Compliance and the limits of RCTs

Matti Sarvimäki

Principles of Empirical Analysis
Lecture 7

Today's learning objectives

- Today's question: How to deal with **imperfect compliance**?
  - some randomized into the treatment group do not get the treament
  - some randomized into the control group get the treament
- Key concepts
  1. compliers, always-takers and never-takers
  2. intention-to-treat (ITT)
  3. first-stage
  4. local average treatment effect (LATE)
  5. average treatment effect on the treated (ATT or TOT)
- We also briefly discuss the limits of RCTs

Today's running example: Moving to Opportunity

- One of the most famous social experiments of all time
  - target group: households with children living in high-poverty public housing projects (primarily minority, single mother families)
  - implemented in 1994-98 in Baltimore, Boston, Chicago, LA, New York

Today's running example: Moving to Opportunity

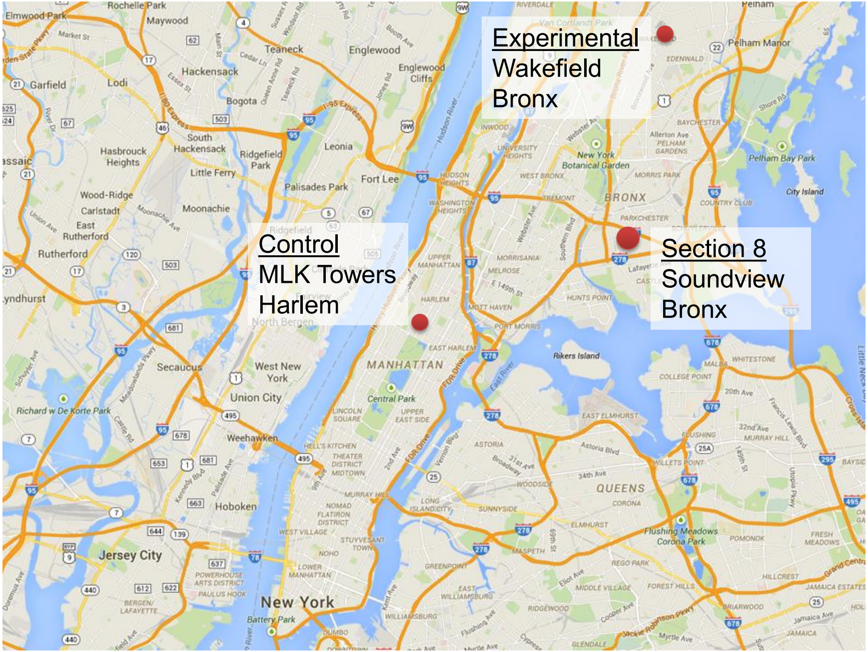
- One of the most famous social experiments of all time
  - target group: households with children living in high-poverty public housing projects (primarily minority, single mother families)
  - implemented in 1994-98 in Baltimore, Boston, Chicago, LA, New York
- Random assignment of 4,600 families into three groups:
  - control: not offered a voucher, stayed in public housing
  - section 8: offered conventional housing vouchers, no restrictions
  - experimental: offered <u>housing vouchers to low-poverty neighborhoods</u>

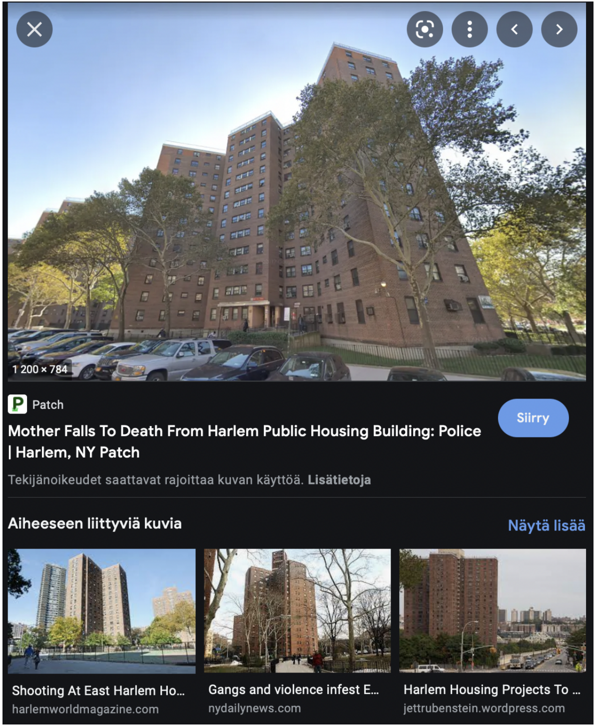
Today's running example: Moving to Opportunity

- One of the most famous social experiments of all time
  - target group: households with children living in high-poverty public housing projects (primarily minority, single mother families)
  - implemented in 1994-98 in Baltimore, Boston, Chicago, LA, New York
- Random assignment of 4,600 families into three groups:
  - control: not offered a voucher, stayed in public housing
  - section 8: offered conventional housing vouchers, no restrictions
  - experimental: offered housing vouchers to low-poverty neighborhoods
- Many families chose not to use the voucher they were offered
  - 48% of experimental group used voucher
  - 66% of Section 8 group used voucher

The MTO parts of these slides draw heavily from lecture 3 of Raj Chetty's excellent course Using Big Data to Solve Economic and Social Problems. I'm also borrowing quite a bit from Tuukka's (also excellent) Urban Economics course.

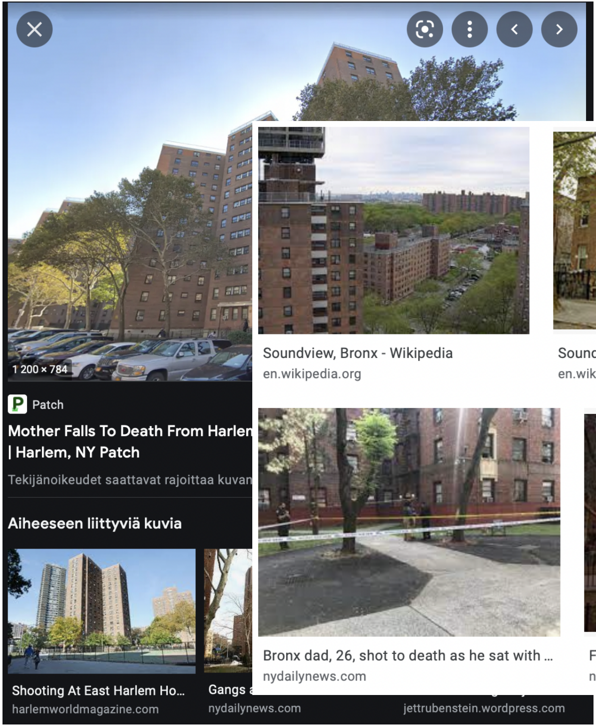
Common MTO Residential Locations in New York

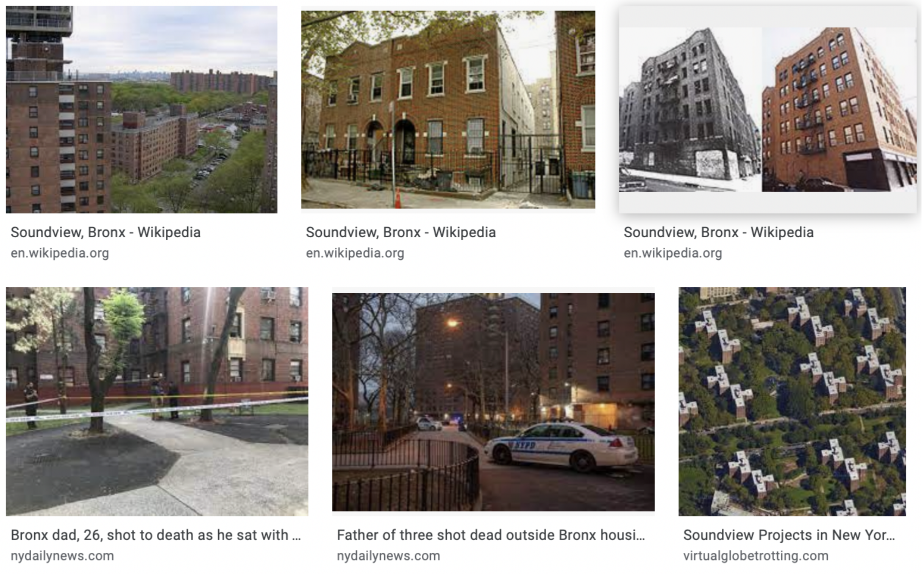




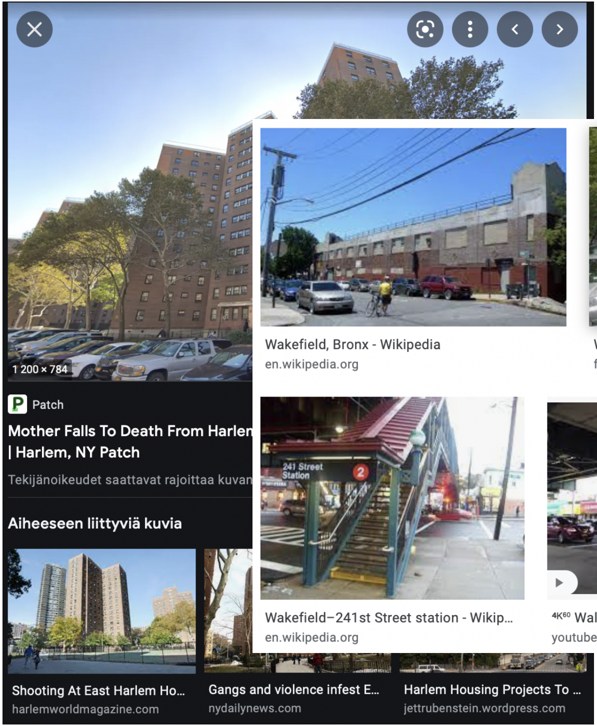
First results from a Google image search, Jan 28th, 2022

