Applied Microeconometrics

Randomized Social Experiments

Derya Uysal

IHS, Vienna

MSc Course IHS/TU Wien

Spring 2014

Basic Idea of Social Experiments Approach

- A program is implemented as a field experiment.
- Individuals can volunteer for the program provided they satisfy certain eligibility criteria.
- The group of volunteers is randomly split into a group of individuals, who actually receive the treatment (treatment group) and a group of individuals for whom treatment is denied (control group).
- The difference in the mean outcome between treatment and control group estimates the average treatment effect on the treated Δ_{TT}.

Social Experiment without Randomization

Consider a regression of Y on D, where $D = 1$ indicates voluntary participation and $D = 0$ voluntary non-participation. This yields the naive comparison between treated and untreated (equivalent to nonexperimental design):

$$
\begin{aligned}
Y &= DY_1 + (1 - D)Y_0 \\
&= \mu_0 + (\mu_1 - \mu_0)D + \omega \\
&= \alpha + \beta * D + \omega
\end{aligned}
$$

$$
\text{where:} \quad \omega = \varepsilon_0 + (\varepsilon_1 - \varepsilon_0) * D
$$

$$
\alpha = \mu_0 \qquad \beta = \mu_1 - \mu_0 = \Delta_{ATE}
$$

Social Experiment without Randomization

Naive Group Comparison:

$$\begin{aligned}\Delta_n &= \mathrm{E}\left[\, Y \middle|\, D = 1\right] - \mathrm{E}\left[\, Y \middle|\, D = 0\right] \\ &= \beta + S_0\end{aligned}$$

where: $S_0 = \mathrm{E}\left[\varepsilon_0 \middle|\, D = 1\right] - \mathrm{E}\left[\varepsilon_0 \middle|\, D = 0\right]$ (Selection Bias)

Social Experiment without Randomization

Least Squares Estimation:

$$\hat{\beta}_{LS} = \frac{\sum_i (D_i - \bar{D})(Y_i - \bar{Y})}{\sum_i (D_i - \bar{D})^2} = \bar{Y}_1 - \bar{Y}_0$$

Since the observations are iid distributed, least squares estimates asymptotically:

$$\begin{aligned} \operatorname*{plim}_{n\to\infty} \hat{\beta}_{LS} &= \operatorname*{plim}_{n\to\infty} \bar{Y}_1 - \bar{Y}_0 \\ &= \mathrm{E}\left[Y \middle| D = 1 \right] - \mathrm{E}\left[Y \middle| D = 0 \right] \\ &\neq \beta \end{aligned}$$

Selection Bias

Recall the selection problem when comparing the mean outcomes for the treated and the untreated:

$$\underbrace{\mathrm{E}[Y|D=1]-\mathrm{E}[Y|D=0]}_{\text{Differences in means}}=\mathrm{E}[Y_1|D=1]-\mathrm{E}[Y_0|D=0]$$
$$=\underbrace{\mathrm{E}[Y_1-Y_0|D=1]}_{\text{ATT}}+\underbrace{\{\mathrm{E}[Y_0|D=1]-\mathrm{E}[Y_0|D=0]\}}_{\text{BIAS}}$$

- Random assignment of units to the treatment forces the selection bias to be zero
- The treatment and control group will tend to be similar along all characteristics (including Y_0)

Identification in Randomized Experiments

- Randomization implies:

  (Y_1, Y_0) independent of D, or $(Y_1, Y_0) \perp D$.

- We have that $\mathrm{E}[Y_0 | D = 1] = \mathrm{E}[Y_0 | D = 0]$ and therefore

  $$\Delta_{ATT} = \mathrm{E}[Y_1 - Y_0 | D = 1] = \mathrm{E}[Y | D = 1] - \mathrm{E}[Y | D = 0]$$

- Also, we have that

  $$\Delta_{ATE} = \mathrm{E}[Y_1 - Y_0] = \mathrm{E}[Y_1 - Y_0 | D = 1] = \mathrm{E}[Y | D = 1] - \mathrm{E}[Y | D = 0]$$

- As a results,

  $$\underbrace{\mathrm{E}[Y | D = 1] - \mathrm{E}[Y | D = 0]}_{\text{Differences in means}} = \Delta_{ATE} = \Delta_{ATT}$$

Identification in Randomized Experiments

- The identification result extends beyond average treatment effects.
- Given random assignment $(Y_1, Y_0) \perp D$:

$$
\begin{aligned}
F_{Y_0}(y) &= \Pr\left[Y_0 \leq y\right] = \Pr\left[Y_0 \leq y \,|D = 0\right] \\
&= \Pr\left[Y \leq y \,|D = 0\right]
\end{aligned}
$$

- Similarly,

$$F_{Y_1}(y) = \Pr\left[Y \leq y \,|D = 1\right]$$

- So effect of the treatment at any quantile, $Q_\theta(Y_1) - Q_\theta(Y_0)$ is identified.
- Randomization identifies the entire marginal distributions of Y_0 and Y_1
- Does not identify the quantiles of the effect: $Q_\theta(Y_1 - Y_0)$ (the difference of quantiles is not the quantile of the difference)

Estimation in Randomized Experiments

- Consider a randomized trial with N individuals. Suppose that the estimand of interest is ATE:

$$\Delta_{ATE} = \mathrm{E}[Y_1 - Y_0] = \mathrm{E}[Y | D = 1] - \mathrm{E}[Y | D = 0]$$

- Using the analogy principle, we construct an estimator:

$$\hat{\Delta}_{ATE} = \bar{Y}_1 - \bar{Y}_0$$

where

$$\bar{Y}_1 = \frac{\sum_i Y_i D_i}{\sum_i D_i} = \frac{1}{N_1} \sum_{i \in D_i = 1} Y_i;$$

$$\bar{Y}_0 = \frac{\sum_i Y_i (1 - D_i)}{\sum_i (1 - D_i)} = \frac{1}{N_0} \sum_{i \in D_i = 0} Y_i$$

with $N_1 = \sum_i D_i$ and $N_0 = N - N_1$

- $\hat{\Delta}_{ATE}$ is an unbiased and consistent estimator of Δ_{ATE}

Testing in Large Samples: Two Sample t-Test

- Notice that:

$$\frac{\hat{\Delta}_{ATE} - \Delta_{ATE}}{\sqrt{\frac{\hat{\sigma}_1^2}{N_1} + \frac{\sigma_0^2}{N_0}}} \xrightarrow{d} N(0,1)$$

where

$$\hat{\sigma}_1^2 = \frac{1}{N_1 - 1} \sum_{i \in D_i = 1} (Y_i - \bar{Y}_1)^2$$

and $\hat{\sigma}_0^2$ is analogously defined.

- In particular, let

$$t = \frac{\hat{\Delta}_{ATE}}{\sqrt{\frac{\hat{\sigma}_1^2}{N_1} + \frac{\sigma_0^2}{N_0}}}.$$

We reject the null hypothesis $H_0 : \Delta_{ATE} = 0$ against the alternative $H_1 : \Delta_{ATE} \neq 0$ at the 5% significance level if $|t| > 1.96$.

Testing in Small Samples: Fisher's Exact Test

- Test of differences in means with large N:

$$H_0 : \mathrm{E}[Y_1] = \mathrm{E}[Y_0], \quad H_1 : \mathrm{E}[Y_1] \neq \mathrm{E}[Y_0]$$

- Fisher's Exact Test with small N:

$$H_0 : Y_{1i} = Y_{0i}, \forall i = 1, \ldots, N, \quad H_1 : \exists i \text{such that} Y_{1i} \neq Y_{0i}$$

- Let Ω be the set of all possible randomization realizations.
- We only observe the outcomes, Y_i , for one realization of the experiment. We calculate $\hat{\Delta} = \bar{Y}_1 - \bar{Y}_0$.
- Under the sharp null hypothesis we can calculate the value that the difference of means would have taken under any other realization, $\hat{\Delta}(\omega)$, for $\omega \in \Omega$.

