The Economics of European Regions: Theory, Empirics, and Policy

Dipartimento di Economia e Management

Co-funded by the Erasmus+ Programme of the European Union

Project funded by

European Commission Erasmus + Programme –Jean Monnet Action

Project number 553280-EPP-1-2015-1-IT-EPPJMO-MODULE

Davide Fiaschi Angela Parenti[1]

November 12, 2019

[1] davide.fiaschi@unipi.it, and aparenti@ec.unipi.it.

The selection problem: an example

Let consider the example in Angrist and Pischke (2009).

- Suppose we want to answer the question "Do hospitals make people healthier?"
- Suppose we are studying a poor elderly population that uses hospital emergency rooms for primary care.
- Some of these patients are admitted to the hospital.
- This sort of care might be not very effective: exposure to other sick patients by those who are vulnerable can have a negative impact.
- However, the answer to the hospital effectiveness seems likely to be yes.

The selection problem: an example (cont.)

To answer the question we need to compare the health status of those who have been to the hospital with the health of those who have not.

Consider the data from National Health Interview Survey (NHIS) 2005, and in particular the questions:

- "During the past 12 months, was the respondent a patient in a hospital overnight?"
- "Would you say your health in general is excellent (5), very good (4), good (3), fair (2), poor (1)?"

Group	Sample Size	Mean Health Status	Std. Error
Hospital	7,774	3.21	0.014
No hospital	90,049	3.93	0.003

The difference in means is 3.93-3.21=0.72 (highly significant)

The selection problem: an example (cont.)

From previous evidence we could conclude that going to hospital makes people sicker.

Problems:

- people who go to the hospital are probably less healthy to begin with;
- even after hospitalization people who have sought medical care are *on average* not as healthy as those who were never hospitalized in the first place

The selection problem: an example (cont.)

- Population of individuals, indexed by $i = 1, ..., N$.
- The treatment indicator W_i takes on the values 0 (no hospital) and 1 (hospital).
- The outcome of interest, Y_i, is a measure of health status.
- For each individual we have two potential outcomes:

$$\begin{cases} Y_i(0) & \text{if } W_i = 0 \\ Y_i(1) & \text{if } W_i = 1 \end{cases}$$

- So $Y_i(0)$ is the health status of an individual had he not gone to the hospital, while $Y_i(1)$ is the individual's health status if he goes.

$\Rightarrow$ the *causal effect* of going to the hospital for individual i is given by $Y_i(1) - Y_i(0)$

The selection problem: an example (cont.)

- For individual i, the realized (and possibly observed) outcome Y_i^{obs} is given by:

$$Y_i^{obs} = Y_i(W_i) = \begin{cases} Y_i(0) & \text{if } W_i = 0 \\ Y_i(1) & \text{if } W_i = 1 \end{cases}$$

- The realized outcome can be rewritten as:

$$\begin{aligned} Y_i^{obs} &= Y_i(0)(1 - W_i) + Y_i(1)W_i \\ &= Y_i(0) + (Y_i(1) - Y_i(0))\, W_i. \end{aligned}$$

The selection problem: an example (cont.)

- *The fundamental problem of causal inference*: it is impossible to observe for the same i both potential outcomes $Y_i(1)$, $Y_i(0)$
$\Rightarrow$ the unit-level causal effect is well defined but cannot be estimated.
- The statistical solution replaces the impossible-to-observe individual causal effect with the possible-to-estimate **average treatment effect**, i.e. the effect of treatment on a random individual:

$$E\left[Y_i(1) - Y_i(0)\right] = E\left[Y_i(1)\right] - E\left[Y_i(0)\right].$$

- However, we cannot observe the outcome in both counterfactual situations.

The selection problem: an example (cont.)

- Alternatively, we can estimate the **average treatment effect on treated**, i.e., the effect of treatment on those who are actually treated:

$$E\left[Y_i(1) - Y_i(0) | W_i = 1\right] = E\left[Y_i(1) | W_i = 1\right] - E\left[Y_i(0) | W_i = 1\right].$$

- Even in this case, we cannot observe the outcome in both counterfactual situations.
- To estimate the average causal effect we have to rely on the **observed difference** in the average health of those who were and were not hospitalized:

$$E\left[Y_i^{obs} | W_i = 1\right] - E\left[Y_i^{obs} | W_i = 0\right]$$

The selection problem: an example (cont.)