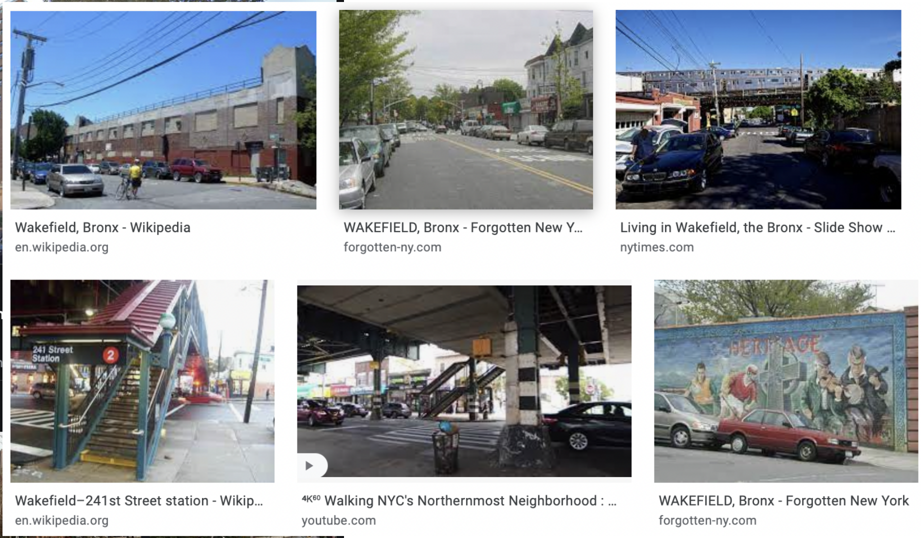
First results from a Google image search, Jan 28th, 2022





First results from a Google image search, Jan 28th, 2022





Early research

MEAN EFFECT SIZES FOR SUMMARY MEASURES OF OUTCOMES[a]

	All Adults
	E – C
	(i)
Economic self-sufficiency	0.017
	(0.031)
Absence of physical health problems	0.012
	(0.024)
Absence of mental health problems	0.079*
	(0.030)
Absence of risky behavior	
Education	
Overall	0.036
	(0.020)

[a] E – C denotes experimental – control

Robust standard errors adjusted for household clustering are in parentheses; * = p-value < 0.05.

Kling, Liebman, Katz (2007): Experimental Analysis of Neighborhood Effects.

Early research

MEAN EFFECT SIZES FOR SUMMARY MEASURES OF OUTCOMES[a]

	All Adults	
	E – C (i)	S – C (ii)
Economic self-sufficiency	0.017 (0.031)	0.037 (0.033)
Absence of physical health problems	0.012 (0.024)	0.019 (0.026)
Absence of mental health problems	0.079* (0.030)	0.029 (0.033)
Absence of risky behavior		
Education		
Overall	0.036 (0.020)	0.028 (0.022)

[a]E – C denotes experimental – control; S – C denotes Section 8 – control. Estimates are the intent-to-treat mean effect sizes,

Robust standard errors adjusted for household clustering are in parentheses; * = p-value < 0.05.

Kling, Liebman, Katz (2007): Experimental Analysis of Neighborhood Effects.

Early research

MEAN EFFECT SIZES FOR SUMMARY MEASURES OF OUTCOMES[a]

	All Adults		All Youth		Female Youth		Male Youth		M – F Youth	
	E – C	S – C	E – C	S – C	E – C	S – C	E – C	S – C	E – C	S–C
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)
Economic	0.017	0.037								
self-sufficiency	(0.031)	(0.033)								
Absence of physical	0.012	0.019	–0.038	–0.020	0.025	0.077	–0.112*	–0.114	–0.138	–0.192*
health problems	(0.024)	(0.026)	(0.038)	(0.040)	(0.053)	(0.055)	(0.053)	(0.061)	(0.076)	(0.084)
Absence of mental	0.079*	0.029	0.102	0.138*	0.267*	0.192*	–0.052	0.054	–0.319*	–0.138
health problems	(0.030)	(0.033)	(0.053)	(0.056)	(0.062)	(0.067)	(0.080)	(0.092)	(0.101)	(0.113)
Absence of risky			–0.023	–0.039	0.142*	0.129*	–0.181*	–0.208*	–0.323*	–0.337*
behavior			(0.043)	(0.050)	(0.053)	(0.059)	(0.062)	(0.071)	(0.080)	(0.092)
Education			0.050	0.028	0.138*	0.056	–0.053	–0.001	–0.191*	–0.057
			(0.041)	(0.047)	(0.065)	(0.068)	(0.047)	(0.060)	(0.080)	(0.090)
Overall	0.036	0.028	0.018	0.018	0.136*	0.109*	–0.099*	–0.078*	–0.235*	–0.187*
	(0.020)	(0.022)	(0.025)	(0.026)	(0.034)	(0.034)	(0.031)	(0.037)	(0.047)	(0.051)

[a] E – C denotes experimental – control; S – C denotes Section 8 – control. Estimates are the intent-to-treat mean effect sizes, from Equation (1), fully interacted with gender in columns (v)–(x) as described in the text. The estimated equations all include site indicators and the baseline covariates listed in Appendix A with those in Table A1 included for adults and those in Tables A1 and A2 included for youth. M – F Youth is male – female difference. Adult economic self-sufficiency: + adult not employed and not on TANF + employed + 2001 earnings – on TANF – 2001 government income. Adult mental health: – distress index – depression symptoms – worrying + calmness + sleep. Adult physical health: – self-reported health fair/poor – asthma attack past year – obesity – hypertension – trouble carrying/climbing. Adult overall includes 15 measures in self-sufficiency, physical health, and mental health. Youth physical health: – self-reported health fair/poor – asthma attack past year – obesity – nonsports injury past year. Youth mental health: – distress index – depression symptoms – anxiety symptoms. Youth risky behavior: – marijuana past 30 days – smoking past 30 days – alcohol past 30 days – ever pregnant or gotten someone pregnant. Youth education: + graduated high school or still in school + in school or working + WJ-R broad reading score + WJ-R broad math score. Youth overall includes 15 measures in physical health, mental health, risky behavior, and education. Sample sizes in the E, S, and C groups are 1,453, 993, and 1,080 for adults and 749, 510, and 548 for youth ages 15–20 on 12/31/2001. Robust standard errors adjusted for household clustering are in parentheses; * = p-value < 0.05.

Kling, Liebman, Katz (2007): Experimental Analysis of Neighborhood Effects.

Early research

- Making sense of the previous table
  - outcomes: indices that aggregate information over multiple measures
    - for example, the index of economic self-sufficiency includes five measures of employment, earnings, and public assistance
  - each index has mean 0 and standard deviation 1

Early research

- Making sense of the previous table
  - outcomes: indices that aggregate information over multiple measures
    - for example, the index of economic self-sufficiency includes five measures of employment, earnings, and public assistance
  - each index has mean 0 and standard deviation 1
- Impacts of being offered an experimental voucher (4–7 years later)
  - no effects on adult economic self-sufficiency or physical health

Early research

- Making sense of the previous table
  - outcomes: indices that aggregate information over multiple measures
    - for example, the index of economic self-sufficiency includes five measures of employment, earnings, and public assistance
  - each index has mean 0 and standard deviation 1
- Impacts of being offered an experimental voucher (4–7 years later)
  - no effects on adult economic self-sufficiency or physical health
    - MDE for economic self-sufficiency: $2.8 \times .031 = .09$sd

Early research

- Making sense of the previous table
  - outcomes: indices that aggregate information over multiple measures
    - for example, the index of economic self-sufficiency includes five measures of employment, earnings, and public assistance
  - each index has mean 0 and standard deviation 1
- Impacts of being offered an experimental voucher (4–7 years later)
  - no effects on adult economic self-sufficiency or physical health
    - MDE for economic self-sufficiency: $2.8 \times .031 = .09$sd
  - improved mental health for adults
  - positive effect on teenage girls
  - negative effect on teenage boys

The most recent results

- Chetty, Hendren, Katz (2016) focus on those moving as children
    - group 1: younger than 13 (average 8.2) at assignment
    - group 2: 13-18 years old (average 15.1) at assignment

The most recent results

- Chetty, Hendren, Katz (2016) focus on those moving as children
  - group 1: younger than 13 (average 8.2) at assignment
  - group 2: 13-18 years old (average 15.1) at assignment
- MTO data linked to 1996–2012 federal income tax returns
  - 4,604 households and 15,892 individuals
    - primary focus on 8,603 children born in or before 1991
  - about 85% of children matched
    - match rates do not differ significantly across treatment groups
    - baseline covariates balanced across treatment groups in matched data

