

政策効果の検証：発展

技術政策学（データ科学編）

土井翔平

2023-06-04

はじめに

政策の効果（因果関係）をデータから明らかにする ⇨ 交絡の除去

- ・ 交絡：原因と結果の両方に影響しているような第三の要因（交絡因子）が存在する状況

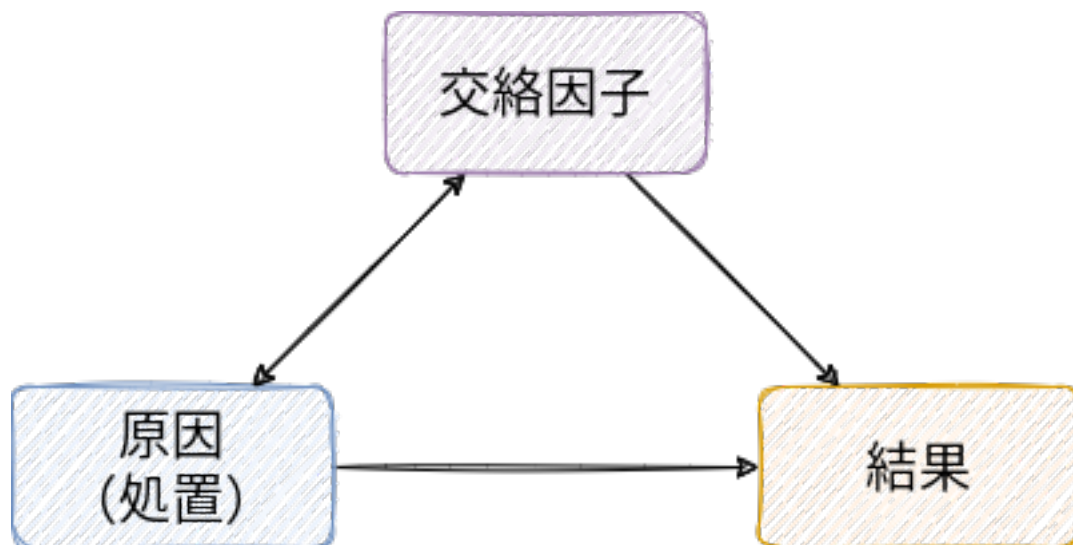


図1: 交絡のイメージ

⇨ 理想としての RCT（および自然実験）／常に実行可能とは限らない

⇨ その他の分析手法の開発

1 統制

RCT のポイント：原因のある（政策を受けた）集団と原因のない（政策を受けなかった）集団の性質がほとんど同じである

⇒ 事後的に同じような集団を作ればよい？

- 性質を一致させるという意味で**統制 (control)** と呼ぶ。

1.1 マッチング

マッチング：原因のある集団とない集団から同じような対象を取り出して比較する。¹

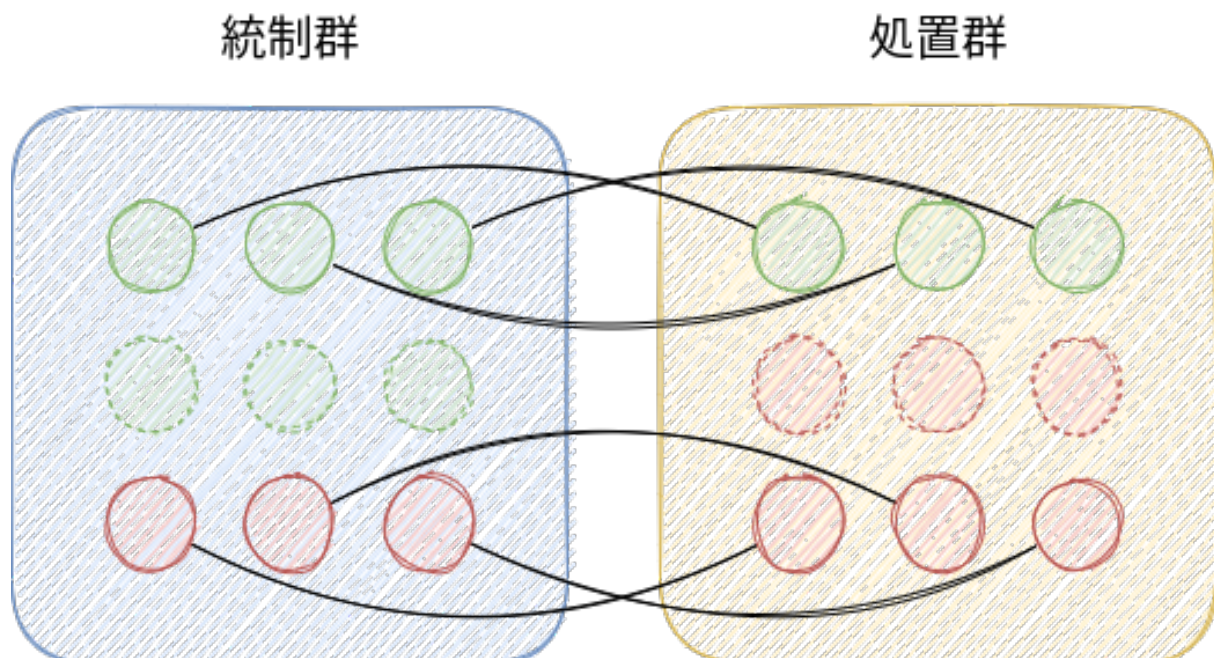


図2: マッチングのイメージ

⇒ 同じような対象を常に取り出せるとは限らない。

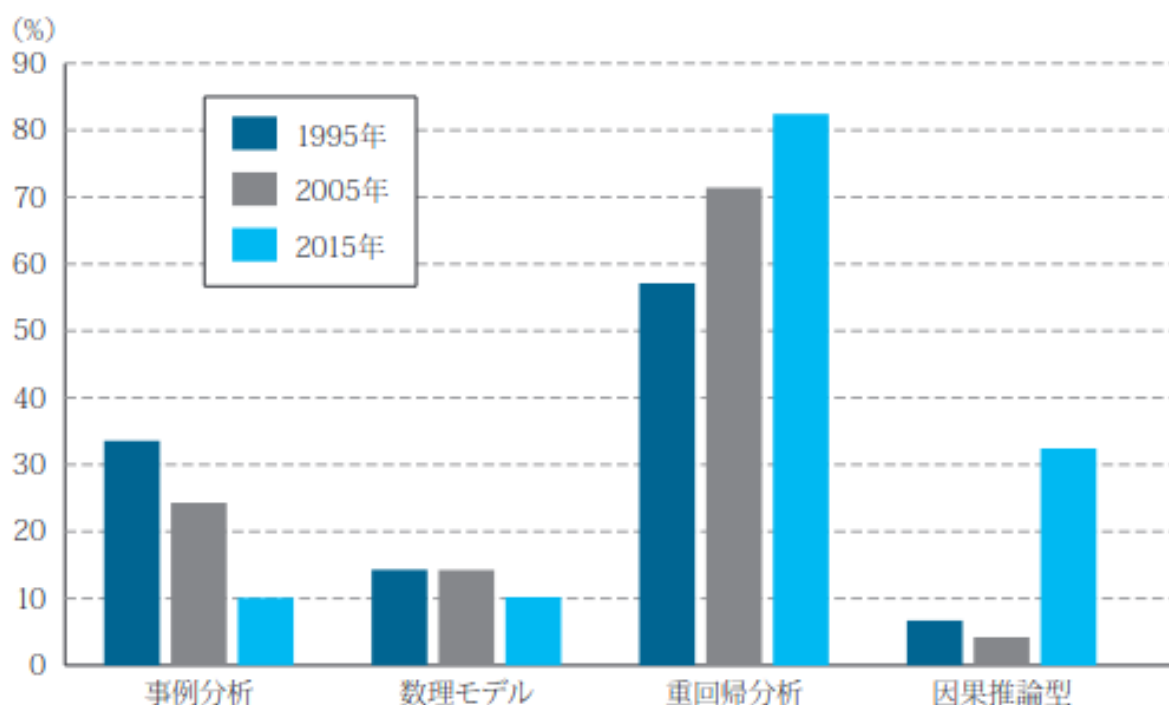
- 1990 年に東京で生まれた生物学上の男性で大学院卒で大学教員として札幌に住んでいるビール好き？

1.2 回帰分析

もっぱら社会科学で広く使われているのは**回帰分析** (粕谷, 2018)

¹ 似たような発想で重み付け (weighting) と呼ばれる手法も用いられる。

図1 主要5誌における実証政治学論文の分析手法



(注) 検討対象は*American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, *World Politics*のうちの国内政治を分析する研究論文(Comparative PoliticsとAmerican Politics分野の論文)、および、*Comparative Political Studies*, *Comparative Politics*に掲載された全ての研究論文である(N= 524)。「因果推論型」は実験型と擬似実験型手法の両方を含む。複数の手法を組み合わせている論文もあるため、4種類の合計が100%とはならない。

(出所) 筆者作成。

図3: 粕谷 (2018)

$$\text{平均寿命} = -9.101 + 8.405 \times \log(\text{一人あたり GDP})$$

- 毎年、一人あたり GDP も平均寿命も増加しているだけでは？

$$\text{平均寿命} = -391.051 + 7.770 \times \log(\text{一人あたり GDP}) + 0.195 \times \text{年}$$

⇒ 回帰分析に入れた特徴量の交絡は取り除くことができる。

- 特定の年において一人あたり GDP が増えた時に平均寿命がどの程度増えるのかを示している。

マッチングも回帰分析も（後述する手法よりも）簡単に交絡を取り除ける。

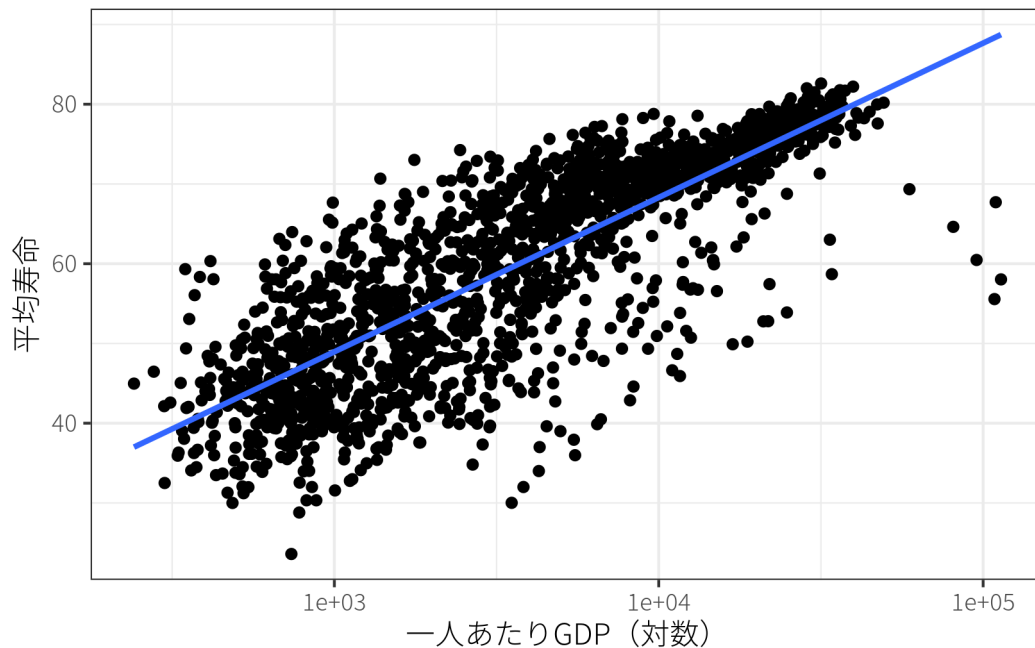


図4: 豊かさと平均寿命

1.3 統制の限界

マッチングも回帰分析も分析に用いた特徴量の交絡のみ取り除く。

⇒ 分析に用いていない（データとして存在しない）特徴量の交絡は取り除くことはできない。

- 平均寿命と一人あたり GDP の交絡因子の候補は？

分析に用いた特徴量以外に交絡因子が存在しないことを証明することは不可能（悪魔の証明）

⇒ もう少し因果関係と言いやすいような状況はないだろうか？

- 経済学、公衆衛生学、政治学.....における因果推論革命、識別革命

2 回帰不連続デザイン

回帰不連続デザイン (regression discontinuity design: RDD) : ある基準を満たすかどうかで、原因の有無が分かるような状況を使ってみる。