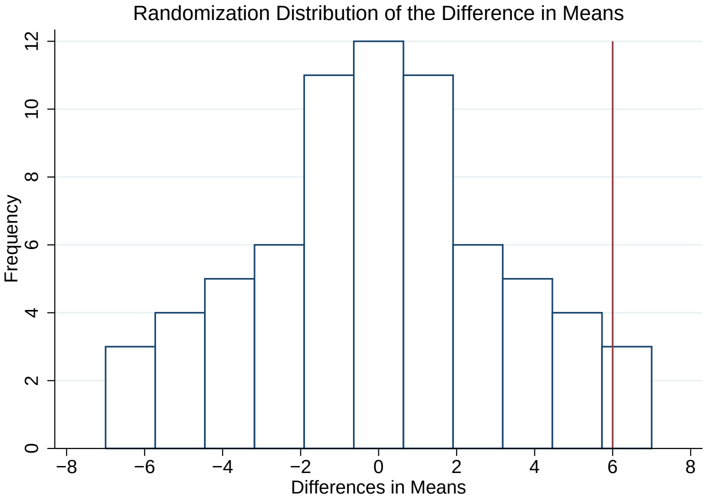
Testing in Small Samples: Fisher's Exact Test

Suppose that we assign 4 individuals out of 8 to the treatment:

Y_i	12	4	6	10	6	0	1	1	
D_i	1	1	1	1	0	0	0	0	$\hat{\Delta} = 6$
									$\hat{\Delta}(\omega)$
$\omega = 1$	1	1	1	1	0	0	0	0	6
$\omega = 2$	1	1	1	0	1	0	0	0	4
$\omega = 3$	1	1	1	0	0	1	0	0	1
$\omega = 4$	1	1	1	0	0	0	1	0	1.5
$\vdots$				$\dots$					$\vdots$
$\omega = 70$	0	0	0	0	1	1	1	1	-6

- The randomization distribution of $\hat{\Delta}$ (under the sharp null hypothesis) is $\Pr\left[\hat{\Delta} \leq z\right] = \frac{1}{70}\sum_{\omega\in\Omega} \mathbb{1}\{\hat{\Delta}(\omega) \leq z\}$
- Now, find $\bar{z} = \inf\{z : \Pr\left[|\hat{\Delta}| > z\right] \leq 0.05\}$
- Reject the null hypothesis, $H_0 : Y_{1i} - Y_{0i} = 0$ for all i, against the alternative hypothesis $H_1 : Y_{1i} - Y_{0i} \neq 0$ for some i, at the 5% significance level if $|\hat{\Delta}| > \bar{z}$

Testing in Small Samples: Fisher's Exact Test



$$\Pr\left[|\hat{\Delta}(\omega)| \geq 6\right] = 0.0857$$

Covariate Balance

- Randomization balances observed but also unobserved characteristics between treatment and control group
- Can check random assignment using so called "balance tests" (e.g., t-tests) to see if distributions of the observed covariates, X, are the same in the treatment and control groups
- Compute normalized differences for each covariate:

$$\Delta_X = \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{S_0^2 + S_1^2}}$$

  where S_d is the sample variance of X_i in the subsample with treatment $D_i = d$.
- Imbens and Rubin discuss rules-of-thumb. Normalized differences above about .25 should raise flags.
- X are pre-treatment variables that are measured prior to treatment assignment (i.e., at "baseline")

Experimental Design: Relative Sample Sizes for Fixed N

- Suppose that you have N experimental subjects and you have to decide how many will be in the treatment group and how many in the control group. We know that:

$$\bar{Y}_1 - \bar{Y}_0 \sim \left(\mu_1 - \mu_0, \frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0} \right).$$

- We want to choose N_1 and N_0, subject to $N_1 + N_0 = N$, to minimize the variance of the estimator of the average treatment effect.
- The variance of $\bar{Y}_1 - \bar{Y}_0$ is:

$$\mathrm{V}\left[\bar{Y}_1 - \bar{Y}_0 \right] = \frac{\sigma_1^2}{pN} + \frac{\sigma_0^2}{(1-p)N}$$

where $p = N_1/N$ is the proportion of treated in the sample.

Experimental Design: Relative Sample Sizes for Fixed N

- Find the value p^* that minimizes $\mathrm{V}\left[\bar{Y}_1 - \bar{Y}_0\right]$:

$$-\frac{\sigma_1^2}{p^{*2}N} + \frac{\sigma_0^2}{(1-p^*)^2 N} = 0$$

- Therefore:

$$\frac{(1-p^*)}{p^*} = \frac{\sigma_0}{\sigma_1},$$

and

$$p^* = \frac{\sigma_1}{\sigma_1 + \sigma_0} = \frac{1}{1 + \sigma_0/\sigma_1}$$

- A "rule of thumb" for the case $\sigma_0 \approx \sigma_1$ is $p^* = 0.5$
- For practical reasons it is sometimes better to choose unequal sample sizes (even if $\sigma_0 \approx \sigma_1$)

Experimental Design: Power Calculations to Choose N

- Recall that for a statistical test:
  - Type I error: Rejecting the null if the null is true.
  - Type II error: Not rejecting the null if the null is false.
- Size of a test is the probability of type I error, usually 0.05.
- Power of a test is one minus the probability of type II error, i.e. the probability of rejecting the null if the null is false.
- Statistical power increases with the sample size.
- But when is a sample “large enough”?
- We want to find N such that we will be able to detect an average treatment effect of size Δ or larger with high probability.

Experimental Design: Power Calculations to Choose N

- Assume a particular value, α, for $\mu_1 - \mu_0$.
- Let $\hat{\Delta} = \bar{Y}_1 - \bar{Y}_0$ and

$$\text{s.e.}(\hat{\Delta}) = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0}}.$$

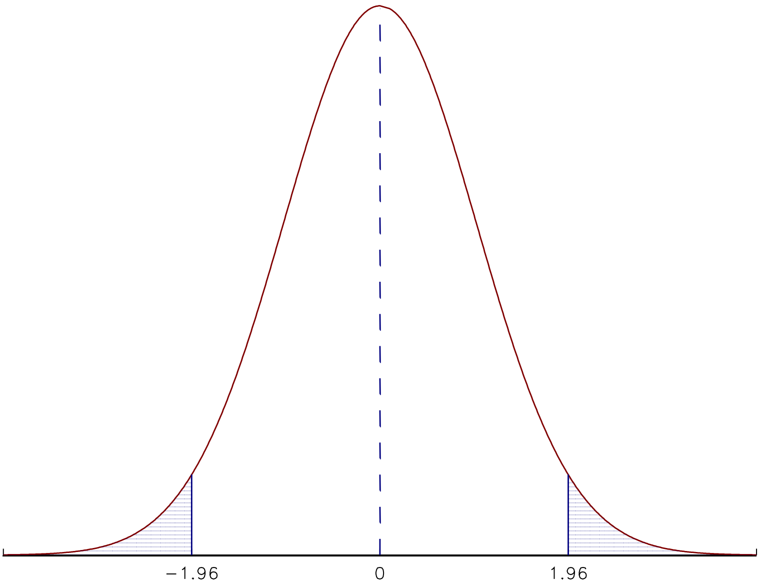
- For a large enough sample, we can approximate:

$$\frac{\hat{\Delta} - \Delta}{\text{s.e.}(\hat{\Delta})} \sim N(0, 1).$$

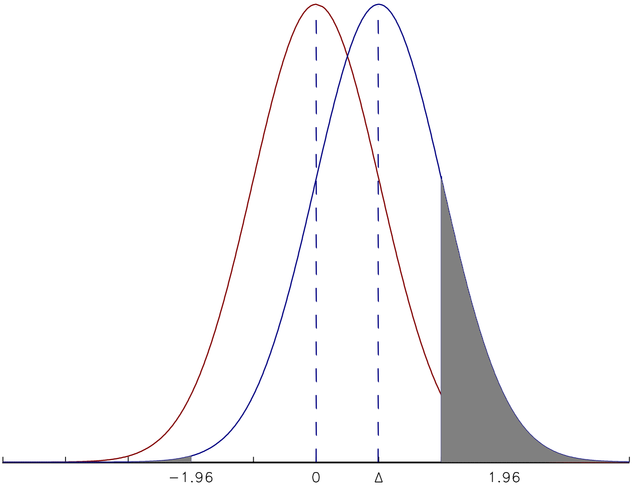
- Therefore, the t-statistic for a test of significance is:

$$t = \frac{\hat{\Delta}}{\text{s.e.}(\hat{\Delta})} \sim N(\frac{\Delta}{\text{s.e.}(\hat{\Delta})}, 1)$$

