$$ATT = \pi ATAT + (1 - \pi) LATE$$

- If there are no always-takers ($\pi = 0$), then (pretty obviously)

$$ATT = LATE$$

- This is known as the Bloom result.

Better LATE than nothing?

- **Is reasonable to think of the LATE as recovering the ATT?**
  - Participation in many (if not all) randomised trials is voluntary among those assigned to receive treatment
  - If no one in the control group has access to the treatment ($P(D_i = 1|Z_i = 0) = 0$), $LATE = ATT$.
  - If almost no one in the control group has access to the treatment ($P(D_i = 1|Z_i = 0) \approx 0$), $LATE \approx ATT$.
  - Hence, sometimes, we can learn about the average causal effect of a treatment among treated individuals

Better LATE than nothing?

- **Is reasonable to think of the LATE as recovering the ATE?**
- The ATE is the average causal effect of a treatment on a randomly chosen individual:

$$\mathbb{E}[Y(1)] - \mathbb{E}[Y_i(0)]$$

- It averages over compliance types
- Even with imperfect compliance, we can learn about the ATE but we can only get a range (bounds) not a number (point).

Numerical example: Siddique (2014)

- The problem is that
  - it is not known how many of the 100 workers who had been randomly assigned to the JTP, but who didn't comply, would have been employed had they taken the program.
  - similarly it is not known how many of the 100 workers who were randomly assigned to the control, but who didn't comply, would have been employed had they not taken the program;
- However we do know that there were 100 subjects did not follow their assignment to treatment.
- We also know that their counterfactual employment rates were between 0% and 100%
- We can use these polar/extreme assumptions on their employment rates to work our how big or small the ATE could be.

Numerical example: Siddique (2014)

Suppose they all would have found jobs if they were on the JT program:

Average outcome: Employment rate	*No Job Training* $(D = 0)$	*Job Training* $(D = 1)$
Control Group $(Z = 0)$	$\frac{140+0}{400+100} = 0.28$	
Treatment Group $(Z = 1)$		$\frac{180+100}{400+100} = 0.56$

$$ATE_{max} = 0.56 - 0.28 = 0.28$$

Numerical example: Siddique (2014)

Suppose that they would have founds jobs anyway:

Average outcome: ***Employment rate***	*No Job Training* $(D = 0)$	*Job Training* $(D = 1)$
Control Group $(Z = 0)$	$\frac{140+100}{400+100} = 0.48$	
Treatment Group $(Z = 1)$		$\frac{180+0}{400+100} = 0.36$

$$ATE_{min} = 0.36 - 0.48 = -0.12$$

Numerical example: Siddique (2014)

- We get the following range

$$ATE \in [-0.12, 0.28]$$

- The range is too wide to determine whether the treatment has a positive or negative causal effect
- Additional assumptions?
- What type of selection should we expect?

RA with perfect compliance

- **Randomisation** of treatment delivery D_i:

$$[Y_i(1), Y_i(0)] \perp\!\!\!\perp D_i$$

- What can we learn from such an experiment?

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \textbf{ATE}$$

RA with imperfect compliance

- **Randomisation** of treatment assignment Z_i:

$$[\{Y_i(d,z); \forall d, z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$$

- What can we learn from such an experiment?

$$\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0] = \textbf{ITT}$$

- Moreover, if we assume **relevance**, **exclusion restriction** and **monotonicity**, then

$$\frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \textbf{LATE}$$

- Finally, if there are no **always-takers**, then

$$\frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \textbf{ATT}$$

Regression format

- **First Stage**: The average causal effect of treatment assignment on treatment delivery:

$$\texttt{regression of } \mathrm{D} \text{ on } \mathrm{Z}$$

- **Reduced form**: The average causal effect of treatment assignment on the outcome of interest (**ITT**):

$$\texttt{regression of } \mathrm{Y} \text{ on } \mathrm{Z}$$

- **2SLS**: The average causal effect of treatment delivery on the outcome of interest among compliers (**LATE**):

$$\texttt{regression of } \mathrm{Y} \text{ on } \widehat{\mathrm{D}}$$

where $\widehat{\mathrm{D}}$ is the predicted D from a

$$\texttt{regression of } \mathrm{D} \text{ on } \mathrm{Z}$$

That is, $\widehat{\mathrm{D}}$ is the variation of D due to Z.

Health Insurance and Health Status

Does Health Insurance improve Health?

- National Health Interview Survey: an annual survey of the U.S. population with detailed information on health and health insurance (HI)
- Measuring health status: "Would you say your health in general is excellent, very good, good, fair, or poor?"
- $Y_i = 1$ if poor health, ..., $Y_i = 5$ if excellent health
- $D_i = 1$ if (private) HI, $D_i = 0$ no HI
- Naive comparison

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0]$$

Health Insurance and Health Status

Table 1: Health and demographic characteristics of insured and uninsured couples

	Husbands			Wives		
	HI	No HI	Difference	HI	No HI	Difference
Health index (1-5)	4.01 [0.93]	3.70 [1.01]	0.31 (0.03)	4.02 [0.92]	3.62 [1.01]	0.40 (0.04)
Education (years)						
Family income (USD)						

Source: Table 1.1. in Angrist and Pischke (2015).