- とある資格（例えば英検）を取るといい給料の仕事に就職しやすくなるのか？
- 基準をぎりぎり満たした人と満たせなかった人はほとんど同じような集団かも？

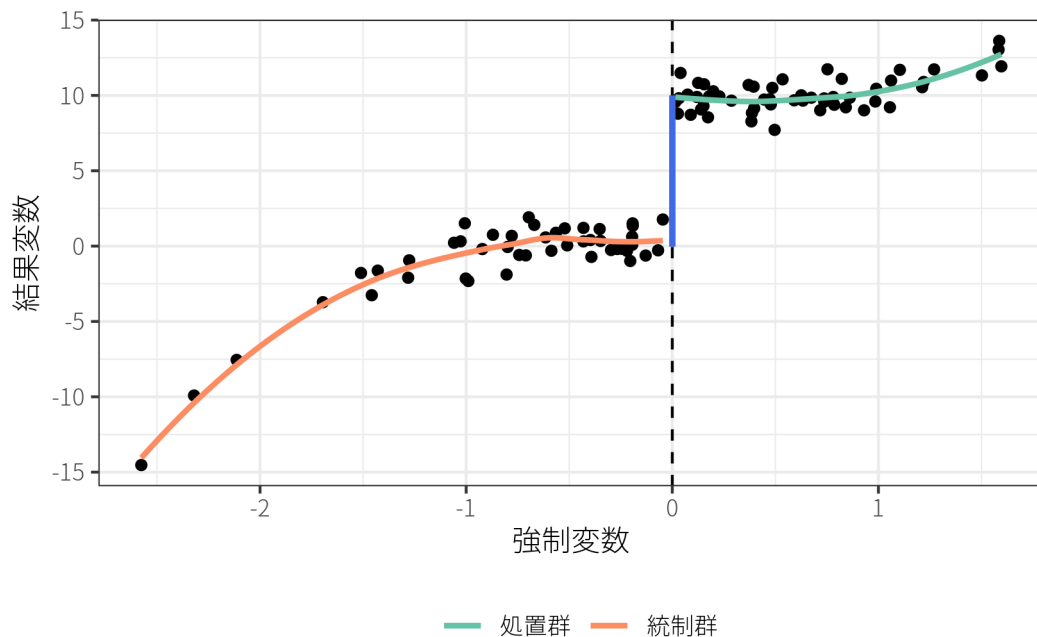


図5: RDD のイメージ

→ もし基準の前後で「ジャンプ」が存在していれば、それは効果と言えそう。

- 医療費負担の引き下げは健康に資するのか？
- 少人数教育は学力の向上に資するのか？
- 女性政治家（市長）は男性と異なるのか？
 - 政治学でよく使われるのは、ギリギリ当選した場合と落選した場合の比較
- 選挙広告は投票率を上げるのか？
 - 地理的境界線もよく使われている。

3 差分の差

3.1 前後即因果の誤謬

前後即因果の誤謬：結果の前に生じたものを原因とみなす誤り

→ ある事象 A が起きた後に事象 B が生じたからといって A が B の原因とは言い切れない

- とある資格（例えば英検）を取るといい給料の仕事に就職しやすくなるのか？
- もちろん、原因の有無で比較しても意味はない。

Figure 2: Age Profile for Outpatient Visits
2A. Overall Outpatient Visits (Log Scale)

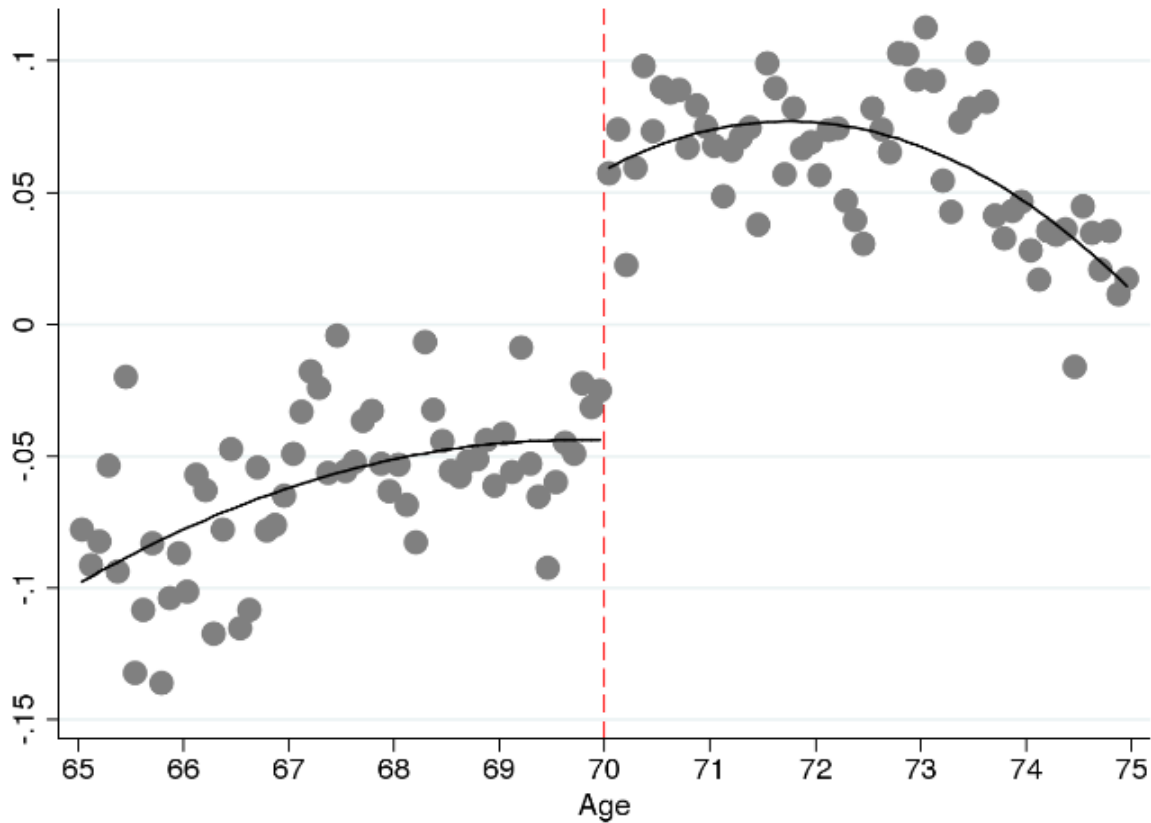


図6: Shigeoka (2014)

3.2 差分の差

なぜ前後比較はダメなのか？→ 時間トレンドがあるから

- ・ 時間トレンド：原因の有無にかかわらず一定の方向へ変化する傾向

差分の差 (difference in differences: DID)：原因のなかったグループで時間トレンドを計算し、原因のあったグループの時間トレンドを除去する。

- ・ 資格を取った人の給料の増加分から資格を取らなかったひとの給料の増加分（時間トレンド）を引く。
- ・ 最低賃金の上昇は雇用を減らすのか？
- ・ 難民の受け入れは賃金や個用を悪化させるのか？
- ・ ヒトラーの演説はナチスへの支持を高めたのか？

Figure 3: Age Profile for Outpatient Visits for Selected Diagnoses (Log Scale)

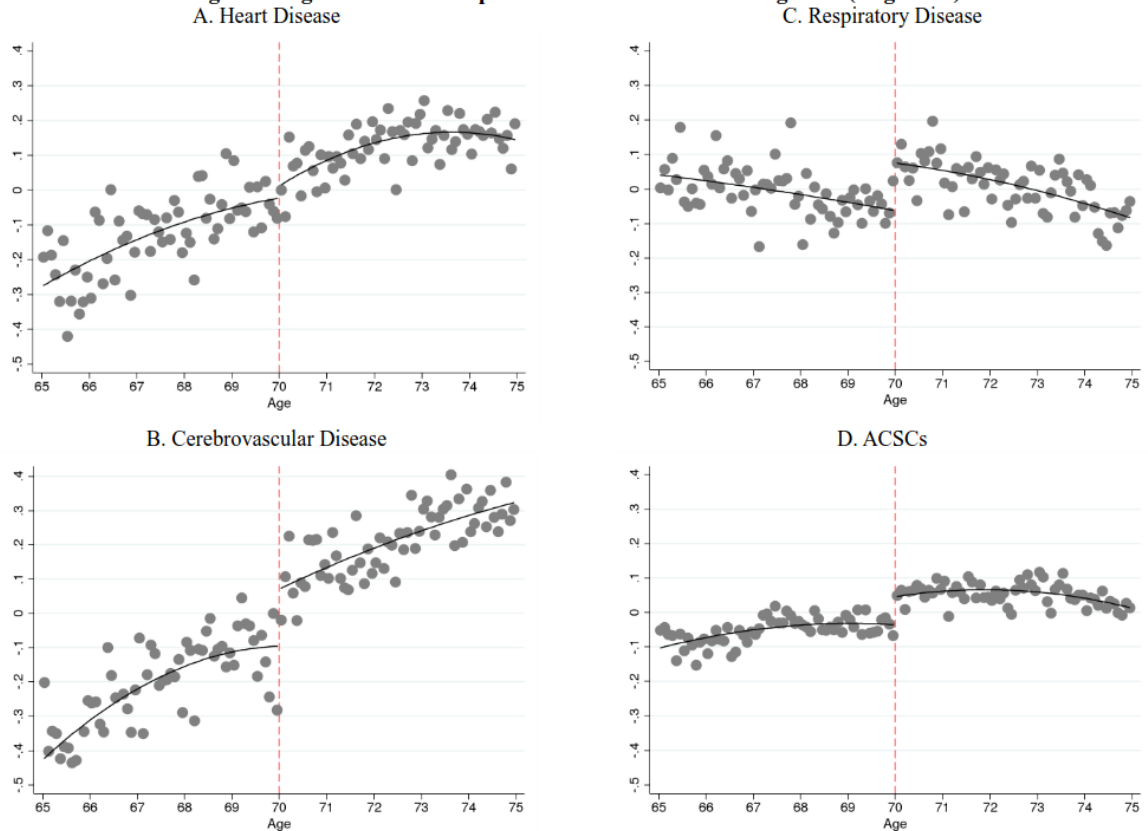


図7: Shigeoka (2014)

3.3 前後比較が可能な状況

時間トレンドが無視可能な状況であれば、前後比較による効果が分かる。

1. 偶然（予期しない）出来事が起こった場合
 - ・ テロリズムは政府への支持を高めるか？
2. 出来事の直前と直後（RDD の応用）を比較する場合
 - ・ 首脳の見聞は好感度を上げるのか？

3.4 合成統制法

合成統制法 (synthetic control method: SCM)：政策が起こった対象が「もし政策を行っていなかった場合」をその他の事例から合成して比較する

