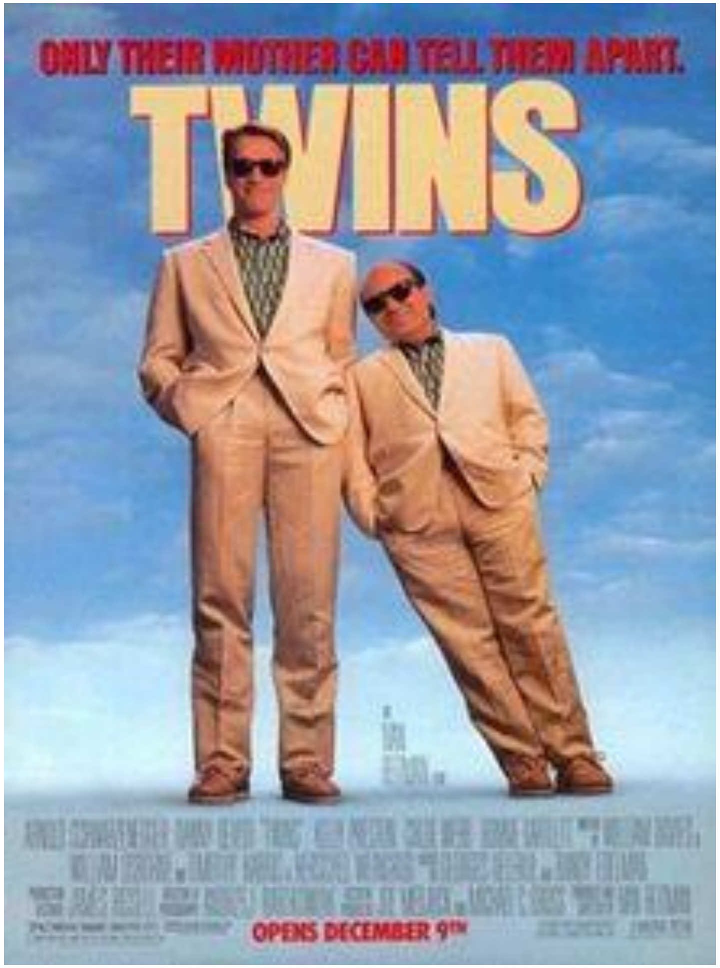
# Lecture 9: Propensity Score Matching

Edmund Malesky, Ph.D.

July 13, 2018

Duke University

# Matching Strategies



# What is Matching

- Tool to improve causal inference by estimating counterfactual
- Constructs artificial comparison group using statistical techniques
  - Assigns one or more nonparticipants to each participant
  - Matches are most similar based on observed characteristics
- Matched nonparticipants are used as the comparison group to estimate counterfactual
- Requires strong assumption: selection only on observables
  - Much Stronger assumption than Diff-in-Diff
  - Impossible to verify, but can assess validity
  - Most serious limitation of matching
- Generally less robust than DD/RDD/Randomized Experiments
  - Use in conjunction, or when others not possible

# Motivation

DEATH RATES PER 1,000 PERSON-YEARS

| Smoking group | Study<br>Canadian | British | U. S. |
|---|---|---|---|
| Non-smokers | 20.2 | 11.3 | 13.5 |
| Cigarettes only | 20.5 | 14.1 | 13.5 |
| Cigars, pipes | 35.5 | 20.7 | 17.4 |

MEAN AGES, YEARS

| Smoking group | Study<br>Canadian | British | U. S. |
|---|---|---|---|
| Non-smokers | 54.9 | 49.1 | 57.0 |
| Cigarettes only | 50.5 | 49.8 | 53.2 |
| Cigars and/or pipe | 65.9 | 55.7 | 59.7 |

Source: Cochran, 1968.

