

Evidence Based Labour Market Policies - Methodological Fundamentals and Innovative Examples

Prof. Dr. Marco Caliendo
(University of Potsdam and Evaluation Office)

caliendo@uni-potsdam.de
www.caliendo.de
www.evaluation-office.de

WAPES World Congress
April 28, 2022

Counterfactual Impact Evaluation

- **Goal:** Estimate the **causal impact** of a certain policy on affected “units”!
 - Central question to answer:
 “What would have happened had the affected units not received the treatment?” (counterfactual outcome)
 - **Causal effect:** Comparison of observed outcome with **counterfactual** situation.
 - **Fundamental Evaluation Problem:** This counterfactual is never observed (for the same unit at the same time)!
- ⇒ Hence, we need to find a **good proxy** from a comparison group!

Selection Bias

- The major problem we we face with ‘standard approaches’ (e.g., before-after-/cross-section-comparisons) is that assignment to treatment and comparison group is **not random**.
- Participants and non-participants might **differ even in absence of the program**.



Treatment Group



Comparison Group

- Hence, simple (mean) comparison are not meaningful because of **selection bias**.

Solving the Selection Problem

- There are a variety of well-established methods to overcome selection bias. **Our focus today:**

Experimental Methods



III. Niklas Elmehed. © Nobel Media.
Abhijit Banerjee
Prize share: 1/3



III. Niklas Elmehed. © Nobel Media.
Esther Duflo
Prize share: 1/3



III. Niklas Elmehed. © Nobel Media.
Michael Kremer
Prize share: 1/3

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2019 was awarded jointly to Abhijit Banerjee, Esther Duflo and Michael Kremer "for their experimental approach to alleviating global poverty."

Quasi-/Non-Experimental Methods



Outline

1 Introduction

2 Evaluation Framework

3 Evaluation Methods

1 Experiments

2 Matching

3 Difference-in-Differences

4 Regression Discontinuity Design

4 Conclusion

Program Evaluation - An Ideal World Scenario (1)

- In an **ideal world**, the evaluator is already involved at early stages of the program design and has influence on the data collected for later evaluation.
- These **stages** include:
 - 1 Defining the program's goals
 - 2 Develop a theory of change
 - 3 Program design
 - 4 Implementation and collection of baseline data
 - 5 Collect final outcome data
 - 6 Counterfactual impact evaluation
- **Process evaluation** (focus on program implementation and operation) and **impact evaluation** should be viewed as complements.
- We can use the information collected in process evaluation to choose amongst alternative evaluation estimators.

Program Evaluation - An Ideal World Scenario (2)

- Important questions which should already be answered at the **design stage**:
 - **Aims and measure of success**:
 - What are the intended effects of the program?
 - How does one measure the success of the program?
 - **Theory of change**:
 - What is the sequence of events that leads to observed outcomes?
 - Which different channels contribute to the success of the program?
 - **Empirical strategy**:
 - What type of evaluation methodology is to be pursued?
 - How will the necessary data be gathered?
 - How can one distinguish which theoretical mechanisms are most important?
- ⇒ In an ideal world, the evaluators have sufficient **time, budget and high-quality-data** at their disposal.

Program Evaluation - The Real World Scenario

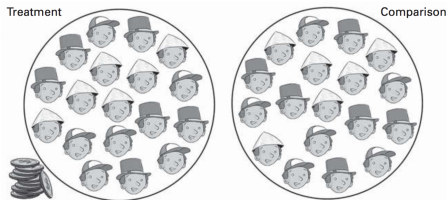
- However, in the **real world** evaluations are often performed under less than optimal circumstances (“**shoestring evaluations**”):

The Constraints Under which Evaluations must be performed			
Time	Budget	Data	Typical Scenario
×			Evaluator is called in late with tight deadline
	×		Difficulties collecting survey data
		×	No baseline data available, sensitive subject with difficult data collection
×	×		Secondary data is available but little time to analyze it
×		×	Little time and no data has been collected survey design limited due to time constraint
	×	×	Evaluator is called in late, deadline not an issue No access to baseline data, budget is tight
×	×	×	Evaluator is called in late with tight deadline and tight budget, no baseline data and no control group has been identified

Source: Bamberger et. al (2004)

Experiments

- **Experiments** assign units from the eligible population **randomly**.
- This guarantees that participation is unrelated to the units' characteristics and balances samples in both observed and unobserved characteristics.



Source: Evaluation in Practice

- Therefore, observed outcome differences between the two groups can be **solely** attributed to the treatment!
- **Estimator**: Simple cross-sectional mean differences in outcome Y .

