Are Job-Training Programs Effective?

Donald B. Rubin

Harvard University

Presentation based on joint work with Fabrizia Mealli, Paolo Frumento, and Barbara Pacini. JASA 2012.

The National Job Corps Study

- Randomized study evaluating the effects of training program on employment and wages
  - Randomization assures fair comparison, in expectation, between treatment groups
  - Sampled youths (n=15,386) were assigned to a job training program (JC) or a control group
  - Only those assigned to JC could enroll in JC
- Post-treatment complications
  - Noncompliance - only 68% immediately enrolled and participated in JC
  - Wages undefined for unemployed
  - Missing outcomes due to nonresponse
  - Use tools of principal stratification and direct likelihood

Potential Outcomes Approach to Causal Inference – Simplest Setting

$T = 1 \Rightarrow$ active treatment (e.g., job training)

$T = 0 \Rightarrow$ control treatment (e.g., no training)

	Y(1)	Y(0)
1		
.		
Units .		
.		
N		
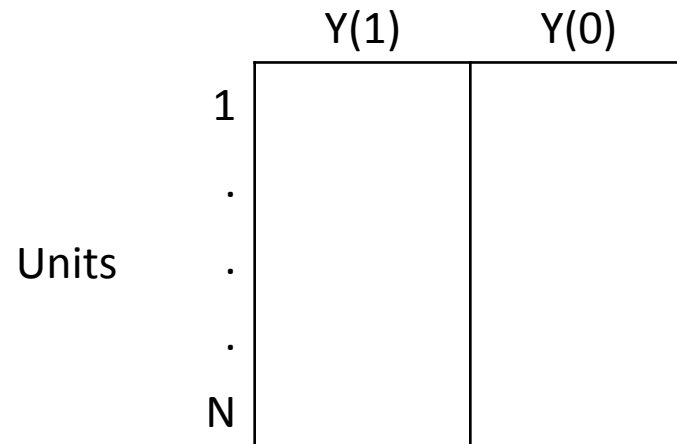
Y(1) = outcomes if exposed to active treatment

Y(0) = outcomes if exposed to control treatment

$\text{Ave}[Y_i(1) - Y_i(0)]$ = Average causal effect of
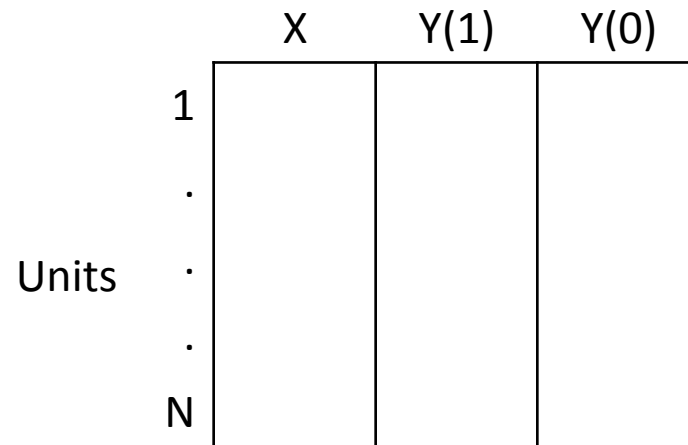
active versus control treatment

Potential Outcomes Approach to Causal Inference – Simplest Setting

- Fundamental problem of causal inference
  - For each i, only $Y_i(1)$ or $Y_i(0)$ can be observed

	Y(1)	Y(0)	T
1	✓	?	1
.	✓	?	1
. (Units)	✓	?	1
.	?	✓	0
.	?	✓	0
N	?	✓	0

- Random assignment of active versus control $\Longrightarrow$ representative sample of $Y_i(1)$ will be compared to representative sample of $Y_i(0)$

Potential Outcomes Approach to Causal Inference – Simplest Setting with Covariates

	X	Y(1)	Y(0)
1			
.			
.			
.			
N			

Units

- Same as before, except includes pretreatment covariates, e.g., age, sex, background education

$$\underset{i:\text{Female}}{\text{Ave}}[Y_i(1) - Y_i(0)] = \text{Average causal effect of active vs control treatment for females}$$

- Randomization still works for females

Potential Outcomes Approach to Causal Inference – Simple Noncompliance with Active Treatment

	D(1)	D(0)	Y(1)	Y(0)
1	1	0		
.	1	0		
Units .	1	0		
.	0	0		
N	0	0		

compliers

noncompliers

D(1) = treatment taken when assigned active treatment
$\quad$ D(1) = 1 $\Rightarrow$ active taken, D(1) = 0 $\Rightarrow$ control taken

D(0) = treatment taken when assigned control treatment
$\quad$ simple setting, always control $\Rightarrow$ D(0) = 0

$$\underset{i:D_i(1)=1}{\text{Ave}}\,[Y_i(1) - Y_i(0)] = \text{Average causal effect for true compliers}$$

