10 – RANDOMIZED CONTROLLED TRIALS

University of Massachusetts Amherst
BE REVOLUTIONARY™

THE PLAN

1. The potential outcomes framework.
2. Selection bias.
3. Randomization.
4. Example 1: The STAR Experiment.
5. Threats to the validity of RCTs.
6. Example 2: The miracle of microfinance? (Banerjee et al, 2015)

THE EFFECT OF HEALTH INSURANCE

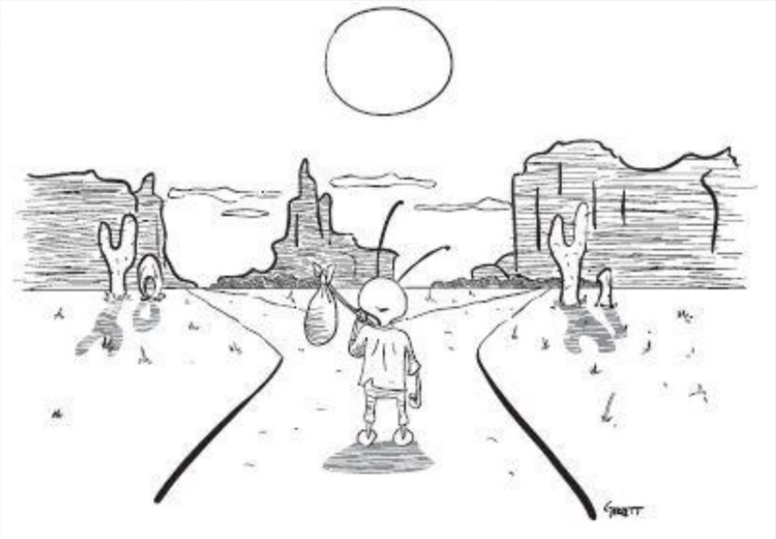
- Does health insurance make people healthier?
- Let's look at the data
  - National Health Interview Survey – NHIS.
  - *Observational* data.
- Is it an apples-to-apples comparison?
  - No *balance* in average characteristics

	Some HI (1)	No HI (2)	Difference (3)
		A. Health	
Health index	4.01 [.93]	3.70 [1.01]	.31 (.03)
		B. Characteristics	
Nonwhite	.16	.17	−.01 (.01)
Age	43.98	41.26	2.71 (.29)
Education	14.31	11.56	2.74 (.10)
Family size	3.50	3.98	−.47 (.05)
Employed	.92	.85	.07 (.01)
Family income	106,467	45,656	60,810 (1,355)
Sample size	8,114	1,281	

10.1 THE POTENTIAL OUTCOMES FRAMEWORK

University *of* Massachusetts Amherst

THE POTENTIAL OUTCOMES FRAMEWORK



- Consider a *binary* treatment
  - getting a covid vaccine VS being unvaccinated
  - having health insurance VS not having it
  - …
- Indicator variable D_i represents treatment status

$$D_i = \begin{cases} 1 \text{ if } i \text{ gets treated} \\ 0 \text{ if } i \text{ not treated} \end{cases}$$

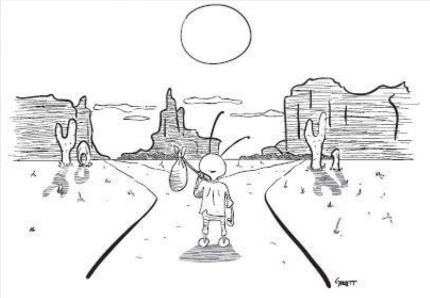
- For each unit *i*, two *potential outcomes*:

$$\text{Potential Outcomes:} \begin{cases} Y_{1i} \text{ if } D_i = 1 \\ Y_{0i} \text{ if } D_i = 0 \end{cases}$$

i's outcome in a world in which *i* gets treated.

i's outcome in a world in which *i* doesn't get treated.