What are we actually measuring if we compare these averages?

$$\underbrace{E\left[Y_i^{obs}|W_i=1\right]-E\left[Y_i^{obs}|W_i=0\right]}_{\textbf{Observed difference in average health}} = \text{substitute } Y_i^{obs}=Y_i(0)+(Y_i(1)-Y_i(0))W_i$$

$$= E\left[Y_i(1)|W_i=1\right]-E\left[Y_i(0)|W_i=0\right] =$$

add and subtract $E[Y_i(0)|W_i=1]$

$$= \underbrace{E\left[Y_i(1)|W_i=1\right]-E\left[Y_i(0)|W_i=1\right]}_{\textbf{Average treatment effect on the treated}} +$$

$$+ \underbrace{E\left[Y_i(0)|W_i=1\right]-E\left[Y_i(0)|W_i=0\right]}_{\textbf{Selection bias}}$$

The selection problem: an example (cont.)

Therefore, if we compare the observed average health we actually measure:

- the **average treatment effect on the treated**, $E[Y_i(1)|W_i = 1] - E[Y_i(0)|W_i = 1]$, which captures the *average causal effect of the hospitalization on those who were hospitalize* $\rightarrow$ the average difference between the health of the hospitalized and what would have happened to *them* had they not been hospitalized;
- the **selection bias**, $E[Y_i(0)|W_i = 1] - E[Y_i(0)|W_i = 0]$, which captures the average difference in the health without hospitalization between those who were and were not hospitalized.

In the example, because the sick are more likely that the healthy to seek treatment, those who were hospitalized have worse value of $Y_i(0) \Rightarrow$ negative selection bias!

Random assignment solves the problem

Random assignment of W_i solves the selection problems given that it makes W_i **independent** of potential outcomes.

$$\begin{aligned}
E\left[Y_i^{obs}|W_i=1\right]-E\left[Y_i^{obs}|W_i=0\right] &= E\left[Y_i(1)|W_i=1\right]-E\left[Y_i(0)|W_i=0\right]= \\
& \quad \text{given } {\scriptstyle W_i \perp Y_i(0)} \\
&= E\left[Y_i(1)|W_i=1\right]-E\left[Y_i(0)|W_i=1\right]= \\
&= E\left[Y_i(1)-Y_i(0)|W_i=1\right]= \\
& \quad \text{given } {\scriptstyle W_i \perp (Y_i(1)-Y_i(0))} \\
&= E\left[Y_i(1)-Y_i(0)\right]
\end{aligned}$$

• Under perfect randomization the treatment and control groups are statistically identical to the entire population.

• The effect of randomly assigned hospitalization on the hospitalized (treated) is the same as the effect of hospitalization on randomly chosen patient (population).

Example of Randomized Experiment: Tennessee Project STAR (Krueger, 1999)

- The Tennessee STAR experiment was designed to estimate the effect of smaller classes on student achievement.
- Many studies of education using non-experimental data suggest that there is little or no link between class size and student learning.
- The observed relation between class size and student achievement can be subject to selection problem: weaker students are often deliberately grouped into smaller classes.
- A *randomized* trial can overcome this problem by ensuring that the students assigned to classes of different size are otherwise comparable $\Rightarrow$ Krueger (1999) re-analyses econometrically the STAR data.

Example of Randomized Experiment: Tennessee Project STAR (cont.)

- The Tennessee Student/Teacher Achievement Ratio (STAR) project was run in the 1980s.
- The study ran for four years and involved 11,600 children.
- The 11,600 students and their teachers were **randomly assigned** to one of three groups (treatments):
  1. Small classes (13-17 students).
  2. Regular classes (22-25 students).
  3. Regular classes (22-25 students) with a full time teacher's aide.
- After the assignment, the design called for students to remain in the same class type for four years.
- Randomization occurred *within* schools (with at least three classes with each grade).