- ・ 経済の自由化は経済成長に貢献するのか？

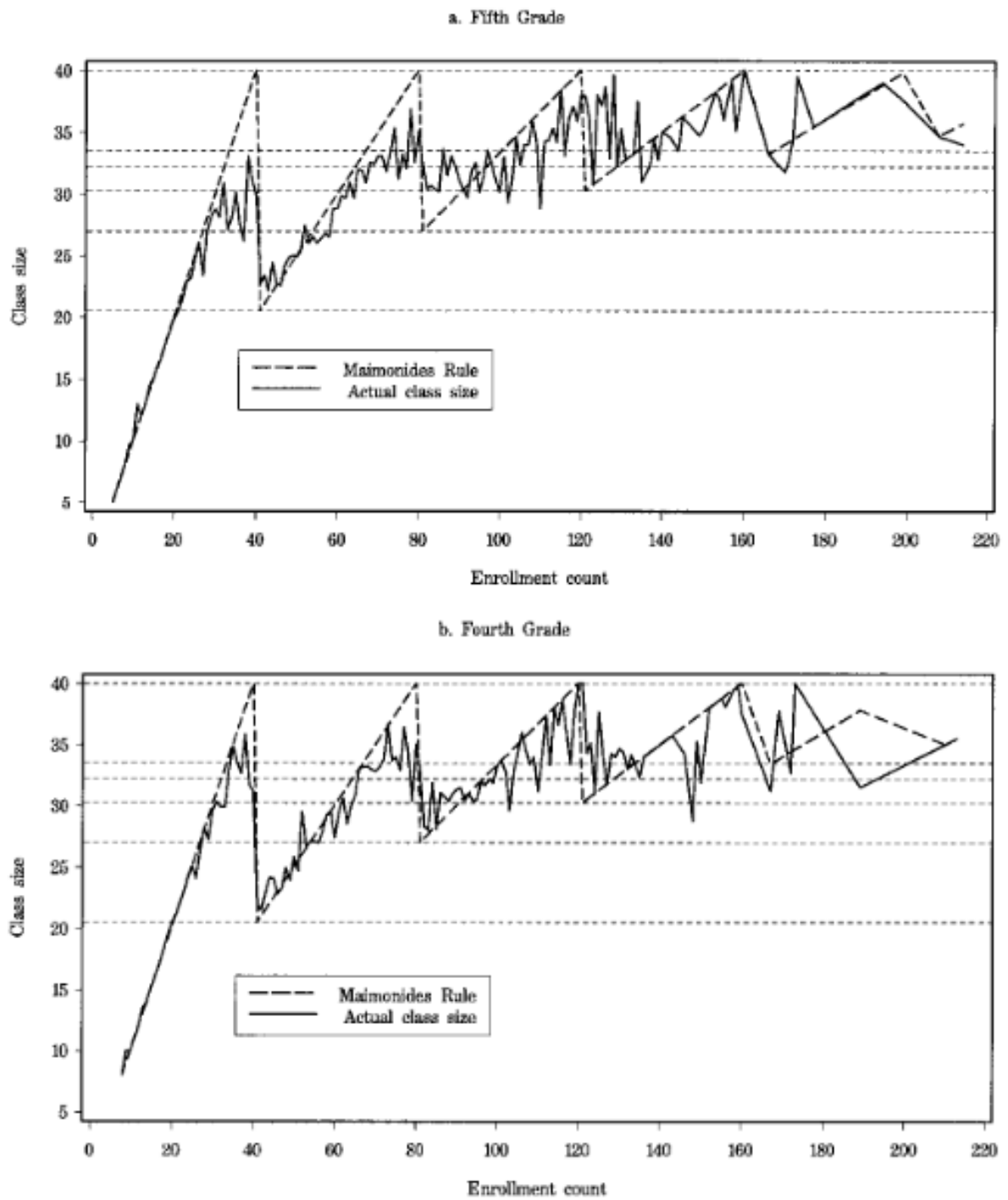


FIGURE I
Class Size in 1991 by Initial Enrollment Count, Actual Average Size and as Predicted by Maimonides' Rule

图8: Angrist and Lavy (1999)

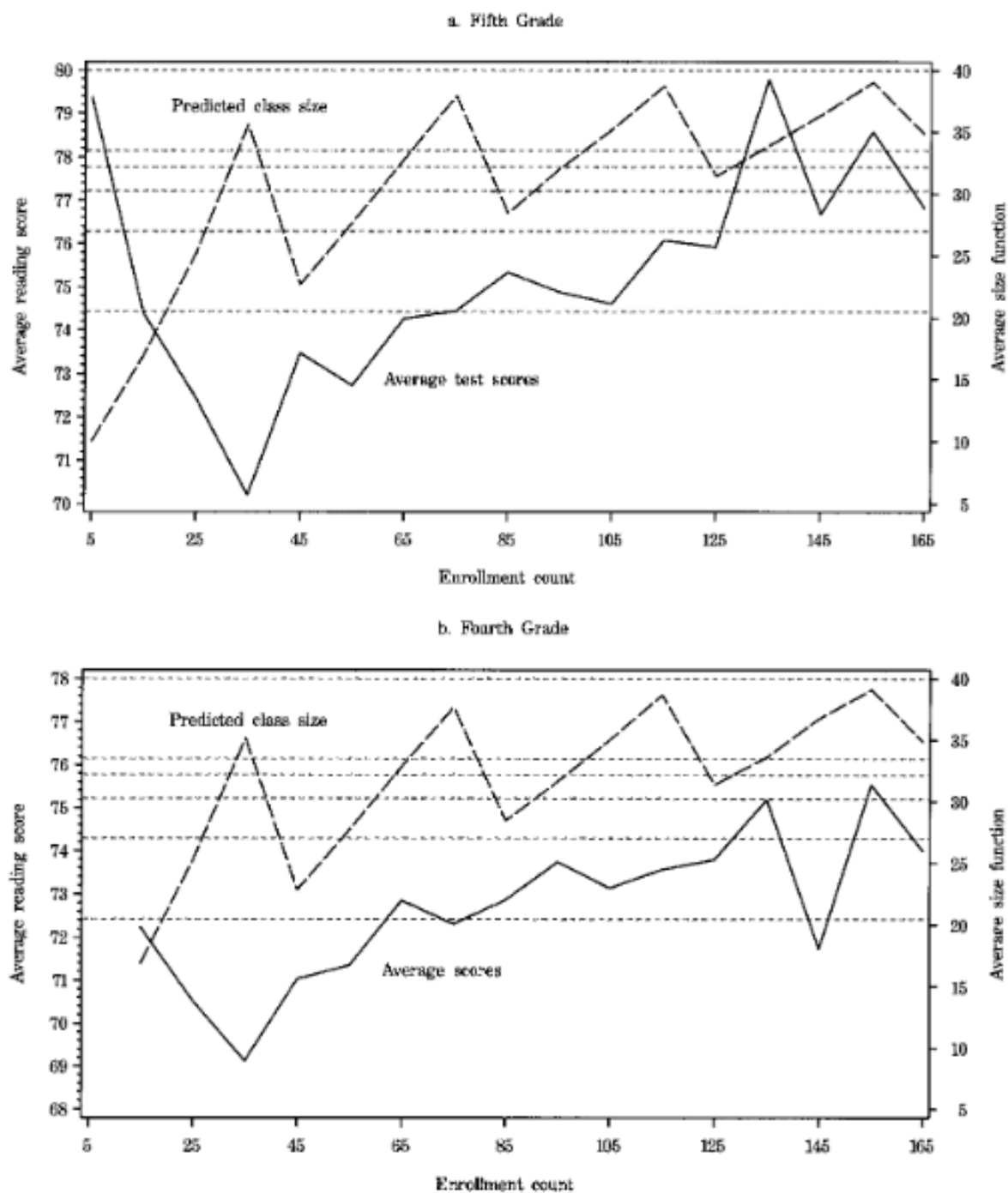


FIGURE II
Average Reading Scores by Enrollment Count, and the Corresponding Average
Class Size Predicted by Maimonides' Rule

Figure 9: Angrist and Lavy (1999)

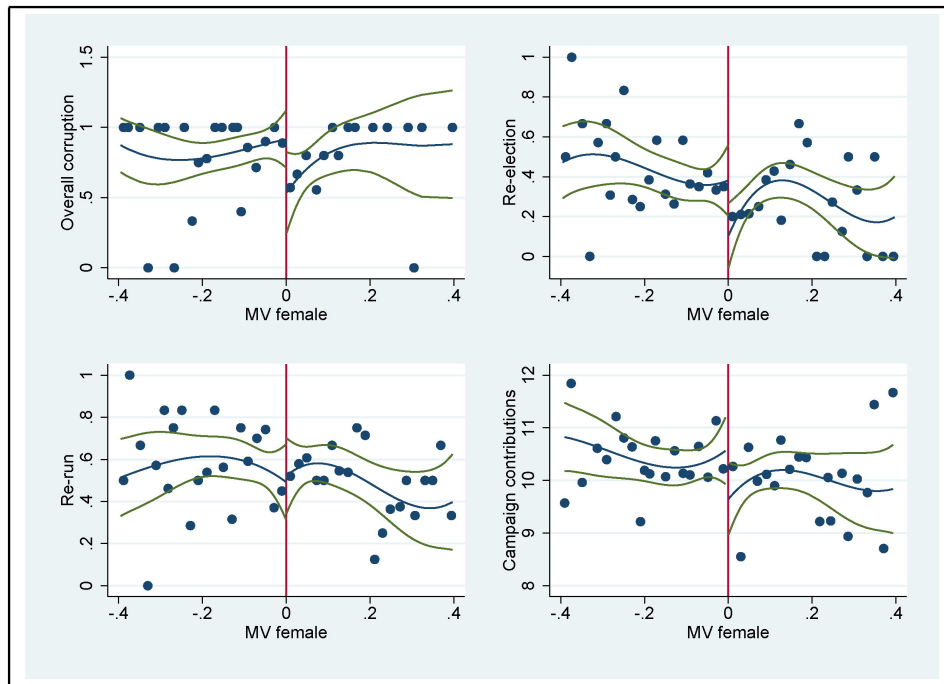


Figure 10: Brollo and Troiano (2016)

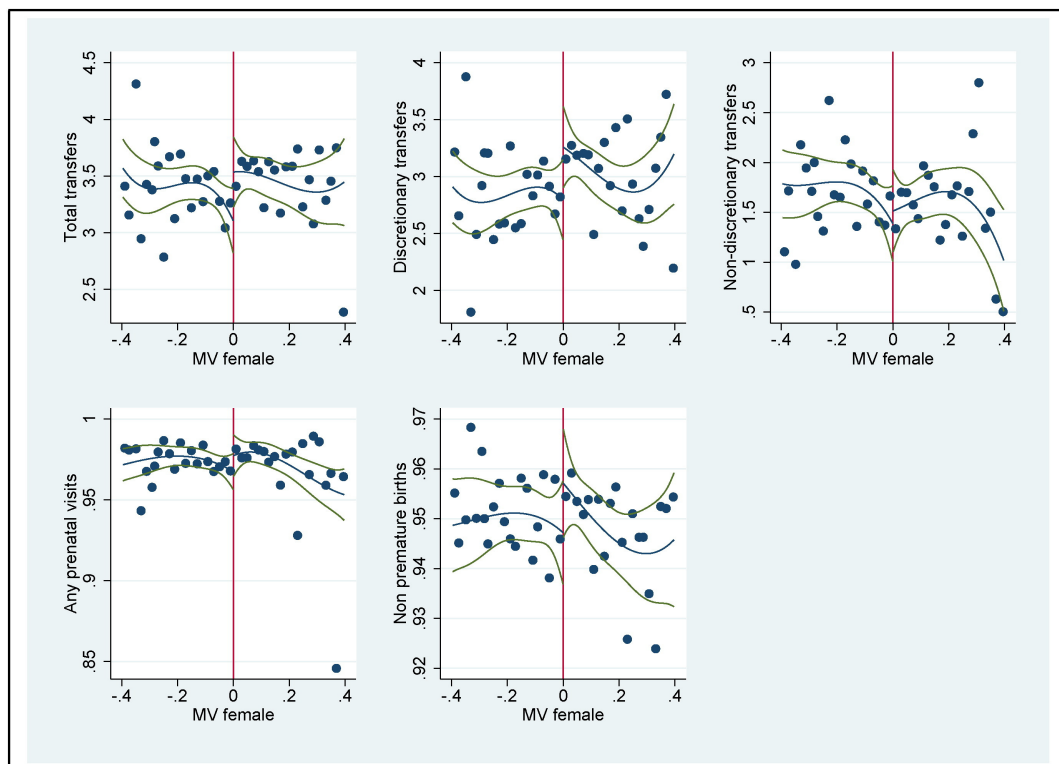


Figure 11: Brollo and Troiano (2016)