THE POTENTIAL OUTCOMES FRAMEWORK



$$\text{Potential Outcomes:} \begin{cases} Y_{1i} \\ Y_{0i} \end{cases}$$

- $Y_{1i} - Y_{0i}$ = *causal effect* of treatment *D* on outcome *Y* for individual *i.*
- $\mathrm{E}(Y_{1i} - Y_{0i})$= *average causal effect* (ATE) in a population.
- $\mathrm{E}(Y_{1i} - Y_{0i}) = Avg(Y_{1i} - Y_{0i}) = \frac{1}{n}\sum_{i=1}^{n}[Y_{1i} - Y_{0i}] = \frac{1}{n}\sum_{i=1}^{n} Y_{1i} - \frac{1}{n}\sum_{i=1}^{n} Y_{0i}$

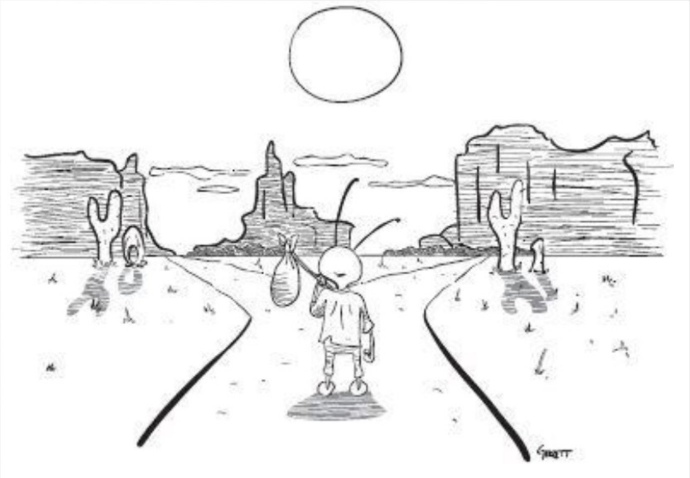
THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

- Estimating $\mathrm{E}(Y_{1i} - Y_{0i})$ from a sample would require observing both Y_{1i} & Y_{0i} for each individual in the sample.
- *The fundamental problem of causal inference*:

  you can't observe both Y_{1i} & Y_{0i} for the same *i*

- What we *can* observe is Y_i

$$Y_i = \begin{cases} Y_{1i} \;\; if\, D_i = 1 \\ Y_{0i} \;\; if\, D_i = 0 \end{cases} = Y_{0i} + D_i(Y_{1i} - Y_{0i})$$



10.2 SELECTION BIAS

University of Massachusetts Amherst

HEALTH INSURANCE & SELECTION BIAS

- Back to our initial Q: does health insurance make people healthier?
- What can we learn from observational data?
- How should we interpret the substantial difference in health index between insured vs. uninsured?

	Some HI (1)	No HI (2)	Difference (3)
			A. Health
Health index	4.01	3.70	.31
	[.93]	[1.01]	(.03)
			B. Characteristics
Nonwhite	.16	.17	−.01
			(.01)
Age	43.98	41.26	2.71
			(.29)
Education	14.31	11.56	2.74
			(.10)
Family size	3.50	3.98	−.47
			(.05)
Employed	.92	.85	.07
			(.01)
Family income	106,467	45,656	60,810
			(1,355)
Sample size	8,114	1,281	

SELECTION BIAS

- Potential Outcomes: $\begin{cases} Y_{1i} \\ Y_{0i} \end{cases}$

- Average causal effect in a population: $\mathrm{E}(Y_{1i} - Y_{0i})$

- Observed Outcome $Y_i = \begin{cases} Y_{1i} & if\, D_i = 1 \\ Y_{0i} & if\, D_i = 0 \end{cases} = Y_{0i} + D_i(Y_{1i} - Y_{0i})$

- What if we compare outcomes for treated vs. untreated individuals?

$$Difference\ in\ group\ means = Average\ Causal\ Effect + Selection\ Bias$$

SELECTION BIAS

- Comparison of observed outcomes for treated vs. untreated:

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$$

- In terms of potential outcomes:

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0)$$

- This can be linked to the average causal effect by rewriting it as follows:

$$= \boxed{E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)} \quad \text{Average causal effect}$$

$$+ \boxed{E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)} \quad \text{Selection bias}$$

SELECTION BIAS

- Suppose the causal effect of treatment is constant (=same for all individuals)

$$Y_{1i} = Y_{0i} + \kappa \quad \rightarrow \quad Y_{1i} - Y_{0i} = \kappa$$

- Then a difference in group means (treated vs untreated) gives

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = \kappa + E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)$$

- *Selection bias* reflects systematic differences between the units in the treated group (D=1) and the units in the control group (D=0).