The most recent results

- Chetty, Hendren, Katz (2016) focus on those moving as children
    - group 1: younger than 13 (average 8.2) at assignment
    - group 2: 13-18 years old (average 15.1) at assignment
- MTO data linked to 1996–2012 federal income tax returns
    - 4,604 households and 15,892 individuals
        - primary focus on 8,603 children born in or before 1991
    - about 85% of children matched
        - match rates do not differ significantly across treatment groups
        - baseline covariates balanced across treatment groups in matched data
- Using administrative data (tax records) is quite new in the US
    - earlier work based typically on survey data
    - in the Nordic countries, we have a long tradition (and much better infrastructure) for using administrative data in research

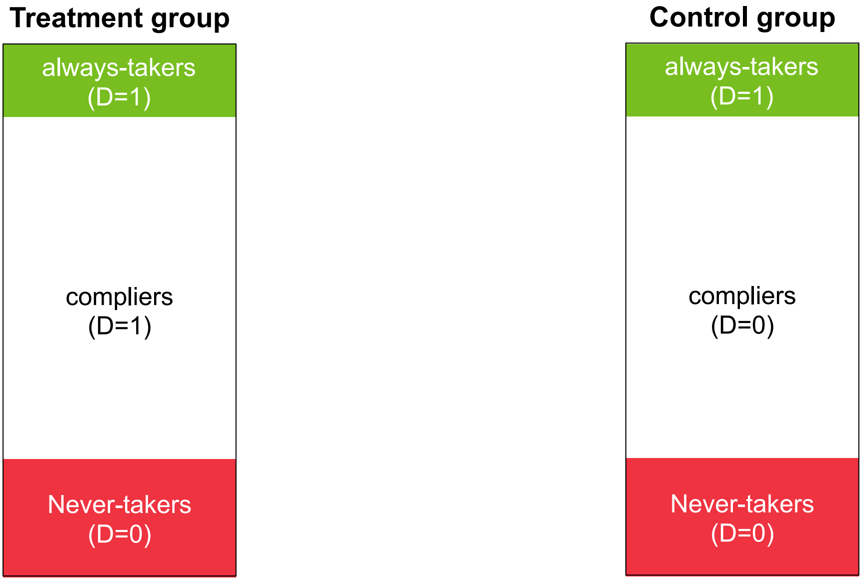
Compliance

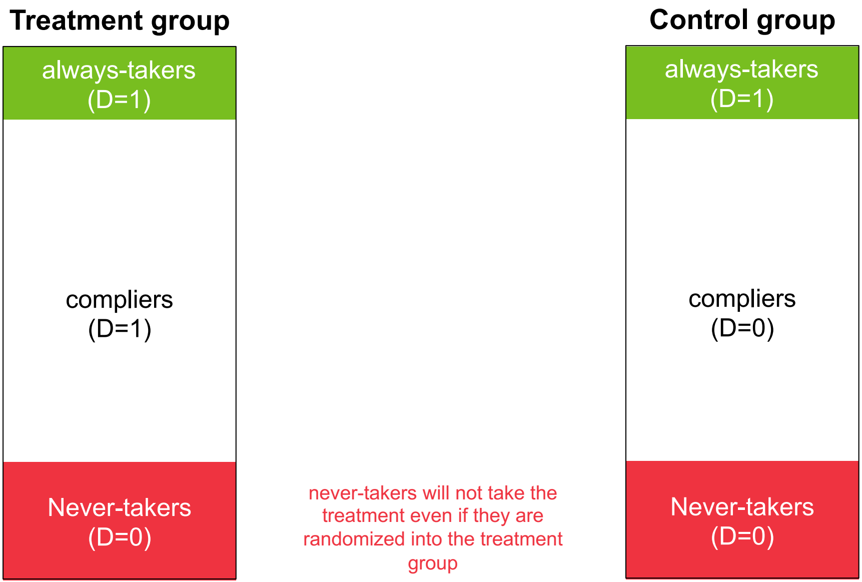
- Often only part of the treatment group actually gets the treament
  - e.g. only 48% of those randomized into the experimental group in MTO chose to use the voucher (column 1 of the previous slide)
  - similarly, 66% of the section 8 group used the voucher

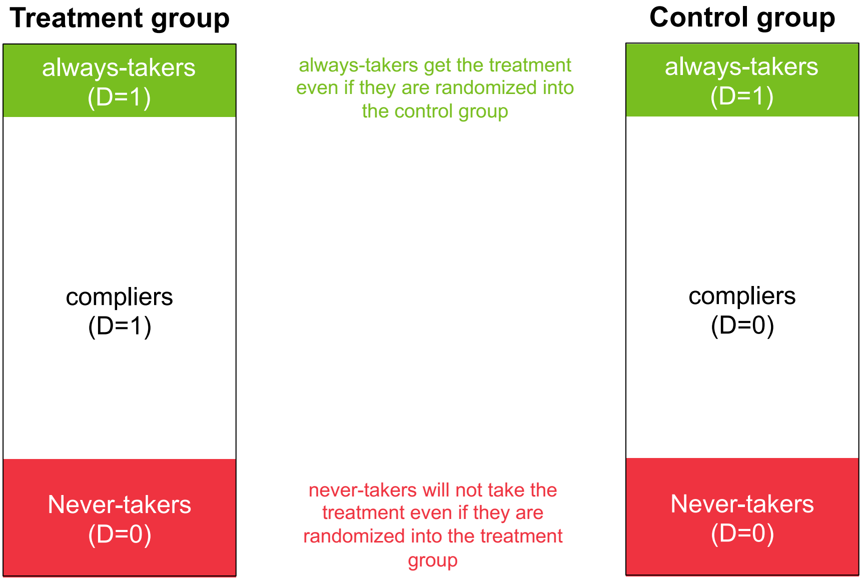
Compliance

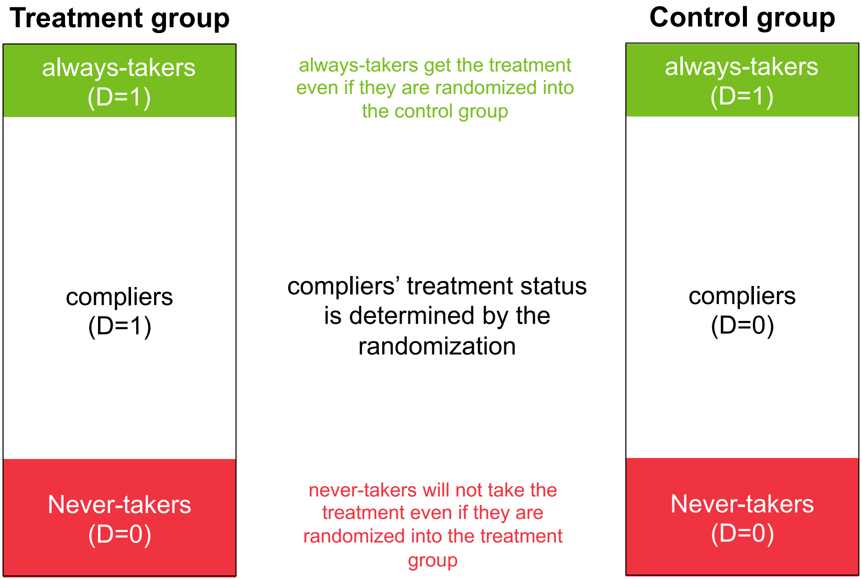
- Often only part of the treatment group actually gets the treament
  - e.g. only 48% of those randomized into the experimental group in MTO chose to use the voucher (column 1 of the previous slide)
  - similarly, 66% of the section 8 group used the voucher
- Compliance *choice* is potentially affected by potential outcomes
  - e.g. those expecting to benefit the least becoming never-takers

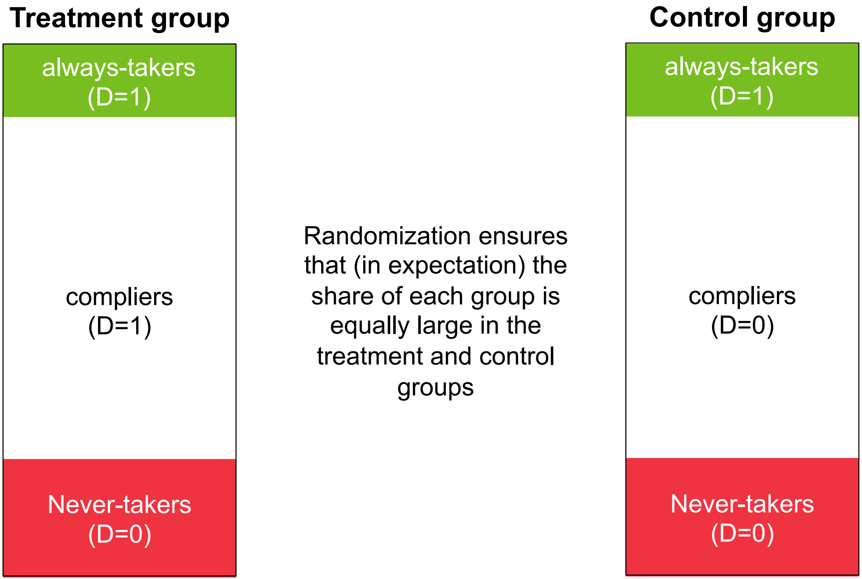
  → comparing those who actually gets the treatment to the entire control group is not a valid comparison


Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)