Probability of Rejection if $\mu_1 - \mu_0 = 0$



Probability of Rejection if $\mu_1 - \mu_0 = \Delta$



Experimental Design: Power Calculations to Choose N

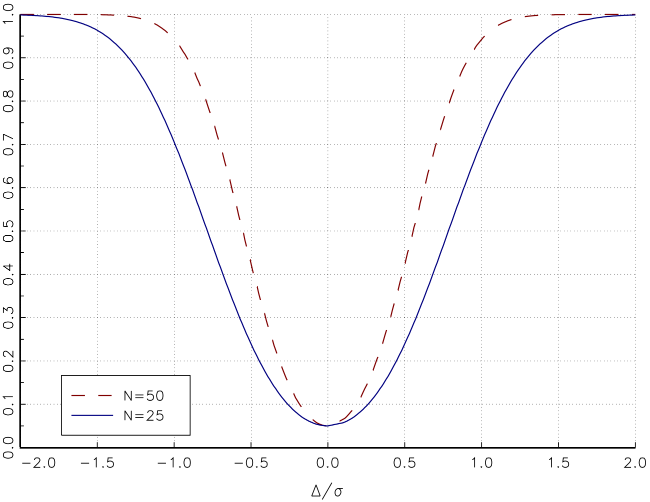
- The probability of rejecting the null $\mu_1 - \mu_0 = 0$ is:

$$
\begin{aligned}
&\Pr\left[|t| > 1.96\right] = \Pr\left[t < -1.96\right] + \Pr\left[t > 1.96\right] \\
&= \Pr\left[t - \frac{\Delta}{s.e.(\hat{\Delta})} < -1.96 - \frac{\Delta}{s.e.(\hat{\Delta})}\right] \\
&+ \Pr\left[t - \frac{\Delta}{s.e.(\hat{\Delta})} > 1.96 - \frac{\Delta}{s.e.(\hat{\Delta})}\right] \\
&= \Phi\left(-1.96 - \frac{\Delta}{s.e.(\hat{\Delta})}\right) + \left(1 - \Phi\left(1.96 - \frac{\Delta}{s.e.(\hat{\Delta})}\right)\right)
\end{aligned}
$$

- Suppose that $p = 1/2$ and $\sigma_1^2 = \sigma_0^2 = \sigma$. Then,

$$
\begin{aligned}
\text{s.e.}(\hat{\Delta}) &= \sqrt{\frac{\sigma^2}{N/2} + \frac{\sigma^2}{N/2}} \\
&= \frac{2\sigma}{\sqrt{N}}.
\end{aligned}
$$

Power Functions with $p = 1/2$ and $\sigma_1 = \sigma_0$



General formula for the power function ($p \neq 1/2$, $\sigma_1 \neq \sigma_0$)

$$\Pr\left[\text{reject } \mu_1 - \mu_0 = 0 \,|\, \mu_1 - \mu_0 = \Delta\right] =$$

$$\Phi\left(-1.96 - \Delta \Big/ \sqrt{\frac{\sigma_1^2}{pN} + \frac{\sigma_0^2}{(1-p)N}}\right)$$

$$+ \left(1 - \Phi\left(1.96 - \Delta \Big/ \sqrt{\frac{\sigma_1^2}{pN} + \frac{\sigma_0^2}{(1-p)N}}\right)\right)$$

To choose N we need to specify:

1. Δ: minimum detectable magnitude of treatment effect
2. Power value (usually 0.80 or higher)
3. σ_1^2 and σ_0^2 (usually $\sigma_1^2 = \sigma_0^2$) (e.g., using previous measures)
4. p: proportion of observations in the treatment group If $\sigma_1^2 = \sigma_0^2$, then the power is maximized by $p = 0.5$

Should We use Regression Adjustment with Randomized Assignment?

- If the treatment D_i is independent of (Y_{1i}, Y_{0i}), then we know that the simply difference in means is an unbiased and consistent estimator of $\Delta_{ATE} = \Delta_{ATT}$. But if we have covariates, should we add them to the regression?
- If we focus on large-sample analysis, the answer is yes, provided the covariates help to predict (Y_{1i}, Y_{0i}). Remember, randomized assignment means D_i is also independent of X_i.
- Consider the case where the treatment effect is constant, so $(Y_{1i} - Y_{0i}) = \Delta_i$ for all i. Then we can write

$$Y_i = Y_{0i} + \Delta D_i = \mu_0 + \Delta D_i + \nu_{i0}$$

  and D_i is independent of Y_{0i} and therefore ν_{i0}.

Should We use Regression Adjustment with Randomized Assignment?

- Simple regression of Y_i on $1, D_i$ is unbiased and consistent for Δ.
- But writing the linear projection

$$Y_{0i} = \alpha_0 + X_i'\beta_0 + u_{0i}$$
$$\mathrm{E}[u_{0i}] = 0, \quad \mathrm{E}[X_i u_{0i}]$$

  we have

$$Y_i = \alpha_0 + \Delta D_i + X_i'\beta_0 + u_{0i}$$

  where, by randomized assignment, D_i is uncorrelated with X_i and u_{0i}.
- So OLS is still consistent for Δ. If $\beta_0 \neq 0$, $\mathrm{V}[u_{0i}] < \mathrm{V}[\nu_{0i}]$, and so adding X_i reduces the error variance.

Should We use Regression Adjustment with Randomized Assignment?

- In fact, under the constant treatment effect assumption and random assignment, the asymptotic variances of the simple and multiple regression estimators are, respectively,

$$\frac{\mathrm{V}\left[\nu_{0i}\right]}{N\rho(1-\rho)}, \frac{\mathrm{V}\left[u_{0i}\right]}{N\rho(1-\rho)}$$

  where $\rho = \Pr\left[D = i = 1\right]$
- The only caveat is that if $\mathrm{E}\left[\, Y_{0i} \middle| X_i \right] \neq \alpha_0 + X_i'\beta_0$ then the OLS estimator of Δ is only guaranteed to be consistent, not unbiased. This distinction can be relevant in small samples (as often occurs in true experiments).

Should We use Regression Adjustment with Randomized Assignment?

- If the treatment effect is not constant, and now we add the linear projection $Y_{1i} = \alpha_1 + X_i'\beta_1 + u_{1i}$, so that

$$\Delta_{ATE} = \Delta = (\alpha_1 - \alpha_0) + \mu_X'(\beta_1 - \beta_0)$$

  we can write

$$\begin{aligned} Y_i &= \alpha_0 + \Delta D_i + X_i'\beta_0 + (X_i - \mu_X)'(\beta_1 - \beta_0) + u_{0i} + D_i(u_{1i} - u_{0i}) \\ &\equiv \alpha_0 + \Delta D_i + X_i'\beta_0 + D_i(X_i - \mu_X)'\delta + u_{0i} + D_i e_i \end{aligned}$$

  with $\delta \equiv \beta_1 - \beta_0$ and $e_i \equiv u_{1i} - u_{0i}$
- Under random assignment of treatment, (e_i, X_i) is independent of D_i, so D_i is uncorrelated with all other terms in the equation.
- OLS is consistent for Δ but it is generally biased unless the equation represents $\mathrm{E}[Y_i | D_i, X_i]$.

Example: The LaLonde (1986) Study

Major Questions:

- How good are non-experimental estimates?
- Do they come close to the estimates from randomized field experiments?
- How robust are the nonexperimental estimators w.r.t. sample choices?

NSW-Program

"The National Supported Work Demonstration (NSW) was a temporary employment program designed to help disadvantaged workers lacking basic job skills move into the labor market by giving them work experience and counseling in a sheltered environment. Unlike other federally sponsored employment and training programs, the NSW program assigned qualified applicants to training positions randomly. Those assigned to the treatment group received all the benefits of the NSW program, while those assigned to the control group were left to fend for themselves."

NSW-Program

- collected in the mid-1979's by the Manpower Demonstration Research Corporation (MDRC)
- US wide, 10 different sites
- AFDC women, ex-drug addicts, ex-criminal offenders, high school drop-out of both sexes
- specific program length between 9 to 18 month depending on target group and site
- NSW counsellor for group of 3-5 people in treatment group
- reasonable enumeration (but lower than market wage)
- background data for treatment and control group

Figure 1: Sample means and standard deviations of pre-training earnings and other characteristics for the NSW AFCD and male participants.