- Systematic differences imply that average outcomes would have differed *even in the absence of treatment*

  > $E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0) \neq 0$

REGRESSION & SELECTION BIAS

- Consider the following OLS regression

$$Y_i = \beta_0 + \beta_1 D_i + u_i$$

  - Does $\hat{\beta}_1$ provide a good estimate of the causal effect of treatment κ?

- We know from Section 4 that $\beta_1 = E(Y|D = 1) - E(Y|D = 0)$

- Therefore $\mathrm{E}(\hat{\beta}_1) = \kappa + E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)$

- This regression is just a comparison of group means, so it conflates the average causal effect of treatment with selection bias.

- Selection bias is another way to say that $corr(D_i, u_i) \neq 0$

10.3 RANDOMIZATION

University *of* Massachusetts Amherst

RANDOMIZATION KILLS SELECTION BIAS

- *Random assignment* of D_i: every individual in the population has the same probability of receiving treatment.

> treated & untreated units come from the same population.

> treated & untreated have same *expected* characteristics.

> $E(Y_{0i}|D_i = 1) = E(Y_{0i}|D_i = 0)$

- *Random assignment eliminates selection bias*

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = \kappa + \cancel{E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)}$$

RANDOMIZATION KILLS SELECTION BIAS

- In a *Randomized Controlled Trial* (RCT), treatment D_i is randomly assigned by the researcher.
- Given randomization, the comparison

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$$

provides an unbiased estimate of the average causal effect.

- With experimental data, the average causal effect can be estimated by running

$$Y_i = \beta_0 + \beta_1 D_i + u_i$$

> $E(\hat{\beta}_0) = E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = \kappa$

REGRESSION ANALYSIS OF RCT DATA

- If the treatment is randomized, the average causal effect of treatment can be estimated through OLS regression

$$Y_i = \beta_0 + \beta_1 D_i + u_i$$

$$\mathrm{E}(\hat{\beta}_1) = E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$$

$$= E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1) = \kappa$$

- Randomization ensures that $\mathrm{corr}(D_i, u_i) = 0$

REGRESSION ANALYSIS OF RCT DATA

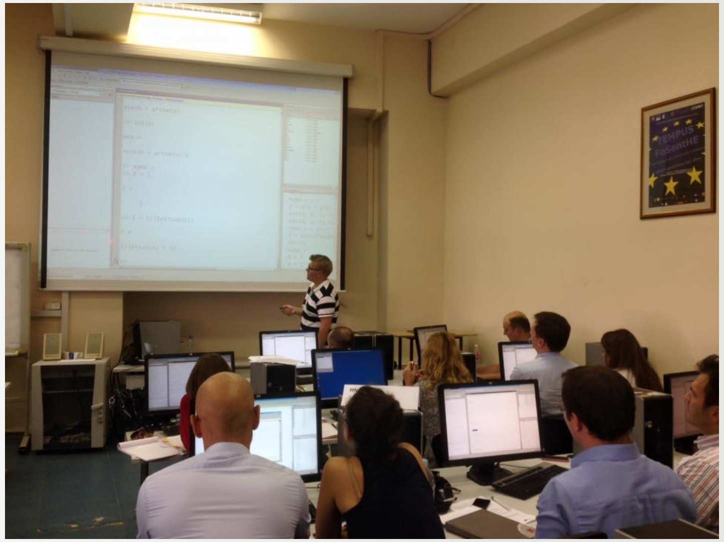
- What if we add control variables?

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 W_{1i} + \cdots + \beta_{1+r} W_{ri} + u_i$$

- With *full randomization*, controls are not needed for unbiasedness & consistency, but can still be useful to increase precision (lower SEs).

- With *randomization based on covariates*, controls are needed to eliminate selection bias.
  - probability of assignment depends on W_i, but X_i is randomly assigned *given* W_i.