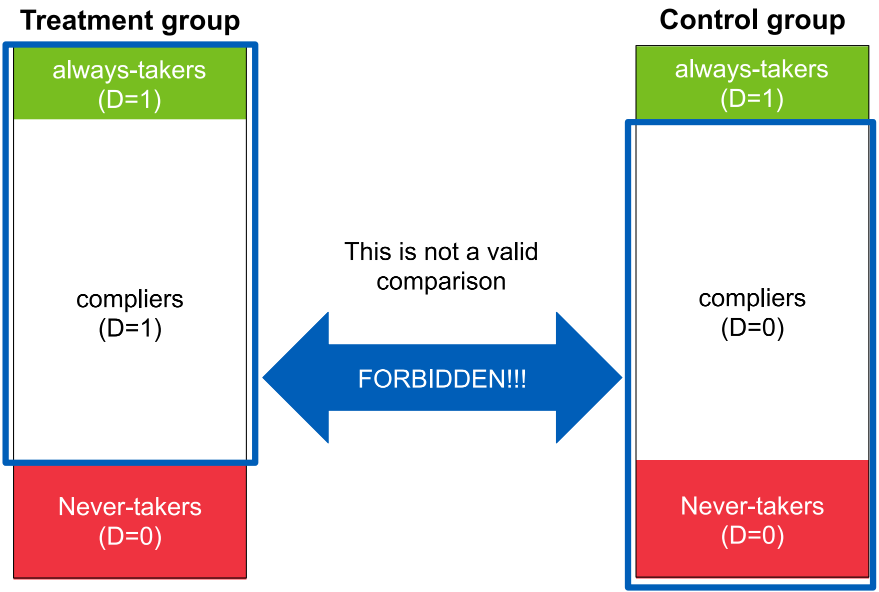

Treatment group
always-takers (D=1)
compliers (D=1)
Never-takers (D=0)
never-takers will not take the treatment even if they are randomized into the treatment group
Control group
always-takers (D=1)
compliers (D=0)
Never-takers (D=0)

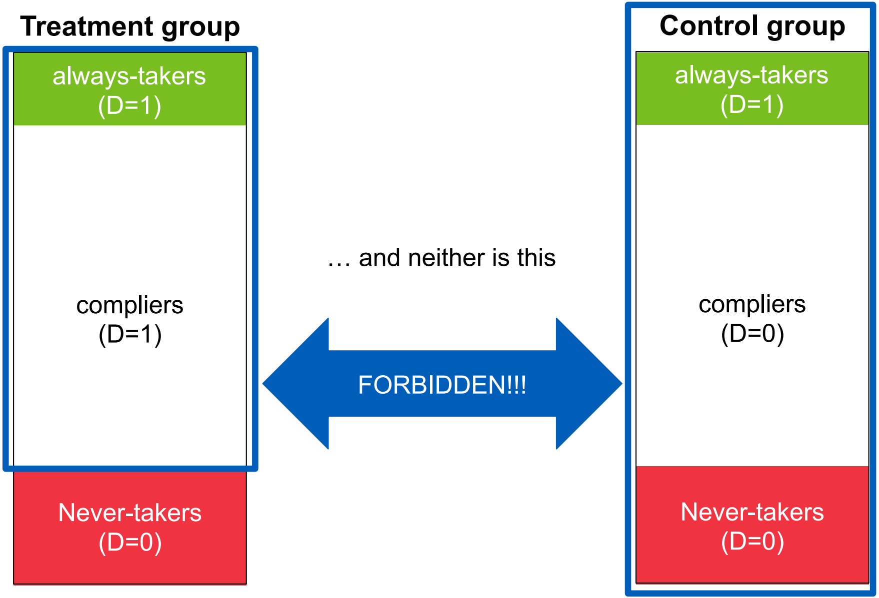

Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
always-takers get the treatment even if they are randomized into the control group
never-takers will not take the treatment even if they are randomized into the treatment group
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)

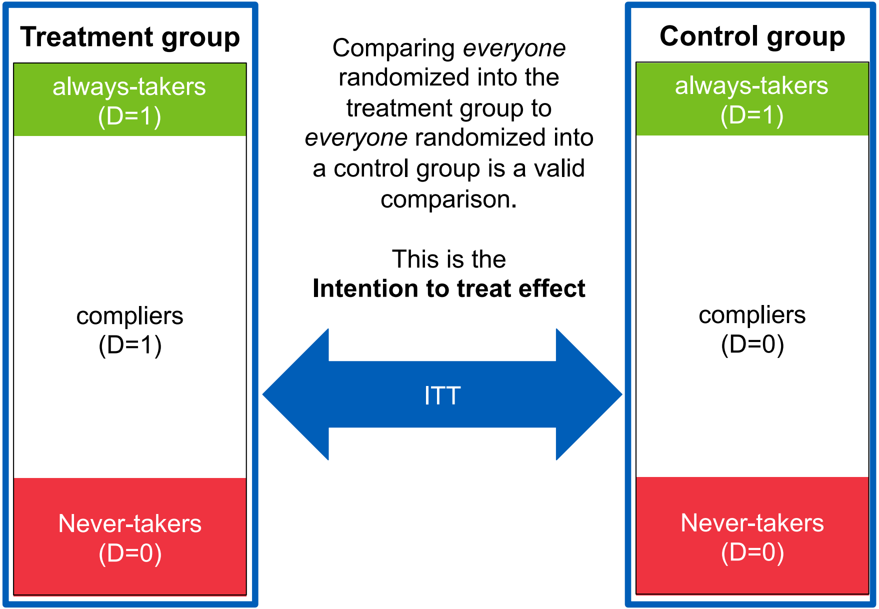

Treatment group
Control group
always-takers (D=1)
compliers (D=1)
Never-takers (D=0)
always-takers get the treatment even if they are randomized into the control group
compliers’ treatment status is determined by the randomization
never-takers will not take the treatment even if they are randomized into the treatment group
always-takers (D=1)
compliers (D=0)
Never-takers (D=0)

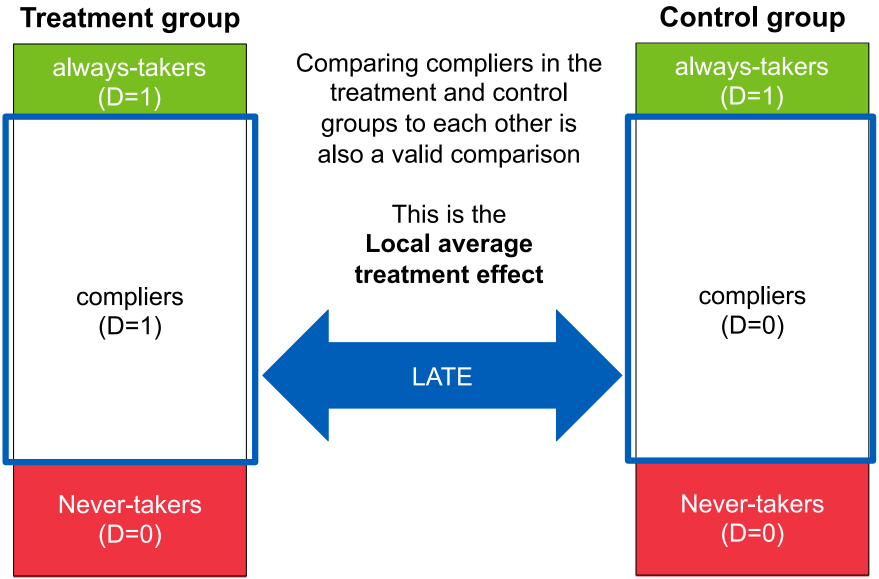


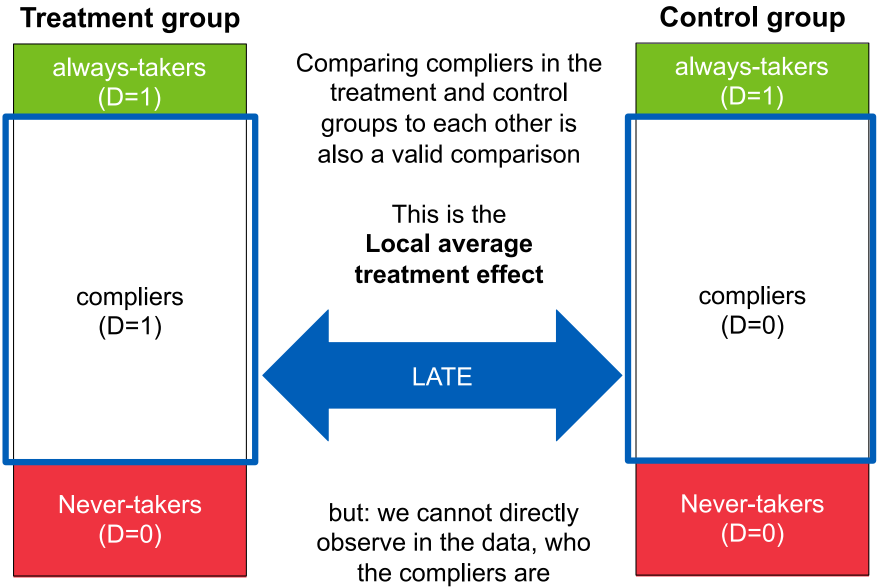
Randomization ensures that (in expectation) the share of each group is equally large in the treatment and control groups


Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
This is not a valid
comparison
FORBIDDEN!!!
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)


Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
… and neither is this
FORBIDDEN!!!
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)


Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
Comparing everyone randomized into the treatment group to everyone randomized into a control group is a valid comparison.
This is the
Intention to treat effect
ITT
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)


Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)
Comparing compliers in the treatment and control groups to each other is also a valid comparison
This is the **Local average treatment effect**
LATE


Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
Comparing compliers in the treatment and control groups to each other is also a valid comparison
This is the
Local average treatment effect
LATE
but: we cannot directly observe in the data, who the compliers are
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)

Treatment group

always-takers
(D=1)

compliers
(D=1)

Never-takers
(D=0)

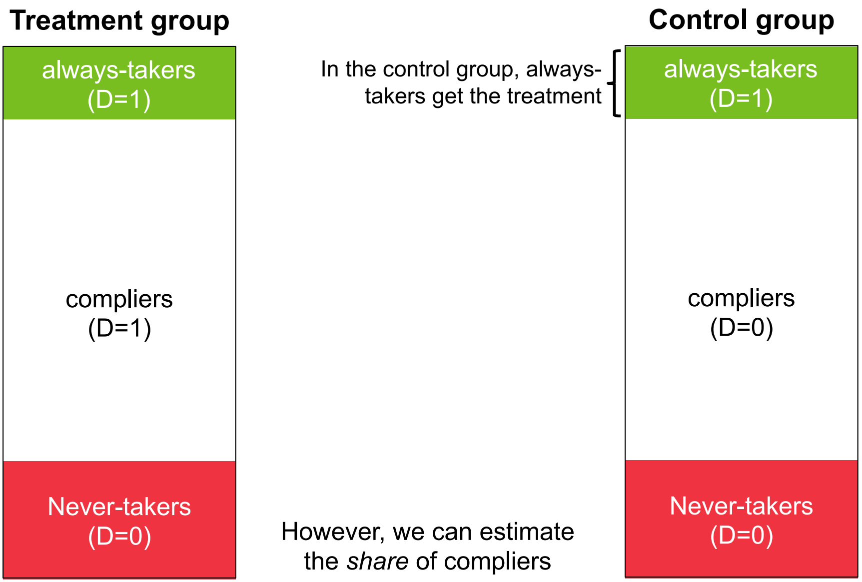
Control group

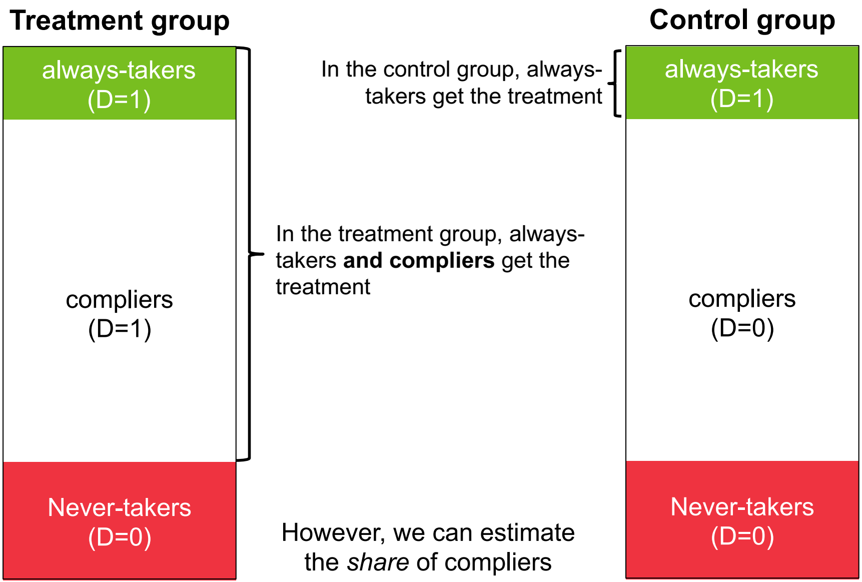
always-takers
(D=1)

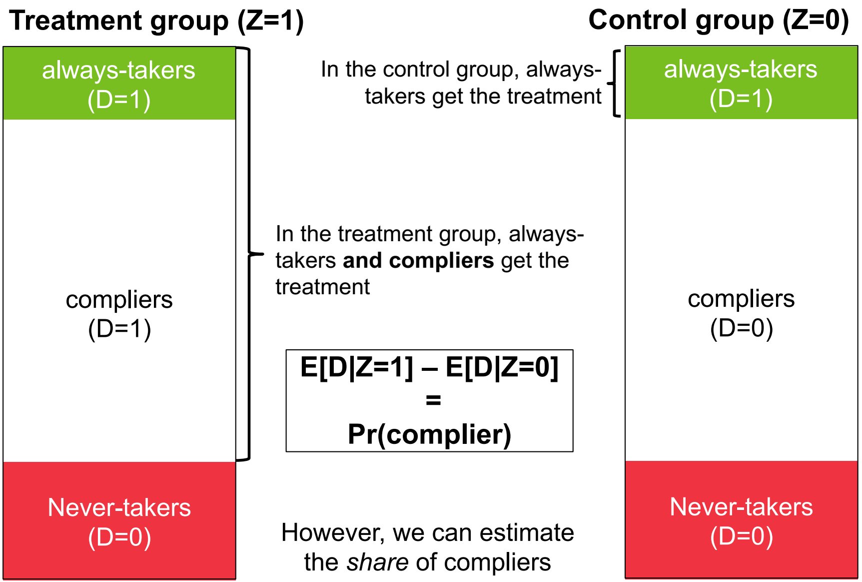
compliers
(D=0)

Never-takers
(D=0)

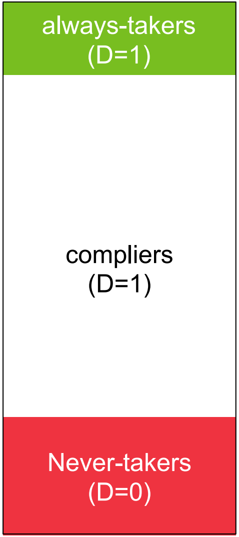
However, we can estimate
the *share* of compliers


Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)
In the control group, always-takers get the treatment
However, we can estimate the *share* of compliers


Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
In the control group, always-takers get the treatment
In the treatment group, always-takers **and compliers** get the treatment
However, we can estimate the *share* of compliers
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)

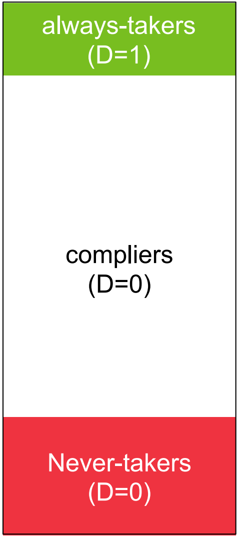

Treatment group (Z=1)
Control group (Z=0)
always-takers (D=1)
compliers (D=1)
Never-takers (D=0)
In the treatment group, always-takers **and compliers** get the treatment
In the control group, always-takers get the treatment
always-takers (D=1)
compliers (D=0)
Never-takers (D=0)
E[D|Z=1] – E[D|Z=0] = Pr(complier)
However, we can estimate the *share* of compliers

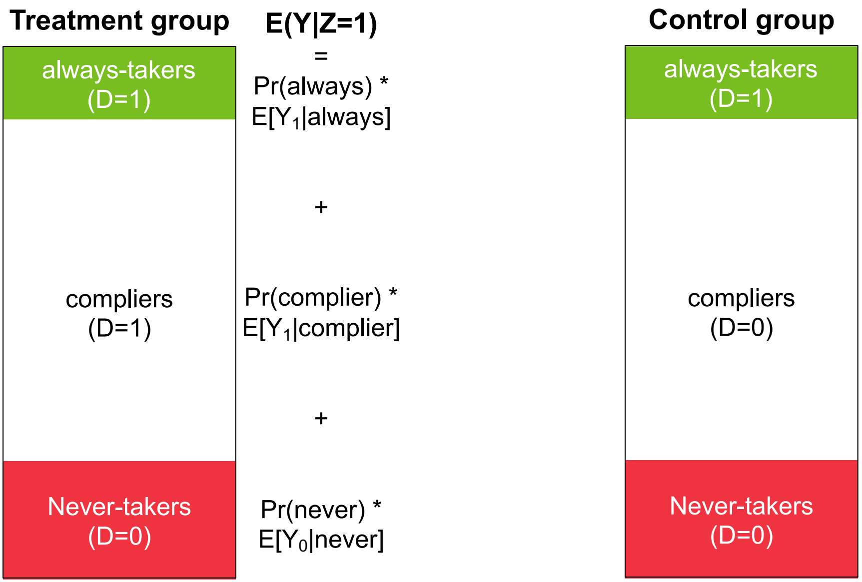
Treatment group (Z=1) **Control group (Z=0)**



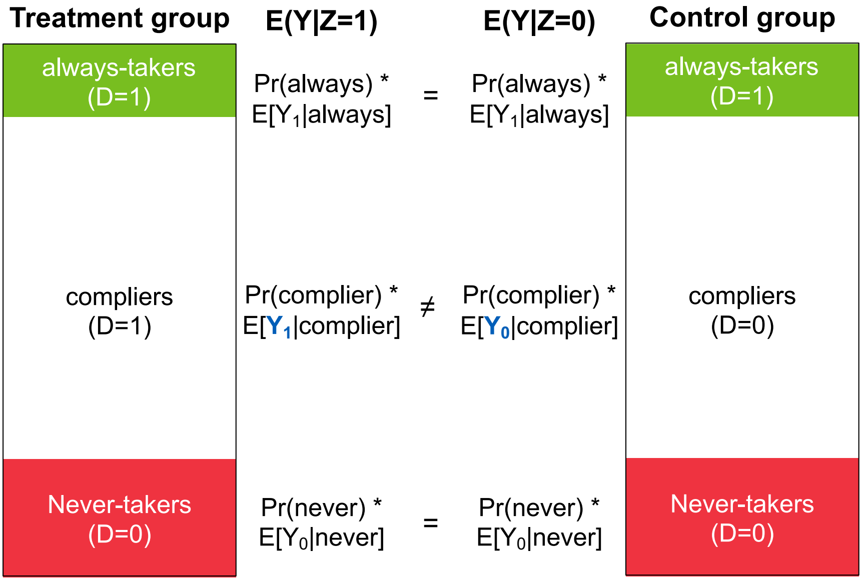
Let’s denote the expected outcome of the treatment group as E[Y|Z=1], where Z denotes randomization status.