	Full National Supported Work Sample			
	AFDC Participants		Male Participants	
Variable	Treatments	Controls	Treatments	Controls
Age	33.37	33.63	24.49	23.99
	(7.43)	(7.18)	(6.58)	(6.54)
Years of School	10.30	10.27	10.17	10.17
	(1.92)	(2.00)	(1.75)	(1.76)
Proportion	.70	.69	.79	.80
High School Dropouts	(.46)	(.46)	(.41)	(.40)
Proportion Married	.02	.04	.14	.13
	(.15)	(.20)	(.35)	(.35)
Proportion Black	.84	.82	.76	.75
	(.37)	(.39)	(.43)	(.43)
Proportion Hispanic	.12	.13	.12	.14
	(.32)	(.33)	(.33)	(.35)
Real Earnings	$393	$395	1472	1558
1 year Before	(1,203)	(1,149)	(2656)	(2961)
Training	[43]	[41]	[58]	[63]
Real Earnings	$854	$894	2860	3030
2 years Before	(2,087)	(2,240)	(4729)	(5293)
Training	[74]	[79]	[104]	[113]
Hours Worked	90	92	278	274
1 year Before	(251)	(253)	(466)	(458)
Training	[9]	[9]	[10]	[10]
Hours Worked	186	188	458	469
2 years Before	(434)	(450)	(654)	(689)
Training	[15]	[16]	[14]	[15]
Month of Assignment	−12.26	−12.30	−16.08	−15.91
(Jan. 78 = 0)	(4.30)	(4.23)	(5.97)	(5.89)
Number of Observations	800	802	2083	2193

Note: The numbers shown in parentheses are the standard deviations and those in the square brackets are the standard errors.(LaLonde, 1986, p.606)

Figure 2: Annual Earnings of NSW Treatments, Controls, and Eight Candidate Comparison Groups from the PSID and the CPS-SSA

			Comparison Group[a,b]							
Year	Treat-ments	Controls	*PSID*-1	*PSID*-2	*PSID*-3	*PSID*-4	*CPS-SSA*-1	*CPS-SSA*-2	*CPS-SSA*-3	*CPS-SSA*-4
1975	$895	$877	7,303	2,327	937	6,654	7,788	3,748	4,575	2,049
	(81)	(90)	(317)	(286)	(189)	(428)	(63)	(250)	(135)	(333)
1976	$1,794	$646	7,442	2,697	665	6,770	8,547	4,774	3,800	2,036
	(99)	(63)	(327)	(317)	(157)	(463)	(65)	(302)	(128)	(337)
1977	$6,143	$1,518	7,983	3,219	891	7,213	8,562	4,851	5,277	2,844
	(140)	(112)	(335)	(376)	(229)	(484)	(68)	(317)	(153)	(450)
1978	$4,526	$2,885	8,146	3,636	1,631	7,564	8,518	5,343	5,665	3,700
	(270)	(244)	(339)	(421)	(381)	(480)	(72)	(365)	(166)	(593)
1979	$4,670	$3,819	8,016	3,569	1,602	7,482	8,023	5,343	5,782	3,733
	(226)	(208)	(334)	(381)	(334)	(462)	(73)	(371)	(170)	(543)
Number of Observations	600	585	595	173	118	255	11,132	241	1,594	87

Note: [a]The Comparison Groups are defined as follows: PSID-1: All female household heads continuously from 1975 through 1979, who were between 20 and 55-years-old and did not classify themselves as retired in 1975; PSID-2: Selects from the PSID-1 group all women who received AFDC in 1975; PSID-3: Selects from the PSID-2 all women who were not working when surveyed in 1976; PSID-4: Selects from the PSID-1 group all women with children, none of whom are less than 5-years-old; CPS-SSA -1: All females from Westat CPS-SSA sample; CPS-SSA -2: Selects from CPS-SSA-1 all females who received AFDC in 1975; CPS-SSA-3: Selects from CPS-SSA-1 all females who were not working in the spring of 1976; CPS-SSA -4: Selects from CPS-SSA-2 all females who were not working in the spring of 1976.

[b] All earnings are expressed in 1982 dollars. The numbers in parentheses are the standard errors. For the NSW treatments and controls, the number of observations refer only to 1975 and 1979. In the other years there are fewer observations, especially in 1978. At the time of the resurvey in 1979, treatments had been out of Supported Work for an average of 20 months.(LaLonde, 1986, p.607)

Figure 3: Annual Earnings of NSW Male Treatments, Controls, and Six Candidate Comparison Groups from the PSID and the CPS-SSA

			Comparison Group[a,b]					
Year	Treatments	Controls	*PSID*-1	*PSID*-2	*PSID*-3	*CPS-SSA*-1	*CPS-SSA*-2	*CPS-SSA*-3
1975	\$3,066	\$3,027	19,056[a]	7,569	2,611	13,650	7,387	2,729
	(283)	(252)	(272)	(568)	(492)	(73)	(206)	(197)
1976	\$4,035	\$2,121	20,267	6,152	3,191	14,579	6,390	3,863
	(215)	(163)	(296)	(601)	(609)	(75)	(187)	(267)
1977	\$6,335	\$3,403	20,898	7,985	3,981	15,046	9,305	6,399
	(376)	(228)	(296)	(621)	(594)	(76)	(225)	(398)
1978	\$5,976	\$5,090	21,542	9,996	5,279	14,846	10,071	7,277
	(402)	(227)	(311)	(703)	(686)	(76)	(241)	(431)
Number of Observations	297	425	2,493	253	128	15,992	1,283	305

Note: [a] The Comparison Groups are defined as follows: PSID-1: All male household heads continuously from 1975 through 1978, who were less than 55-years-old and did not classify themselves as retired in 1975; PSID-2: Selects from the PSID-1 group all men who were not working when surveyed in the spring of 1976; PSID-3: Selects from the PSID-1 group all men who were not working when surveyed in either spring of 1975 or 1976; CPS-SSA-1: All males based on Westat's criteria, except those over 55-years-old; CPS-SSA-2: Selects from CPS-SSA-1 all males who were not working when surveyed in March 1976; CPS-SSA-3: Selects from the CPS-SSA-1 unemployed males in 1976 whose income in 1975 was below the poverty level.

[b] All earnings are expressed in 1982 dollars. The numbers in parentheses are the standard errors. The number of observations refer only to 1975 and 1978. In the other years there are fewer observations. The sample of treatments is smaller than the sample of controls because treatments still in Supported Work as of January 1978 are excluded from the sample, and in the young high school target group there were by design more controls than treatments. (LaLonde, 1986, p.608)

Figure 4: Earnings Comparisons and Estimated Training Effects for the NSW AFCD Participants Using Comparison Groups from the PSID And The CPS-SSA[a,b]

		NSW Treatment Earnings Less Comparison Group Earnings				Difference in Differences: Difference in Earnings Growth 1975–79 Treatments Less Comparisons		Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–79		Controlling for All Observed Variables and Pre-Training Earnings	
	Comparison Group Earnings Growth 1975–79	Pre-Training Year, 1975		Post-Training Year, 1979							
Name of Comparison Group[d]	(1)	Unadjusted (2)	Adjusted[c] (3)	Unadjusted (4)	Adjusted[c] (5)	Without Age (6)	With Age (7)	Unadjusted (8)	Adjusted[c] (9)	Without AFDC (10)	With AFDC (11)
Controls	2,942	−17	−22	851	861	833	883	843	864	854	–
	(220)	(122)	(122)	(307)	(306)	(323)	(323)	(308)	(306)	(312)	
PSID-1	713	−6,443	−4,882	−3,357	−2,143	3,097	2,657	1746	1,354	1664	2,097
	(210)	(326)	(336)	(403)	(425)	(317)	(333)	(357)	(380)	(409)	(491)
PSID-2	1,242	−1,467	−1,515	1,090	870	2,568	2,392	1,764	1,535	1,826	–
	(314)	(216)	(224)	(468)	(484)	(473)	(481)	(472)	(487)	(537)	
PSID-3	665	−77	−100	3,057	2,915	3,145	3,020	3,070	2,930	2,919	–
	(351)	(202)	(208)	(532)	(543)	(557)	(563)	(531)	(543)	(592)	
PSID-4	928	−5,694	−4,976	−2,822	−2,268	2,883	2,655	1,184	950	1,406	2,146
	(311)	(306)	(323)	(460)	(491)	(417)	(434)	(483)	(503)	(542)	(652)
CPS-SSA-1	233	−6,928	−5,813	−3,363	−2,650	3,578	3,501	1,214	1,127	536	1,041
	(64)	(272)	(309)	(320)	(365)	(280)	(282)	(272)	(309)	(349)	(503)
CPS-SSA-2	1,595	−2,888	−2,332	−683	−240	2,215	2,068	447	620	665	–
	(360)	(204)	(256)	(428)	(536)	(438)	(446)	(468)	(554)	(651)	
CPS-SSA-3	1,207	−3,715	−3,150	−1,122	−812	2,603	2,615	814	784	−99	1,246
	(166)	(226)	(325)	(311)	(452)	(307)	(328)	(305)	(429)	(481)	(720)
CPS-SSA-4	1,684	−1,189	−780	926	756	2,126	1,833	1,222	952	827	–
	(524)	(249)	(283)	(630)	(716)	(654)	(663)	(637)	(717)	(814)	