- Randomization still works for compliers

Potential Outcomes Approach to Causal Inference – Simple Noncompliance with Active Treatment: Observed Data

Units		D(1)	D(0)	Y(1)	Y(0)	T	
	1	1	0	✓	?	1	} complier status observed
	.	1	0	✓	?	1	
	.	?	0	?	✓	0	} complier status missing
	.	?	0	?	✓	0	
	.	0	0	✓	?	1	} noncomplier status observed
	N	?	0	?	✓	0	} noncomplier status missing

- **Compliers**
  - For individuals assigned treatment (T=1), D(1)=1 & D(0)=0
  - For individuals assigned control (T=0), D(1)=? because true compliance under treatment is unknown & D(0)=0
- **Noncompliers**
  - For individuals assigned treatment (T=1), D(1)=0 & D(0)=0
  - For individuals assigned control (T=0), D(1)=? because true compliance under treatment is unknown & D(0)=0
- **Randomization still works for compliers**

Key Idea: Principal Stratification
(Frangakis and Rubin, 2002)

- Stratify on values of post-treatment intermediate outcome

- Convert $D_i(1)$, $D_i(0)$ into stratification variable
  - True complier "c" if $D_i(1)=1$
  - Noncomplier "n" if $D_i(1)=0$

- Idea works more generally

- Outgrowth of economics idea of instrumental variables, but more general

Intermediate Outcome - Employment

- Employed (yes, no) at a given time post-treatment is an important outcome, but is also needed to define principal strata for "final" outcomes, Y, describing attributes of possible employment, such as wages, retirement plan benefits, etc., which are not well-defined if unemployed
- Principal strata are defined by employment status
   - EE = employed whether assigned to training or not
   - EU = employed if trained, unemployed if not trained
   - UE = unemployed if trained, employed if not trained
   - UU = unemployed whether assigned to training or not
- Causal effects of training on Y only well-defined for EE
- UE empty? Reservation wage issue – Lock-in issue

Causal Effects of Training within Principal Strata

- Principal strata are defined by compliance with assignment to job training <u>and</u> by employment status
  - c&EE, c&EU, c&UE, c&UU
  - n&EE, n&EU, n&UE, n&UU
- By assumption (exclusion restriction on employment), we rule out n&EU and n&UE
  - If assignment does not affect entry into training, assignment cannot affect employment status
  - Also assume exclusion for attributes of employment, Y
- Causal effects of T on Y are only well-defined for c&EE and n&EE principal strata (no effect on Y in n&EE by exclusion restriction on Y)

Not Done Yet with Needed Principal Strata!

- Indicators for missing information about employment status and wages
  - M(1) and M(0), each indicating missing or not
- Do not make exclusion restriction on M
  - e.g., males could have $M_i(1) \Longrightarrow$ not missing if assigned training, but $M_i(0) \Longrightarrow$ missing if assigned control
- But do assume missing at random (MAR)
  - A nuisance, not of scientific interest

Causal Estimands Involving Assignment – define all estimands before defining estimators

Effect of assignment to be trained on being:
- job-trained

   $Pr(c)$ = proportion compliers

- employed

   $Pr(c\&EU) - Pr(c\&UE)$ (by exclusion on employment)

- employed for compliers

   $[Pr(c\&EU) - Pr(c\&UE)]/Pr(c)$

Causal Estimands Involving Wage Effects

- For the always employed

  $\text{Ave}[Y_i(1) - Y_i(0)|c\&EE \text{ or } n\&EE]$

- For the always employed compliers

  $\text{Ave}[Y_i(1) - Y_i(0)|c\&EE]$

- By exclusion on employment and wages, for the always employed noncompliers

  $\text{Ave}[Y_i(1) - Y_i(0)|n\&EE] = 0$

Descriptive Estimands

- Relative sizes of principal strata
  - c&EE, c&EU, c&UE, c&UU, n&EE, n&UU

- Distributions of X within principal strata

- Done with definition of estimands

- Now can worry about methods of estimation

Method of Analysis

- Direct likelihood at each of three post-treatment points in time

- Search for parsimonious model to help guide policy

- Needs scientific judgement

- Direct likelihood analyses support the conclusion that all six scientifically relevant principal strata exist

- Bayesian? Choice of prior daunting

Percent within Principal Strata by Time Period

Principal Stratum	c&EE	c&EU	c&UE	c&UU	n&EE	n&UU
Week 52	24	3	5	40	13	16
Week 130	29	7	5	30	14	15
Week 208	38	4	3	26	16	12

- For compliers, % EE increases in time, and % UU decreases
- For noncompliers, EE remains fairly stable
- Causal effect of training on employment negative at one year
  – Economists' "lock-in effect" during the period of training