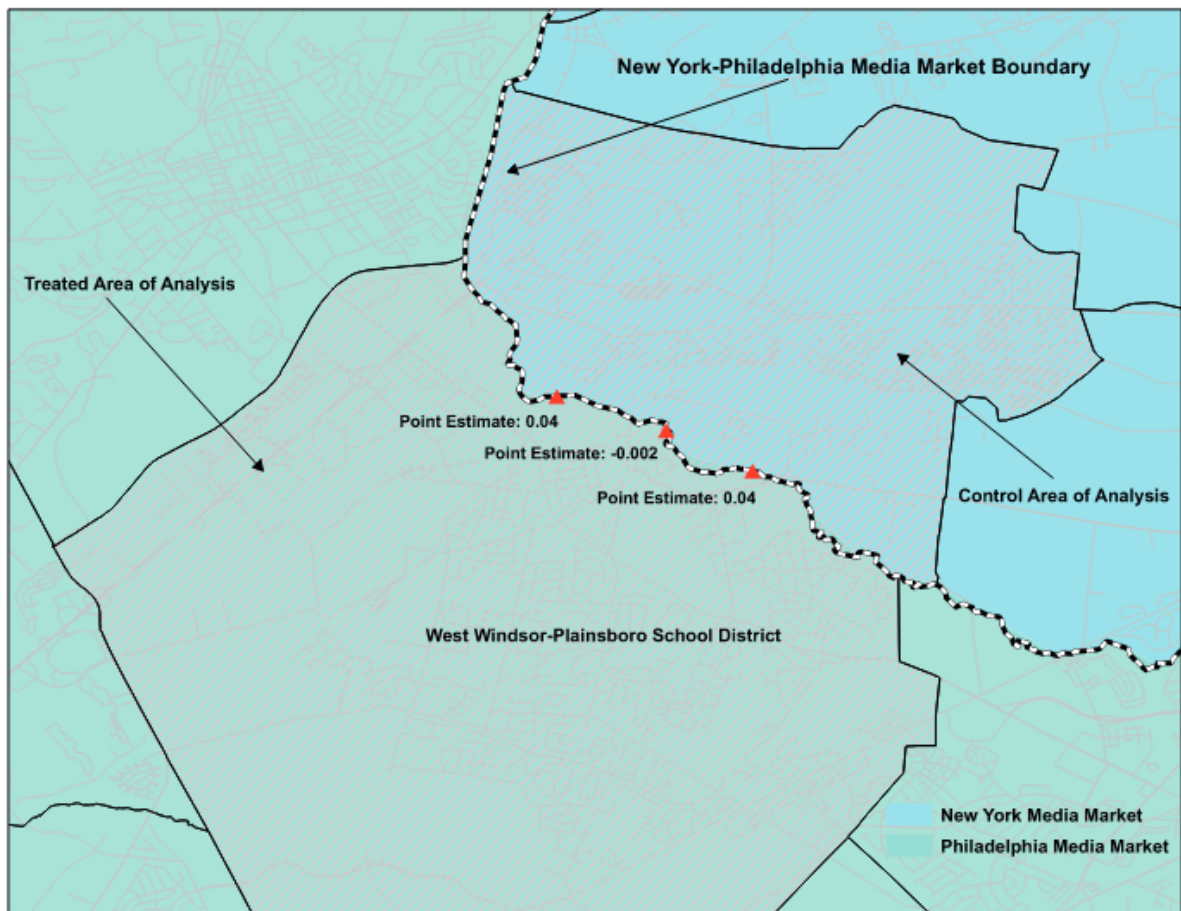


図12: Keele and Titiunik (2015)

4 操作変数法

操作変数法 (instrumental variable: IV)：原因のみに影響し、結果には直接影響しない要因（操作変数）を用いて交絡を除去する

- とある資格（例えば英検）を取るといい給料の仕事に就職しやすくなるのか？
 - とある資格を取るための補助をランダムに行う ⇨ その補助を受けて資格を取れるかどうかはランダム
- 民主主義は経済成長に貢献するのか？
- ナチスはプロパガンダによって支持を獲得できたのか？

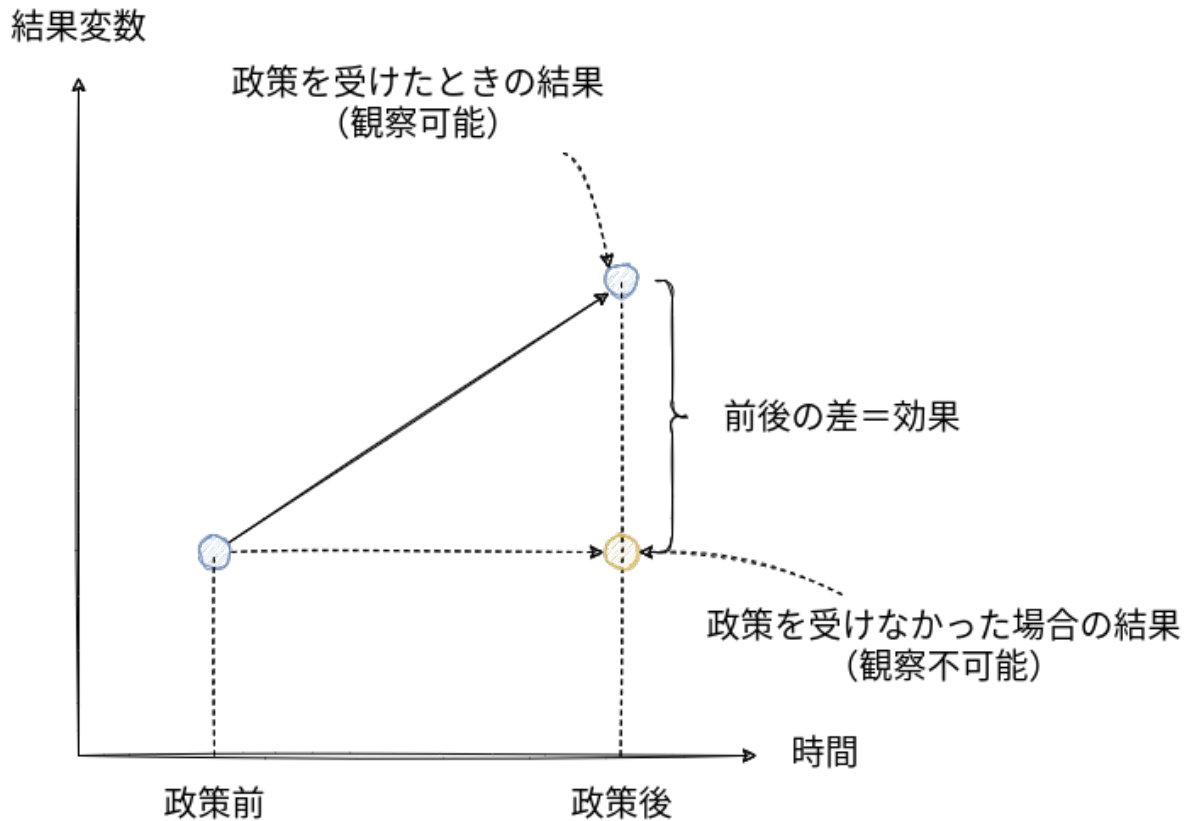


図13: 前後比較のイメージ

5 因果推論の注意点

5.1 サンプルの代表性

無作為化比較試験：無作為に処置を割り当て → 効果を推定

- 内的妥当性 (internal validity)：手元にあるデータの中で正しく因果推論できる

無作為抽出 (random sampling)：特定の集団から一部を無作為に取り出すこと

- 外的妥当性 (external validity)：分析結果が分析に用いたデータ以外にも当てはまる
 - (サンプルの) 代表性：サンプルにおける属性（性別や年齢など）の割合が本当に知りたい集団と似ている。
 - たとえ実験ではなくても世論調査などをする場合は無作為抽出は必要

無作為割り当てができていても無作為抽出をしていなければ、分析結果が元々の集団に当てはまるかは分からない。

- もちろん、無作為抽出でも別の集団については当てはまるか分からない。

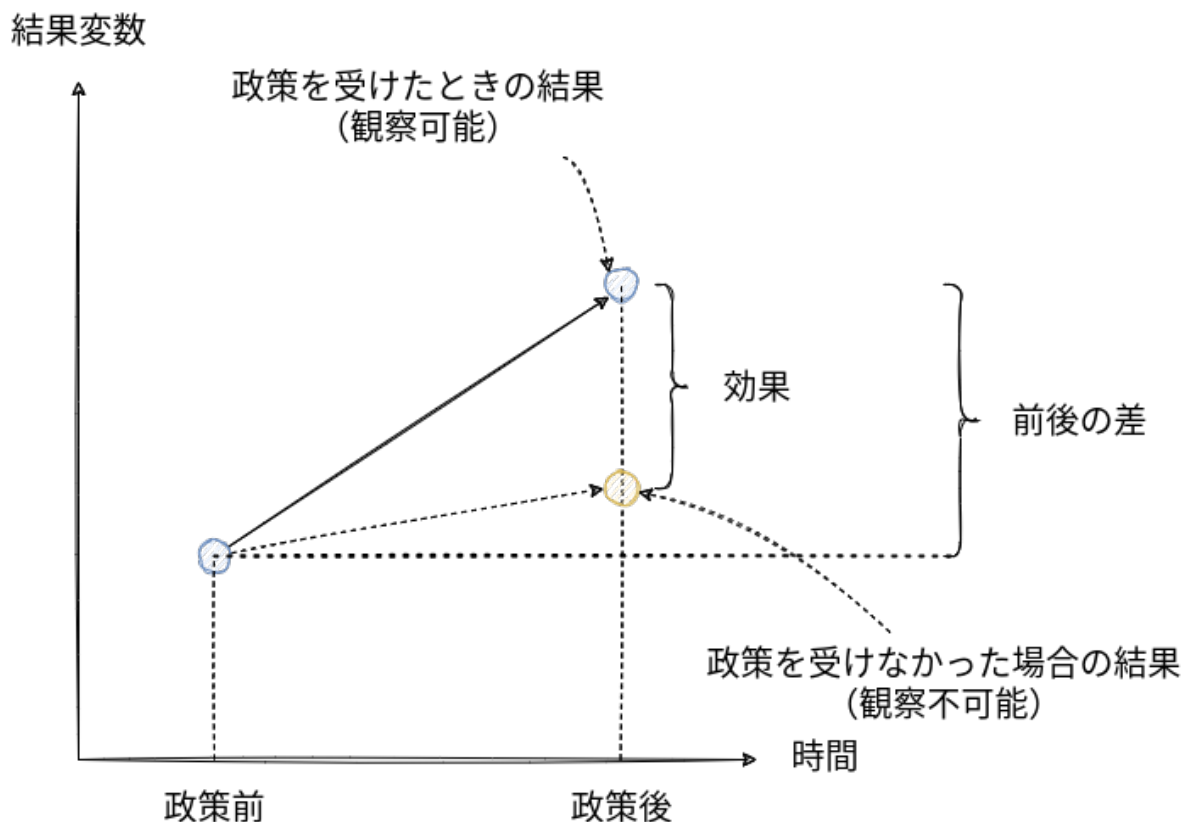


図14: 前後即因果の誤謬のイメージ

オンラインのサンプルは市民を代表しているのか？

- (スペインとアメリカでは) Twitter ユーザーは男性、都市部の住民、政治的に極端な人が多い、あるいは多くのツイートをしている ([Barberá and Rivero, 2015](#))
- (アメリカでは) 調査会社のサンプルに比べてクラウドソーシングの参加者の属性は偏っている ([Weinberg et al., 2014](#))
- ただし、RCT の結果はどちらでも同じような傾向を持つ ([Weinberg et al., 2014](#))