Note: Standard deviations are reported in [brackets].

Note: Standard errors are reported in (parentheses).

Health Insurance and Health Status

- Individuals with HI tend to report to be healthier

$$\widehat{\mathbb{E}}[Y_i^{husband}|D_i = 1] - \widehat{\mathbb{E}}[Y_i^{husband}|D_i = 0] = 4.01 - 3.70 = 0.31$$

$$\widehat{\mathbb{E}}[Y_i^{wife}|D_i = 1] - \widehat{\mathbb{E}}[Y_i^{wife}|D_i = 0] = 4.02 - 3.62 = 0.40$$

- The $\widehat{\mathbb{E}}$ denotes sample average.

Health Insurance and Health Status

Table 1: Health and demographic characteristics of insured and uninsured couples

	Husbands			Wives		
	HI	No HI	Difference	HI	No HI	Difference
Health index (1-5)	4.01	3.70	0.31	4.02	3.62	0.40
	[0.93]	[1.01]	(0.03)	[0.92]	[1.01]	(0.04)
Education (years)	14.31	11.56	2.75	14.44	11.80	2.64
			(0.10)			(0.11)
Family income (USD)	106,467	45,656	60,811	106,212	46,385	59,827
			(1,355)			(1,406)

Source: Table 1.1. in Angrist and Pischke (2015).

Note: Standard deviations are reported in [brackets].

Note: Standard errors are reported in (parentheses).

Health Insurance and Health Status

- In the US, individuals with HI: richer and more educated
- Control for education and family income using regression

$$Y_i = \alpha + \beta D_i + \gamma_1 education_i + \gamma_2 income_i + \epsilon_i$$

- Is this regression control for enough? Can we ever find a regression which controls for enough using observational data?
  - Many risk factors are not observed/observable, so it cannot be included in the regression
  - Controlling for family income and education is unlikely to overcome selection bias
  - Randomly assigning health insurance coverage?

Randomising Health Insurance

- RAND Health Insurance Experiment
- OREGON Health Insurance Experiment

Randomising Health Insurance

RAND Health Insurance Experiment

- 1974-1981
- It provided Health Insurance to more than 5800 individuals (2000 households)
- Random assignment of different HI plans:
  - from full coverage ("free care"): zero price
  - ... to almost no coverage: full price
- All analyses use some sort of aggregation of similar plans
- Experimental plans enrollment: 3-5 years
- Cost of this social experiment $\sim$ \$ 300 million

Randomising Health Insurance

RAND Health Insurance Experiment

1. What is the effect of HI on health status?
2. What is the effect of HI on the use of medical care?

Randomising Health Insurance

RAND Health Insurance Experiment

- 6 HI plans: 1 free HI care plan + other 5 HI plans
- Random allocation to one of the 6 HI plans

Randomising Health Insurance

RAND Health Insurance Experiment

- **What is the effect of HI on the use of medical care?**
- Proxy outcome measure of medical care use: spending (Y)
- Regression Y on HI plan variables $P1, ..., P5$

$$Y_i = \alpha + \beta_1 P_{1i} + \beta_2 P_{2i} + \beta_3 P_{3i} + \beta_4 P_{4i} + \beta_5 P_{5i} + \epsilon_i$$

where

  - $P_{1i} = 1$ if HI plan 1, and 0 if other HI plan; ...; $P_{5i} = 1$ if HI plan 5, and 0 if other HI plan.

- Why only 5 HI plan dummy variables plus a constant if there are 6 HI plans? Answer: Frisch-Waugh-Lovell

Randomising Health Insurance

- If a categorical variable contains C categories, we may include $C-1$ categories and a constant term
- If a categorical variable contains C categories, we may include C categories but no constant term
- Otherwise, "dummy variable trap" (Stock and Watson, 2011)
- By convention, the constant term is retained so the R^2 makes sense, in which case one of the categories is excluded:

$$Y_i = \alpha + \beta_1 P_{1i} + \beta_2 P_{2i} + \beta_3 P_{3i} + \beta_4 P_{4i} + \beta_5 P_{5i} + \epsilon_i$$

- What is the interpretation of α?

$$\mathbb{E}[Y_i | P_{6i} = 1] = \alpha$$

- What is the interpretation of β_1?

$$\mathbb{E}[Y_i | P_{1i} = 1] - \mathbb{E}[Y_i | P_{6i} = 1] = (\alpha + \beta_1) - \alpha = \beta_1$$

