# Official Micro Data, Causal Inference and Evidence-Based Policy Making

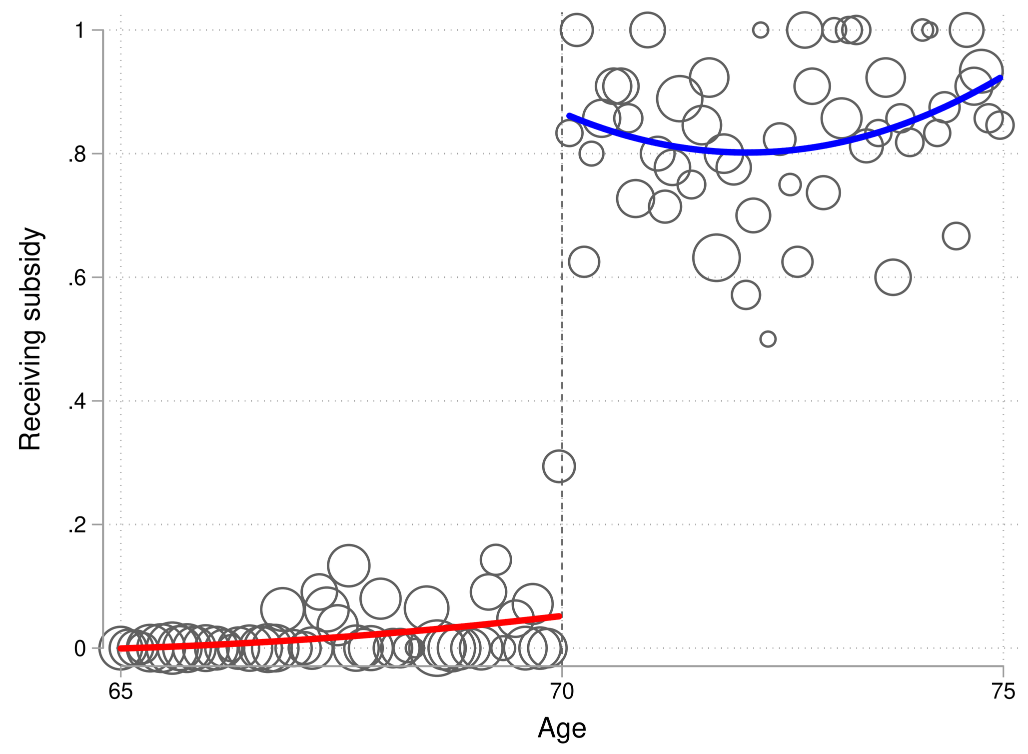
Junchao Zhang
zhang@ism.ac.jp

## Why is EBPM important?

- Before EBPM become widespread, policy makers actually do not know whether they are doing the right things.
- However, policy resources are not unlimited. Governments have to allocate resources among different fields to maximize the "economic pie".
- From the cross-sectional perspective, policy decision making implies trade-offs between different objectives/sub-populations.
  → Efficiency or equality?
  → Quantity or quality?
  → Aging or declining birthrates? etc.
- Whether policy interventions have causal effects on people's outcomes? Do these policies improve the economic and social well-being of people?

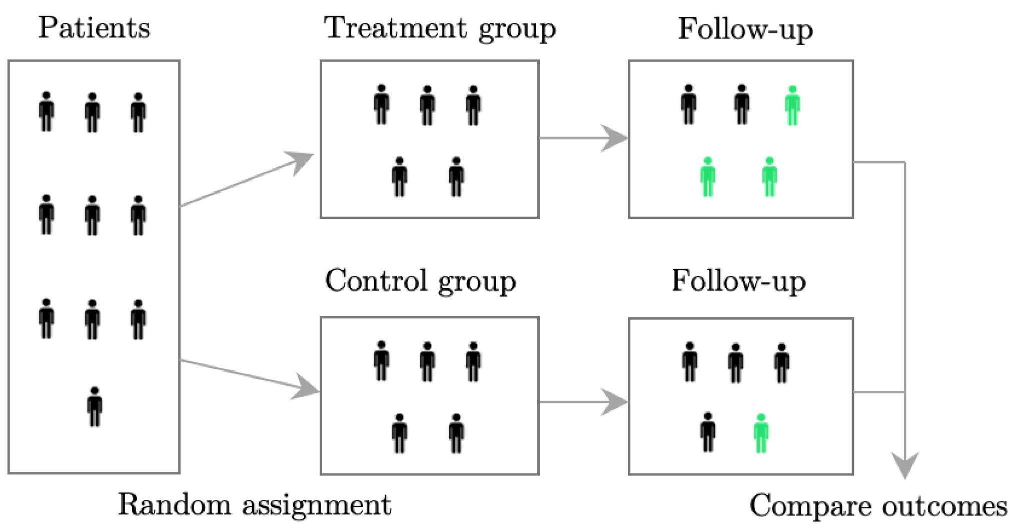
## Why do we need high-quality micro data?

- For instance, regression discontinuity design strongly relies on the continuous running variable, otherwise we can not clearly observe the cutoff point to distinguish treatment and control groups.




## Challenges in policy studies

- In hard sciences, a randomized controlled trial(RCT) is the gold standard for estimating treatment effects.

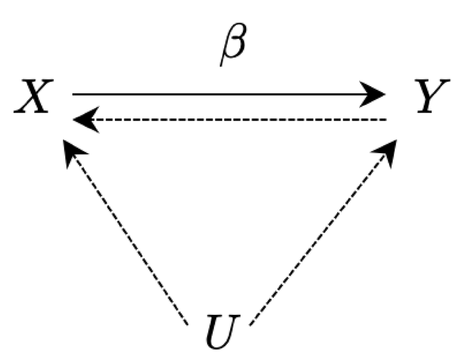



- We practically meet the following issues in policy studies.
  → RCT is not allowed because of law and ethics. Randomization is not implemented well.
  → We can not clearly define treatment and control groups using general survey data.
- Who are exposed to the policy intervention and who are not?