5.2 不均一効果

政策の効果はあらゆる集団に対して同じとは限らない。

- 年齢や性別、学歴、経済状況.....などによって効果は変わりうる。

⇒ 条件ごとに集団を分割して効果を推定する。

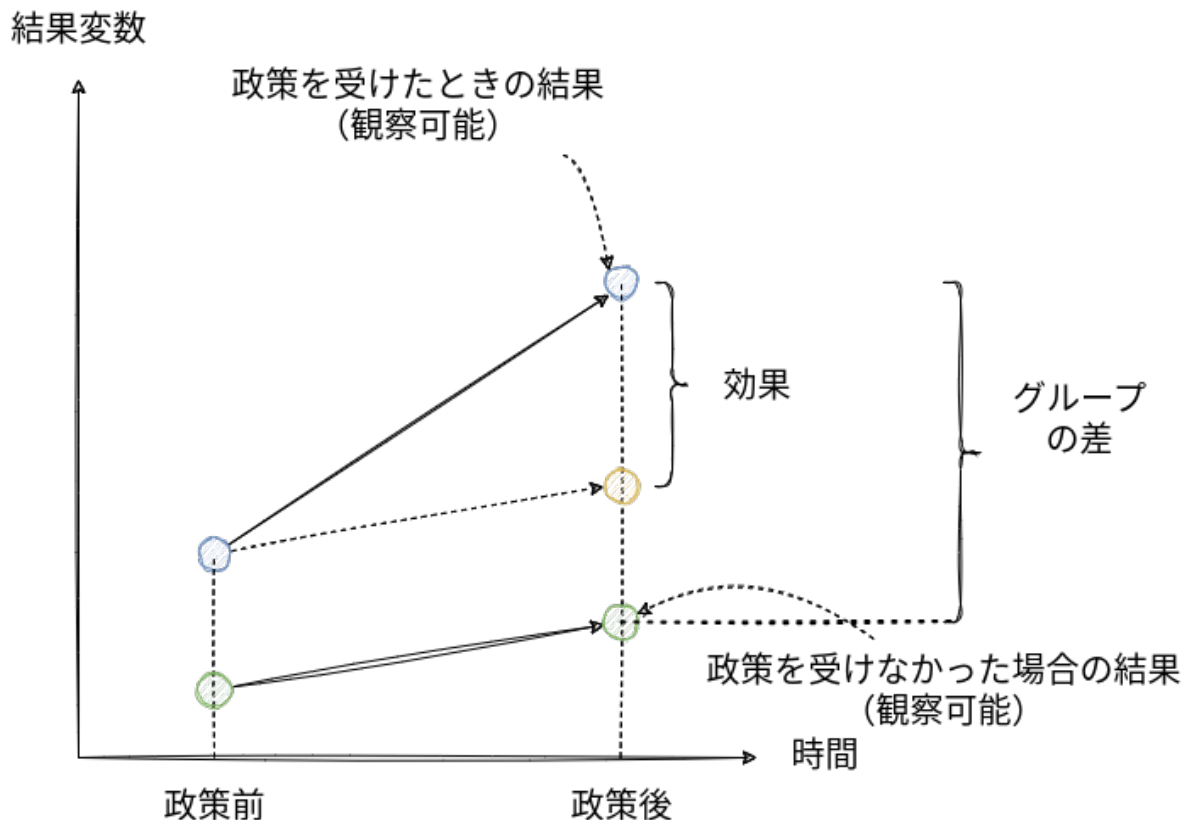


図15: グループ間比較のイメージ

5.3 一般均衡効果

RCT では集団全体から一部を取り出して、政策の有無を決定する。

⇒ 実際に政策を受けるのは全体から見るとごく一部

政策として集団全体に実施した場合は、RCT 通りの結果にならないかもしれない。

⇒ 集団全体における効果（一般均衡効果）が生じる。

- ・ 職業訓練や教育が賃金を上げるとしても、全員がプログラムを受けるとその効果は相殺？
- ・ 少人数に SNS をやらせても意味はない？

効果が波及する場合：RCT は適切に政策の効果を推定できない。

6 実証分析への示唆

実際に因果推論しなくても、実証分析をする上で重要な点はある（と信じている）

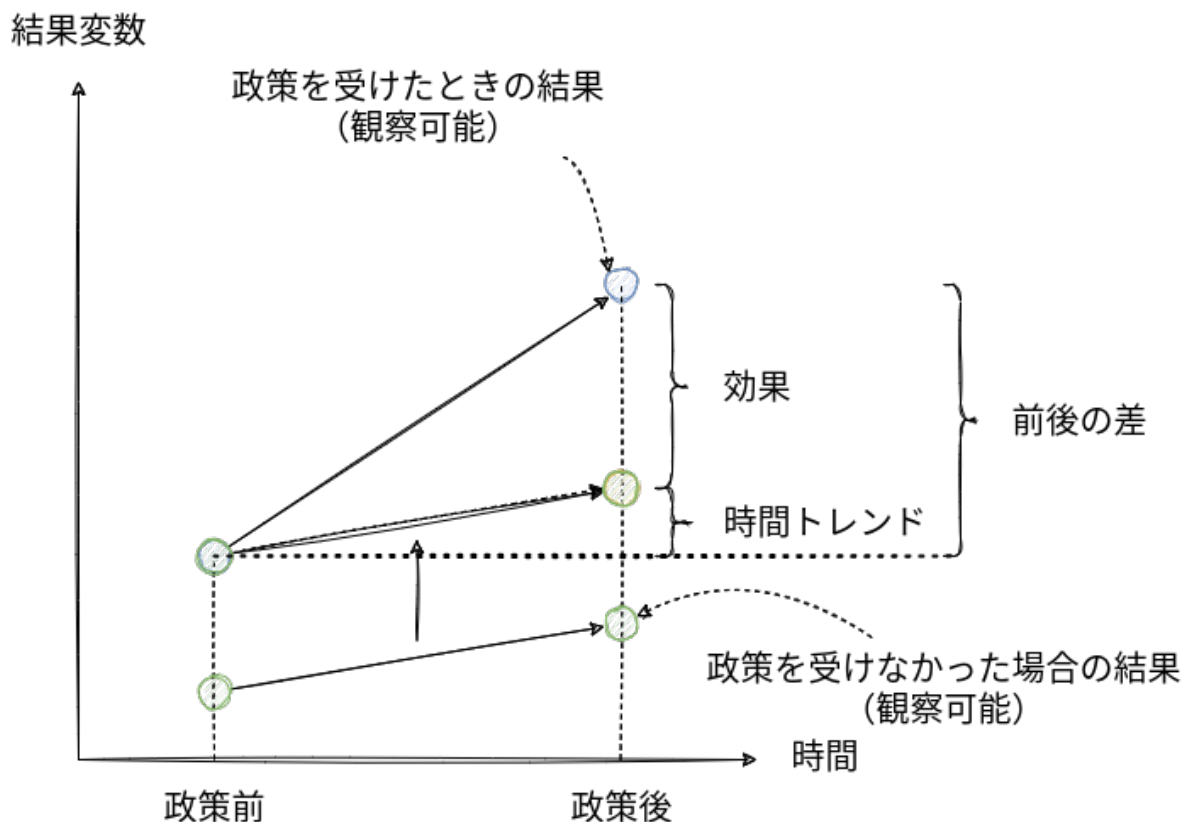


図16: グループ間比較のイメージ

6.1 計量分析

とりあえず実際のデータに触れてみるという姿勢は大事

関係分野のデータ分析の研究に触れてみるのも大事

- 案外、反直観的な（予想と異なる）結果が得られることはある。
- 政治学や経済学の一般向けの本 ([Banerjee and Duflo, 2020](#); [北村, 2022](#))

データ分析だからといって鵜呑みにするのも危険

- どのようなデータと手法を用いているのかを確認
- 「データの読み方」を学ぶ ([菅原, 2022](#); [荻原, 2023](#); [筒井, 2023](#))

6.2 事例分析

事例分析をする際にも効果を示す際には注意が必要 ([久米, 2013](#); [伊藤, 2022](#))

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)
4. Change in mean FTE employment, balanced sample of stores ^c	– 2.28 (1.25)	0.47 (0.48)	2.75 (1.34)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0 ^d	– 2.28 (1.25)	0.23 (0.49)	2.51 (1.35)

Figure 17: Card and Krueger (1994)

Table 3. Logarithms of Real Hourly Earnings of Workers Age 16–61 in Miami and Four Comparison Cities, 1979–85.

Group	1979	1980	1981	1982	1983	1984	1985
<i>Miami:</i>							
Whites	1.85 (.03)	1.83 (.03)	1.85 (.03)	1.82 (.03)	1.82 (.03)	1.82 (.03)	1.82 (.05)
Blacks	1.59 (.03)	1.55 (.02)	1.61 (.03)	1.48 (.03)	1.48 (.03)	1.57 (.03)	1.60 (.04)
Cubans	1.58 (.02)	1.54 (.02)	1.51 (.02)	1.49 (.02)	1.49 (.02)	1.53 (.03)	1.49 (.04)
Hispanics	1.52 (.04)	1.54 (.04)	1.54 (.05)	1.53 (.05)	1.48 (.04)	1.59 (.04)	1.54 (.06)
<i>Comparison Cities:</i>							
Whites	1.93 (.01)	1.90 (.01)	1.91 (.01)	1.91 (.01)	1.90 (.01)	1.91 (.01)	1.92 (.01)
Blacks	1.74 (.01)	1.70 (.02)	1.72 (.02)	1.71 (.01)	1.69 (.02)	1.67 (.02)	1.65 (.03)
Hispanics	1.65 (.01)	1.63 (.01)	1.61 (.01)	1.61 (.01)	1.58 (.01)	1.60 (.01)	1.58 (.02)

Figure 18: Card (1990)

Table 4. Unemployment Rates of Individuals Age 16–61 in Miami and
Four Comparison Cities, 1979–85.
(Standard Errors in Parentheses)