Note: [a] The columns above present the estimated training effect for each econometric model and comparison group. The dependent variable is earnings in 1979. Based on the experimental data, an unbiased estimate of the impact of training presented in col. 4 is $851. The first three columns present the difference between each comparison group's 1975 and 1979 earnings and the difference between the pre-training earnings of each comparison group and the NSW treatments.

[b] Estimates are in 1982 dollars. The numbers in parentheses are the standard errors.

[c] The exogenous variables used in the regression adjusted equations are age, age squared, years of schooling, high school dropout status, and race.

[d] See Table 2 for definitions of the comparison groups. (LaLonde, 1986, p.609)

Figure 5: Earnings Comparisons and Estimated Training Effects for the NSW Male Participants Using Comparison Groups from the PSID And The CPS-SSA[a,b]

Name of Comparison Group[d]	Comparison Group Earnings Growth 1975–78	NSW Treatment Earnings Less Comparison Group Earnings				Difference in Differences: Difference in Earnings Growth 1975–78 Treatments Less Comparisons		Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–78		Controlling for All Observed Variables and Pre-Training Earnings
		Pre-Training Year, 1975		Post-Training Year, 1978						
		Unadjusted	Adjusted[c]	Unadjusted	Adjusted[c]	Without Age	With Age	Unadjusted	Adjusted[c]	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Controls	$2,063	$39	$ − 21	$886	$798	$847	$856	$897	$802	$662
	(325)	(383)	(378)	(476)	(472)	(560)	(558)	(467)	(467)	(506)
PSID-1	$2,043	− $15,997	− $7,624	− $15,578	− $8,067	$425	− $749	− $2,380	− $2,119	− $1,228
	(237)	(795)	(851)	(913)	(990)	(650)	(692)	(680)	(746)	(896)
PSID-2	$6,071	− $4,503	− $3,669	− $4,020	− $3,482	$484	− $650	− $1,364	− $1,694	− $792
	(637)	(608)	(757)	(781)	(935)	(738)	(850)	(729)	(878)	(1024)
PSID-3	($3,322	($455	$455	$697	− $509	$242	− $1,325	$629	− $552	$397
	(780)	(539)	(704)	(760)	(967)	(884)	(1078)	(757)	(967)	(1103)
CPS-SSA-1	$1,196	− $10,585	− $4,654	− $8,870	− $4,416	$1,714	$195	− $1,543	− $1,102	− $805
	(61)	(539)	(509)	(562)	(557)	(452)	(441)	(426)	(450)	(484)
CPS-SSA-2	$2,684	− $4,321	− $1,824	− $4,095	− $1,675	$226	− $488	− $1,850	− $782	− $319
	(229)	(450)	(535)	(537)	(672)	(539)	(530)	(497)	(621)	(761)
CPS-SSA-3	$4,548	$337	$878	− $1,300	$224	− $1,637	− $1,388	− $1,396	$17	$1,466
	(409)	(343)	(447)	(590)	(766)	(631)	(655)	(582)	(761)	(984)

Note: [a] The columns above present the estimated training effect for each econometric model and comparison group. The dependent variable is earnings in 1978. Based on the experimental data an unbiased estimate of the impact of training presented in col. 4 is $886. The first three columns present the difference between each comparison group's 1975 and 1978 earnings and the difference between the pre-training earnings of each comparison group and the NSW treatments.

[b] Estimates are in 1982 dollars. The numbers in parentheses are the standard errors.

[c] The exogenous variables used in the regression adjusted equations are age, age squared, years of schooling, high school dropout status, and race.

[d] See Table 3 for definitions of the comnarison arouns. (LaLonde, 1986, p.610)

Example: Job Training Partnership Act (JTPA)

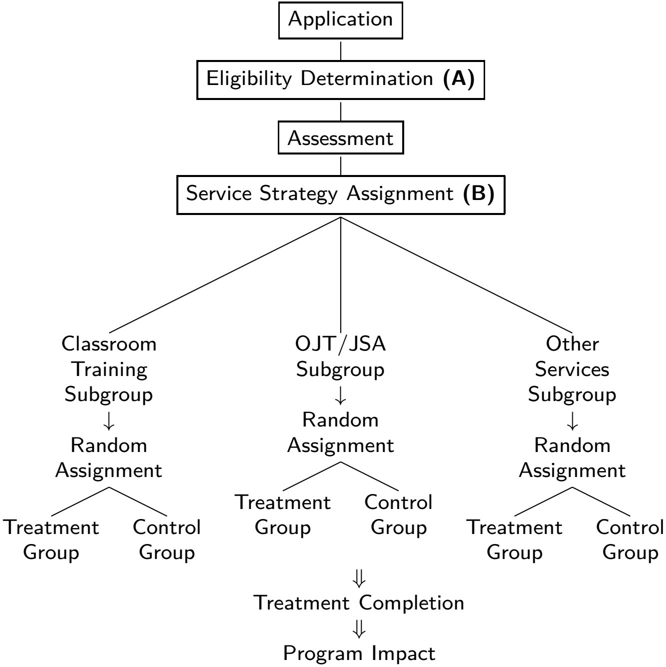
- Largest randomized training evaluation ever undertaken in the U.S.; started in 1983 at 649 sites throughout the country
- Sample: Disadvantaged persons in the labor market (previously unemployed or low earnings)
- D: Assignment to one of three general service strategies
  - classroom training in occupational skills
  - on-the-job training and/or job search assistance
  - other services (eg. probationary employment)
- Y: earnings 30 months following assignment
- X: Characteristics measured before assignment (age, gender, previous earnings, race, etc.)

Example: Job Training Partnership Act (JTPA)

Exhibit 5: Impacts of Total 30-Month Earnings: Assignees an Enrollees, by target group

	Mean Earnings		Impact per assignee		
	Treatments Group (1)	Control Group (2)	In dollars (3)	As a (2) percent of (2)	Impact per enrollee in dollars
Adult Woman	\$13,417	\$12,241	\$1,178***	9.6%	\$1,837***
Adult Men	19,474	18,496	987*	5.3	1,599*
Female Youth	10,241	10,106	135	1.3	20
Male youth non-arrestees	15,786	16,375	0.589	-3.6	-868
Male youth arrestees					
Using Survey data	14,633	18,842	-4,209**	22.3	-6,804**
Using scaled UI data	14,148	14,152	-4	0	-6

Sources: Estimates based on Frst and Second Follow-up Survey responses and earnings data from state unemployment insurance (UI) agencies
Sample size: adult women, 6,102; adult men, 5,102; female youths, 2,657; male youth non-arrestees, 1,704; male youth arrestees, 416.
* Statistically significant at the 10% level, ** 5% level, *** 1% level (two tailed test).