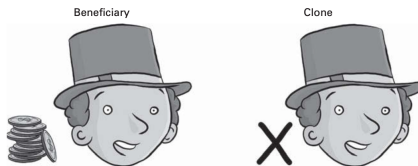
Example Experiment: Job Displacement

Crépon et al. (2013): Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment

- **Research Question:** What is the direct and indirect (displacement) effect of job placement assistance on the labor market outcomes of young, educated job seekers in France?
- **Treatment:** Participation in a job placement program.
- **Data:** Administrative and Survey data, gathered in 2007-2010.
 $N \approx 22,000$.
- **Method:** Randomised controlled field experiment. Unemployed, educated youths were randomly assigned to treatment or control group.
- **Results:** There is a positive effect for treated youths on finding employment, while untreated youths are less likely to find a job (displacement effect).

Matching

- **Idea**: Choose for each participant, one (or many) **statistical twins** from the sample of non-participants.
- They should be identical in **all relevant characteristics**! This is a very strong requirement and requires informative data.



Source: Evaluation in Practice

- Similar to an experiment, this leads to a balanced sample.
- **Estimator**: Simple cross-sectional mean differences in the outcome on the **matched sample**.
- **Curse of dimensionality**: If the number of relevant characteristics is large, it may be very difficult to find an exact match!
- **One solution**: Propensity-score matching summarizes all information in one index and choose the closest non-participant in terms of that index.

Better Data Helps A Lot!

- Implementing a matching approach in a credible way is not easy.
Better data helps a lot!
- Often, the estimates can be improved by **combining several data sources**:
 - Individual- and firm-level data are often available from **administrative records** at low cost (e.g. through national employment agencies).
 - Regional/country-level data are provided by (inter-) **national statistics** agencies.
- **Examples**:
 - IZA Evaluation Data Set combines administrative and survey data allowing to enrich the admin data with information on **“usually unobserved”** characteristics.
 - DellaVigna et al. (2022) conduct a text-message-based survey **twice a week** for four months.

Example Matching: Start-Up Subsidies

Caliendo/Künn/Weißenberger (2016): Personality traits and the evaluation of start-up subsidies

- **Research Question:** Are start-up subsidies for the unemployed an effective active labour market program? And do omitted personality traits pose a threat to the reliability of the matching estimates?
- **Treatment:** Unemployed individuals willing to set-up a business obtain monthly transfers for up to 15 months.
- **Data:** Combination of administrative and survey data. $N \approx 1,300$.
- **Source of selection bias:** Participants self-select into the program; participants differ in their characteristics from non-participants!
- **Method:** Matching participants and non-participants based on a large set of characteristics and pre-treatment outcomes.
- **Results:**
 - Positive effects on employment probabilities and income.
 - Results are robust to the inclusion of usually unobserved personality traits!

Difference-in-Differences (1)

- **Difference-in-Differences** (DiD) set-ups often exploit some kind of “natural experiment” that occurs because of some policy change, where one group of units is affected by the treatment and one group is unaffected.
 - **For example:** One state raises the minimum wage, but the neighbouring state does not.
- **Important:** DiD assumes parallel time trends (PTT) for treatment and control group in absence of the treatment and allows for different pre-treatment levels (“baseline bias”).
- **Validity of the PTT:**
 - Inspecting the **similarity of pre-treatment trends** provides some indication on the likelihood that the PTT assumption holds.
 - Significantly different pre-treatment trends cast serious doubt on the reliability of estimates.

Difference-in-Differences (2)

- **Intuition of the DiD Estimator:** Combine before-after estimates for the treatment and the control group.
 - By comparing changes within groups, we implicitly control for **time-constant unobserved factors**.
 - By comparing these changes across groups, we also control for **time-trends in outcomes**.
- **Estimator:** Compare the **changes** in the treatment group over time with the changes in the comparison group over time.

Example DiD - Minimum Wages (1)

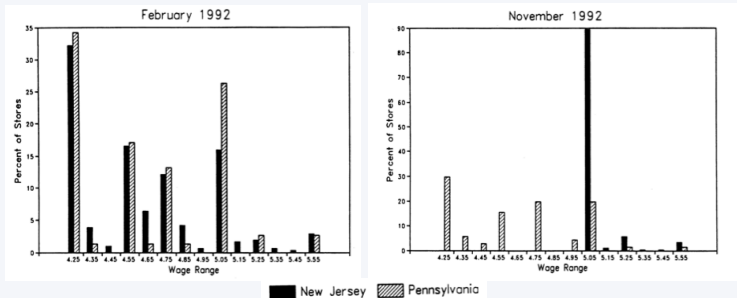
Card/Krüger (1994): Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania

- **Research Question:** Impact of minimum wage increase on low-wage employment?
- **Treatment:** Rise of minimum wage from \$4.25 to \$5.05 per hour in New Jersey in April 1992.
- **Data:** Survey data on wages and employment for $N = 410$ fast food restaurants in New Jersey and Pennsylvania.
- **Source of selection bias:** Unaffected restaurants in New Jersey may serve to different customers and offer more pricey meals.
- **Method:** Compare the evolution of full-time employment in fast-food restaurants in NJ and neighboring state PA.