## What can we do with offical micro data?

- To estimate causal effects, we need detailed information on $Z$(e.g. exact date of birth, place of residence, etc.) to define treatment status $X$. Without $Z$, we can not tell whether one was exposed to a specific policy intervention.
- Advantages of official micro data
  → Raw data with detailed information on $Z$
  → Large sample size
- With offical micro data, we can simply do
  → Causal studies without RCT
  → More sub-sample analysis
  → GIS analysis, etc.
  that are useful for evidence-based policy making.

## Identification Problem

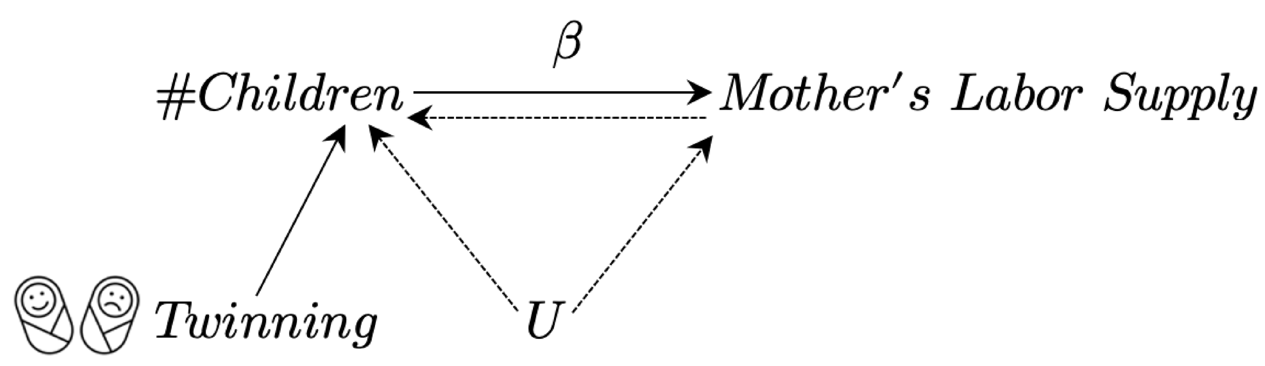



- Using RCT(if randomization is properly implemented), we can simply estimate the average treatment effects by comparing the outcomes between treatment and control groups, or by linear regression.

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

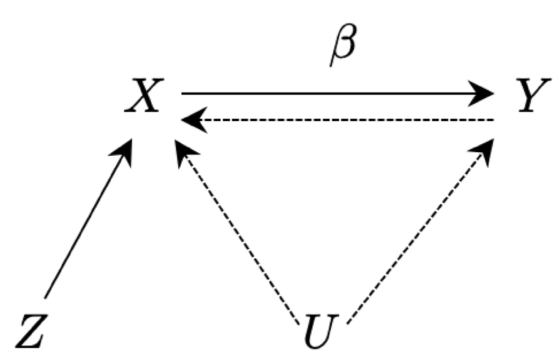
- However, $Cov(x_i, u_i) = 0$ condition is probably not satisfied in most cases of policy studies. Treatment variable $x_i$ is not independent to the error term $u_i$. With confounder $U$, $\beta_1$ reflects simple correlation rather than causality.

## Case study: #Children and labor suppy




- Data: Population census of Japan(2005, 2010, and 2015) that covers 100% population (other countries usually offer only 1-5% sample)
- Strategy: We use twinning as the instrument for number of children, which induces exogenous increase in number. Note that this strategy relies on twinning that needs very large sample size.

## Identification strategy




- Empirically, we need an exogenous variable $Z$, which can only affect $Y$ through $X$, to identity the causal parameter $\beta$ .
- $Z$ should randomly assign people into treatment and control.
- Identification of quasi-experiment design relies on rare events(sudden policy changes, weather events, natural disasters, etc.).
  → Regression discontinuity design
  → Difference-in-difference
  → Instrumental variable, etc.

## Comparison of OLS and causal estimates

**Table 1.** Effects of number of children on maternal labor supply by birth parity and time since last child birth.

| | Unconditioned | | No more than 3 years | | No more than 1 year | | No more than 3 months | |
|---|---|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | OLS | IV | OLS | IV |
| **Birth parity: 1** | -0.003 *** | 0.000 | -0.005 *** | -0.047 *** | -0.027 *** | -0.031 *** | -0.056 *** | 0.004 |
| **Birth parity: 2** | -0.027 *** | -0.002 | 0.012 *** | 0.004 | 0.015 *** | 0.009 | -0.008 *** | -0.001 |
| **Birth parity: 3** | -0.037 *** | 0.050 ** | 0.002 | 0.066 ** | 0.012 ** | 0.026 | 0.010 | 0.072 |

Notes: All specifications control for age, age squared, education attainment, husband's education attainment, husband's labor force participation, co-residence with elder parents, and prefecture dummies. In all panels, upper bounds on the number of children are not imposed. *** p<0.01, **p<0.05, * p<0.1. Robust standard errors are not reported because of space constraint.

- OLS and causal estimates are quite different in significance and magnitude, which have different implications for policy decision making.
- Negative impacts of children on maternal labor supply are time-varying.
- The impacts also vary across birth parity. First birth has large negative effects, however, second and third birth has very few effect.
- Governments should target who will benefit the most from childcare subsidies.