Randomising Health Insurance

Effects on Utilisation

Table 2

Plans' Effects on Utilization

	Total spending[a]		*Inpatient spending*		*Outpatient spending*	
	Share with any (1)	*Spending in $* (2)	*Share with any* (3)	*Spending in $* (4)	*Share with any* (5)	*Spending in $* (6)
Constant (Free Care Plan, $N = 6{,}840$)	0.931 (0.006)	2,170 (78)	0.103 (0.004)	827 (60)	0.930 (0.006)	1,343 (35)
25% Coinsurance ($N = 2{,}361$)	−0.079 (0.015)	−648 (152)	−0.022 (0.009)	−229 (116)	−0.078 (0.015)	−420 (62)
Mixed Coinsurance ($N = 1{,}702$)	−0.053 (0.015)	−377 (178)	−0.018 (0.009)	21 (141)	−0.053 (0.016)	−398 (70)
50% Coinsurance ($N = 1{,}401$)	−0.100 (0.019)	−535 (283)	−0.031 (0.009)	4 (265)	−0.100 (0.019)	−539 (77)
Individual Deductible ($N = 4{,}175$)	−0.124 (0.012)	−473 (121)	−0.006 (0.007)	−67 (98)	−0.125 (0.012)	−406 (52)
95% Coinsurance ($N = 3{,}724$)	−0.170 (0.015)	−845 (119)	−0.024 (0.007)	−217 (91)	−0.171 (0.016)	−629 (50)
p-value: all differences from Free Care = 0	< 0.0001	< 0.0001	0.0008	0.1540	< 0.0001	< 0.0001

Figure: Aron-Dine, Einav, and Finkelstein (2013)

Randomising Health Insurance

Effects on Utilisation: *F* test is back

- Last row of the table contains a $p - value$ for the test all differences from Free Care $= 0$
- This is the $p - value$ of the $F - statistic$ for the following hypothesis test:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_1 : \beta_j \neq 0 \text{ for at least one } j = 1, ..., 5$$

Randomising Health Insurance

Effects on Utilisation: *F* test is back

- The $F - statistic$ is given by

$$F = \frac{\frac{SSR_R - SSR_U}{r}}{\frac{SSR_U}{n-K-1}} \sim F_{r,\infty}$$

  for a large sample, where SSR_R is the sum of squared residuals from the restricted regression, and SSR_U is the sum of squared residuals from the unrestricted regression, r is the number of restrictions under the H_0, and k is the number of slopes in the unrestricted regression.

- The $p - value$ of the $F - statistic$ can be computed using the large-sample approximation to its distribution as

$$p - value = Pr[F_{r,\infty} > F^{act}]$$

Randomising Health Insurance

Findings from the RAND Health Insurance Experiment

- Subjects assigned to more generous insurance plans used substantially more health care
- Hospital inpatient admissions were less sensitive to price than was outpatient care: admissions decisions are usually made by doctors
- This extra care and expense did not seem to make them much healthier.

Randomising Health Insurance

Internal validity checks

- Non-random assignment to plans?
- Differential participation in the experiment across treatment arms? People offered better plans are more likely to participate...
- Differential reporting (of medical care utilization) across treatment arms? People with better plans are more likely to file claims to get reimbursed...

The incentive for filing claims was to get reimbursed, an so the filing incentive was weaker for participants enrolled in higher coinsurance rate plans (or their providers) than for those enrolled in lower coinsurance rate plans or the free care plan [those with more coverage more incentives to report].

Randomising Health Insurance

External validity?

- Participants of the RAND HIE were middle class
- Financial liability for health-care costs was limited under every treatment
- Effects of expanding Medicaid to cover the currently uninsured?
- Uninsured Americans are (on average) younger, less educated, poorer and less likely to be working than RAND HIE participants. The value of extra health in such a group might be different than that for middle class families that participated in the RAND HIE

Randomising Health Insurance

Back to the present

- **Medicaid** in the US: social health care program for families and individuals with limited resources
- The **Affordable Care Act** (aka *Obamacare*) expanded both eligibility for and federal funding of Medicaid

Randomising Health Insurance

Oregon Health Trial

The state of Oregon recently offered Medicaid to thousands of randomly chosen people in a publicly announced health insurance lottery

- Oregon Health Insurance Experiment
- $\sim$ 90,000 individuals participated in a lottery to apply for OHP (Oregon Health Plan) Standard
- OHP Standard: a Medicaid program for low-income, uninsured, able-bodied adults who are not eligible for other public insurance in Oregon(Medicare: for persons 65+ and for disabled; Children's Health Insurance Program: poor children: Medicaid for poor children, pregnant women, or other specific, categorically eligible population).8 random lottery drawings between March and September 2008% persons who were selected won the opportunity (for themselves and any household member) to apply for OHP Standard.