# Curse of Multidimensionality

| Treated units | | | |
|---|---|---|---|
| Age | Gender | Months unemployed | Secondary diploma |
| 19 | 1 | 3 | 0 |
| 35 | 1 | 12 | 1 |
| 41 | 0 | 17 | 1 |
| 23 | 1 | 6 | 0 |
| 55 | 0 | 21 | 1 |
| 27 | 0 | 4 | 1 |
| 24 | 1 | 8 | 1 |
| 46 | 0 | 3 | 0 |
| 33 | 0 | 12 | 1 |
| 40 | 1 | 2 | 0 |

| Untreated units | | | |
|---|---|---|---|
| Age | Gender | Months unemployed | Secondary diploma |
| 24 | 1 | 8 | 1 |
| 38 | 0 | 2 | 0 |
| 58 | 1 | 7 | 1 |
| 21 | 0 | 2 | 1 |
| 34 | 1 | 20 | 0 |
| 41 | 0 | 17 | 1 |
| 46 | 0 | 9 | 0 |
| 41 | 0 | 11 | 1 |
| 19 | 1 | 3 | 0 |
| 27 | 0 | 4 | 0 |

- So many things to compare, what is most important for matching?
- Can compare participants/nonparticipants sharing observables
- But with many variables, hard to find good match
- Often tough to find two identical households
- Propensity score matching solves this problem

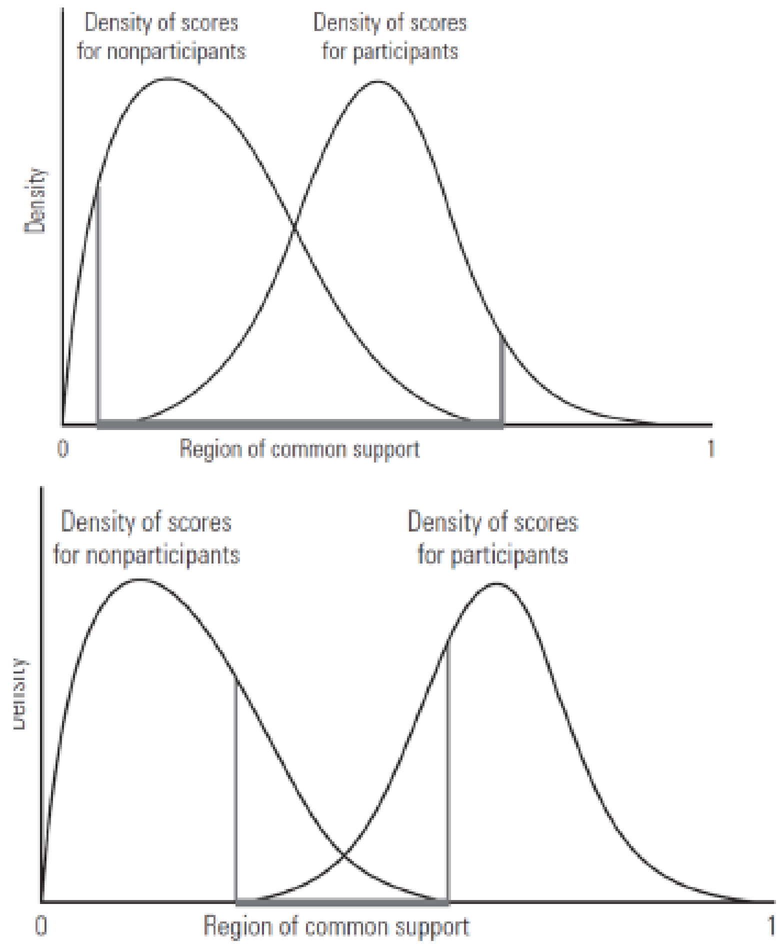
# Propensity Score Matching (PSM)

- Matches on *probability of participation* in intervention, based on observables
- *Propensity Score, or P(X):* probability that unit will participate in program based on observable characteristics
  - Single # summarizes all observables influencing participation
- PSM matches participants to nonparticipants with "closest" P(X)
- Validity of PSM depends on two key assumptions
  1. **Conditional independence:** $(Y_i^T, Y_i^C) \perp T_i \mid X_i$
  2. **Common support:** $0 < P(T_i = 1 \mid X_i)$

1. **Conditional independence:** given set of observable covariates X that are not affected by treatment, potential outcomes Y are independent of (orthogonal to) treatment assignment T
2. **Common Support:** Uptake of program entirely based on observables

# PSM & Common Support

- Common support ensures participants have nonparticipants with “close” P(X)
- Lack of common support appears in tails of distributions
- Larger sample of eligible nonparticipants helps matching
- Poor common support can induce bias in matching estimator
  - E.g., if no matches may drop nonrandom subset of participants