This is just the weighted average of the expectations among the always takers, compliers and never-takers in the treatment group, where the weights correspond to the shares of each group.




Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
E(Y|Z=1)
=
Pr(always) * E[Y_1|always]
+
Pr(complier) * E[Y_1|complier]
+
Pr(never) * E[Y_0|never]
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)

Treatment group	**E(Y|Z=1)**	**E(Y|Z=0)**	**Control group**
always-takers (D=1)	= Pr(always) * $E[Y_1\|always]$	= Pr(always) * $E[Y_1\|always]$	always-takers (D=1)
	+	+	
compliers (D=1)	Pr(complier) * $E[Y_1\|complier]$	Pr(complier) * $E[Y_0\|complier]$	compliers (D=0)
	+	+	
Never-takers (D=0)	Pr(never) * $E[Y_0\|never]$	Pr(never) * $E[Y_0\|never]$	Never-takers (D=0)


Treatment group
E(Y|Z=1)
E(Y|Z=0)
Control group
always-takers (D=1)
Pr(always) * E[Y_1|always]
=
Pr(always) * E[Y_1|always]
always-takers (D=1)
compliers (D=1)
Pr(complier) * E[Y_1|complier]
≠
Pr(complier) * E[Y_0|complier]
compliers (D=0)
Never-takers (D=0)
Pr(never) * E[Y_0|never]
=
Pr(never) * E[Y_0|never]
Never-takers (D=0)

Treatment group

always-takers
(D=1)

compliers
(D=1)

Never-takers
(D=0)

E(Y|Z=1) - E(Y|Z=0)

=

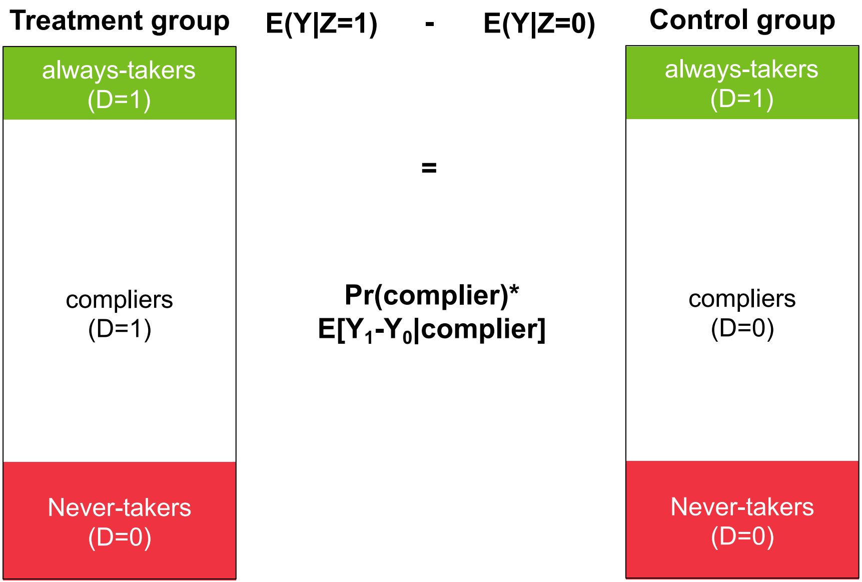
Pr(complier) * E[Y_1|complier] - Pr(complier) * E[Y_0|complier]

Control group

always-takers
(D=1)

compliers
(D=0)

Never-takers
(D=0)


Treatment group
E(Y|Z=1)
-
E(Y|Z=0)
Control group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
=
Pr(complier)*
E[Y_1-Y_0|complier]
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)

Treatment group

always-takers
(D=1)

compliers
(D=1)

Never-takers
(D=0)

$$\frac{\Pr(\text{complier}) * E[Y_1 - Y_0 | \text{complier}]}{\Pr(\text{complier})} = E[Y_1 - Y_0 | \text{complier}]$$

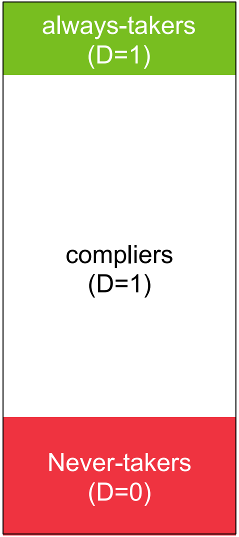
Control group

always-takers
(D=1)

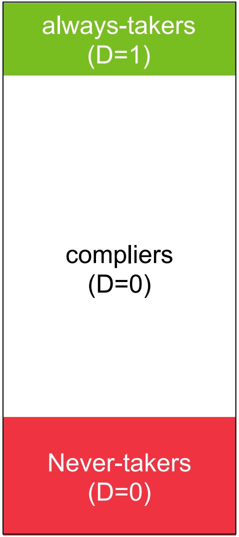
compliers
(D=0)

Never-takers
(D=0)

Treatment group (Z=1)



Control group (Z=0)



$$\frac{E[Y|Z=1] - E[Y|Z=0]}{Pr(complier)} = E[Y_1 - Y_0|complier]$$

Treatment group (Z=1)

always-takers
(D=1)

compliers
(D=1)

Never-takers
(D=0)

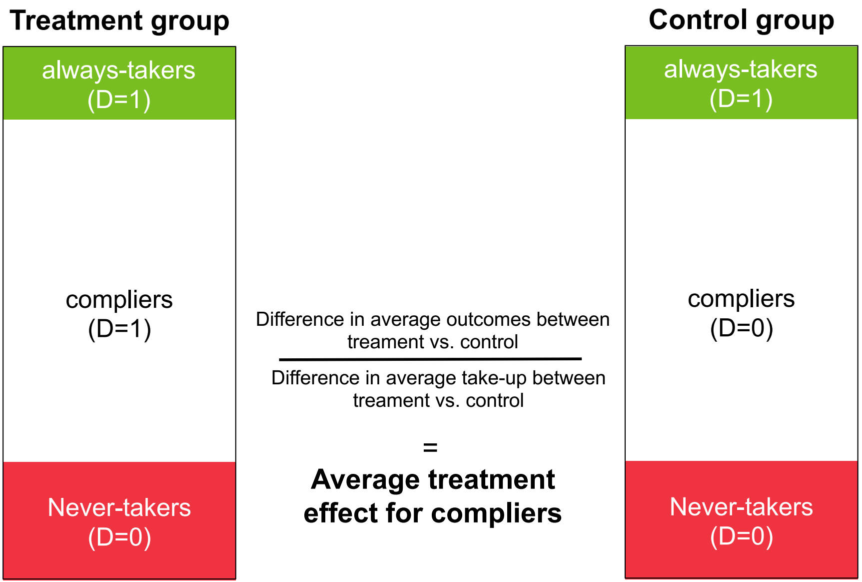
Control group (Z=0)

always-takers
(D=1)

compliers
(D=0)

Never-takers
(D=0)

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} = E[Y_1 - Y_0|\text{complier}]$$


Treatment group
always-takers
(D=1)
compliers
(D=1)
Never-takers
(D=0)
Difference in average outcomes between treament vs. control
Difference in average take-up between treament vs. control
=
Average treatment effect for compliers
Control group
always-takers
(D=1)
compliers
(D=0)
Never-takers
(D=0)

Wald estimator

- We just derived the **Wald estimator**

$$\beta_{LATE} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]}$$

  - Y is the outcome
  - Z is a 0/1 indicator for being randomized into the treatment group
  - D is a 0/1 indicator for actually receiving the treatment

Abraham Wald

- **Wald, A.** (1940): The Fitting of Straight Lines if Both Variables Are Subject to Error. *Annals of Mathematical Statistics* 11(3): 284–300.