Randomising Health Insurance
Oregon Health Trial

- $\sim$ 35,000 individuals won the lottery ($\sim$ 30,000 households) and 30% of them successfully enrolled
- Eligibility requirements: 19-64, Oregon residents who were U.S. citizens or legal immigrants, ineligible for other public insurance, uninsured for the previous 6 months, income $<$ federal poverty line, and assets $<$ \$ 2,000. Persons who were randomly selected in the lottery were sent an application. Those who completed it and met the eligibility criteria were enrolled in the plan: poverty status and submit paper power within 45 days.
- OHP Standard provides: comprehensive medical benefits and low monthly premiums (\$0 to \$20, based on income)

Randomising Health Insurance
Oregon Health Trial

1. Does expanding Medicaid will increase health care utilization?
   - Example: Admissions?
2. Does expanding Medicaid will improve health?
   - Example: Mental health?
3. Does expanding Medicaid will insure people against risk of catastrophic expenditures?
   - Example: Medical bills?

Randomising Health Insurance
Oregon Health Trial

- **ITT**: effect of winning the lottery on the outcome Y
- It can be estimated by regressing Y on $LOTTERY$

$$Y = \alpha_0 + \alpha_1 LOTTERY + e$$

$$LOTTERY = \begin{cases} 1 & \text{if winning the lottery to apply for OHP} \\ 0 & \text{if not winning the lottery to apply for OHP} \end{cases}$$

- α_1 is the average causal effect of winning the lottery:
  the average effect of being able to apply for OHP Standard

Randomising Health Insurance
Oregon Health Trial

- **LATE**: effect of health coverage on the outcome Y among those covered because of winning the lottery
- It can be estimated by dividing the **ITT** by the difference in compliance rates under four assumptions: **independence**, **relevance**, **exclusion restriction** and **monotonicity**.
- Difference in compliance rates: difference in health insurance status between lottery winners and lottery losers
- Difference in compliance rates: β_1
- β_1 is obtained by regressing *INSURANCE* on *LOTTERY*

$$INSURANCE = \beta_0 + \beta_1 LOTTERY + u$$

- $\frac{\alpha_1}{\beta_1}$ is the average causal effect of health insurance among those insured because of winning the lottery.

Randomising Health Insurance
Oregon Health Trial

1. **Independence**? Are lottery winners and lottery losers similar?
2. **Relevance**? Are lottery winners more likely to get insured?
3. **Exclusion restriction**?
   - The event winning (or losing) the lottery may have direct effects on the outcomes under study
   - Individuals who apply for public health insurance may also be encouraged to apply for other public programs for which they are eligible
4. **Monotonicity**? Likely to be satisfied (no defiers): "winning the lottery pushes individuals in one direction"

Randomising Health Insurance
Oregon Health Trial

- Do we care about ITT or LATE?
  - ITT captures the effect of expanded access to Medicaid
  - LATE captures the impact of health insurance among those who became insured because of the expansion of Medicaid

Randomising Health Insurance

Oregon Health Trial

Table 2: Hospital Utilization (Admissions)

	Control Mean	ITT	LATE
Admissions through ER	0.048	0.0018	0.0070
		(0.0016)	(0.0062)
Admissions not through ER	0.029	0.0041	0.016
		(0.0013)	(0.0051)

Source: Table 4 in Finkelstein et al. (2012).

Note: Standard errors are reported in (parentheses).

Randomising Health Insurance
Oregon Health Trial

Table 3: Financial Strain (Medical Debt)

	Control Mean	ITT	LATE
Any medical collection	0.281	−0.016	−0.064
		(0.0040)	(0.016)
Amount owed in medical collections	1,999	−99	−390
		(45)	(177)

Source: Table 7 in Finkelstein et al. (2012).

Note: Standard errors are reported in (parentheses).

Randomising Health Insurance
Oregon Health Trial

- Higher utilization
- Lottery winners were insulated from the financial consequences of poor health: insurance against risk
- Health status?
  - No improvements on physical health
  - Improvements on mental health (less financial strain?)

Randomising Health Insurance
Oregon Health Trial

- HI appears to worsen health in the observational estimates
- Individuals who select HI coverage are in poorer health (and therefore demand more medical care) than those who are uninsured: {Adverse Selection}