# Steps to Implement PSM

1. Use comparable surveys of participants & nonparticipants
2. Pool samples & estimate probability of individual participating based on observables – i.e., propensity score, or P(X)
   – Specifically, we use a an adapted version of the OLS regression model that you had in section 1 of the class. There are two differences:
      i. The dependent variable (treatment) =1 if participant, and =0 if non-participant.
      ii. We use a *logit* or *probit* regression to estimate probability of participation for each member of the treatment group, based on observable.
3. Restrict sample to common support
4. Sort data by propensity score – P(X).
   - For each participant, locate nonparticipant(s) with similar P(X)
5. Compare Y (DV) for participants & their twins (matched comparison units).
6. Difference of average outcomes = effect on participants
7. Mean of individual impacts = estimated average treatment effect

# Multiple Techniques for PSM

Various techniques for matching participants and nonparticipants

1. nearest neighbor matching
2. caliper & radius matching
3. stratification & interval matching
4. kernel & local linear matching
5. genetic matching
6. Entropy balancing

While they vary in flavor and precision, they all generate pretty much the same matches.

# Getting PSM Right

- PSM only useful when observables believed to affect participation
  - Depends on targeting rules for intervention and factors for self-selection
  - Impossible to prove
  - Must understand context of selection; use surveys to evaluate
- Only as good as background characteristics used
  - The more data to match with the better; many Xs crucial
- Beware of ex-post matching
  - Matching must be done using baseline characteristics
  - Danger with ex-post surveys: participation may affect Xs
- Can combine matching with other methods, such as Diff-in-Diff
- Addresses selection bias due to time-invariant unobservables

# PSM vs Randomization

- Randomization does not require the *untestable* assumption of independence conditional on observables
- PSM requires large samples and good data:

1. Ideally, the same data source is used for participants and non-participants
2. Participants and non-participants have access to similar institutions and markets, and
3. The data include *X* variables capable of identifying program participation and outcomes.

# Back to the HISP Example

**Table 7.1 Estimating the Propensity Score Based on Observed Characteristics**

| **Dependent Variable: *Enrolled* = 1** | |
|---|---|
| **Explanatory variables / characteristics** | **Coefficient** |
| Head of household's age (years) | −0.022** |
| Spouse's age (years) | −0.017** |
| Head of household's education (years) | −0.059** |
| Spouse's education (years) | −0.030** |
| Head of household is female = 1 | −0.067 |
| Indigenous = 1 | 0.345** |
| Number of household members | 0.216** |
| Dirt floor = 1 | 0.676** |
| Bathroom = 1 | −0.197** |
| Hectares of land | −0.042** |
| Distance to hospital (km) | 0.001* |
| Constant | 0.664** |

*Source:* Authors.

*Note:* Probit regression. The dependent variable is 1 if the household enrolled in HISP, and 0 otherwise. The coefficients represent the contribution of each listed explanatory variable / characteristic to the probability that a household enrolled in HISP.

* Significant at the 5 percent level; ** Significant at the 1 percent level.

Source: Gertler et al., 2011.

# Health Insurance Subsidy Example

**Table 7.2 Case 7—HISP Impact Using Matching (Comparison of Means)**

| | Enrolled | Matched comparison | Difference | *t*-stat |
|---|---|---|---|---|
| Household health expenditures | 7.8 | 16.1 | −8.3 | −13.1 |

**Table 7.3 Case 7—HISP Impact Using Matching (Regression Analysis)**

| | Multivariate linear regression |
|---|---|
| Estimated impact on household health expenditures | −8.3**<br>(0.63) |

*Source:* Authors.

*Note:* Standard errors are in parentheses.

** Significant at the 1 percent level.

Source: Gertler et al., 2011.

# Jalan and Ravillion (2003)

- Each year, 4 million children under 5 die from diarrhea
  - Main cause: unsafe drinking water
- Paper examines effect of piped water in India
  - 1.5 MM child deaths/year due to diseases related to poor water
  - Highest number in world
- Finds lower prevalence/duration of diarrhea if piped water
- But health gains bypass families in poverty or w/ poorly educated mother
- Need complementary inputs, such as knowing to boil & store safely