Group	1979	1980	1981	1982	1983	1984	1985
<i>Miami:</i>							
Whites	5.1 (1.1)	2.5 (0.8)	3.9 (0.9)	5.2 (1.1)	6.7 (1.1)	3.6 (0.9)	4.9 (1.4)
Blacks	8.3 (1.7)	5.6 (1.3)	9.6 (1.8)	16.0 (2.3)	18.4 (2.5)	14.2 (2.3)	7.8 (2.3)
Cubans	5.3 (1.2)	7.2 (1.3)	10.1 (1.5)	10.8 (1.5)	13.1 (1.6)	7.7 (1.4)	5.5 (1.7)
Hispanics	6.5 (2.3)	7.7 (2.2)	11.8 (3.0)	9.1 (2.5)	7.5 (2.1)	12.1 (2.4)	3.7 (1.9)
<i>Comparison Cities:</i>							
Whites	4.4 (0.3)	4.4 (0.3)	4.3 (0.3)	6.8 (0.3)	6.9 (0.3)	5.4 (0.3)	4.9 (0.4)
Blacks	10.3 (0.8)	12.6 (0.9)	12.6 (0.9)	12.7 (0.9)	18.4 (1.1)	12.1 (0.9)	13.3 (1.3)
Hispanics	6.3 (0.6)	8.7 (0.6)	8.3 (0.6)	12.1 (0.7)	11.8 (0.7)	9.8 (0.6)	9.3 (0.8)

図19: Card (1990)

- ・ 差異法：同じような事例だが、原因の有無が異なるような事例を比較する
 - 理想としては原因がランダムに生じている状況
 - 原因が生じた直前と直後の比較
 - 時間トレンドの確認
- ・ 一致法：異なる事例だが、原因と結果が生じている事例を比較する
 - 有力な原因を提案できるが、因果関係・効果を主張するのは難しい
- ・ 反実仮想：仮に原因がなかったらどうなっていたのかを明らかにして、原因であることを説得する
- ・ 過程追跡：原因から結果での出来事の連鎖を辿って、因果関係・プロセスを示す
- ・ 逸脱事例：通説では説明できない事例を提示して、通説を修正する

参考文献

- Acemoglu, Daron, Simon Johnson, and James A Robinson (2001) “The colonial origins of comparative development: An empirical investigation,” *American economic review*, Vol. 91, No. 5, pp. 1369–1401.
- Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya (2015) “Radio and the Rise of the Nazis in Prewar Germany,” *The Quarterly Journal of Economics*, Vol. 130, No. 4, pp. 1885–1939.
- Angrist, Joshua D and Victor Lavy (1999) “Using Maimonides’ rule to estimate the effect of class

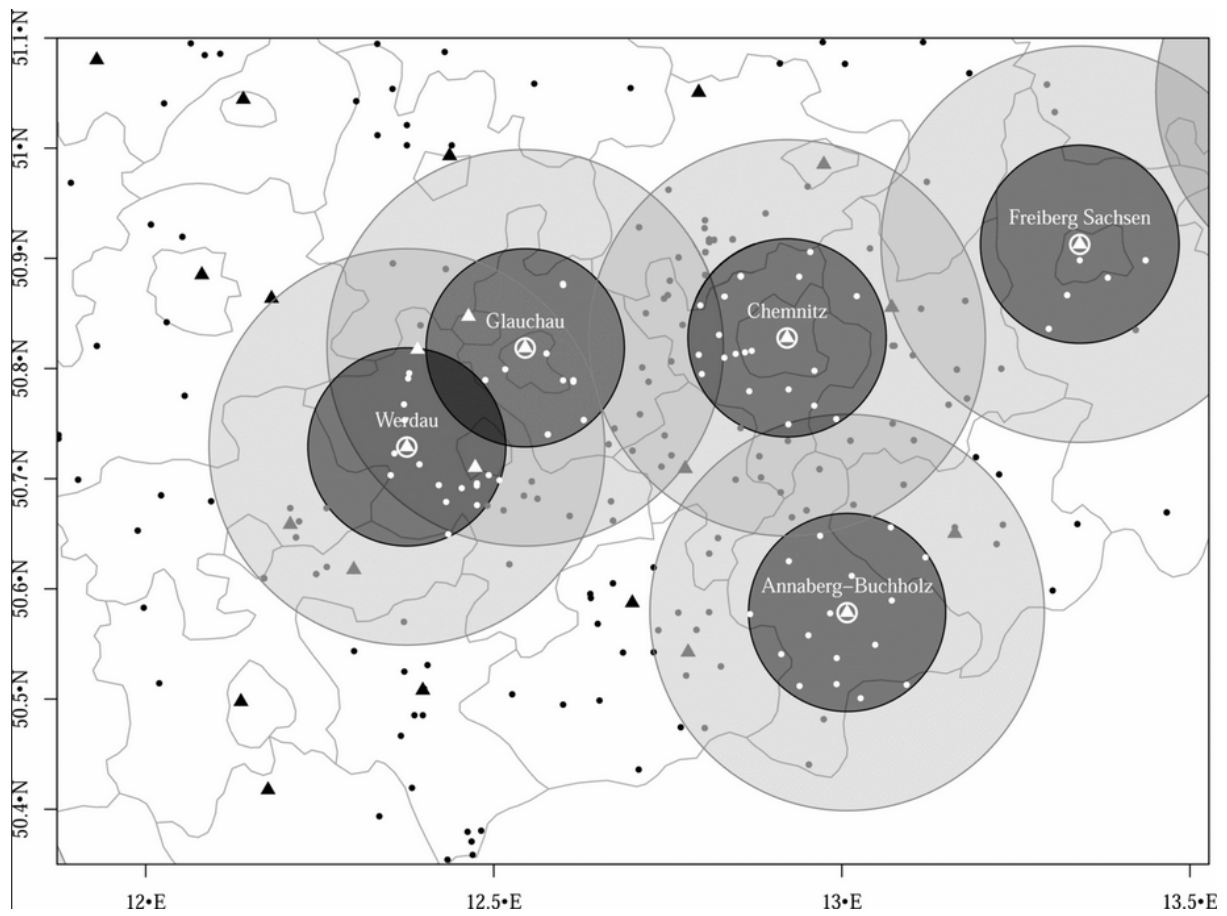


Figure 20: Selb and Munzert (2018)

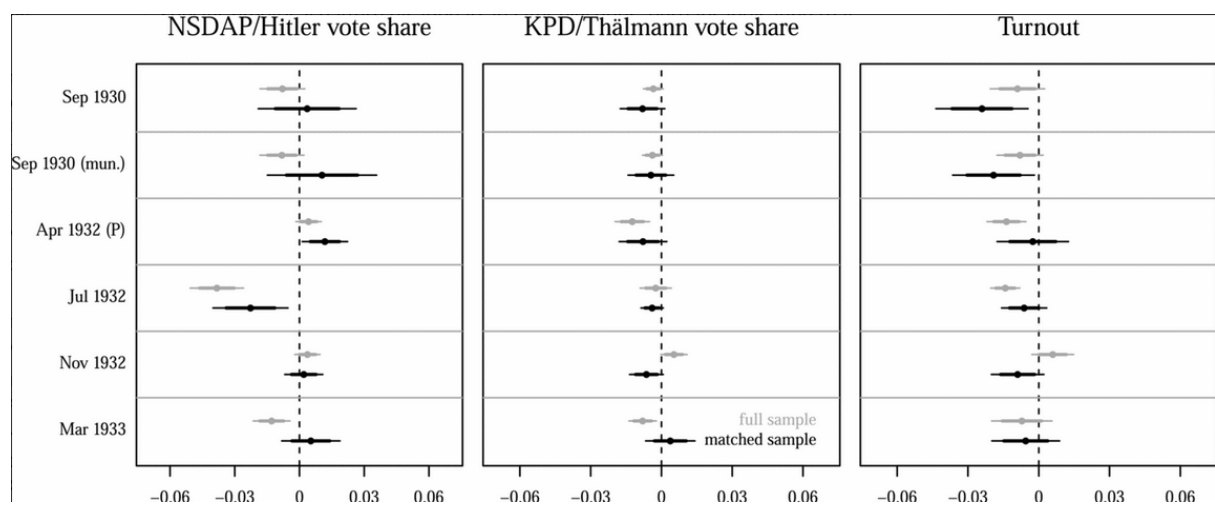


Figure 21: Selb and Munzert (2018)

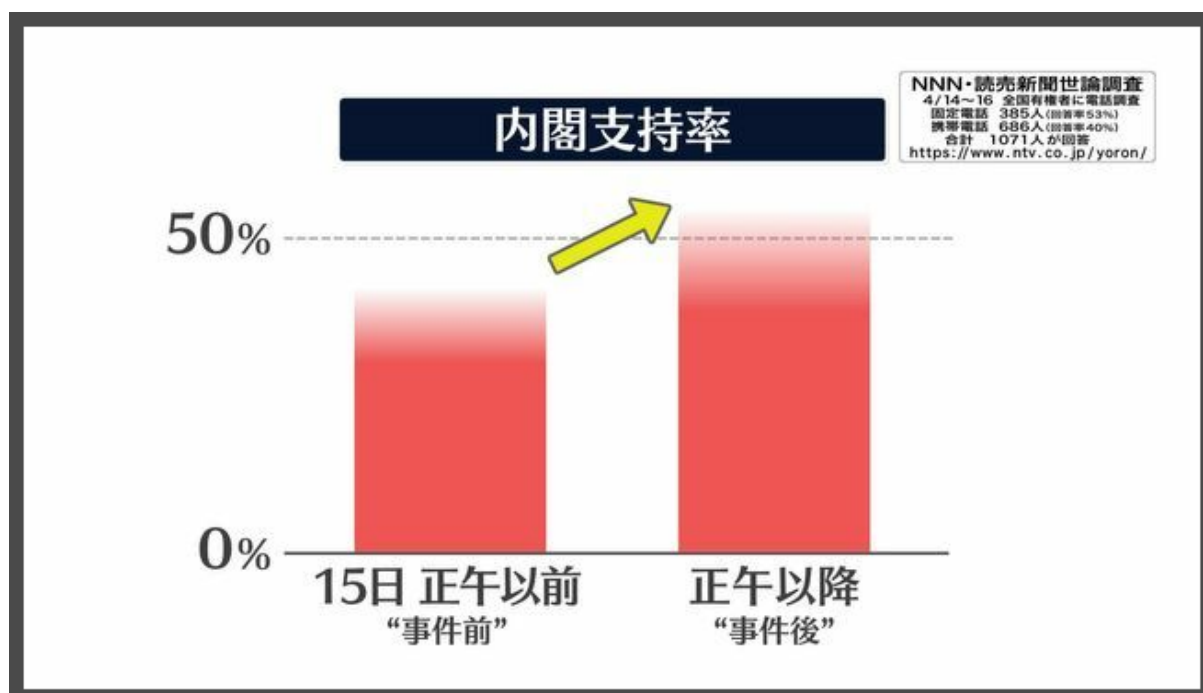


図22: NNN 読売世論調査

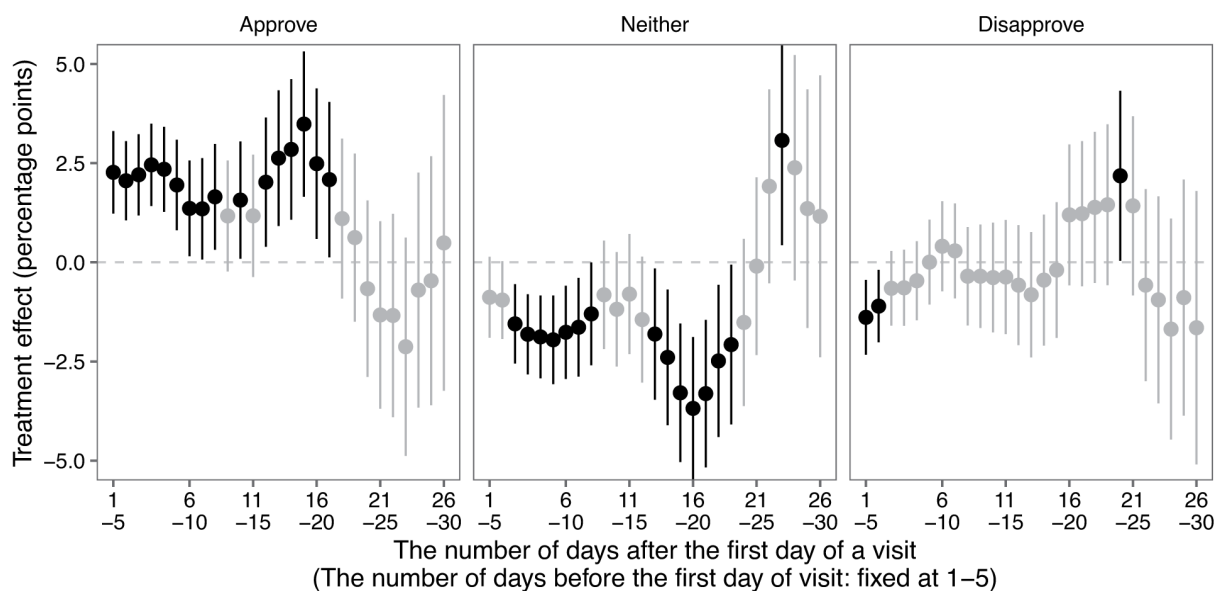


図23: Goldsmith et al. (2021)