Regression Analysis of Experiments

- Because randomization eliminates selection bias, one could simply compare mean outcomes of treatment and control group to obtain the causal effect of the treatment.
- Nonetheless, it is often useful to analyze experimental data with regression analysis.
- Suppose that the treatment effect, τ, is *constant* (i.e. the treatment affects everyone by the same magnitude), so that for each drawn i, $\tau = Y_i(1) - Y_i(0)$.
- Further, assume that $Y_i(0) = \alpha + \epsilon_i$, where $\epsilon_i = Y_i(0) - E\left[Y_i(0)\right]$ is the *residual* capturing the *unobservables* affecting the response in the *absence* of treatment.

Regression Analysis of Experiments (cont.)

- Then, given that the observed outcome is defined as:

$$Y_i^{obs} = Y_i(0) + (Y_i(1) - Y_i(0))\, W_i$$

we can rewrite it as:

$$Y_i^{obs} = \underbrace{\alpha}_{E[Y_i(0)]} + \underbrace{\tau}_{Y_i(1)-Y_i(0)} W_i + \underbrace{\epsilon_i}_{Y_i(0)-E[Y_i(0)]} \tag{1}$$

- Regression (1) could therefore be estimated to obtain the causal effect of treatment W.

Regression Analysis of Experiments (cont.)

- The conditional expectations of (1) with respect to the two treatment status $W_i = 1$ and $W_i = 0$ are:

$$\begin{aligned} E\left[Y_i^{obs}|W_i = 1\right] &= \alpha + \tau + E\left[\epsilon_i|W_i = 1\right] \\ E\left[Y_i^{obs}|W_i = 0\right] &= \alpha + E\left[\epsilon_i|W_i = 0\right] \end{aligned}$$

- So that, their difference is equal to:

$$E\left[Y_i^{obs}|W_i = 1\right] - E\left[Y_i^{obs}|W_i = 0\right] = \underbrace{\tau}_{\text{Treatment effect}} + \underbrace{E\left[\epsilon_i|W_i = 1\right] - E\left[\epsilon_i|W_i = 0\right]}_{\text{Selection bias}}$$

- Under randomized experiment $W_i \perp Y_i(0)$.
- In regression model this is equivalent to $W_i \perp \epsilon_i \Rightarrow$ no selection bias!

Regression Analysis of Experiments (cont.)

• To evaluate experimental data one may want to add additional controls (pre-treatment variables) in the regression:

1. Covariates commonly serve to make estimates more precise by explaining some of the variation in outcomes.
   Including controls thus reduces residual variance and therefore lowers the standard errors of the regression estimates.
2. Allow to evaluate causal effect of the treatment on subgroups.
3. Allow a *conditional random assignment* on observable.
   The independence of assignment mechanism and potential outcomes is more plausible (in the STAR example at the school level).

$\Rightarrow$ Eq. (1) becomes $Y_i^{obs} = \alpha + \tau W_i + \beta' X_i + \epsilon_i$.

Under randomization $W_i \perp Y_i(0)|X_i$, so that $W_i \perp \epsilon_i|X_i$.

Do the treatment and control groups "looked similar"?

• Does the randomization successfully balanced subject's characteristics across the different treatment groups?
$\Rightarrow$ Unfortunately STAR experiment does not provide pre-treatment test scores.

• Nonetheless, if the students were successfully randomly assigned between class types, one would expect those assigned to small- and regular-size classes to look *similar* along other measurable dimensions.

Do the treatment and control groups "looked similar"?