RANDOMIZATION BASED ON COVARIATES: EXAMPLE



- Treatment: mandatory (vs optional) econometrics course.
- Outcome: post-graduation earnings.
- Treatment is randomized *except that* econ majors are more likely to receive treatment than non-econ majors.
- → selection bias if econ majors have different expected earnings.
- Controlling for binary variable W (=1 for econ majors) eliminates bias.

10.4 THE STAR EXPERIMENT

University *of* Massachusetts Amherst

THE STAR EXPERIMENT

- 4-year study, $12 million
- 80 schools in Tennessee.
- Students randomly assigned to 3 groups
    1. regular class (22 – 25 students)
    2. regular class + aide
    3. small class (13 – 17 students)
- regular class students re-randomized after first year to regular or regular + aide
- Y = Stanford Achievement Test scores

THE STAR EXPERIMENT

- Regression model:

$$Y_i = \beta_0 + \beta_1 SmallClass_i + \beta_2 RegAide_i + u_i$$

  - $SmallClass_i = 1$ if in a small class
  - $RegAide_i = 1$ if in regular class with aide
  - SEs clustered by school.

ESTIMATED EFFECTS

TABLE 13.1 Project STAR: Differences Estimates of Effect on Standardized Test Scores of Class Size Treatment Group

	Grade			
Regressor	K	1	2	3
Small class	13.90 (4.23) [5.48, 22.32]	29.78 (4.79) [20.24, 39.32]	19.39 (5.12) [9.18, 29.61]	15.59 (4.21) [7.21, 23.97]
Regular-sized class with aide	0.31 (3.77) [−7.19, 7.82]	11.96 (4.87) [2.27, 21.65]	3.48 (4.91) [−6.31, 13.27]	−0.29 (4.04) [−8.35, 7.77]
Intercept	918.04 (4.82)	1039.39 (5.82)	1157.81 (5.29)	1228.51 (4.66)
Number of observations	5786	6379	6049	5967

The regressions were estimated using the Project STAR public access data set described in Appendix 13.1. The dependent variable is the student's combined score on the math and reading portions of the Stanford Achievement Test. Standard errors, clustered at the school level, appear in parentheses, and 95% confidence intervals appear in brackets.

ADDING CONTROL VARIABLES

TABLE 13.2 Project STAR: Differences Estimates with Additional Regressors for Kindergarten

Regressor	(1)	(2)	(3)	(4)
Small class	13.90 (4.23) [5.48, 22.32]	14.00 (4.25) [5.55, 22.46]	15.93 (4.08) [7.81, 24.06]	15.89 (3.95) [8.03, 23.74]
Regular-sized class with aide	0.31 (3.77) [−7.19, 7.82]	−0.60 (3.84) [−8.25, 7.05]	1.22 (3.64) [−6.04, 8.47]	1.79 (3.60) [−5.38, 8.95]
Teacher's years of experience		1.47 (0.44) [0.60, 2.34]	0.74 (0.35) [0.04, 1.45]	0.66 (0.36) [−0.05, 1.37]
Boy				−12.09 (1.54)
Free lunch eligible				−34.70 (2.47)
Black				−25.43 (4.52)
Race other than black or white				−8.50 (12.64)
School indicator variables?	no	no	yes	yes
$\bar{R}^2$	0.01	0.02	0.22	0.28
Number of observations	5786	5766	5766	5748

The regressions were estimated using the Project STAR public access data set described in Appendix 13.1. The dependent variable is the student's combined test score on the math and reading portions of the Stanford Achievement Test. All regressions include an intercept (not reported). The number of observations differs in the different regressions because of some missing data. Standard errors, clustered at the school level, appear in parentheses, and 95% confidence intervals appear in brackets.

HOW BIG ARE THESE ESTIMATED EFFECTS?