Randomising Health Insurance
Oregon Health Trial

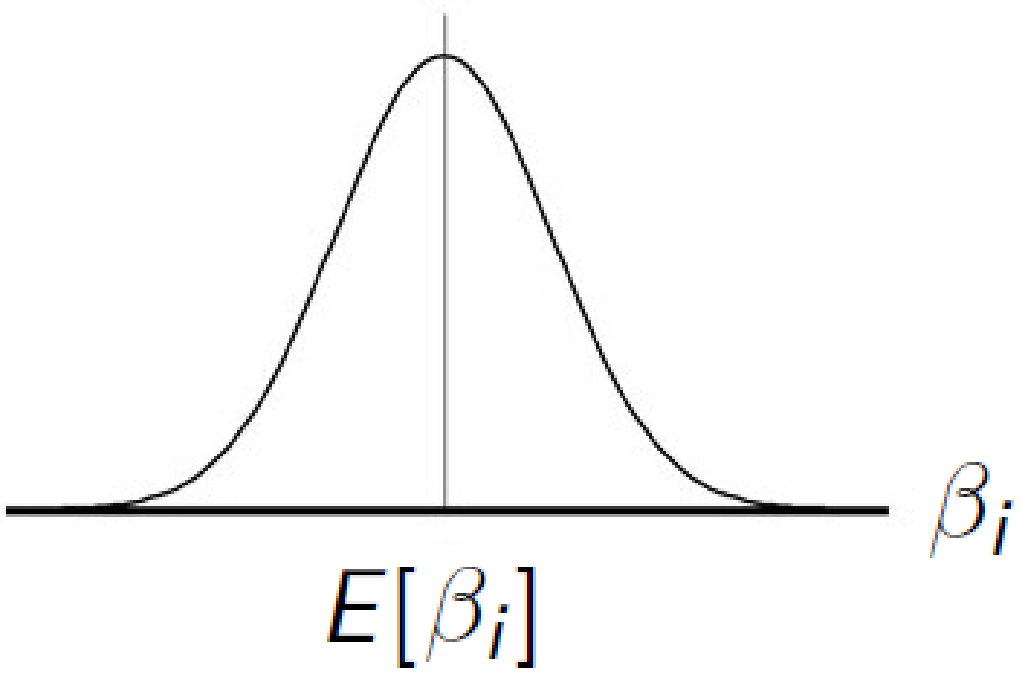
- External validity?
  - The low-income uninsured population in Oregon differs from the overall population in the US
  - Very short-term effect
  - Partial equilibrium effects the effects of covering a small number of people, holding constant the rest of the health care system--how about supply-side responses by the health care sector?
  - Medicaid coverage may have different effects for persons who seek insurance through the lottery than that for the general population affected by coverage mandates.

Applied Micro - Randomised Controlled Trails

- Two important problems are
  1. compliance (non-random assignment)
  2. heterogeneity in treatment effects

Heterogeneous Treatment Effects

- A treatment can have different effects for different individuals
- $\beta_i \equiv Y_i(1) - Y_i(0)$
- For instance, β_i could be normally distributed:



- $\mathbb{E}[\beta_i]$ is the **Average Treatment Effect** (**ATE**)

Heterogeneous Treatment Effects

- Registration of RCTs
- Heterogeneous causal effects
- The Queen of Katwe and test scores (Emma Riley)
  - Did the randomisation work as intended?
  - Average causal effects
  - Heterogeneous causal effects

Registration of Randomised Controlled Trials

- Data mining can be a problem even in randomised studies.
- Most studies look at more than one outcome measure and often several subgroups.
- If conventional significance levels are used for the individual tests, there is a high chance of a Type I error; (incorrectly rejecting a null hypothesis of no effect).
- In other words, studies of this kind are likely to find statistically significant effects on at least one of the outcomes, or for one of the subgroups, even in the absence of such effects, simply as a result of sampling variation.

Registration of Randomised Controlled Trials

- Assume that we are testing m null hypotheses for m independent variables, representing either different subgroups or different outcome variables.
- Let α be the probability of a Type I error for each of these tests.
- If the null hypotheses are all true, then the overall probability of at least one Type I error is $1–(1–\alpha)^m$.
- So, if $\alpha = 0.05$, then this is 0.10 if $m = 2$, and 0.40 if $m = 10$.

Registration of Randomised Controlled Trials

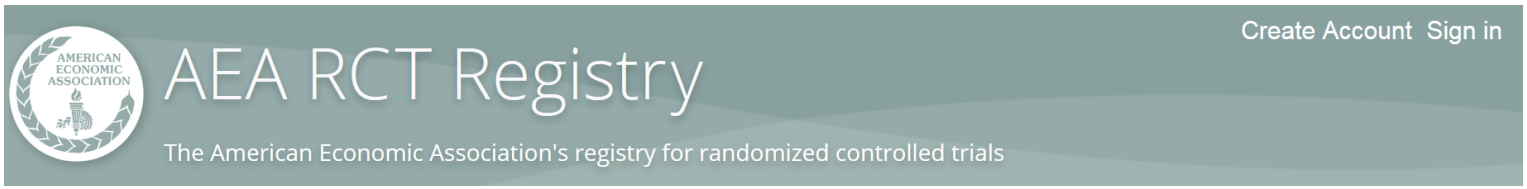
- At the same time, studies are prone to various forms of more or less (un)conscious data mining.
- Researchers typically report only on the subgroups and outcomes where there are significant results or omit insignificant findings.
- This could be out of habit or editorial/publication bias or generally the social/institutional structure which surrounds research.