Variable	Small	Regular	Regular/Aide	Joint P-Value[a]
A. Students who entered STAR in kindergarten[b]				
1. Free lunch[c]	.47	.48	.50	.09
2. White/Asian	.68	.67	.66	.26
3. Age in 1985	5.44	5.43	5.42	.32
4. Attrition rate[d]	.49	.52	.53	.02
5. Class size in kindergarten	15.1	22.4	22.8	.00
6. Percentile score in kindergarten	54.7	49.9	50.0	.00
B. Students who entered STAR in first grade				
1. Free lunch	.59	.62	.61	.52
2. White/Asian	.62	.56	.64	.00
3. Age in 1985	5.78	5.86	5.88	.03
4. Attrition rate	.53	.51	.47	.07
5. Class size in first grade	15.9	22.7	23.5	.00
6. Percentile score in first grade	49.2	42.6	47.7	.00
C. Students who entered STAR in second grade				
1. Free lunch	.66	.63	.66	.60
2. White/Asian	.53	.54	.44	.00
3. Age in 1985	5.94	6.00	6.03	.66
4. Attrition rate	.37	.34	.35	.58
5. Class size in third grade	15.5	23.7	23.6	.01
6. Percentile score in second grade	46.4	45.3	41.7	.01
D. Students who entered STAR in third grade				
1. Free lunch	.60	.64	.69	.04
2. White/Asian	.66	.57	.55	.00
3. Age in 1985	5.95	5.92	5.99	.39
4. Attrition rate	NA	NA	NA	NA
5. Class size in third grade	16.0	24.1	24.4	.01
6. Percentile score in third grade	47.6	44.2	41.3	.01

- Differences in these characteristics *not* significant.
- Class sizes are significantly lower in small classrooms $\Rightarrow$ the experiment succeeded in creating the desired variation.
- Some significant differences for students who entered STAR in first, second or third grade $\Rightarrow$ because random assignment was *only valid within schools*, these differences suggest the importance of controlling for **school effects**.

Conditioning on school effects: school dummies

TABLE II
P-VALUES FOR TESTS OF WITHIN-SCHOOL DIFFERENCES AMONG SMALL, REGULAR, AND REGULAR/AIDE CLASSES

Variable	Grade entered STAR program			
	K	1	2	3
1. Free lunch	.46	.29	.58	.18
2. White/Asian	.66	.28	.15	.21
3. Age	.38	.12	.48	.40
4. Attrition rate	.01	.07	.58	NA
5. Actual class size	.00	.00	.00	.00
6. Percentile score	.00	.00	.46	.00

Each *p*-value is for an *F*-test of the null hypothesis that assignment to a small, regular, or regular/aide class has no effect on the outcome variable in that grade, conditional on school of attendance.

All rows except 4 pertain to the first grade in which the student entered the STAR program. The attrition rate in row 4 measures whether the student ever left the sample after initially being observed.

None of the three background variables displays a statistically significant association with class-type assignment at the 10% level
$\Rightarrow$ random assignment produced relatively similar groups in each class size, on average.

Regression in Krueger (1999)

Krueger estimates the following econometric model:

$$Y_{ics} = \beta_0 + \beta_1 SMALL_{cs} + \beta_2 Reg/A_{cs} + \beta_3 X_{ics} + \alpha_s + \epsilon_{isc}$$

where:

- Y_{ics} = average percentile score on the SAT test of student i in class c at school s,
- $SMALL_{cs}$ = dummy variable indicating whether the student was assigned to a small class,
- Reg/A_{cs} = dummy variable indicating whether the student was assigned to a regular-size class with an aide,
- X_{ics} = a vector of observed student and teacher covariates (e.g., gender),
- α_s = School FE (independence between class-size assignment and other variables is only valid within schools),
- ϵ_{ics} = error term ($\epsilon_{ics} = \mu_{cs} + \epsilon_{ics} \rightarrow$ clustered SE at class-school level).