- Put on same basis by dividing by std. dev. of *Y*
- Units are now standard deviations of test scores

TABLE 13.3 Estimated Class Size Effects in Units of Standard Deviations of the Test Score Across Students

	Grade			
Treatment Group	**K**	**1**	**2**	**3**
Small class	0.19 (0.06)	0.33 (0.05)	0.23 (0.06)	0.21 (0.06)
Regular-sized class with aide	0.00 (0.05)	0.13 (0.05)	0.04 (0.06)	0.00 (0.06)
Sample standard deviation of test scores (s_Y)	73.75	91.25	84.08	73.27

The estimates and standard errors in the first two rows are the estimated effects in Table 13.1, divided by the sample standard deviation of the Stanford Achievement Test for that grade (the final row in this table), computed using data on the students in the experiment. Standard errors, clustered at the school level, appear in parentheses.

COMPARISON WITH MA & CA OBSERVATIONAL STUDIES

Estimated Effects of Reducing the Student–Teacher Ratio by 7.5 SDs

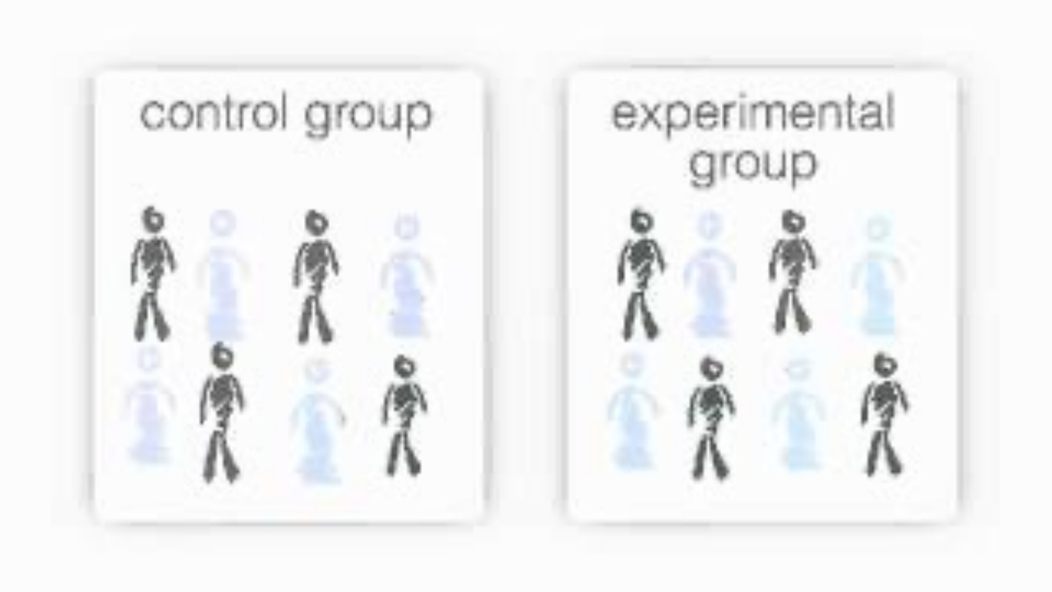
Study	Effect	Change in Student–Teacher Ratio	Standard Deviation of Test Scores Across Students	Estimated Effect	95% Confidence Interval
STAR (grade K)	–13.90** (2.45)	Small class vs. regular class	73.8	0.19** (0.03)	(0.13, 0.25)
California	– 0.73** (0.26)	–7.5	38.0	0.14** (0.05)	(0.04, 0.24)
Massachusetts	– 0.64* (0.27)	–7.5	39.0	0.12* (0.05)	(0.02, 0.22)

10.5 THREATS TO VALIDITY OF RCT$_S$

WHAT MAKES AN RCT CONVINCING?

THREATS TO INTERNAL VALIDITY OF A RCT

- Failure to randomize.
- Deviations from treatment protocol.
- Attrition.
- Experimental effects.
- Spillover effects.
- Small sample size.



CHECKING FOR BALANCE

a) Comparison of sample averages of pre-treatment characteristics & outcomes.