# PSM in Practice

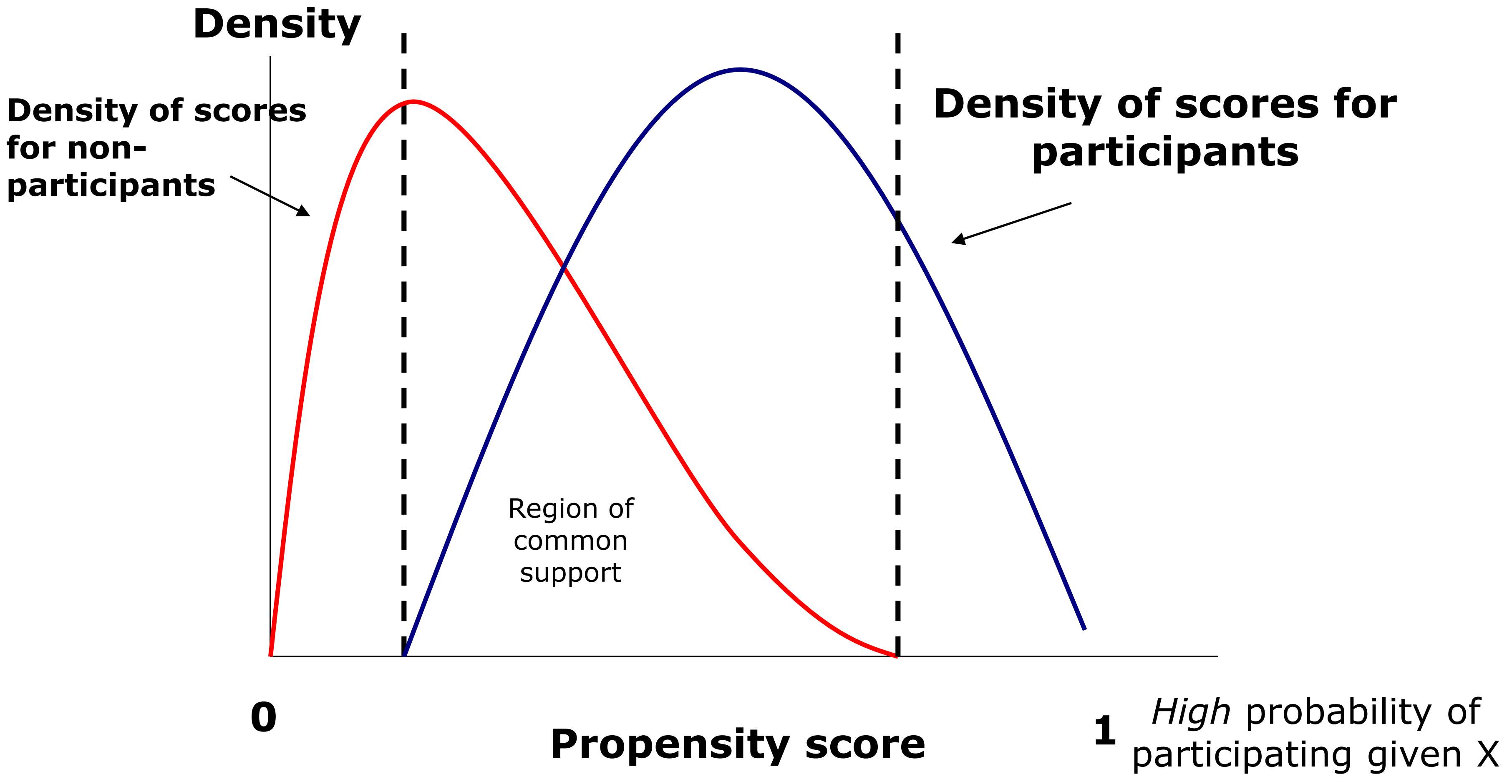
- To estimate the propensity score, authors used:
- Village level characteristics
  - Including: Village size, amount of irrigated land, schools, infrastructure (bus stop, railway station)
- Household variables
  - Including: Ethnicity / caste / religion, asset ownership (bicycle, radio, thresher), educational background of HH members
- Are there variables which can not be included?
  - Only using cross-section, so no variables influenced by project