Example DiD - Minimum Wages (2)

Card/Krüger (1994): Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania

- Descriptive comparison of pre- and post-treatment wages.



Example DiD - Minimum Wages (3)

Card/Krüger (1994): Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania

- Calculating the sample averages yields (s.e. in parentheses):

Variable	Stores by state		
	PA (i)	NJ (ii)	Difference, NJ – PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	– 2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	– 0.14 (1.07)
3. Change in mean FTE employment	– 2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

Source: Card/Krueger (1994), p. 780

Regression Discontinuity Designs

- Many programs operate with some **eligibility cut-off** with respect to some index.
- For example, workers above a certain age receive **unemployment benefits** for a longer duration.
- The (sharp) RDD compares average outcomes of units just below and just above the cut-off.
- The difference gives an estimate of the **local average treatment effects** of the program for people **at the cut-off**.

Example RDD: Unemployment Benefits (1)

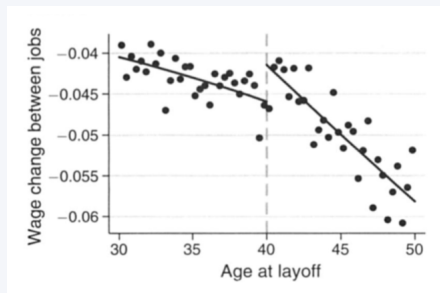
Nekoei/Weber (2017): Does Extending Unemployment Benefits Improve Job Quality?

- **Research Question:** Impact of extended unemployment benefit duration on post-unemployment wages?
- **Treatment:** On August 1 1989 the Austrian government increased the potential unemployment benefit duration from 30 to 39 weeks for individuals 40 and older at the time of job separation.
- **Data:** Administrative data on wages and employment for a sample of $N \approx 1.7$ million involuntary job separations.
- **Source of selection bias:** Unaffected unemployed individuals may have less labour market experience and different education levels.
- **Method:** Compare re-employment wages of individuals slightly younger and older than 40 years when losing their job.

Example RDD: Unemployment Benefits (2)

- Graphically, the treatment effect is equal to the vertical jump in the outcome variable at the cut off.

Nekoei/Weber (2017): Does Extending Unemployment Benefits Improve Job Quality?



Source: Fig. 3

- **Interpretation:** Extension of UB eligibility increases re-employment wages by about 0.5 percent for individuals at the cut off.

Conclusions

- To ensure a **successful program evaluation** ...
 - ... **involve the evaluator** as early as possible.
 - ... **plan the evaluation method** while designing the project.
- **Important:** Better data helps a lot!
 - The **combination of different data sources** (e.g. administrative and survey data) can be helpful in many situations (but may also take time)!
 - It may be useful to think about **collecting new data** specifically for the program evaluation.

“You can do anything. But you can’t do everything and you certainly can’t do everything at once.”

Thank you for your attention!

Appendix

Better Data Helps A Lot - Example

DellaVigna/Heining/Schmieder/Trenkle (2022): Evidence on Job Search Models from a Survey of Unemployed Workers in Germany

- **Research Question:** What mechanisms drive the job finding rate of unemployed?
- **Potential Mechanisms:** Does the decline in job finding rate in the initial months reflect workers discouragement, or the fact that more able workers get jobs faster?
- **Data:** SMS based survey data (twice a week) on job search intensity. $N \approx 7,000$.
- **Method:** Ordinary Least Squares.
- **Results:**
 - Job search intensity stable in the first months and increases shortly before the unemployment insurance expires.
 - No evidence of discouragement of the job search in the first months.

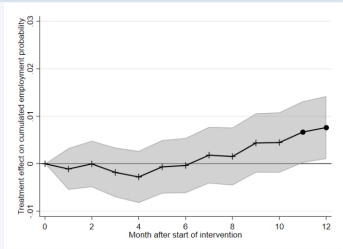
Innovative Example Experiment: Transparency of the Welfare System (1)

Cairo/Mahlstedt (2021): Transparency of the Welfare System and Labor Market Outcomes of Unemployed Workers

- **Research Question:** How does information transparency regarding the risk of benefit reduction affect the labor market outcomes of unemployed workers?
- **Treatment:** Personalised online tool that informs benefit recipients about their personal risk of a benefit reduction.
- **Data:** Countrywide randomised controlled trial and administrative data from Danish social security records for a sample of $N \approx 47,000$ unemployed individuals.
- **Method:** Randomised controlled field experiment. Unemployed individuals are randomly assigned to treatment or control group.