Registration of Randomised Controlled Trials

- The “Replication Crisis” in medicine and psychology (esp Social Psychology which has been at the centre of much of it) holds lessons for Economics.
- See for example Ioannidis, J., Stanley, T.and H. Doucouliagos: (2017-10-01). "The Power of Bias in Economics Research". *Economic Journal*. 127 (605): F236–F265

Registration of Randomised Controlled Trials

- The American Economic Association set up a registry for randomised controlled trials
- `https://www.socialscienceregistry.org/`



- AEA RCT Registry currently lists 1230 studies with locations in 106 countries (May 22, 2017)

Registration of Randomised Controlled Trials

Pre-Analysis Plan (PAP): Mitigates concerns of data mining and specification searching. It should include: by setting out in advance exactly the specifications that will be run and with which variables. It should include:

- The outcomes you'll look at
- How you'll construct variables
- The specification you'll use
- The regressions you'll run
- The hypotheses you'll test

All this must be done **before** you do any analysis on the data.

Linking regression with potential outcomes

Regression with experimental data: constant causal effects

- The observed outcome Y_i can be written as

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) = Y_i(0) + (Y_i(1) - Y_i(0)) D_i$$

- We assumed that $\beta_i \equiv Y_i(1) - Y_i(0) = \beta \ \ \forall i$.
- Hence we can rewrite the observed outcome as follows

$$Y_i = \alpha + \beta D_i + \eta_i$$

where

$$\alpha \equiv \mathbb{E}(Y_i(0))$$

$$\eta_i \equiv (Y_i(0) - \mathbb{E}(Y_i(0))$$

Linking regression with potential outcomes

Regression with experimental data: heterogeneity

- The same treatment might have different causal effects
- Different individuals respond differently
- Individuals are characterised by X_i (e.g., age)
- We allow for heterogeneous causal effects across X_i
- Heterogeneous causal effects regression

$$Y_i = \alpha + \beta_0 X_i + \delta D_i + \gamma D_i X_i + \eta_i$$

- Heterogeneous causal effects?

$$H_0 : \gamma = 0 \quad \text{(the effect of D does not vary with X)}$$

$$H_1 : \gamma \neq 0 \quad \text{(the effect of D varies with X)}$$

- Where does this regression equation come from?

Linking regression with potential outcomes

Derivation

- Let

$$Y_i(0) = \alpha + \beta_0 X_i + \eta_i$$

$$Y_i(1) = \alpha + \beta_1 X_i + \delta + \eta_i$$

- Hence the causal effect is given by

$$Y_i(1) - Y_i(0) = (\beta_1 - \beta_0) X_i + \delta$$

- Remember that the observed outcome can be written as

$$Y_i = Y_i(0) + (Y_i(1) - Y_i(0)) D_i$$

- Thus

$$Y_i = \mathbb{E}[Y_i(0)|X_i] + ((\beta_1 - \beta_0) X_i + \delta) D_i + (Y_i(0) - \mathbb{E}[Y_i(0)|X_i])$$

Linking regression with potential outcomes

Derivation (con't)

- We have

$$Y_i = \mathbb{E}[Y_i(0)|X_i] + ((\beta_1 - \beta_0)X_i + \delta)D_i + (Y_i(0) - \mathbb{E}[Y_i(0)|X_i])$$

- Now

$$\mathbb{E}[Y_i(0)|X_i] = \mathbb{E}[\alpha + \beta_0 X_i + \eta_i|X_i] = \alpha + \beta_0 X_i$$

- Hence

$$Y_i = \alpha + \beta_0 X_i + ((\beta_1 - \beta_0)X_i + \delta)D_i + \eta_i$$

- The regression with heterogeneous effects across X_i is

$$Y_i = \alpha + \beta_0 X_i + \delta D_i + (\beta_1 - \beta_0)D_i X_i + \eta_i$$

- And finally

$$Y_i = \alpha + \beta_0 X_i + \delta D_i + \gamma D_i X_i + \eta_i$$

where $\gamma = (\beta_1 - \beta_0)$.

How does a role model affect education?

- The following is a study conducted by Emma Riley

  https://emmaalriley.wordpress.com/research/

- Role models improve educational attainment through increasing aspirations and self-belief
- Role models have been used to address the lack of women in the sciences and academic positions
- Role models can overcome stereotypes about academic performance even with very brief exposure where $\gamma = (\beta_1 - \beta_0)$.

Queen of Katwe

How does a role model affect education?

What is the treatment?

- **Queen of Katwe (treatment):** true story of the rise of a teenage Ugandan girl from the slums of Kampala to become a Chess Master.
- Role model characteristics:
  - Similar background and age to students
  - Gendered role models matter for women but not men
  - Counter-stereotype (a woman can do well at a "man's" game)
  - Addresses low aspirations and lack of belief in yourself
    - perseverance and hard work
    - over-coming hardship
    - shaping your own life
    - importance of education for success in life

Miss Peregrine's Home for Peculiar Children

How does a role model affect education?