# Estimating Propensity Score for Access to Clean Water

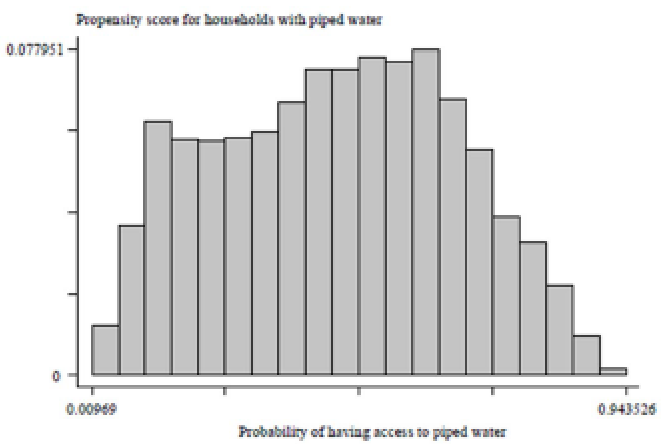
Logit regression for piped water

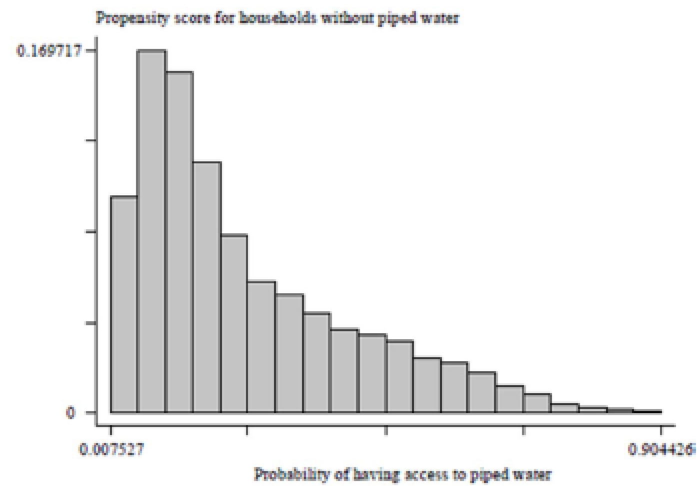
| | Coefficient | t-statistic |
|---|---|---|
| *Village variables* | | |
| Village size (log) | 0.08212 | 4.269 |
| Proportion of gross cropped area which is irrigated: > 0.75 | −0.04824 | −1.185 |
| Proportion of gross cropped area which is irrigated: 0.5–0.75 | 0.19399 | 4.178 |
| Whether village has a day care center | −0.07249 | −2.225 |
| Whether village has a primary school | −0.08136 | −1.434 |
| Whether village has a middle school | −0.09019 | −2.578 |
| Whether village has a high school | 0.26460 | 7.405 |
| Female to male students in the village | 0.10637 | 3.010 |
| Female to male students for minority groups | −0.07661 | −2.111 |
| Main approachable road to village: pucca road | 0.19441 | 3.637 |
| jeepable/kuchha road | −0.00163 | −0.033 |
| Whether bus-stoop is within the village | 0.11423 | 2.951 |
| Whether railway station is within the village | 0.00920 | 0.179 |
| Whether there is a post-office within the village | 0.02193 | 0.550 |
| Whether the village has a telephone facility | 0.33059 | 9.655 |
| Whether there is a community TV center in the village | 0.09859 | 2.661 |
| Whether there is a library in the village | −0.04153 | −1.116 |
| Whether there is a bank in the village | 0.19084 | 4.655 |
| Whether there is a market in the village | 0.31690 | 6.092 |
| Student teacher ratio in the village | 0.00242 | 5.295 |
| *Household variables* | | |
| Whether household belongs to the Scheduled Tribe | −0.21288 | −4.203 |
| Whether household belongs to the Scheduled Caste | −0.01045 | −0.288 |
| Whether it is a Hindu household | −0.24195 | −1.709 |
| Whether it is a Muslim household | −0.21631 | −1.427 |
| Whether it is a Christian household | 0.40367 | 2.426 |
| Whether it is a Sikh household | −0.86645 | −4.531 |
| Household size | 0.00337 | 0.571 |
| Utilization of landholdings: used for cultivation? | 0.17109 | 1.914 |
| Whether the house belongs to the household | −0.18988 | −2.854 |
| Whether the household owns other property | 0.00181 | 0.044 |
| Whether the household has a bicycle | −0.26514 | −8.243 |
| Whether the household has a sewing machine | 0.01183 | 0.252 |
| Whether the household owns a thresher | −0.05790 | −0.577 |
| Whether the household owns a winnower | 0.21842 | 1.820 |
| Whether the household owns a bullock-cart | −0.25900 | −5.430 |
| Whether the household owns a radio | 0.01036 | 0.251 |
| Whether the household owns a TV | 0.08095 | 1.335 |
| Whether the household owns a fan | 0.01336 | 0.321 |
| Whether the household owns any livestock | −0.07780 | −2.339 |
| [illegible] | [illegible] | [illegible] |

Source: Jalan & Ravallion, 2003.