Application
Eligibility Determination (A)
Assessment
Service Strategy Assignment (B)
Classroom Training Subgroup
Random Assignment
Treatment Group
Control Group
OJT/JSA Subgroup
Random Assignment
Treatment Group
Control Group
Other Services Subgroup
Random Assignment
Treatment Group
Control Group
Treatment Completion
Program Impact

Assessing the Case for Social Experiments (Heckman & Smith (1995), J.Ec.Persp.'95)

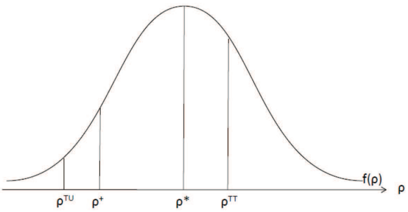
Reservations against LaLonde's Study

- richer data improve nonexperimental estimates
- exploitation of longitudinal information
- more advanced nonexperimental estimators available
- battery of specification test reduce bias of nonexperimental estimates (Heckman&Hotz JASA '89)
- experience: linear approaches can be ruled out (e.g. DiD)
- LaLonde's results cannot be generalized

Potential Problems when Running Experiments (Heckman & Smith (1995), J.Ec.Persp.'95)

1. Randomization Bias
   "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 as it normally operates." (Heckman & Smith (1995), J.Ec.Persp., p.99)
   - existence on selection based randomization may effect participants' behavior positively or negatively
   - change of participants' behavior due to threat of denial
   - Randomization may reduce overall participation (Kramer & Shapiro, JAMA '85 for drug trials)
   - rarely additional nonexperimental evidence
   - Little empirical evidence on randomization bias (exception JTPA)

Potential Problems when Running Experiments



- $\rho_i = Y_{1i} - Y_{0i}$ is treatment effect of individual i.
- ρ^* is the average treatment effect.
- ρ^+ is the cutoff value above which people participate in the experiment.
- ρ^{TT} is the treatment effect on the treated which is measured in the experiment
- ρ^{TU} is the treatment effect on the untreated which is not measured as those people would not participate.

Figure 6: Percent of Training Centers Citing Specific Concerns about Participating in the Experiment

Concern	*Percent of Training Centers Citing the Concern*
(1) Ethical and Public Relations Implications of:	
(a) Random Assignment in Social Programs	61.8
(b) Denial of Services to Controls	54.4
(2) Potential Negative Effect of Creation of a Control Group on Achievement of Client Recruitment Goals	47.8
(3) Potential Negative Impact on Performance Standards	25.4
(4) Implementation of the Study When Service Providers Do Intake	21.1
(5) Objections of Service Providers to the Study	17.5
(6) Potential Staff Administrative Burden	16.2
(7) Possible Lack of Support by Elected Officials	15.8
(8) Legality of Random Assignment and Possible Grievances	14.5
(9) Procedures for Providing Controls With Referrals to Other Services	14.0
(10) Special Recruitment Problems for Out-of-School Youth	10.5

Source: Based on the responses of 228 JTPA training centers connected about possible participation in the National JTPA Study (Doolittle and Traeger, 1990, Table 2.1, p. 34)
Notes: Concerns noted by fewer than 5 percent of the training centers are not listed. Percentages may add to more than 100 because training centers could raise more than one concern. (Heckman & Smith, 1995, p. 101)

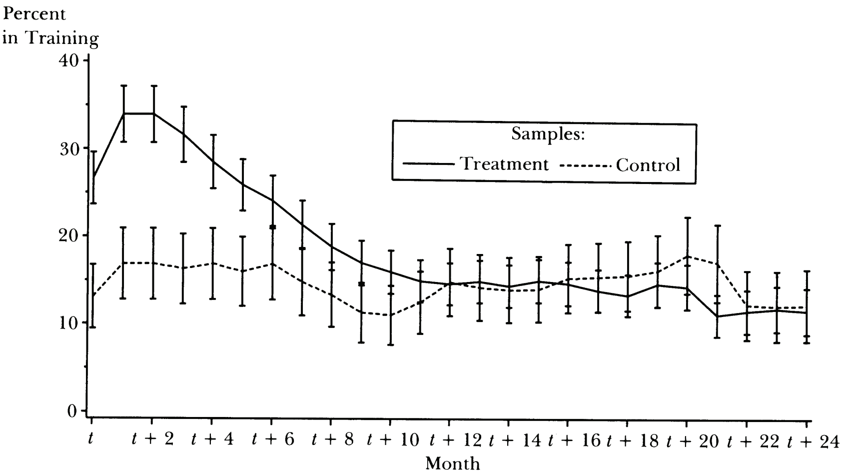
Potential Problems when Running Experiments

2. Administrative Limitations on Social Experiments
   - voluntary participation of training sites
   - complicated administrative process until final placement to specific program
   - often multiple and sequential treatments
   - alternative points at which randomization can take place have different merits (Heckman & Smith '93)
   - complexity if welfare programs
   - non-random selection by officers in charge of the program
   - uneven attrition particular if pre-treatment phase and/or treatment phase is long an

Potential Problems when Running Experiments

3. Substitution Bias
   "Substitution bias arises when members of an experimental control group gain access to close substitutes for the experimental treatment, like similar services offered by other providers or the same service offered under different funding arrangements." (Heckman & Smith (1995), J.Ec.Persp., p.105)
   - Control group outcome is different from untreated state
   - Control group members may seek substitutes for treatment.
   - This would bias estimated treatment effects downwards.
   - Can also occur if the experiment frees up resources that can now be concentrated on the control group.
   - Example JTPA: 32 % of control group received treatment from other resources during treatment period, 48 % of the treated really received treatment (self reports)
   - Arises often in clinical trial if treatment was denied

Figure 1: Controls and Treatments Percent in School or Training, Female Youth



Note: 1. Month 't' is the month of random assignment for the controls and treatments. Bars indicate confindence bands.
2. The monthly training information graphed here is derived from self-reported data on spells of schooling and training (Heckman & Smith, 1995, p. 107)

Potential Problems when Running Experiments

4. Hawthorne Effect

   Do participants behave differently because they know that the program is an experiment?

   see Burtles&Orr (1986), Heckman&Smith (1995)

5. The Ethical Issue

   - Can we deny treatment randomly for those who are eligible? Do we have the right to deny a potentially beneficial treatment?
   - Can we collect confidential data from those who are not treated?

   But given rationing is necessary due to expenditure restriction: A random assignment is a fair (the best ?) way of rationing.

Potential Problems when Running Experiments

6. Little Evidence on Many Questions of Interest
   - What are the effects of factors such as subsidies, local labor markets, advertising, gender, race on the participation decision ?
   - How do administrative rules effect participation?
   - What are the effects of factors such as subsidies, local labor markets, advertising, gender, race on the drop out decision ?
   - What are the effects and costs of various treatments ?
   - How does the length of the program effect treatment ?
   - How does the treatment work for other non-eligible groups ?

Potential Problems when Running Experiments

7. Other Possible Issues
   - The long delays often associated with experimental evaluations
   - Attrition from experimental samples (Hausman & Wise, 1985)
   - The inability of small-scale experiments to predict general equilibrium effects or to produce results that can be extrapolated to other populations (Zellner & Rossi, 1986)

Example of Large Randomized Experiment: Tennessee Project STAR

The Effect of Class Size on Educational Achievement:

- Krueger (1999) econometrically re-analyses a randomized experiment of the effect of class size on student achievement.
- The project is known as Tennessee Student/Teacher Achievement Ratio (STAR) and was run in the 1980s.
- 11,600 students and their teachers were randomly assigned to one of three groups
  - Small classes (13-17) students.
  - Regular classes (22-25) students.
  - 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.

Regression Analysis of Experiments

Conditional Treatment Effects in Terms of a Regression Model, cont'd

- With randomization 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.
- To see this, let's start with the assumption of constant treatment effects (i.e. the treatment affects everyone by the same magnitude). $Y_{1i} - Y_{01} = \Delta$
- We can therefore rewrite $Y_i = Y_{0i} + (Y_{1i} - Y_{01})D_i$ as:

$$Y_i = \underbrace{\alpha}_{=\mathrm{E}[Y_{01}]} + \underbrace{\Delta}_{=(Y_{1i}-Y_{01})} D_i + \underbrace{\eta_i}_{=Y_{01}-\mathrm{E}[Y_{01}]} \qquad (1)$$

where η_i is random part of Y_{01}.

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

Regression Analysis of Experiments

- The conditional expectation of (1) with treatment status switched on and off gives:

$$\begin{aligned} E[Y_i | D_i = 1] &= \alpha + \Delta + E[\eta_i | D_i = 1] \\ E[Y_i | D_i = 0] &= \alpha + E[\eta_i | D_i = 0] \end{aligned}$$

- So that

$$E[Y_i | D_i = 1] - E[Y_i | D_i = 0] = \underbrace{\Delta}_{\text{Treatment Effect}} + \underbrace{E[\eta_i | D_i = 1] - E[\eta_i | D_i = 0]}_{\text{Selection Bias}}$$

- In the STAR experiment D_i (being in a small class) is randomly assigned and therefore the selection bias disappears.