Regression Results Kindergarten

Explanatory variable	OLS: actual class size (1)	(2)	(3)	(4)
			A. Kindergarten	
Small class	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
Regular/aide class	.12	.29	.53	.31
	(2.23)	(1.13)	(1.09)	(1.07)
White/Asian (1 = yes	—	—	8.35	8.44
			(1.35)	(1.36)
Girl (1 = yes)	—	—	4.48	4.39
			(.63)	(.63)
Free lunch (1 = yes)	—	—	−13.15	−13.07
			(.77)	(.77)
White teacher	—	—	—	−.57
				(2.10)
Teacher experience	—	—	—	.26
				(.10)
Master's degree	—	—	—	−.51
				(1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.25	.31	.31

- Small classes score is about **5 pp higher** than regular-size classes.
- Students in regular/aide class perform **as those** in regular-size classes.
- If class size were truly randomly assigned, including additional exogenous variables would not significantly alter the coefficient on the class-size dummies.
- In fact, including covariates seems to have a **very modest effect** on the class-size coefficients conditional on school effects.
- The student characteristics add **considerable explanatory power**.
- The teacher characteristics have notably **weak explanatory power**.

Regression on 1st/2nd/3rd Grade

B. First grade

Small class	8.57 (1.97)	8.43 (1.21)	7.91 (1.17)	7.40 (1.18)
Regular/aide class	3.44 (2.05)	2.22 (1.00)	2.23 (0.98)	1.78 (0.98)
White/Asian (1 = yes)	—	—	6.97 (1.18)	6.97 (1.19)
Girl (1 = yes)	—	—	3.80 (.56)	3.85 (.56)
Free lunch (1 = yes)	—	—	−13.49 (.87)	−13.61 (.87)
White teacher	—	—	—	−4.28 (1.96)
Male teacher	—	—	—	11.82 (3.33)
Teacher experience	—	—	—	.05 (0.06)
Master's degree	—	—	—	.48 (1.07)
School fixed effects	No	Yes	Yes	Yes
R^2	.02	.24	.30	.30

C. Second grade

Small class	5.93 (1.97)	6.33 (1.29)	5.83 (1.23)	5.79 (1.23)
Regular/aide class	1.97 (2.05)	1.88 (1.10)	1.64 (1.07)	1.58 (1.06)
White/Asian (1 = yes)	—	—	6.35 (1.20)	6.36 (1.19)
Girl (1 = yes)	—	—	3.48 (.60)	3.45 (.60)
Free lunch (1 = yes)	—	—	−13.61 (.72)	−13.61 (.72)
White teacher	—	—	—	.39 (1.75)
Male teacher	—	—	—	1.32 (3.96)
Teacher experience	—	—	—	.10 (.06)
Master's degree	—	—	—	−1.06 (1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.22	.28	.28

D. Third grade

Small class	5.32 (1.91)	5.58 (1.22)	5.01 (1.19)	5.00 (1.19)
Regular/aide class	−.22 (1.95)	−.16 (1.12)	−.33 (1.11)	−.75 (1.07)
White/Asian (1 = yes)	—	—	6.12 (1.45)	6.11 (1.44)
Girl (1 = yes)	—	—	4.16 (.66)	4.16 (.65)
Free lunch (1 = yes)	—	—	−13.02 (.81)	−12.96 (.81)
White teacher	—	—	—	.64 (1.75)
Male teacher	—	—	—	−7.42 (2.80)
Teacher experience	—	—	—	.04 (.06)
Master's degree	—	—	—	1.10 (1.15)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.17	.22	.23

The gap in average performance is about **8 pp** in first grade, and **5-6 pp** in second and third grade.

Problem 1: Attrition

A common problem in randomized experiments

- Sample **attrition** is a feature of longitudinal or panel data in which individual observations drop out from the study over time.
- In this case, half of students who were present in kindergarten were missing in at least one subsequent year.
- If attrition is random and affects the treatment and control groups in the same way the estimates would remain unbiased.
- Here the attrition is likely to be **non-random**: especially good students from large classes may have enrolled in private school upon learning their class-type assignments
  $\Rightarrow$ (positive) selection bias problem!

Problem 1: Attrition (cont.)

A common problem in randomized experiments

- If the students originally assigned to regular classes who left the sample had higher test scores, on average, than students assigned to small classes who also left the sample
  $\Rightarrow$ the small class effects will be **biased upwards**.
- Krueger addresses this concern by imputing test scores for students who exited the sample.
- He assignes the student's most recent test percentile to that student in years when the student was absent from the sample.
- Finally, he re-estimates the model including students with imputed test scores.