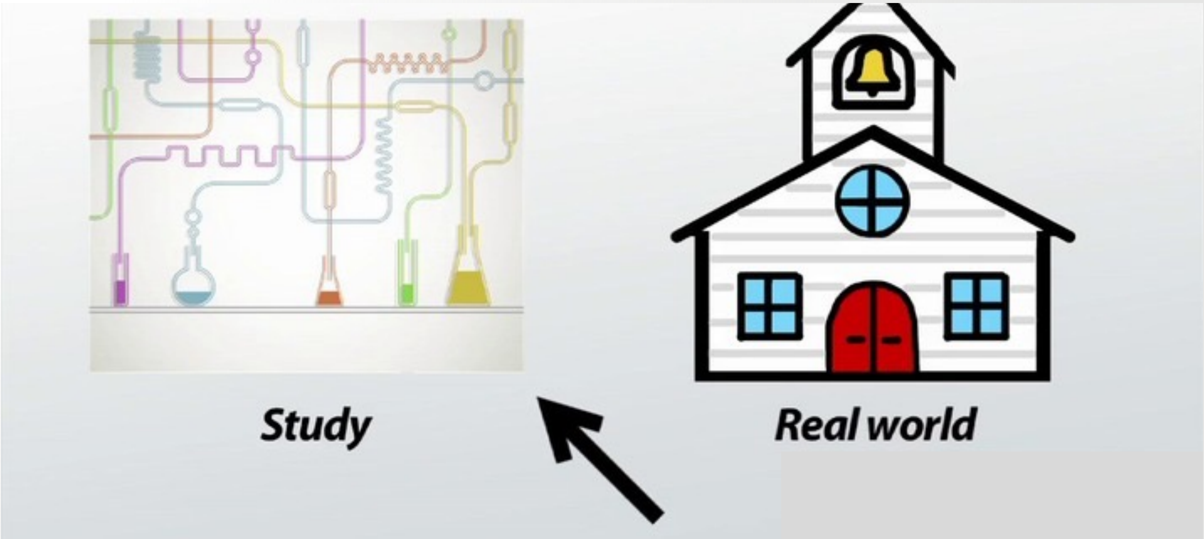
	Treatment (1)	Control (2)	Difference (3)
Panel A. Teacher attendance			
School open	0.66	0.64	0.02
			(0.11)
	41	39	80
Panel B. Student participation (random check)			
Number of students present	17.71	15.92	1.78
			(2.31)
	27	25	52
Panel C. Teacher qualifications			
Teacher test scores	34.99	33.54	1.44
			(2.02)
	53	54	107

b) Regression of treatment indicator on pre-treatment covariates:

$$D_i = \beta_0 + \beta_1 W_{1i} + \cdots + \beta_n W_{ni} + u_i$$

THREATS TO EXTERNAL VALIDITY OF A RCT

- Nonrepresentative sample.
- Nonrepresentative program or policy.
- Scaling-up (“general equilibrium”) effects.