- **Miss Peregrine's Home for Peculiar Children (Control):**
- adventure story about a group of children fighting against monsters that threaten their home
- chosen as also featured teenage characters
- not set in Uganda, with a strong role model or featuring an education context

How does a role model affect education?

- **Location:** Central Kampala, Uganda
- **Sample:** 1,450 secondary school students in years sitting national exams (S4 and S6)
  - 50% female and 50% male
  - 49% S4 and 51% S6
- **Schools:** 13 schools participated
- **Timing:** Screenings were held from Friday 7th till Tuesday 11th October 2016
  - Two screenings at 11am and 2pm
  - S4 exams started 19th October (1 week later)
  - S6 exams started 14th November (1 month later)

How does a role model affect education?

- Randomisation at the cinema straight before the movie
- Individual level randomisation
- Students lined up outside the cinema
- Were given a ticket out a dark, shuffled bag
- Immediately directed to next room depending on ticket
- No mingling after getting a ticket
- Registered
  - Name
  - School
  - Age
  - Gender
- Perfect compliance: $(Z = D)$

How does a role model affect education?

Did randomisation work as intended?

	Control		**Treatment**			
	mean	sd	mean	sd	difference	p-value
Age	17.28	1.25	17.25	1.23	0.03	(0.76)
Female	0.51	0.50	0.51	0.50	0.00	(0.61)
No. of subjects	9.73	0.62	9.68	0.60	0.04	(0.34)
Observations	344		391		735	

- The p-value is the probability of drawing a statistic at least as adverse to the H_0 as the one that is actually computed in the sample assuming that H_0 is true. For a two-sided test, p-value $= 2\Phi(-t_{act})$.
- The p-value is the *smallest significance level* at which you can reject the H_0.

How does a role model affect education?

Did randomisation work as intended?

	Control		**Treatment**			
	mean	sd	mean	sd	difference	p-value
Age	19.09	1.24	19.00	1.13	0.09	(0.31)
Female	0.47	0.50	0.50	0.50	-0.03	(0.40)
Maths/science	0.33	0.47	0.30	0.46	0.02	(0.53)
Observations	341		369		710	

How does a role model affect education?
Regression with experimental data

Regressions of the type:

$$Y_{i,s} = \alpha + \beta D_{i,s} + \gamma \mathbf{X}_{i,s} + \epsilon_{i,s}$$

where i indexes student at school s and:

- $Y_{i,s}$ denotes the (standardised) exam outcome of interest
- $D_{i,s}$ is an indicator variable equal to one for if the student saw the movie The Queen of Katwe
- $\mathbf{X}_{i,s}$ is a vector of individual characteristics: age and gender (and number of subjects for S4 and Maths/science for S6)
- $\epsilon_{i,s}$ is a random error term

Why do we include control variables?

- To improve the precision of our estimates

How does a role model affect education?
Regression with experimental data

Outcomes at S4
Based on GCSEs in the UK: Students sit between 7 and 11 papers, 6 being compulsory. Maximum score is 8, minimum is zero.

1. **Exam score aggregate**: aggregate score composed of exam scores across all subjects taken by a student
2. **Core exam score** : composed of exam score in the 6 mandatory subjects taken by all students
3. **Individual subject grade**: Score achieved in Maths and English

All subject scores standardized to have mean zero standard deviation one in the control group.

How does a role model affect education?

Regression with experimental data

Treatment effects on S4 test scores

	(1) Total score	(2) Core score	(3) Maths	(4) English
Treatment	0.00	-0.01	0.13*	-0.04
	(0.07)	(0.07)	(0.07)	(0.07)
Age	-0.07**	-0.07**	-0.09***	-0.09***
	(0.03)	(0.03)	(0.03)	(0.03)
Female	-0.27***	-0.25***	-0.23***	-0.10
	(0.06)	(0.06)	(0.07)	(0.06)
No. subjects	0.08	0.06	-0.04	0.07
	(0.06)	(0.06)	(0.06)	(0.06)
Observations	733	733	733	733
R-squared	0.33	0.31	0.19	0.26

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

How does a role model affect education?

Regression with experimental data

Impact of treatment on Maths grade

	(1) Fail
Treatment	-0.11***
	(0.03)
Age	0.04***
	(0.01)
Female	0.08***
	(0.03)
No. subjects	0.01
	(0.02)
Mean in control	0.27
Observations	733
R-squared	0.17

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

How does a role model affect education?
Regression with experimental data

Investigation of heterogeneous treatment effects by

- Students' characteristics (age, gender and number of subjects)
- Schools' characteristics (rank and fees)

Regressions of the type:

$$Y_{i,s} = \alpha + \beta D_{i,s} + \gamma X_{i,s} + \delta D_{i,s} X_{i,s} + \epsilon_{i,s}$$

where i indexes student at school s and:

- $X_{i,s}$ is a **student** or **school** characteristic
- $\delta \neq 0$: heterogeneous causal effects along $X_{i,s}$

If the world is heterogeneous, $X_{i,s}$ can be used to investigate heterogeneous causal effects!