Regression results imputing test scores

Krueger (1999). Table VI. Exploration of effect of attrition.

Grade	Actual test data: Coefficient on small class dum.	Actual test data: Sample size	Actual and imputed test data: Coefficient on small class dum.	Actual and imputed test data: Sample size
K	5.32 (.76)	5900	5.32 (.76)	5900
1	6.95 (.74)	6632	6.30 (.68)	8328
2	5.59 (.76)	6282	5.64 (.65)	9773
3	5.58 (.79)	6339	5.49 (.63)	10919

The reported coefficient on small class dummy is relative to regular classes. Standard errors are in parentheses.

Non-random attrition *does not appear* to bias the estimated class size effects.

Problem 2: Students switched between classes after random assignment

- Approximately 10% of students switched between small and regular classes between grades (primarily because of behavioural problems or parental complaints).
- Furthermore, some students and their families naturally relocate during the school year.
- Students moved between treatment and control groups.
- These non-random transitions could compromise the experimental results:
  $\Rightarrow$ if the movement between class types was associated with student characteristics (e.g., students with stronger academic backgrounds more likely to move into small classes), these transitions would bias a simple comparison of outcomes across class types.

Transitions between Grade 1-Grade 2 and Grade 2-Grade 3

Krueger (1999).Table IV. Transitions between class-size in adjacent grades.

B. First grade to second grade

	Second grade			
First grade	Small	Regular	Reg/aide	All
Small	1435	23	24	1482
Regular	152	1498	202	1852
Aide	40	115	1560	1715
All	1627	1636	1786	5049

C. Second grade to third grade

	Third grade			
Second grade	Small	Regular	Reg/aide	All
Small	1564	37	35	1636
Regular	167	1485	152	1804
Aide	40	76	1857	1973
All	1771	1598	2044	5413

Number of students in each type class.

Problem 2: Students switched between classes after random assignment (cont.)

To address this potential problem, Krueger use *initial assignment* (here initial assignment to small or regular classes) as an **instrument** for actual assignment.

$$\begin{aligned} CS_{ics} &= \pi_0 + \pi_1 S_{ios} + \pi_2 R_{ios} + \pi_3 X_{ics} + \delta_s + \tau_{ics} \quad (2) \\ Y_{ics} &= \beta_0 + \beta_1 CS_{ics} + \beta_2 X_{ics} + \alpha_s + \epsilon_{isc} \quad (3) \end{aligned}$$

where:

- CS_{ics} = actual number of students in the class,
- S_{ios} = dummy variable indicating assignment to a small class the first year the student is observed in the experiment,
- R_{ios} = dummy variable indicating assignment to a regular class the first year the student is observed in the experiment.

Problem 2: Students switched between classes after random assignment (cont.)

- In the test score equation (2) only variation in class size due to *initial assignment* to a regular or small class is used to provide variation in actual class size.
- Due to the random assignment of initial class type, the instrumental variable should be uncorrelated with ϵ_{isc}.
- If attending a small class has a beneficial effect on students' test scores, β_1 would be negative: $\Rightarrow$ the smaller the class size, the higher the average test score!

Problem 2: Students switched between classes after random assignment (cont.)

Explanatory variable	OLS: actual class size				Reduced form: initial class size			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
B. First grade								
Small class	8.57	8.43	7.91	7.40	7.54	7.17	6.79	6.37
	(1.97)	(1.21)	(1.17)	(1.18)	(1.76)	(1.14)	(1.10)	(1.11)
Regular/aide class	3.44	2.22	2.23	1.78	1.92	1.69	1.64	1.48
	(2.05)	(1.00)	(0.98)	(0.98)	(1.12)	(0.80)	(0.76)	(0.76)
White/Asian (1 = yes)	—	—	6.97	6.97	—	—	6.86	6.85
			(1.18)	(1.19)			(1.18)	(1.18)
Girl (1 = yes)	—	—	3.80	3.85	—	—	3.76	3.82
			(.56)	(.56)			(.56)	(.56)
Free lunch (1 = yes)	—	—	−13.49	−13.61	—	—	−13.65	−13.77
			(.87)	(.87)			(.88)	(.87)
White teacher	—	—	—	−4.28	—	—	—	−4.40
				(1.96)				(1.97)
Male teacher	—	—	—	11.82	—	—	—	13.06
				(3.33)				(3.38)
Teacher experience	—	—	—	.05	—	—	—	.06
				(0.06)				(.06)
Master's degree	—	—	—	.48	—	—	—	.63
				(1.07)				(1.09)
School fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
R^2	.02	.24	.30	.30	.01	.23	.29	.30