10.6 THE MIRACLE OF MICROFINANCE? (BANERJEE ET AL, 2015)

University *of* Massachusetts Amherst

The Miracle of Microfinance? Evidence from a Randomized Evaluation

Abhijit Banerjee

Esther Duflo

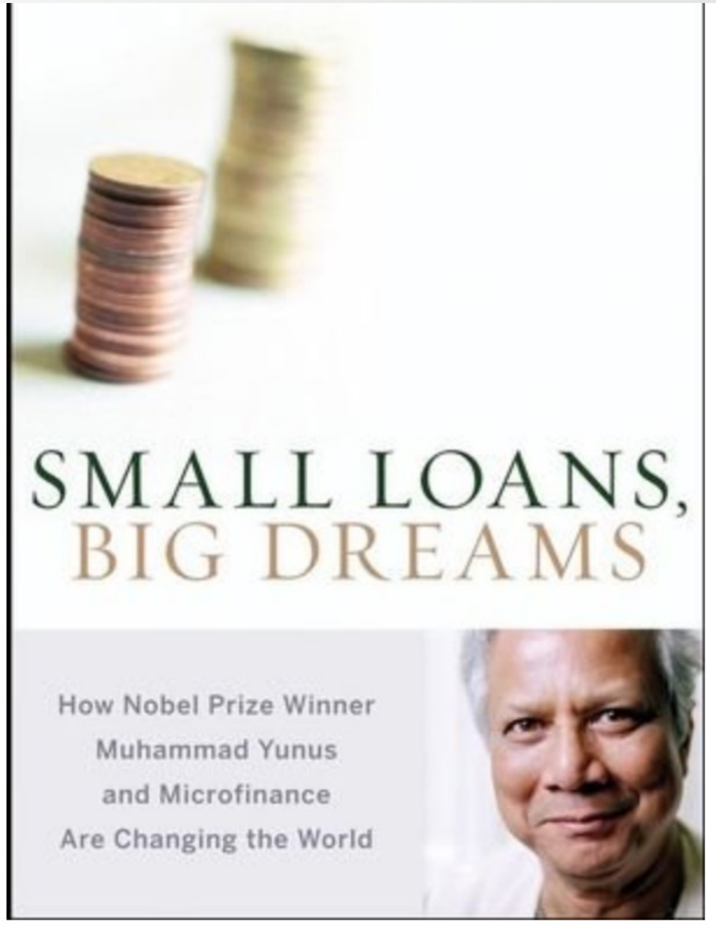
Rachel Glennerster

Cynthia Kinnan

AMERICAN ECONOMIC JOURNAL: APPLIED ECONOMICS
VOL. 7, NO. 1, JANUARY 2015
(pp. 22-53)

THE MIRACLE OF MICROFINANCE?

- Microfinance: small loans to low-income households & small businesses who banks wouldn't lend to.
- A cure for poverty and underdevelopment?
  - 2006 Nobel Peace Prize
- But how do we assess its effects?



THE HYDERABAD MICROFINANCE EXPERIMENT

- 104 poor neighborhoods in Hyderabad, India.
- 52 randomly selected for opening of MFI (*Spandana*) branch.

Surveyed random samples of households in three waves:

1. ≈ 2,800 before the program (baseline).
2. ≈ 6,800 15/18 month after program start.
3. Same 6,800 re-interviewed 3 years after program start.

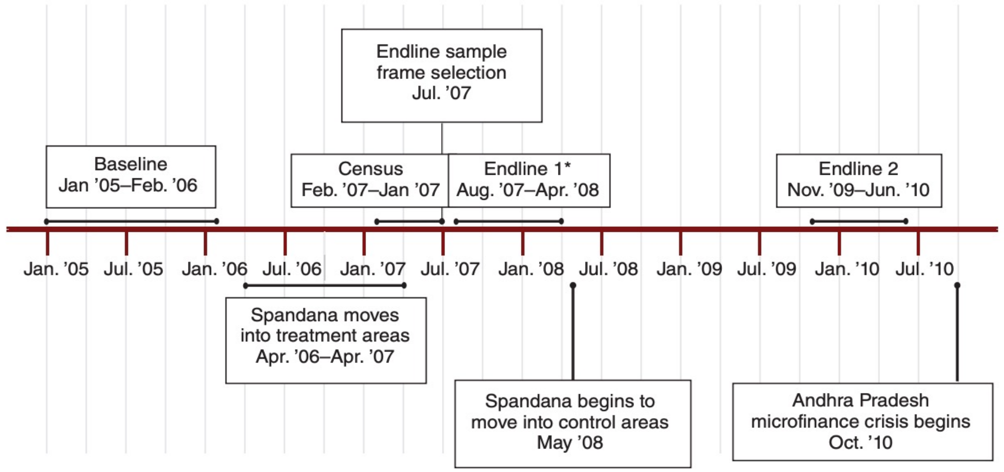


FIGURE 1. TIMELINE OF INTERVENTION AND DATA COLLECTION

Note: No treatment area was surveyed for endline 1 until at least one year had elapsed from the start of Spandana lending in that area.