Why Include Additional Controls?

- To evaluate experimental data one may want to add additional controls in the regression. Instead of estimating equation (1) one would estimate:

$$Y_i = \alpha + \Delta D_i + X_i'\gamma + \eta_i$$

- There are 2 main reasons for including additional controls in the regression model.
  - Conditional random assignment. sometimes randomization is done conditional on some observables (here at the school level).
  - Additional controls increase precision. Although the control variables X_i are uncorrelated with D_i they may have substantial explanatory power for Y_i. Including controls thus reduces residual variance and therefore lowers the standard errors of the regression estimates.

Regression in Krueger (1999)

- Krueger estimates the following econometric model:

$$Y_{ics} = \beta_0 + \beta_1 \text{SMALL}_{cs} + \beta_2 \text{REG/A}_{cs} + \beta_3 X_{ics} + \alpha_s + \varepsilon_{isc}$$

  - Y_{ics} =percentile score.
  - SMALL_{cs} =Indicator whether student was assigned to a small class.
  - REG/A_{cs} =Indicator whether student was assigned to a regular class with aide.
  - α =School FE. Because random assignment occurred within schools.

Regression Results Kindergarten

- Krueger estimates the following econometric model:

$$Y_{ics} = \beta_0 + \beta_1 \text{SMALL}_{cs} + \beta_2 \text{REG/A}_{cs} + \beta_3 X_{ics} + \alpha_s + \varepsilon_{isc}$$

  - Y_{ics} =percentile score.
  - SMALL_{cs} =Indicator whether student was assigned to a small class.
  - REG/A_{cs} =Indicator whether student was assigned to a regular class with aide.
  - α =School FE. Because random assignment occurred within schools.

Regression Results Kindergarten

Explanatory Variables	OLS: Actual Class Size			
	A. Kindergarten			
Small Class	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
Regular/aide class	0.12	0.29	0.53	0.31
	(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
			(0.63)	(0.63)
Free lunch (1=yes)	-	-	-13.15	-13.07
			(0.77)	(0.77)
White teacher	-	-	-	-0.57
				(2.10)
Teacher Experience	-	-	-	0.26
				(0.10)
Master's degree				-0.51
				(1.06)
School FE	No	Yes	Yes	Yes
R^2	0.01	0.25	0.31	0.31

Regression Results First Grade

Explanatory Variables	OLS: Actual Class Size B. First Grade			
Small Class	8.57	8.43	7.91	7.4
	(1.97)	(1.21)	(1.17)	(1.18)
Regular/aide class	3.44	2.22	2.23	1.78
	(2.05)	(1.00)	(0.98)	(0.98)
White/Asian (1=yes)	-	-	6.97	6.97
			(1.18)	(1.19)
Girl (1=yes)	-	-	3.8	3.85
			(0.56)	(0.56)
Free lunch (1=yes)	-	-	-13.49	-13.61
			(0.87)	(0.87)
White teacher	-	-	-	-4.28
				(1.96)
Male Teacher	-	-	-	11.82
				(3.33)
Teacher Experience				0.05
				(0.06)
Master's degree				0.48
				(1.07)
School FE	No	Yes	Yes	Yes
R^2	0.02	0.24	0.3	0.3

Problem 1: Attrition

A Common Problem in Randomized Experiments:

- 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 schools creating a selection bias problem.
- Krueger addresses this concern by imputing test scores (from their earlier test scores) for all children who leave the sample and then reestimates the model including students with imputed test scores.

Regression Results Imputing Test Scores to Address Attrition

	Actual test data		Actual and imputed test data	
Grade	Coefficient on small class dum.	Sample size	Coefficient on small class dum.	Sample size
K	5.32	5900	5.32	5900
	(0.76)		(0.76)	
1	6.95	6632	6.3	8328
	(0.74)		(0.68)	
2	5.59	6282	5.64	9773
	(0.76)		(0.65)	
3	5.58	6339	5.49	10919
	(0.79)		(0.63)	

- Non-random attrition hardly biases the results.

Problem 2: Students changed Classes After Random Assignment

Example Transitions between Grades 1 and Grade 2:

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

Problem 2: Students changed Classes After Random Assignment

- Subjects moved between treatment and control groups.
- A common solution to this problem is to use initial assignment (here initial assignment to small or regular classes) as an instrument for actual assignment (more on Instrumental Variable methods in lecture 2).
- 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.

Problem 2: Students changed Classes After Random Assignment

Explanatory	OLS: actual class size				Reduced form: initial class size			
Variable	(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
			(0.56)	(0.56)			(0.56)	(0.56)
Free lunc (1=yes)			-13.49	-13.61			-13.65	-13.77
			(0.87)	(0.87)			(0.88)	(0.87)
White teacher				-4.28				-4.40
				(1.96)				(1.97)
Male teacher				11.82				13.06
				(3.33)				(3.38)
Teacher experience				0.05				0.06
				(0.06)				(0.06)
Master's degree				0.48				0.63
				(1.07)				(1.09)
School FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
R^2	0.02	0.24	0.30	0.30	0.01	0.23	0.29	0.30

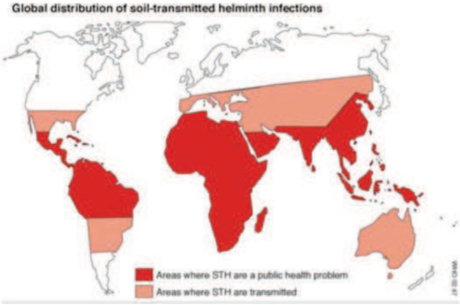
How to Write a Good Experimental Paper?[1]

- Advising on how to write good papers is very difficult in any setup.
- It seems to become more difficult to publish experimental papers which simply randomize a certain treatment and evaluate its effect.
- A promising avenue for experimental papers seem to be the ones that combine experimental data with economic theory:
  1. Discriminating between important theories.
  2. First obtain "reduced form" results of a causal effect and then use structural econometrics to disentangle economic mechanisms.
  3. Use an experiment to estimate externalities or other market failures.
- In the following we will review another paper with particularly nice links between empirics and theory.

[1]The following part is mostly based on lecture slides by Fabian Waldinger

Miguel and Kremer (2004) - Worms

- Miguel and Kremer study the impact of a treatment against intestinal worms in primary school in rural Kenya.
- Intestinal worms affect one in four people worldwide and are particularly prevalent among school-age children in developing countries.
- Prevalence of Intestinal Worms



Estimating Externalities

- Studies in which treatment is randomized at the individual level may potentially obtain biased treatment effects because of externalities:
  1. Externalities within treated schools.
  2. Externalities across schools from treatment to control schools.
- While they investigate both types of externalities they only have experimental variation to identify cross-school externalities (within school externalities have to be evaluated differently).
- They evaluate a Kenyan programme where randomization occurred at the school level which allows them to look at externalities.
- They evaluate the effect of the de-worming treatment on health, school absenteeism, and test scores.

Estimating Externalities

- Overall 75 schools were treated in 3 groups. The health intervention was phased in sequentially:
  1. Group 1 schools: received free deworming treatment in 1998 and 1999.
  2. Group 2 schools: in 1999.
  3. Group 3 school: in 2001.