Density
Density of scores for non-participants
Density of scores for participants
Region of common support
0
Propensity score
1
*High* probability of participating given X

# Common Support Assumption




Source: Jalan & Ravallion, 2003.

# Potential Unobserved Factors

- The behavioral factors – importance put on sanitation and behavioral inputs – are also likely correlated with whether a HH has piped water
- However, there are no behavioral variables in data: water storage, soap usage, latrines
  - These are unobserved factors NOT included in propensity score

# Results of Clean Water

Impacts of piped water on diarrhea prevalence and duration for children under five

| | Prevalence of diarrhea | | Duration of illness | |
|---|---|---|---|---|
| | Mean for those with piped water (st. dev.) | Impact of piped water (st. error) | Mean for those with piped water (st. dev.) | Impact of piped water (st. error) |
| Full sample | 0.0108 | −0.0023* | 0.3254 | −0.0957* |
| | (0.046) | (0.001) | (1.650) | (0.021) |
| *Stratified by household income per capita (quintiles)* | | | | |
| 1 (poorest) | 0.0155 | 0.0032* | 0.4805 | 0.0713 |
| | (0.055) | (0.001) | (2.030) | (0.053) |
| 2 | 0.0136 | 0.0007 | 0.4170 | 0.0312 |
| | (0.051) | (0.001) | (1.805) | (0.051) |
| 3 | 0.0083 | −0.0039* | 0.2636 | −0.1258* |
| | (0.038) | (0.001) | (1.418) | (0.042) |
| 4 | 0.0100 | −0.0036* | 0.3195 | −0.1392* |
| | (0.044) | (0.001) | (1.703) | (0.048) |
| 5 | 0.0076 | −0.0068* | 0.1848 | −0.2682* |
| | (0.042) | (0.001) | (1.254) | (0.036) |
| *Stratified by highest education level of a female member* | | | | |
| Illiterate | 0.0131 | −0.0000 | 0.3588 | −0.0904* |
| | (0.053) | (0.001) | (1.710) | (0.036) |
| At most primary school educated | 0.0112<br>(0.045) | −0.0015<br>(0.001) | 0.3502<br>1.739) | −0.0465<br>(0.036) |
| At most matriculation educated | 0.0074<br>(0.038) | −0.0065*<br>(0.001) | 0.2573<br>(1.476) | −0.1708*<br>(0.039) |
| Higher secondary or more | 0.0050<br>(0.027) | −0.0080*<br>(0.002) | 0.1880<br>(1.158) | −0.2077*<br>(0.076) |

*Indicates significance at the 5% level or lower.

Source: Jalan & Ravallion, 2003.

# Impact of Piped Water on Diarrhea

Child-health impacts of piped water by income and education

| | Illiterate | | At most primary | | At most matriculation | | Higher secondary or more | |
|---|---|---|---|---|---|---|---|---|
| | Prevalence of diarrhea | Duration of illness | Prevalence of diarrhea | Duration of illness | Prevalence of diarrhea | Duration of illness | Prevalence of diarrhea | Duration of illness |
| 1 (poorest quintile) | 0.0100* | 0.1028 | 0.0010 | 0.0548 | −0.0118* | −0.1091 | Small Sample | |
| | (0.002) | (0.089) | (0.002) | (0.094) | (0.003) | (0.132) | | |
| 2 | 0.0057* | 0.0777 | 0.0013 | 0.1061 | −0.0121* | −0.2580* | Small Sample | |
| | (0.003) | (0.083) | (0.002) | (0.083) | (0.002) | (0.087) | | |
| 3 | −0.0038* | −0.1503* | −0.0008 | 0.0056 | −0.0069* | −0.1659* | Small Sample | |
| | (0.002) | (0.069) | (0.002) | (0.081) | (0.002) | (0.059) | | |
| 4 | −0.0062* | −0.2224* | −0.0041* | −0.1691 | 0.0008 | −0.0186 | Small Sample | |
| | (0.002) | (0.097) | (0.002) | (0.070) | (0.003) | (0.091) | | |
| 5 | −0.0075* | −0.2932* | −0.0051* | −0.2435* | −0.0063* | −0.2578* | −0.010* | −0.2637* |
| | (0.000) | (0.045) | (0.002) | (0.075) | (0.002) | (0.008) | (0.003) | (0.085) |

*Note*: Figures in parentheses are the respective standard errors.
*Indicates significance at 5% or lower.