How does a role model affect education?

Regression with experimental data

Heterogeneous Treatment effects on maths scores by school characteristics

	(1) Maths score	(2) Fail maths
Treatment	0.03	-0.07**
	(0.08)	(0.03)
Treatment × below median number of subjects	0.36**	-0.15**
	(0.15)	(0.06)
Below median subjects	-0.18	0.12***
	(0.12)	(0.05)
Mean below median subjects	0.28	
Observations	733	733
R-squared	0.19	0.22

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

How does a role model affect education?

Regression with experimental data

Heterogeneous Treatment effects on maths, stadardised test scores by school characteristics

	(1) Maths score	(2) Fail maths	(3) Maths score	(4) Fail maths
Treatment	0.15	-0.17***	0.18**	-0.17***
	(0.10)	(0.04)	(0.08)	(0.03)
Treatment × top 500	-0.05	0.12**		
	(0.14)	(0.06)		
Top 500	0.94***	-0.52***		
	(0.23)	(0.09)		
Treatment × high fees			-0.15	0.18***
			(0.15)	(0.06)
High fees			1.31***	-0.65***
			(0.23)	(0.09)
Mean top 500	0.46			
Mean high fees				0.31
Observations	734	734	734	734
R-squared	0.19	0.21	0.19	0.21

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Outcomes S6
Based on A-levels in the UK:

1. **Principal score**: 3 main subjects scored A-F, with A scoring 6 points
2. **General paper and subsidiary paper score**: 2 papers, one in maths or computer, taken by all students scored 8-0. 2 or below is a fail.
3. **Total exam score**: Principal scores + one point per subsidiary paper if passed

All subject scores standardized to have mean zero standard deviation one in the control group.

How does a role model affect education?

Regression with experimental data

Impact on S6 standardised test scores

	(1) Overall score	(2) Principal subjects	(3) subsidiary subjects	(4) uni pass
Treatment	0.15**	0.15*	0.08	0.05*
	(0.07)	(0.08)	(0.06)	(0.03)
Age	-0.08**	-0.07**	-0.06**	-0.02*
	(0.03)	(0.03)	(0.03)	(0.01)
Female	-0.02	0.03	-0.12*	-0.03
	(0.07)	(0.07)	(0.06)	(0.03)
maths/science	0.17**	0.01	0.65***	-0.31***
	(0.08)	(0.09)	(0.07)	(0.04)
Observations	710	710	710	710
R-squared	0.21	0.16	0.38	0.19

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

How does a role model affect education?
Regression with experimental data

Summary of findings

- A one-off exposure to a role model improved students exam performance
- Reduction in students failing maths at S4
  - Main effect coming from students taking less subjects than the median
  - Largest effect seen for students at poorest performing/less expensive schools
  - 0.15 improvement in overall test score at S6
  - more likely to make the grades to get into public university
- Extremely costs effective
  - The total cost for the screening was \$2 for transport and \$3 for the cinema ticket=\$5

RCTs in the Real World^{TM}

Summary

- RCTs are the gold standard for causal inference
- But people make for
  - unreliable experimental subjects, and
  - unreliable experimenters.
- When you are assessing causal claims you have to be aware of thes effects and interpret accordingly.
  - LATE etc helps with the former
  - Pre-analysis plans can allow you to uncover heterogeneous treatment effects without cherry-picking results.

Reading

- Angrist, J.D. and J-S. Pischke (2015) *Mastering 'Metrics: The Path from Cause to Effect*, Princeton University Press. Chapter 1.
- Angrist, J.D. and J-S. Pischke (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press. Chapters 1, 2 and 4.
- Aron-Dine, A., L. Einav, and A. Finkelstein. (2013). "The RAND Health Insurance Experiment, Three Decades Later." *Journal of Economic Perspectives*, 27 (1): 197-222.
- Siddique, Z. (2014) "Randomized control trials in an imperfect world." *IZA World of Labor*, 2014. 110.
- Finkelstein, A., S. Taubman, B. Wright. M. Bernstein, J. Gruber, J. Newhouse, H. Allen, K. Baicker, Oregon Health Study Group (2012) "The Oregon Health Insurance Experiment: Evidence from the First Year" *The Quarterly Journal of Economics*, Volume 127, Issue 3, 1, 1057–1106.
- Riley, E. (2017), "Increasing students' aspirations: the impact of Queen of Katwe on students' educational attainment", CSAE Working Paper WPS/2017 - 13