Estimated Means of Covariates within Principal Strata Week 52

Principal Stratum	c&EE	c&EU	c&UE	c&UU	n&EE	n&UU
Percent in Stratum	24	3	5	40	13	16
Female	0.4	0.3	0.3	0.4	0.4	0.5
Age at baseline	19.0	19.0	19.3	18.4	19.5	18.8
White	0.3	0.4	0.3	0.2	0.3	0.2
With a Partner	0.1	0.0	0.0	0.0	0.1	0.1
Has children	0.2	0.1	0.1	0.2	0.2	0.3
Education	10.2	10.2	10.1	9.9	10.6	9.9
Ever arrested	0.2	0.3	0.3	0.2	0.3	0.3
Mother's education	11.7	11.7	11.6	11.4	11.6	11.5
Father's education	11.7	12.1	11.7	11.4	11.6	11.5
Household income > $6000	0.6	0.6	0.6	0.5	0.6	0.5
Person income > $6000	0.1	0.1	0.1	0.1	0.1	0.1
Have job	0.3	0.3	0.3	0.1	0.3	0.2
Had job, previous year	0.8	0.8	0.7	0.6	0.8	0.6
Months in Job, previous year	5.0	5.1	5.1	2.8	5.6	3.1
Earnings, previous year	3890	4112	4380	1974	4781	2508

Estimated Means of Covariates within Principal Strata
Week 130

Principal Stratum	c&EE	c&EU	c&UE	c&UU	n&EE	n&UU
Percent in Stratum	29	7	5	30	14	15
Female	0.4	0.3	0.2	0.5	0.4	0.5
Age at baseline	19.0	18.9	19.0	18.4	19.4	18.9
White	0.3	0.4	0.4	0.2	0.3	0.2
With a Partner	0.1	0.1	0.1	0.0	0.1	0.1
Has children	0.2	0.1	0.2	0.2	0.2	0.3
Education	10.2	10.1	10.1	9.8	10.5	10.0
Ever arrested	0.2	0.3	0.3	0.3	0.3	0.3
Mother's education	11.5	11.5	11.7	11.5	11.6	11.6
Father's education	11.5	11.8	12.0	11.5	11.5	11.6
Household income > $6000	0.5	0.6	0.6	0.5	0.6	0.5
Person income > $6000	0.1	0.1	0.1	0.1	0.1	0.1
Have job	0.3	0.3	0.2	0.2	0.3	0.2
Had job, previous year	0.7	0.7	0.7	0.6	0.8	0.6
Months in Job, previous year	4.4	5.0	4.2	2.9	5.0	3.4
Earnings, previous year	3221	4112	3756	2063	4291	2776

Estimated Means of Covariates within Principal Strata
Week 208

Principal Stratum	c&EE	c&EU	c&UE	c&UU	n&EE	n&UU
Percent in Stratum	38	4	4	26	16	12
Female	0.4	0.3	0.3	0.5	0.4	0.5
Age at baseline	18.9	18.4	18.5	18.5	19.3	18.9
White	0.3	0.4	0.4	0.2	0.3	0.2
With a Partner	0.1	0.0	0.1	0.0	0.1	0.1
Has children	0.2	0.1	0.2	0.2	0.2	0.2
Education	10.2	10.0	9.9	9.8	10.4	9.9
Ever arrested	0.2	0.3	0.4	0.3	0.3	0.3
Mother's education	11.6	11.7	11.6	11.4	11.6	11.6
Father's education	11.5	12.0	11.4	11.5	11.6	11.5
Household income > $6000	0.5	0.7	0.7	0.4	0.6	0.5
Person income > $6000	0.1	0.1	0.1	0.0	0.1	0.1
Have job	0.3	0.3	0.2	0.1	0.3	0.2
Had job, previous year	0.7	0.7	0.7	0.5	0.7	0.6
Months in Job, previous year	4.5	4.2	4.2	2.7	4.8	3.3
Earnings, previous year	3355	3667	3622	1884	4068	2731

Estimated Average Hourly Wages for Those Employed in Dollars within Principal Strata by Time Period

Principal Stratum	c&EE(1)	c&EE(0)	c&EU(1)	c&UE(0)	n&EE
Week 52	5.8	5.5	7.3	6.8	6.5
Week 130	6.7	6.4	9.2	7.2	7.9
Week 208	7.8	7.5	9.3	9.0	9.0

- Estimated causal effect on wages for always employed compliers is approximately 0.3 for all time periods (4-5% increase relative to wages with no JC)

- Always employed compliers, whether trained or not, have lower hourly wages than the sometimes employed compliers (c&EU or c&UE) or the always employed noncompliers (n&EE)

Final Conclusions for This Job Training Program

- In long run, for compliers, negligible effect on employment status

- For always employed compliers, minor positive effect on wages at all time periods

- Background characteristics of individuals differ substantially across principal strata

- Suggests need for more targeted programs

- Even if evaluation is based on randomized experiment, difficult to analyze correctly