Source: Jalan & Ravallion, 2003.

# Impact of Water Privatization on Child Mortality

| | Full Sample (1) | Full Sample (2) | Full Sample (3) | Using Observations on Common Support (4) | Using Observations on Common Support (5) | Using Observations on Common Support (6) | Kernel Matching on Common Support[a] (7) |
|---|---|---|---|---|---|---|---|
| Private water services (=1) | −.334 | −.320 | −.283 | −.540 | −.541 | −.525 | −.604 |
| | (.169)** | (.170)* | (.170)* | (.177)*** | (.178)*** | (.178)*** | (.168)*** |
| | [.157]** | [.163]** | [.162]* | [.191]*** | [.198]*** | [.195]*** | |
| | \|.195\|* | \|.203\| | \|.194\| | \|.261\|** | \|.274\|** | \|.266\|** | |
| %Δ in mortality rate | −5.3 | −5.1 | −4.5 | −8.6 | −8.6 | −8.4 | −9.7 |
| Other covariates: | | | | | | | |
| Real GDP per capita | | .007 | .009 | | .005 | .006 | |
| | | (.005) | (.006) | | (.006) | (.006) | |
| | | [.006] | [.006] | | [.007] | [.007] | |
| | | \|.007\| | \|.007\| | | \|.007\| | \|.008\| | |
| Unemployment rate | | −.555 | −.636 | | −.778 | −.836 | |
| Income inequality | | 5.171 | 5.085 | | 2.932 | 3.052 | |
| | | (2.868)* | (2.880)* | | (2.907) | (2.926) | |
| | | [3.468] | [3.445] | | [3.314] | [3.289] | |
| | | \|3.696\| | \|3.691\| | | \|3.833\| | \|3.838\| | |
| Public spending per capita | | −.028 | −.035 | | −.068 | −.070 | |
| | | (.038) | (.038) | | (.039)* | (.039)* | |
| | | [.055] | [.055] | | [.059] | [.059] | |
| | | \|.054\| | \|.055\| | | \|.049\| | \|.050\| | |
| Local government by Radical party (=1) | | | .482 | | | .166 | |
| | | | (.267)* | | | (.284) | |
| | | | [.281]* | | | [.301] | |
| | | | \|.288\|* | | | \|.365\| | |
| Local government by Peronist party (=1) | | | −.202 | | | −.168 | |
| | | | (.191) | | | (.193) | |
| | | | [.202] | | | [.230] | |
| | | | \|.254\| | | | \|.309\| | |
| $R^2$ | .1227 | .1256 | .1272 | .1390 | .1415 | .1420 | |
| Observations | 4,732 | 4,597 | 4,597 | 3,970 | 3,870 | 3,870 | 3,970 |

Source: Galiani et al, 1995.

# Lessons on Matching Methods

- Typically used when neither randomization, RD or other quasi experimental options are not possible
  - Case 1: no baseline. Can do ex-post matching
  - Dangers of ex-post matching:
    - Matching on variables that change due to participation (i.e., endogenous)
    - What are some variables that won't change?
- Matching helps control only for OBSERVABLE differences, not unobservable differences

# More Lessons on Matching Methods

- Matching becomes much better in combination with other techniques, such as:
  - Exploiting baseline data for matching and using difference-in-difference strategy
  - If an assignment rule exists for project, can match on this rule
- Need good quality data
  - Common support can be a problem if two groups are very different

| Design | When to use | Advantages | Disadvantages |
|---|---|---|---|
| Randomization | Whenever feasible<br>When there is variation at the individual or community level | Gold standard<br>Most powerful | Not always feasible<br>Not always ethical |
| Randomized Encouragement Design | When an intervention is universally implemented | Provides exogenous variation for a subset of beneficiaries | Only looks at sub-group of sample<br>Power of encouragement design only known ex post |
| Regression Discontinuity | If an intervention has a clear, sharp assignment rule | Project beneficiaries often must qualify through established criteria | Only look at sub-group of sample<br>Assignment rule in practice often not implemented strictly |
| Difference-in-Differences | If two groups are growing at similar rates<br>Baseline and follow-up data are available | Eliminates fixed differences not related to treatment | Can be biased if trends change<br>Ideally have 2 pre-intervention periods of data |
| Matching | When other methods are not possible | Overcomes observed differences between treatment and comparison | Assumes no unobserved differences (often implausible) |