3. Consistent Estimates of the Parameters α, β, σ_ϵ, σ_η. For the sake of simplicity we assume that N is even. We consider the expression

$$(1) \qquad a_1 = \frac{(x_1 + \cdots + x_m) - (x_{m+1} + \cdots + x_N)}{N}; \qquad a_2 = \frac{(y_1 + \cdots + y_m) - (y_{m+1} + \cdots + y_N)}{N},$$

where $m = N/2$. As an estimate of α we shall use the expression

$$(2) \qquad a = \frac{a_2}{a_1} = \frac{(y_1 + \cdots + y_m) - (y_{m+1} + \cdots + y_N)}{(x_1 + \cdots + x_m) - (x_{m+1} + \cdots + x_N)}.$$

Abraham Wald, 1902–1950

Guido Imbens, 1963–

2021 Nobel Memorial Price "for their methodological contributions to the analysis of causal relationships"

Joshua Angrist, 1960–

ITT, LATE and TOT

- Back to the Wald estimator

$$\beta_{LATE} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]}$$

  - Y is the outcome
  - Z is a 0/1 indicator for being randomized into the treatment group
  - D is a 0/1 indicator for actually receiving the treatment

ITT, LATE and TOT

- Back to the Wald estimator

$$\beta_{LATE} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]}$$

  - Y is the outcome
  - Z is a 0/1 indicator for being randomized into the treatment group
  - D is a 0/1 indicator for actually receiving the treatment
- Components of the Wald estimator
  - the numerator is the intention to treat effect (ITT)
  - the denominator is the share of compliers (first-stage)
  - $\beta_{LATE} = \mathbb{E}[Y_1 - Y_0|\text{complier}]$ is the local average treatment effect
    - the impact of receiving the treatment for the *compliers*
    - may differ from the impact on never-takers and always-takers

ITT, LATE and TOT

- Back to the Wald estimator

$$\beta_{LATE} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]}$$

  - Y is the outcome
  - Z is a 0/1 indicator for being randomized into the treatment group
  - D is a 0/1 indicator for actually receiving the treatment
- Components of the Wald estimator
  - the numerator is the intention to treat effect (ITT)
  - the denominator is the share of compliers (first-stage)
  - $\beta_{LATE} = \mathbb{E}[Y_1 - Y_0|\text{complier}]$ is the local average treatment effect
    - the impact of receiving the treatment for the *compliers*
    - may differ from the impact on never-takers and always-takers
- This is one version of the **instrumental variables (IV)** estimators
  - you'll see more in later courses (no need to get this now)

ITT, LATE and TOT

- Sometimes ITT is the most relevant estimate
  - in the context of the MTO, it is the impact of *offering* housing vouchers
  - this is arguably the most relevant effect given that offering vouchers is likely to be the relevant policy (rather than forcing everyone to move)

ITT, LATE and TOT

- Sometimes ITT is the most relevant estimate
  - in the context of the MTO, it is the impact of *offering* housing vouchers
  - this is arguably the most relevant effect given that offering vouchers is likely to be the relevant policy (rather than forcing everyone to move)
- Sometimes LATE is more relevant
  - in MTO, it is the impact of living in better neighborhoods
  - potentially informative for policy discussion on whether we should invest in improving existing neighborhoods ("place-making policies")

ITT, LATE and TOT

- Sometimes ITT is the most relevant estimate
  - in the context of the MTO, it is the impact of *offering* housing vouchers
  - this is arguably the most relevant effect given that offering vouchers is likely to be the relevant policy (rather than forcing everyone to move)
- Sometimes LATE is more relevant
  - in MTO, it is the impact of living in better neighborhoods
  - potentially informative for policy discussion on whether we should invest in improving existing neighborhoods ("place-making policies")
- LATE informs us only about the impact on compliers
  - usefulness depends on how representative the compliers are
  - when there are no never-takers LATE = TOT (average treatment effect on the treated)
    - this is the case in MOT, thus the "TOT"s in the tables

ITT, LATE and TOT

- Sometimes ITT is the most relevant estimate
  - in the context of the MTO, it is the impact of *offering* housing vouchers
  - this is arguably the most relevant effect given that offering vouchers is likely to be the relevant policy (rather than forcing everyone to move)
- Sometimes LATE is more relevant
  - in MTO, it is the impact of living in better neighborhoods
  - potentially informative for policy discussion on whether we should invest in improving existing neighborhoods ("place-making policies")
- LATE informs us only about the impact on compliers
  - usefulness depends on how representative the compliers are
  - when there are no never-takers LATE = TOT (average treatment effect on the treated)
    - this is the case in MOT, thus the "TOT"s in the tables
- We now have everything we need to understand the MTO results

TABLE 3—IMPACTS OF MTO ON CHILDREN'S INCOME IN ADULTHOOD

	Individual earnings 2008–2012 ($)	
	ITT (2)	TOT (4)
Panel A. Children < age 13 at random assignment		
Exp. versus control	1,624.0**	3,476.8**
	(662.4)	(1,418.2)
Observations	8,420	8,420
Control group mean	11,270.3	11,270.3

$$\boxed{\begin{array}{c}\text{TOT}\\=\\\text{ITT / First stage}\\=\\\$1{,}624/.467\\=\\\$3{,}476.8\end{array}}$$

Chetty, Hendren, Katz (2016): The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment

TABLE 3—IMPACTS OF MTO ON CHILDREN'S INCOME IN ADULTHOOD

	W-2 earnings ($) 2008–2012	Individual earnings 2008–2012 ($)		
	ITT	ITT	ITT w/ controls	TOT
	(1)	(2)	(3)	(4)
Panel A. Children < age 13 at random assignment				
Exp. versus control	1,339.8**	1,624.0**	1,298.9**	3,476.8**
	(671.3)	(662.4)	(636.9)	(1,418.2)
Sec. 8 versus control	687.4	1,109.3	908.6	1,723.2
	(698.7)	(676.1)	(655.8)	(1051.5)
Observations	8,420	8,420	8,420	8,420
Control group mean	9,548.6	11,270.3	11,270.3	11,270.3

TOT
=
ITT / First stage
=
$1,624/.467
=
$3,476.8

Chetty, Hendren, Katz (2016): The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment

Table 3—Impacts of MTO on Children's Income in Adulthood

	W-2 earnings ($) 2008–2012 ITT (1)	Individual earnings 2008–2012 ($) ITT (2)	ITT w/ controls (3)	TOT (4)
Panel A. Children < age 13 at random assignment				
Exp. versus control	1,339.8**	1,624.0**	1,298.9**	3,476.8**
	(671.3)	(662.4)	(636.9)	(1,418.2)
Sec. 8 versus control	687.4	1,109.3	908.6	1,723.2
	(698.7)	(676.1)	(655.8)	(1051.5)
Observations	8,420	8,420	8,420	8,420
Control group mean	9,548.6	11,270.3	11,270.3	11,270.3
Panel B. Children age 13–18 at random assignment				
Exp. versus control	−761.2	−966.9	−879.5	−2,426.7
	(870.6)	(854.3)	(817.3)	(2,154.4)
Sec. 8 versus control	−1,048.9	−1,132.8	−1,136.9	−2,051.1
	(932.5)	(922.3)	(866.6)	(1,673.7)
Observations	11,623	11,623	11,623	11,623
Control group mean	13,897.1	15,881.5	15,881.5	15,881.5

Chetty, Hendren, Katz (2016): The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment

Table 3—Impacts of MTO on Children's Income in Adulthood

	W-2 earnings ($) 2008–2012 ITT	Individual earnings 2008–2012 ($)			Individual earnings ($)		Employed (%) 2008–2012 ITT	Hhold. inc. ($) 2008–2012 ITT	Inc. growth ($) 2008–2012 ITT
		ITT	ITT w/ controls	TOT	Age 26 ITT	2012 ITT			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Children < age 13 at random assignment									
Exp. versus control	1,339.8**	1,624.0**	1,298.9**	3,476.8**	1,751.4*	1,443.8**	1.824	2,231.1***	1,309.4**
	(671.3)	(662.4)	(636.9)	(1,418.2)	(917.4)	(665.8)	(2.083)	(771.3)	(518.5)
Sec. 8 versus control	687.4	1,109.3	908.6	1,723.2	551.5	1,157.7*	1.352	1,452.4**	800.2
	(698.7)	(676.1)	(655.8)	(1051.5)	(888.1)	(690.1)	(2.294)	(735.5)	(517.0)
Observations	8,420	8,420	8,420	8,420	1,625	2,922	8,420	8,420	8,420
Control group mean	9,548.6	11,270.3	11,270.3	11,270.3	11,398.3	11,302.9	61.8	12,702.4	4,002.2
Panel B. Children age 13–18 at random assignment									
Exp. versus control	−761.2	−966.9	−879.5	−2,426.7	−539.0	−969.2	−2.173	−1,519.8	−693.6
	(870.6)	(854.3)	(817.3)	(2,154.4)	(795.4)	(1,122.2)	(2.140)	(11,02.2)	(571.6)
Sec. 8 versus control	−1,048.9	−1,132.8	−1,136.9	−2,051.1	−15.11	−869.0	−1.329	−936.7	−885.3
	(932.5)	(922.3)	(866.6)	(1,673.7)	(845.9)	(1213.3)	(2.275)	(11,85.9)	(625.2)
Observations	11,623	11,623	11,623	11,623	2,331	2,331	11,623	11,623	11,623
Control group mean	13,897.1	15,881.5	15,881.5	15,881.5	13,968.9	16,602.0	63.6	19,169.1	4,128.1

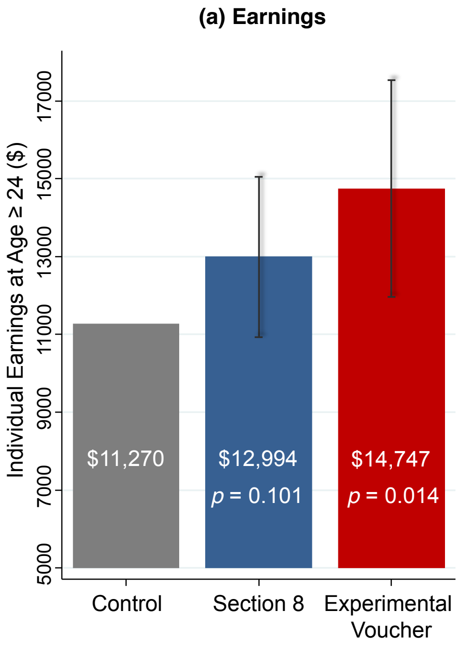
Chetty, Hendren, Katz (2016): The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment

Table 3—Impacts of MTO on Children's Income in Adulthood

	W-2 earnings ($) 2008–2012 ITT (1)	Individual earnings 2008–2012 ($) ITT (2)	ITT w/ controls (3)	TOT (4)	Individual earnings ($) Age 26 ITT (5)	2012 ITT (6)	Employed (%) 2008–2012 ITT (7)	Hhold. inc. ($) 2008–2012 ITT (8)	Inc. growth ($) 2008–2012 ITT (9)
Panel A. Children < age 13 at random assignment									
Exp. versus control	1,339.8**	1,624.0**	1,298.9**	3,476.8**	1,751.4*	1,443.8**	1.824	2,231.1***	1,309.4**
	(671.3)	(662.4)	(636.9)	(1,418.2)	(917.4)	(665.8)	(2.083)	(771.3)	(518.5)
Sec. 8 versus control	687.4	1,109.3	908.6	1,723.2	551.5	1,157.7*	1.352	1,452.4**	800.2
	(698.7)	(676.1)	(655.8)	(1051.5)	(888.1)	(690.1)	(2.294)	(735.5)	(517.0)
Observations	8,420	8,420	8,420	8,420	1,625	2,922	8,420	8,420	8,420
Control group mean	9,548.6	11,270.3	11,270.3	11,270.3	11,398.3	11,302.9	61.8	12,702.4	4,002.2
Panel B. Children age 13–18 at random assignment									
Exp. versus control	−761.2	−966.9	−879.5	−2,426.7	−539.0	−969.2	−2.173	−1,519.8	−693.6
	(870.6)	(854.3)	(817.3)	(2,154.4)	(795.4)	(1,122.2)	(2.140)	(11,02.2)	(571.6)
Sec. 8 versus control	−1,048.9	−1,132.8	−1,136.9	−2,051.1	−15.11	−869.0	−1.329	−936.7	−885.3
	(932.5)	(922.3)	(866.6)	(1,673.7)	(845.9)	(1213.3)	(2.275)	(11,85.9)	(625.2)
Observations	11,623	11,623	11,623	11,623	2,331	2,331	11,623	11,623	11,623
Control group mean	13,897.1	15,881.5	15,881.5	15,881.5	13,968.9	16,602.0	63.6	19,169.1	4,128.1

Chetty, Hendren, Katz (2016): The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment

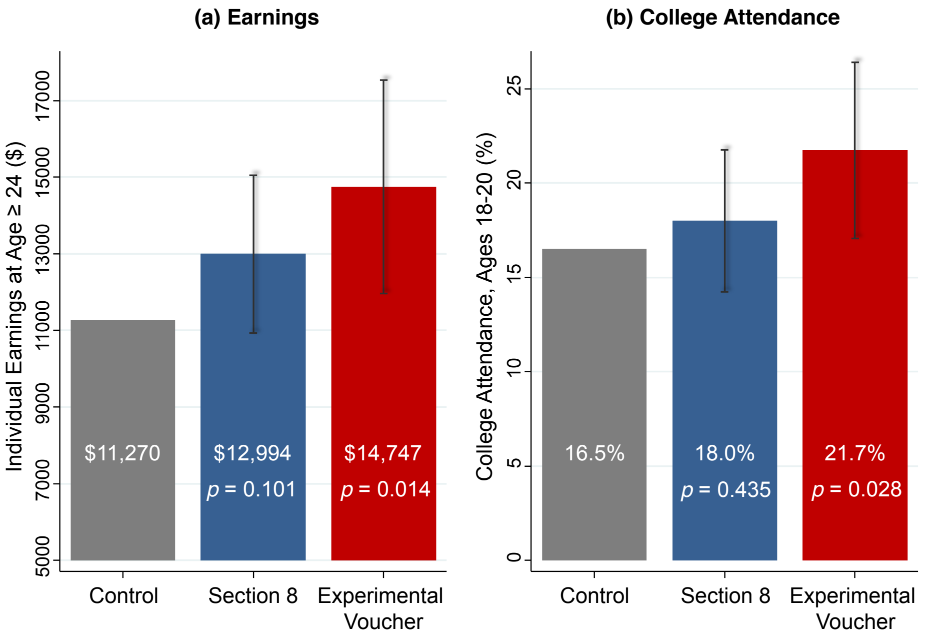
Impacts of MTO on Children Below Age 13 at Random Assignment



95% confidence intervals =
[ToT-1.96*SE, ToT+1.96*SE] =
[$698, $6,255]
(for the experimental group)

Average income for the experimental group = baseline + ToT = $11,270 + $3,477 = $14,747

Impacts of MTO on Children Below Age 13 at Random Assignment



Take-aways from the MTO experiment

- Strong evidence on the existence of neighborhood effects
    - might seem obvious, but hard evidence on them is scarce

Take-aways from the MTO experiment

- Strong evidence on the existence of neighborhood effects
  - might seem obvious, but hard evidence on them is scarce
- Putting the effects into a context
  - the average income of the participants remains far below average (even though it is much higher in comparison to the control group)
  - external validity: would the effects be similar also in other contexts?
    - you'll discuss these points in more depth with Tuukka

Take-aways from the MTO experiment

- Strong evidence on the existence of neighborhood effects
  - might seem obvious, but hard evidence on them is scarce
- Putting the effects into a context
  - the average income of the participants remains far below average (even though it is much higher in comparison to the control group)
  - external validity: would the effects be similar also in other contexts?
    - you'll discuss these points in more depth with Tuukka
- Methodological lesson: how to deal with partial compliance
  - manipulation of the *likelihood* of being treated can take us a long way
  - but: important to think about who the *compliers* are

- Some folks are *really* excited about randomization

Limits of RCTs

- Some folks are *really* excited about randomization, typically for a good reason: RCTs are often the best way the evaluate the impact of "treatments"
  - simple and transparent
    $\rightarrow$ everyone can understand the results
  - requires less (untestable) assumptions than the alternative approaches

Limits of RCTs

Introduction

The parachute is used in recreational, voluntary sector, and military settings to reduce the risk of orthopaedic, head, and soft tissue injury after gravitational challenge, typically in the context of jumping from an aircraft. The perception that parachutes are a successful intervention is based largely on anecdotal evidence. Observational data have shown that their use is associated with morbidity and mortality, due to both failure of the intervention[1 2] and iatrogenic complications.[3] In addition, "natural history" studies of free fall indicate that failure to take or deploy a parachute does not inevitably result in an adverse outcome.[4] We therefore undertook a systematic review of randomised controlled trials of parachutes.

Source: Smith GC, Pell JP. (2003): Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. BMJ 327: 1459-61

- Some folks are *really* excited about randomization, typically for a good reason: RCTs are often the best way the evaluate the impact of "treatments"
  - simple and transparent
    $\rightarrow$ everyone can understand the results
  - requires less (untestable) assumptions than the alternative approaches
- So, why don't we always run an RCT?
  - ethical and practical limitations
  - fundamental limitations

Ethical and practical limitations of RCTs

- Experiments should not knowingly harm anyone
  - but we still need to understand the effect of potentially harmful things

Ethical and practical limitations of RCTs

- Experiments should not knowingly harm anyone
  - but we still need to understand the effect of potentially harmful things
- Meaningful experiments are sometimes very expensive
  - on the other hand, policy and business mistakes can also be very costly
    $\rightarrow$ even large investments in experimentation can be justified

Ethical and practical limitations of RCTs

- Experiments should not knowingly harm anyone
  - but we still need to understand the effect of potentially harmful things
- Meaningful experiments are sometimes very expensive
  - on the other hand, policy and business mistakes can also be very costly
    $\rightarrow$ even large investments in experimentation can be justified
- The relevant time horizon may be very long
  - sometimes many decades!

Ethical and practical limitations of RCTs

- Experiments should not knowingly harm anyone
  - but we still need to understand the effect of potentially harmful things
- Meaningful experiments are sometimes very expensive
  - on the other hand, policy and business mistakes can also be very costly
    $\rightarrow$ even large investments in experimentation can be justified
- The relevant time horizon may be very long
  - sometimes many decades!
- Hawthorne and John Henry Effects
  - the evaluation itself may push people to change their behavior
  - likely less of a problem with administrative data and long follow-up periods (subjects not reminded about being evaluated)

Fundamental limitations of RCTs

- Spillovers / general equilibrium (GE) effects
  - the treatment also affects the control group $\rightarrow$ cannot use the control group to infer what would have happened without the treatment
  - the GE effects may be the main value of some treatments
    - RCTs never capture economy-wide GE effects
    - some examples of measuring more limited spillovers with RCTs exist

Fundamental limitations of RCTs

- Spillovers / general equilibrium (GE) effects
  - the treatment also affects the control group $\rightarrow$ cannot use the control group to infer what would have happened without the treatment
  - the GE effects may be the main value of some treatments
    - RCTs never capture economy-wide GE effects
    - some examples of measuring more limited spillovers with RCTs exist
- Scarcity of potential observations
  - some treatments affect entire countries or even the whole world
  - we'll never have experimental designs for these treatments

Summary

- Methodological take-away: how to deal with partial compliance
  - manipulation of the *likelihood* of being treated can take us a long way
  - but: important to think about who the *compliers* are
- RCTs are a powerful tool, but they also have important limitations
  - alternative 1: quasi-experimental approaches (rest of this course)
  - alternative 2: "structural" methods