FIGURE 1.—GDP TRENDS, TREATED COUNTRY VERSUS SYNTHETIC CONTROL—ASIA

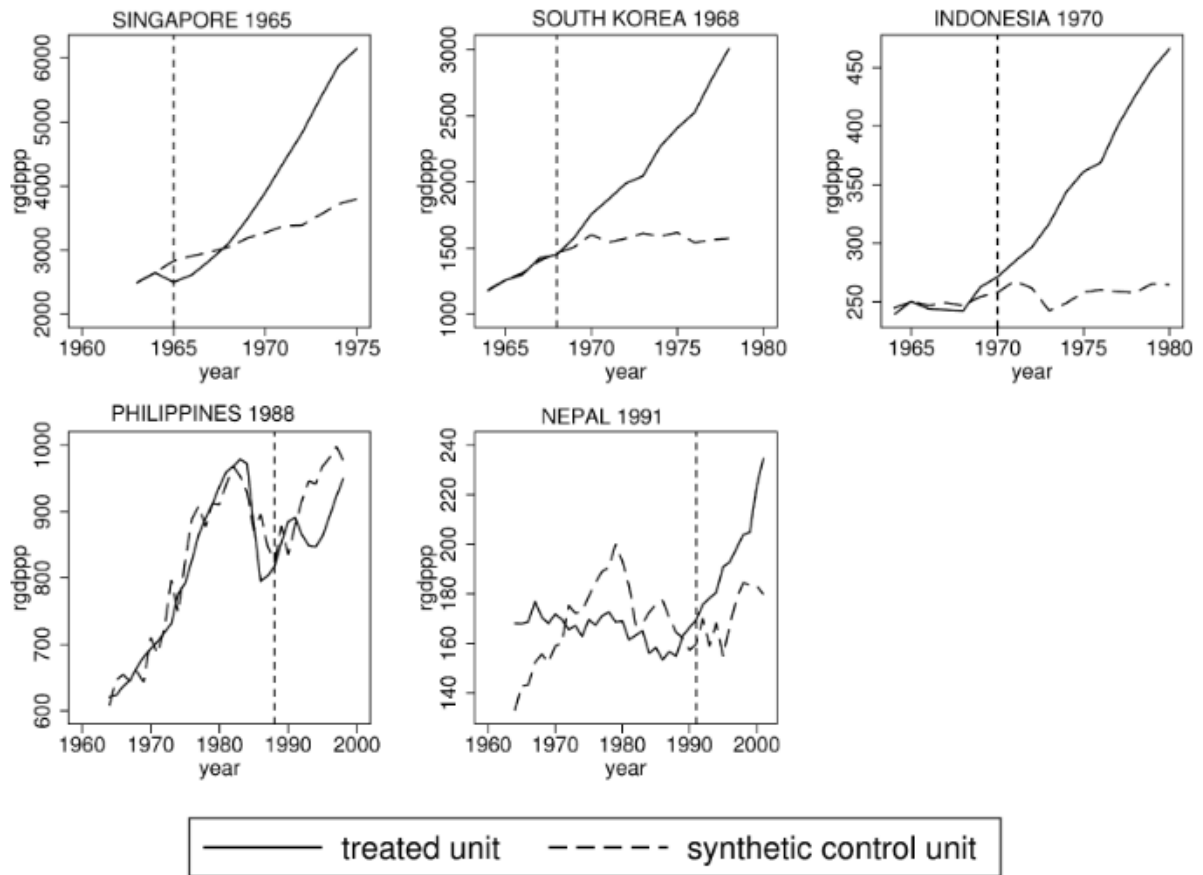


図24: Billmeier and Nannicini (2013)

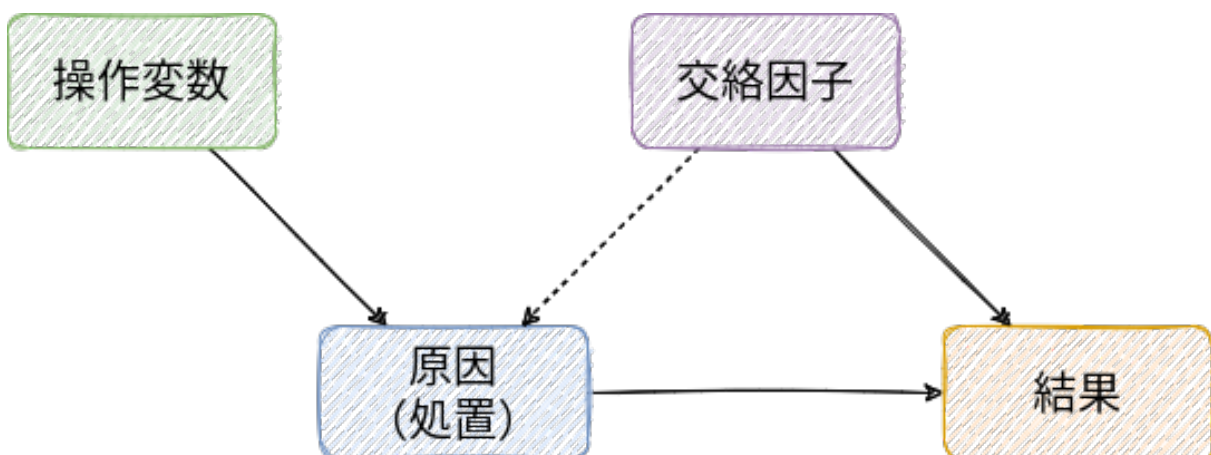


図25: 操作変数のイメージ

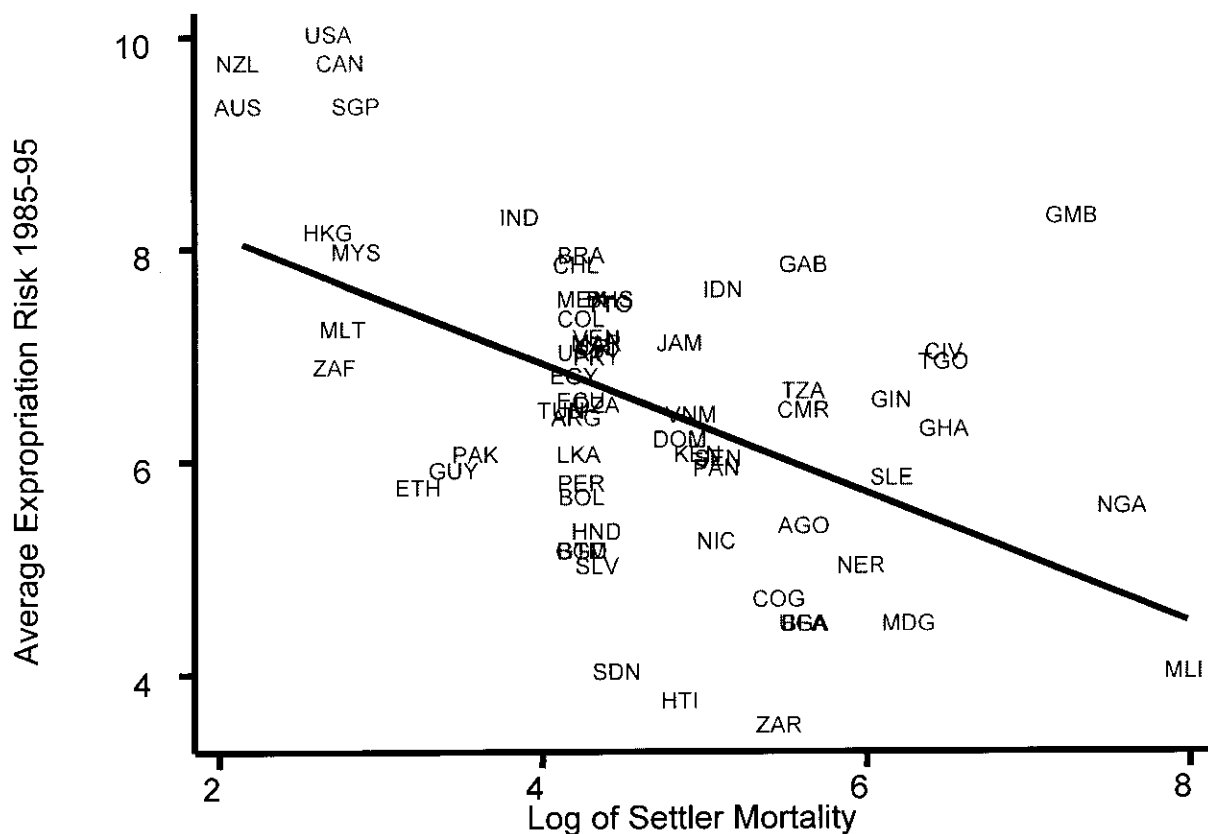


図26: Acemoglu et al. (2001)

size on scholastic achievement,” The Quarterly journal of economics, Vol. 114, No. 2, pp. 533–575.

Banerjee, Abhijit V. and Esther Duflo (2020) 『絶望を希望に変える経済学: 社会の重大問題をどう解決するか』, 日経 BP 日本経済新聞出版本部 and 日経 BP マーケティング (発売).

Barberá, Pablo and Gonzalo Rivero (2015) “Understanding the political representativeness of Twitter users,” Social Science Computer Review, Vol. 33, No. 6, pp. 712–729.

Billmeier, Andreas and Tommaso Nannicini (2013) “Assessing economic liberalization episodes: A synthetic control approach,” Review of Economics and Statistics, Vol. 95, No. 3, pp. 983–1001.

Brollo, Fernanda and Ugo Troiano (2016) “What happens when a woman wins an election? Evidence from close races in Brazil,” Journal of Development Economics, Vol. 122, pp. 28–45.

Card, David (1990) “The impact of the Mariel boatlift on the Miami labor market,” Ilr Review, Vol. 43, No. 2, pp. 245–257.