Innovative Example Experiment: Transparency of the Welfare System (2)

Cairo/Mahlstedt (2021): Transparency of the Welfare System and Labor Market Outcomes of Unemployed Workers



Source: Cairo/Mahlstedt (2021) - Figure 3

– Findings:

- Access to the online tool reduces the likelihood of experiencing a benefit cut by about 5%.
- Personal information increases job finding rates by 3%, total number of working hours by 7%, and labor earnings by 6%.

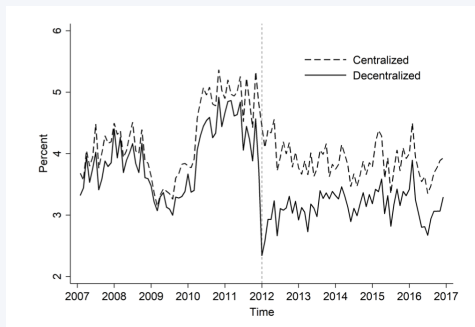
Innovative Example DiD: Public Employment Services (1)

Mergele/Weber (2020): Public employment services under decentralization: Evidence from a natural experiment

- **Research Question:** Does the decentralization of public employment services increase job placements among the unemployed?
- **Potential Mechanisms:** Decentralized offices are better informed about the local job economy, but it could reduce labor market mobility across regions.
- **Treatment:** Partial devolution of job centers to the district level in 2012.
- **Data:** Administrative data on unemployment welfare recipients and vacancies for a sample of $N = 334$ districts.
- **Method:** Compare the development of job matches in districts with and without decentralization of job centers.

Innovative Example DiD: Public Employment Services (2)

Mergele/Weber (2020): Public employment services under decentralization: Evidence from a natural experiment



Source: Mergele/Weber (2020)

Figure 2: average aggregate monthly job-finding rates by job center type

- **Findings:** job-center decentralization reduced job finding by approximately 10% within five years.

Innovative Example RDD: EU Funds

Becker et al. (2010): Going NUTS: the effect of EU structural funds on regional performance

- **Research Question:** What is the causal effect of EU structural funds on regional performance, e.g. economic growth?
- **Treatment:** Regions with a per capita GDP level below 75% of the EU average qualify for structural funds transfers from the central EU budget.
- **Data:** Regional data on GDP and employment for a sample of $N = 3,301$ regions.
- **Source of selection bias:** Regions who qualify for the funds are per design systematically different from non-eligible regions (“poor” vs. “rich” regions).
- **Method:** Compare regions with a per capita GDP level below and above the 75% of the EU average.
- **Results:**
 - Transfers raise the per capita GDP by about 1.8%.
 - There are no significant employment effects.

References:

- World Bank Repository Impact Evaluation:
www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice
- Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2013) "Do labor market policies have displacement effects? Evidence from a clustered randomized experiment." *The Quarterly Journal of Economics* 128.2, 531-.580.
- Caliendo, M., Künn, S., and Weißenberger, M. (2016): "Personality traits and the evaluation of start-up subsidies." *European Economic Review*, Vol. 86, pp. 87 – 108.
- Card, D., and Krueger, A. B. (1994): "Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania", *American Economic Review*, Vol. 84, No. 4, pp. 772 –793.
- Nekoei, A., and Weber, A. (2017): "Does Extending Unemployment Benefits Improve Job Quality?" *American Economic Review* 107.2, 527-61.
- DellaVigna, S., Heining, J., Schmieder, J., Trenkle, S. (2022): "Evidence on Job Search Models from a Survey of Unemployed Workers in Germany" *forthcoming in Quarterly Journal of Economics*.

Further Examples

Experiments

- van den Berg, G. J., Hofmann, B., Stephan, G., and Uhlenhorff, A. (2021) "Mandatory integration agreements for unemployed job seekers: a randomized controlled field experiment in Germany." Available at SSRN 3762896.
- Cairo, S., and Mahlstedt, R. (2021). "Transparency of the Welfare System and Labor Market Outcomes of Unemployed Workers." No. 14940. *Institute of Labor Economics (IZA)*.
- Duflo, E., Dupas, P., and Kremer M. (2011): "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review*, 101(5), 1739-74.

Matching

- Caliendo, M., and Künn, S. (2011). "Start-up subsidies for the unemployed: Long-term evidence and effect heterogeneity." *Journal of Public Economics*, Vol. 95, No. 3-4, pp. 311 – 331.

DiD

- Mergele, L., and Weber, W. (2020): "Public employment services under decentralization: Evidence from a natural experiment." *Journal of Public Economics* 182, 1 – 26.

RDD

- Becker, S. O., Egger, P. H., and Von Ehrlich, M. (2010): "Going NUTS: The effect of EU Structural Funds on regional performance", *Journal of Public Economics*, Vol. 94, No. 9 – 10, pp. 578-590.