- Krueger reports reduced form results where he uses initial assignment and not current status as explanatory variable.
- In Kindergarten OLS and reduced form are the same because students remained in their initial class for at least one year.
- From grade 1 onwards OLS (column 1-4) and reduced form (columns 5-8) are different.

Other potential problems when running experiments

1. **Randomization bias**:
occurs when random assignment causes the type of persons participating in a program to differ from the type that would participate in the program in the absence of experiment.

Doolittle and Traeger, 1990, p. 121: "... Expanded recruitment efforts needed to generate the control group draw in additional applicants who are not identical to the people previously served."

The assumption of no randomization bias is unnecessary under the *alternative assumption* of **homogeneous treatment effect**:
the mean impact of treatment on participants is then the same for persons participating in the presence and in the absence of an experiment.

Other potential problems when running experiments: Heterogeneous treatment effect

Heterogeneous treatment effect: people selecting to take part in the randomized trial may have different returns compared to the population average.

	Boys	Girls
Small	4.18	1.28
	(1.11)	(1.13)
Cumulative years in small class	.60	.92
	(.56)	(.54)
Sample size	12,576	11,773

	Free lunch	Not on free lunch
Small	3.14	2.85
	(1.10)	(1.12)
Cumulative years in small class	.94	.55
	(.59)	(.51)
Sample size	12,064	12,285

	Black	White
Small	3.84	2.58
	(1.29)	(1.02)
Cumulative years in small class	1.04	.66
	(.68)	(.48)
Sample size	8,150	16,069

	Inner city	Metropolitan	Towns	Rural
Small	3.74	2.92	3.09	2.58
	(1.68)	(1.55)	(2.83)	(1.23)
Cumulative years in small class	1.71	.57	−1.35	1.03
	(.90)	(.83)	(1.50)	(.56)
Sample size	5,154	5,906	1,872	11,417

Other potential problems when running experiments: "Hawthorne" effects and "John Henry" effects

2. **"Hawthorne" effects:**
   occurs when people behave differently because they are part of an experiment.

   If these "Hawthorne" effects operate *differently* on treatment and control groups they may introduce biases.

   If people from the control group behave differently these effects are called **"John Henry" effects**.

"Hawthorne" and "John Henry" effects in Krueger (1999)

- "Hawthorne" effects: teachers in small classes responded to the fact that they were part of an experiment, rather than a true causal effect of small classes themselves.
- "John Henry" effects: teachers in regular classes provided greater than normal effort to demonstrate that they could overcome the bad luck of being assigned more students.

$\Rightarrow$ They could limit the *external validity* of the results of the STAR experiment.

Krueger examines the relationship between class size (number of students in the class) and student achievement (average SAT test score)*just* among students assigned to regular-size classes (pooled across all grade levels).
$\Rightarrow$ not much evidence of either Hawthorne or John Henry effects.

Other potential problems when running experiments: Substitution bias

3. **Substitution bias:**
   arises when control group members gain access to close *substitutes* for the treatment, like similar services offered by other providers or the same service offered under different funding arrangement.

   In the presence of substitution bias, control group outcomes no longer correspond to the untreated state.

   The mean difference in outcomes between the treatment and control groups no longer provides an estimate of the mean impact of treatment on the treated.

References

- Angrist, J. D., and Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- Heckman, J. J., and Smith, J. A. (1995). Assessing the case for social experiments. *The Journal of Economic Perspectives*, 9(2), 85-110.
- Imbens, G. W., and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 114(2), 497-532.