Card, David and Alan B Krueger (1994) “Minimum Wages and Employment: A Case Study of the

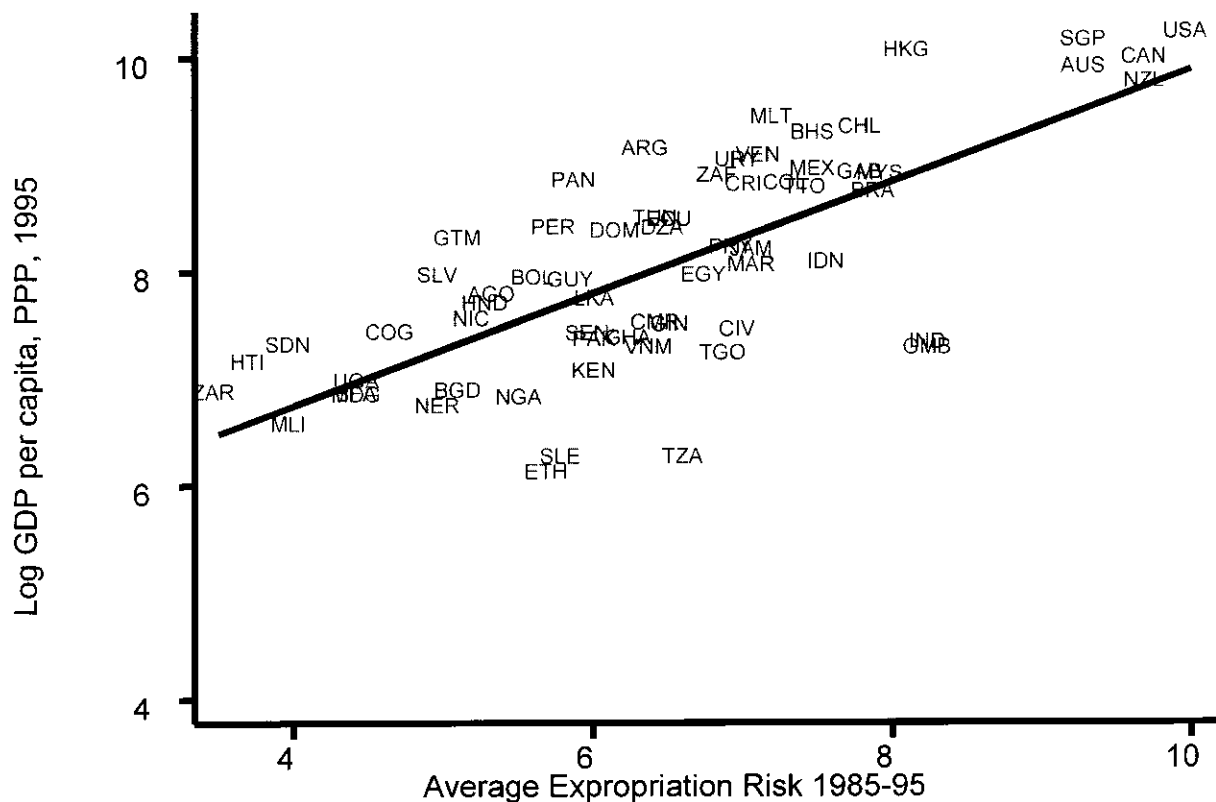


Figure 27: Acemoglu et al. (2001)

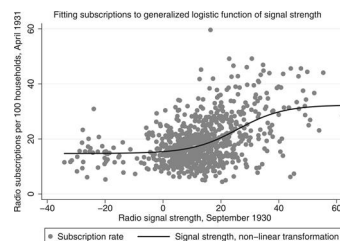


Figure 28: Adena et al. (2015)

Fast-Food Industry in New Jersey and Pennsylvania,” *The American Economic Review*, pp. 772–793.

Goldsmith, Benjamin E, Yusaku Horiuchi, and Kelly Matush (2021) “Does public diplomacy sway foreign public opinion? Identifying the effect of high-level visits,” *American Political Science Review*, Vol. 115, No. 4, pp. 1342–1357.

Keele, Luke J and Rocio Titiunik (2015) “Geographic boundaries as regression discontinuities,” *Political Analysis*, Vol. 23, No. 1, pp. 127–155.

Selb, Peter and Simon Munzert (2018) “Examining a most likely case for strong campaign effects:

TABLE IV
RADIO AND AN INCREASE IN NAZI VOTE SHARE

	(1)	(2)	(3)	(4)
	Change in vote share of the Nazi party since previous elections			
Election dates	Sep 1930 (change from May 1928)		Mar 1933 (change from Nov 1932)	
Panel A: reduced-form estimation				
Radio signal strength	-0.061*** [0.022]		0.045** [0.020]	
Radio signal strength, nonlinear transformation		-0.217*** [0.071]		0.128* [0.071]
Region fixed effects	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes
Observations	958	958	918	918
Panel B: OLS and IV results				
Date for the subscription rate variable	Apr 1931		Apr 1933	
Specification	OLS	IV	OLS	IV
Radio subscription rate, %	-0.086* [0.045]	-0.347*** [0.095]	0.032* [0.017]	0.222* [0.117]
Region fixed effects	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes
Observations	857	855	853	851
F-statistic for the exclusion of the instrument		50.19		20.30
Panel C: Altonji-Elder-Taber tests				
Index of observables (predicted signal strength)	-0.037 [0.070]		-0.047 [0.049]	
Index of observables (predicted nonlinear signal strength)		0.014 [0.210]		-0.348* [0.200]
Population, fifth-order polynomial	Yes	Yes	Yes	Yes
Observations	958	958	918	918
R ²	0.432	0.432	0.573	0.576

Figure 29: Adena et al. (2015)

Table 1: Comparison of demographic characteristics of Knowledge Networks and Mechanical Turk samples

	Weighted KN	Unweighted KN	MT
Female	51.6	49.5	61.3
Age			
18-30	22.8	14.5	50.9
31-45	24.6	21.6	32.4
46-60	28.2	33.4	14.5
61-95	24.4	30.6	2.2
Race			
White, non-Hispanic	72.8	77.8	76.1
Black, non-Hispanic	11.6	8.3	6.2
Hispanic	4.3	8.7	7.4
Other, non-Hispanic	10.1	2.5	7.3
2+ races reported	1.2	2.7	2.9
Education			
Less than high school	9.8	5.5	1.1
High school	31.8	27.3	10.5
Some college	29.7	29.5	41.3
College degree or higher	28.7	37.8	47.1
Marital status			
Married/living with partner	59.8	67.6	50.8
Never married	22.2	16.8	39.8
Divorced/separated	13.0	11.4	8.8
Widowed	5.1	4.2	0.7
Region			
Northeast	18.4	18.0	18.9
Midwest	22.9	23.5	24.1
South	36.5	34.3	35.2
West	22.2	24.2	21.9
Household head	79.0	84.0	69.4
Household income in thousands	60.2	69.0	49.2
SD		(43.2)	(33.9)
N		2,087	1,349

Note: Chi-square test of differences between unweighted KN and MT samples significant for all variables ($p < .001$), except region.

图30: Weinberg et al. (2014)

Hitler’s speeches and the rise of the Nazi party, 1927–1933,” American Political Science Review, Vol. 112, No. 4, pp. 1050–1066.

Shigeoka, Hitoshi (2014) “The effect of patient cost sharing on utilization, health, and risk protection,” American Economic Review, Vol. 104, No. 7, pp. 2152–2184.

Weinberg, Jill D, Jeremy Freese, and David McElhattan (2014) “Comparing data characteristics and results of an online factorial survey between a population-based and a crowdsource-

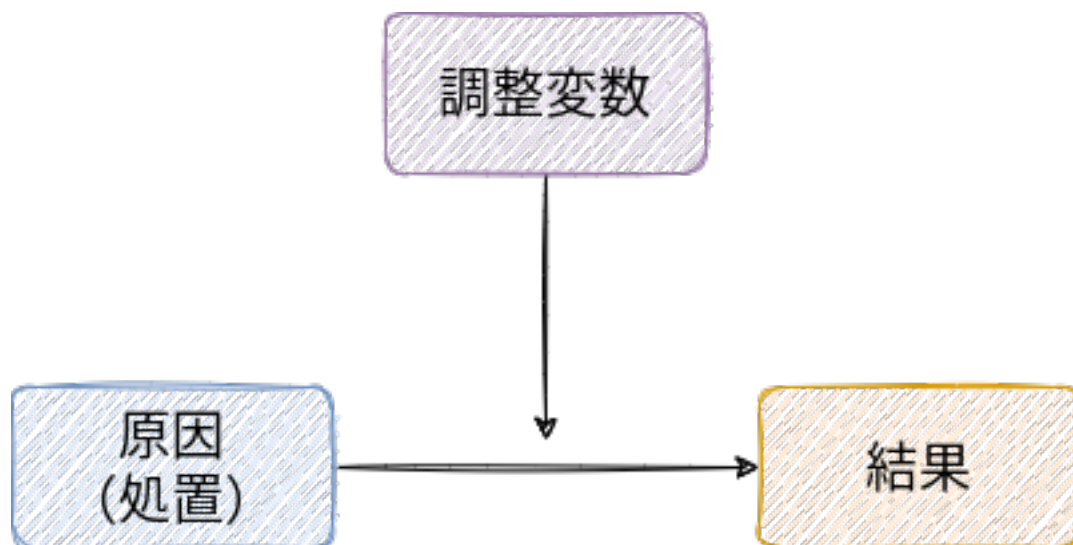


図31: 条件付き効果

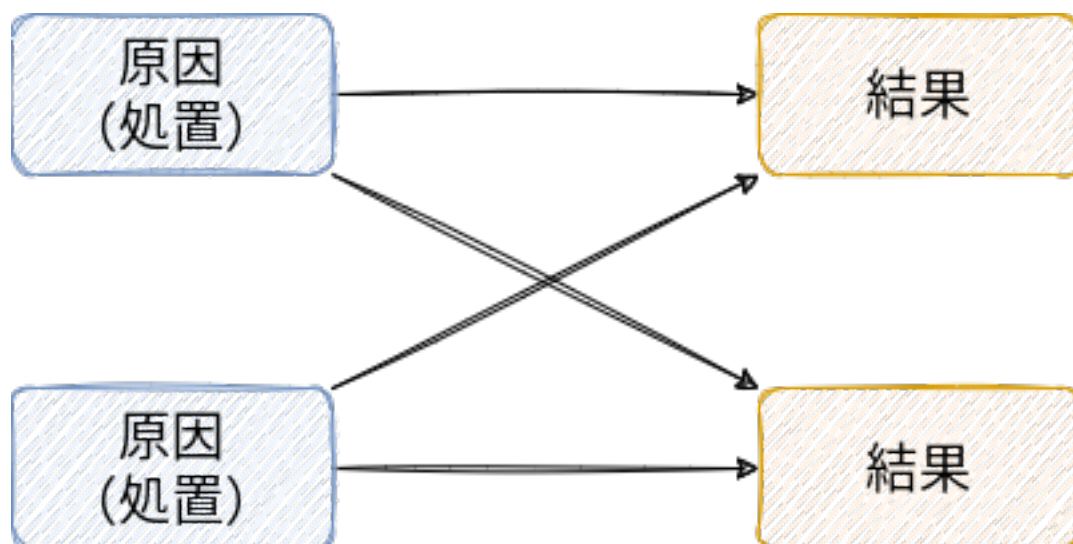


図32: 効果の波及

recruited sample.,” Sociological Science, Vol. 1.

久米郁男 (2013) 『原因を推論する: 政治分析方法論のすゝめ』, 有斐閣.

伊藤修一郎 (2022) 『政策リサーチ入門: 仮説検証による問題解決の技法』, 東京大学出版会, 第増補版版.

北村周平 (2022) 『民主主義の経済学: 社会変革のための思考法』, 日経 BP.

筒井淳也 (2023) 『数字のセンスを磨く: データの読み方・活かし方』, 光文社新書, 光文社.

粕谷祐子 (2018) 「政治学における「因果推論革命」の進行 (特集変わる世界、変わる研究 – ディシプリン/トピック編)」, 『アジ研ワールド・トレンド』, 第 269 巻, 70–71 頁, 03 月.

萩原和樹 (2023) 『データ思考入門』, 講談社現代新書, 講談社.

菅原琢 (2022) 『データ分析読解の技術』, 中公新書ラクレ, 中央公論新社.