Year	Treatment	Control
1998	1	2;3
1999	1;2	3

- Treatment schools received half yearly (or yearly for different worms) treatment and medical education of how to avoid worm infection.
- Even in treated schools not all children received treatment mostly because of school absence on the treatment day.
- Because some students switched schools they use initial assignment to evaluate the programme (intention-to-treat; equivalent to RF in Krueger, 1999)

Summary Statistics

	Group 1 (25 schools)	Group 2 (25 schools)	Group 3 (25 schools)	Group 1- Group 3	Group 2- Group 3
Panel A: Pre-school to Grade 8					
Male	0.53	0.51	0.52	0.01	-0.01
				(0.02)	(0.02)
Proportion girls $<$ 13 years, and all boys	0.89	0.89	0.88	0.00	0.01
				(0.01)	(0.01)
Grade progression (=Grade-(Age-6))	-2.1	-1.9	-2.1	-0.0	0.1
				(0.1)	(0.1)
Year of birth	1986.2	1986.5	1985.8	0.4^{**}	0.8^{***}
				(0.2)	(0.2)
Panel B: Grades 3 to 8					
Attendance recorded in school registers (during the four weeks prior to the pupil survey)	0.973	0.963	0.969	0.003	-0.006
				(0.004)	(0.004)
Access to latrine at home	0.82	0.81	0.82	0	-0.01
				(0.03)	(0.03)
Have livestock (cows, goats, pigs, sheep) at home	0.66	0.67	0.66	0	-0.01
				(0.03)	(0.03)
Weight-for-age Z-score (low scores denote undernutrition)	-1.39	-1.4	-1.44	0.05	0.04
				(0.05)	(0.05)
Blood in stool (self-reported)	0.26	0.22	0.19	0.07^{**}	0.03
				(0.03)	(0.03)
Sick often (self-reported)	0.1	0.1	0.08	0.02^{**}	0.02^{**}
				(0.01)	(0.01)
Malaria/fever in past week (self-reported)	0.37	0.38	0.4	-0.03	-0.02
				(0.03)	(0.03)
Clean (observed by field workers)	0.6	0.66	0.67	-0.07^{**}	-0.01
				(0.03)	(0.03)

Despite random assignment: Group 1 schools seem to be slightly worse off before treatment (would underestimate treatment effects).

Summary Statistics - Intestinal Infections

	Prevalence of infection	Prevalence of moderate-heavy infection	Average infection intensity, in eggs per gram (s.e)
Hookworm	0.77	0.15	426 (1055)
Roundworm	0.42	0.16	2337 (5156)
Schistosomiasis, all schools	0.22	0.07	91 (413)
Schistosomiasis, schools$<$ 5 km from Lake Victoria	0.8	0.39	487 (879)
Whipworm	0.55	0.1	161 (470)
At least one infection	0.92	0.37	
Born since 1985	0.92	0.4	
Born before 1985	0.91	0.34	
Female	0.91	0.34	
Male	0.93	0.38	
At least two infections	0.31	0.1	
At least three infections	0.28	0.01	

Econometric Specification

- Randomization of deworming treatment across schools allows estimation of the overall effect of the programme by comparing treatment and comparison schools even if there are within-school externalities. (To decompose direct deworming and within school externality they must rely on non-experimental methods).
- Externalities also occur across schools because children from the same farm often attend different schools.
- They estimate cross-school externalities by taking advantage of variation in the local density of treatment schools introduced by randomization.

$$Y_{itj} = \alpha + \beta_1 \text{Treatment(Year1)}_{it} + \beta_2 \text{Treatment(Year2)}_{it} + X'_{ijt}\delta + \sum_d \left(\gamma_d N^T_{dit}\right) + \sum_d \left(\phi_d N_{dit}\right) + u_i + e_{ijt} \qquad (2)$$

X_{ijt} is a vector of control variables (to increase statistical precision).
N^T_{dit} is the number of pupils randomly assigned to treatment.
N_{dit} is the number of pupils at distance d from school i and year t.

Econometric Specification

- $Y_{itj} = \alpha + \beta_1 \text{Treatment(Year1)}_{it} + \beta_2 \text{Treatment(Year2)}_{it} + X'_{ijt}\delta$
$$+ \sum_d \left(\gamma_d N^T_{dit}\right) + \sum_d \left(\phi_d N_{dit}\right) + u_i + e_{ijt} \qquad (2)$$
- γ_d measures the extent of cross-school externalities
- β_1 captures direct effect of deworming + within school externalities on untreated children in treated schools for year 1
- $\beta_1 + \sum_d \left(\gamma_d \bar{N}^T_{dit}\right)$ is the average effect of deworming treatment on overall infection prevalence in treatment schools in year 1 (including cross school externalities from other treated schools).
- If the authors were just after the total programme effect in treated schools and cross school externalities they could simply estimate equation (2).

Econometric Specification - Within-School Externalities

- But the authors also want to quantify within school externalities (on pupils in treated schools who do not get the treatment).
- Given that the authors do not have within school randomization they cannot estimate the within school externality using experimental variation.
- They use a nice feature of the experimental setup (plus some additional assumptions) to quantify within school externalities.
- They use the following idea:
  Using data from the first year they compare the difference in health outcomes for the following pupils:
  - pupils in treated schools (group 1) who do not take up treatment in year 1 to
  - pupils in control schools (group 2) who will not take up treatment in year 2

Econometric Specification - Within-School Externalities

- To see this more formally: call D a dummy which indicates whether an individual takes up treatment if it is offered in his school.
- Assuming covariates X are the same in group 1 and group 2 schools we now focus on group 1 and 2 schools only.
- Group 1 gets treatment in 1998 group 2 in 1999; $T = 1$ if treated in 1998.
- $E\left[Y_{ijt} \mid T = 1, D = 0\right] - E\left[Y_{ijt} \mid T = 0, D = 0\right]$
$= \beta_1 + \sum_d \left[E\left[N^T \middle| T = 1, D = 0\right] - E\left[N^T \middle| T = 0, D = 0\right]\right]$
$+ \sum_d \phi_d \left[E\left[N \mid T = 1, D = 0\right] - E\left[N \mid T = 0, D = 0\right]\right]$
$+ \left[E\left[\varepsilon \mid T = 1, D = 0\right] - E\left[\varepsilon \mid T = 0, D = 0\right]\right]$
- β_1 is the within school externality effect.
- The second and third term are due to differing local densities of primary schools between treatment and control and will be close to 0 (they also control for those differences).
- The last term will also be close to 0.

Econometric Specification - Within-School Externalities

- Group 1 and Group 2 pupils who missed their first year of treatment can therefore be used to obtain an estimate of within school externalities.
- The authors can estimate both within-school and cross school externalities using equation (3):

$$
\begin{aligned}
Y_{ijt} = {} & \alpha + \beta_1 T_{1it} + b_1 D_{it} + b_2 (T_{1it} \times D_{it}) + X'_{ijt}\delta \\
& + \sum_d \left(\gamma_d N^T_{dit} \right) + \sum_d (\phi_d N_{dit}) + u_i + e_{ijt}
\end{aligned}
\tag{3}
$$

β_1 is the within-school externality effect on the untreated.
$\beta_1 + b_2$ is within-school externality + direct effect on the treated.
γ_d is the cross-school externality.

Results - No Within-School Externalities

	Any moderate-heavy helminth infection, 1999	
	(1)	(2)
Indicator for Group 1 (1998 Treatment) School	-0.25***	-0.12*
	(0.05)	(0.07)
Group 1 pupils within 3 km (per 1000 pupils)	-0.26***	-0.26***
	(0.09)	(0.09)
Group 1 pupils within 3-6 km (per 1000 pupils)	-0.14**	-0.13**
	(0.06)	(0.06)
Total pupils within 3 km (per 1000 pupils)	0.11***	0.11***
	(0.04)	(0.04)
Total pupils within 3-6 km (per 1000 pupils)	0.13**	0.13**
	(0.06)	(0.06)
Received first year of deworming treatment, when offered (1998 for Group 1, 1999 for Group 2)		-0.06*
		(0.07)
(Group 1 Indicator)*Received treatment, when offered		-0.14*
		(0.07)
	Equation (2) no within-school externality	

Results - Including Within-School Externalities

	Any moderate-heavy helminth infection, 1999	
	(1)	(2)
Indicator for Group 1 (1998 Treatment) School	-0.25***	-0.12*
	(0.05)	(0.07)
Group 1 pupils within 3 km (per 1000 pupils)	-0.26***	-0.26***
	(0.09)	(0.09)
Group 1 pupils within 3-6 km (per 1000 pupils)	-0.14**	-0.13**
	(0.06)	(0.06)
Total pupils within 3 km (per 1000 pupils)	0.11***	0.11***
	(0.04)	(0.04)
Total pupils within 3-6 km (per 1000 pupils)	0.13**	0.13**
	(0.06)	(0.06)
Received first year of deworming treatment, when offered (1998 for Group 1, 1999 for Group 2)		-0.06*
		(0.07)
(Group 1 Indicator)*Received treatment, when offered		-0.14*
		(0.07)
		Equation (3) no within-school externality