CHECKING FOR BALANCE

TABLE 1A—BASELINE SUMMARY STATISTICS

	Control group			Treatment − control	
	Obs. (1)	Mean (2)	SD (3)	Coeff. (4)	*p*-value (5)
Household composition					
Number members	1,220	5.038	(1.666)	0.095	0.303
Number adults (>=16 years old)	1,220	3.439	(1.466)	−0.011	0.873
Number children (<16 years old)	1,220	1.599	(1.228)	0.104	0.098
Male head	1,216	0.907	(0.290)	−0.012	0.381
Head's age	1,216	41.150	(10.839)	−0.243	0.676
Head with no education	1,216	0.370	(0.483)	−0.008	0.787
Access to credit					
Loan from Spandana	1,213	0.000	(0.000)	0.007	0.195
Loan from other MFI	1,213	0.011	(0.103)	0.007	0.453
Loan from a bank	1,213	0.036	(0.187)	0.001	0.859
Informal loan	1,213	0.632	(0.482)	0.002	0.958
Any type of loan	1,213	0.680	(0.467)	0.002	0.942
Amount borrowed from (in Rs)					
Spandana	1,213	0	(0.000)	69	0.192
Other MFI	1,213	201	(2,742)	170	0.568
Bank	1,213	7,438	(173,268)	−5,420	0.279
Informal loan	1,213	28,460	(65,312)	−570	0.856
Total	1,213	37,892	(191,292)	−5,879	0.343
Self-employment activities					
Number of activities	1,220	0.320	(0.682)	−0.019	0.579
Number of activities managed by women	1,220	0.145	(0.400)	−0.007	0.750
Share of HH activities managed by women	295	0.488	(0.482)	−0.006	0.904
Businesses					
Revenue/month (Rs)	295	15,991	(53,489)	4,501	0.539
Expenses/month (Rs)	295	3,617	(26,144)	641	0.751
Investment/month (Rs)	295	385	(3,157)	14	0.959
Employment (employees)	295	0.169	(0.828)	0.255	0.148
Self-employment (hours per week)	295	76.315	(66.054)	−4.587	0.414
Businesses (all households)					
Revenue/month (Rs)	1,220	3,867	(27,147)	904	0.626
Expenses/month (Rs)	1,220	875	(12,933)	116	0.812
Investment/month (Rs)	1,220	93	(1,559)	−0.098	0.999
Employment (employees)	1,220	0.041	(0.413)	0.057	0.166
Self-employment (hours per week)	1,220	18.453	(46.054)	−1.801	0.400
Consumption (per household per month)					
Total consumption (Rs)	1,220	4,888	(4,074)	270	0.232
Nondurables consumption (Rs)	1,220	4,735	(3,840)	252	0.235
Durables consumption (Rs)	1,220	154	(585)	18	0.531
Asset index	1,220	1.941	(0.829)	0.027	0.669

Notes: Unit of observation: household. Standard errors of differences, clustered at the area level, in parentheses. Sample includes all households surveyed at baseline. Informal lender includes moneylenders, loans from friends/family, and buying goods/services on credit from seller. Asset index is calculated on a list of 40 home durable goods. Each asset is given a weight using the coefficients of the first factor of a principal component analysis. The index, for a household i, is calculated as the weighted sum of standardized dummies equal to 1 if the household owns the durable good.

Source: Baseline household survey

REGRESSION ANALYSIS

- Regression for estimating the effects of micro-credit:

$$y_{ia} = \beta_0 + \beta_1 Treat_a + \gamma_1 W_{1a} + \cdots + \gamma_n W_{na} + u_{ia}$$

  - y_{ia} = outcome of interest for household *i* in area *a.*
  - $Treat_a$ = binary variable for living in a treated area.
  - $W_{1a}, W_{2a}, \ldots, W_{na}$ = control variables (to increase precision).
  - SEs clustered at the area level.
  - $\hat{\beta}_1$ estimates the *average causal effect* of microcredit access on *y*.

RESULTS

- Probability of receiving MFI loan higher by 8.4pp (+46%) in treatment areas.
  - 42% in treatment areas
  - 33% in control areas
- More investment in (existing) small businesses & durable goods.
- No effect on new businesses creation.
- No effect on economic and/or human development!
  - No effect on living standards (consumption).
  - No increase in investment in children's education.
  - No change in health.

POTENTIAL THREATS TO VALIDITY

- **Internal Validity**
  - Attrition & selective migration.
  - Baseline households different from 1^{st} & 2^{nd} wave households.
  - Some microfinance was available also in control areas.
  - Experiment estimates the effect of expanded & easier access to microcredit, not of introducing microcredit where there is none.
- **External Validity**
  - Context of very high economic growth.
  - For-profit microfinance model (unlike Yunus' Grameen Bank).
  - BUT results replicated in other settings (Morocco, Bosnia-Herzegovina, Mexico, Mongolia, Ethiopia)