

# 政策効果の検証：基礎

技術政策学（データ科学編）

土井翔平

2023-06-05

## はじめに

近年、**証拠に基づく政策立案** (evidence-based policy making: EBPM) の重要性が主張されている。

- 行政官の経験や勘に頼らない意思決定
- policy-based evidence making の回避
- [Google Trends](#) の傾向

↪ 証拠とはなにか？

1. 政策の効果：ある政策によってどの程度、目標のアウトカムは変化したのか？
  - 政策効果と**政策評価**とは似て非なるもの。
  - 政策評価：目標をどの程度、実現したのか？
  - 例：訪日観光客  $n$  万人という目標の実現と、観光政策によって訪日観光客がどの程度増えたのかは別
2. **ロジックモデル**：政策資源の投入から政策成果までの論理的繋がりを可視化し、KPI を定めたもの。
  - あくまで、論理を可視化するもので、それ自体が証拠ではない。

↪ この授業では証拠＝政策効果として議論する。

### 警告

講師は政策評価の専門家ではないので、他の授業（例えば、政策評価論や行政学系のもの）では異なる説明があると思われる。どちらが正しいというものではないことに留意。

- [統計的因果推論の例](#)
- 統計的**因果推論** (statistical causal inference)：原因と結果の関係（効果）を統計的に分析する
  - マーケティングなどでも役に立つ
  - 事例研究をするときにも（データを使わなくても）役に立つ（かも）

# 1 交絡

## 1.1 ワクチンの効果

データ=証拠ではない！

- ・新型コロナワクチンを例に

ワクチン接種者の方が重症者 (sever cases) が多い？

Age	Population (%)		Severe cases		Efficacy vs. severe disease
	Not Vax %	Fully Vax %	Not Vax	Fully Vax	
All ages			214	301	<b>Vax don't work!</b>

図1: Israeli data: How can efficacy vs. severe disease be strong when 60% of hospitalized are vaccinated?

67.5 の人はワクチンを打っていれば重症化しなかった？

Age	Population (%)		Severe cases		Efficacy vs. severe disease
	Not Vax %	Fully Vax %	Not Vax per 100k	Fully Vax per 100k	
All ages	1,302,912 18.2%	5,634,634 78.7%	214 16.4	301 5.3	<b>67.5%</b>

図2: Israeli data: How can efficacy vs. severe disease be strong when 60% of hospitalized are vaccinated?

- ・有効性： $(16.4 - 5.3)/16.4 \approx 67.5\%$

世代で分けると有効性が変わる？

ワクチンを接種するかどうかは（パンデミック初期は）重症化のしやすさに影響を受けていた。

↪ 原因（政策）の有無で結果の違いが生じていても、効果とは言えない！

Age	Population (%)		Severe cases		Efficacy vs. severe disease
	Not Vax %	Fully Vax %	Not Vax per 100k	Fully Vax per 100k	
All ages	1,302,912 18.2%	5,634,634 78.7%	214 16.4	301 5.3	<b>67.5%</b>
<50	1,116,834 23.3%	3,501,118 73.0%	43 3.9	11 0.3	<b>91.8%</b>
>50	186,078 7.9%	2,133,516 90.4%	171 91.9	290 13.6	<b>85.2%</b>

図3: Israeli data: How can efficacy vs. severe disease be strong when 60% of hospitalized are vaccinated?

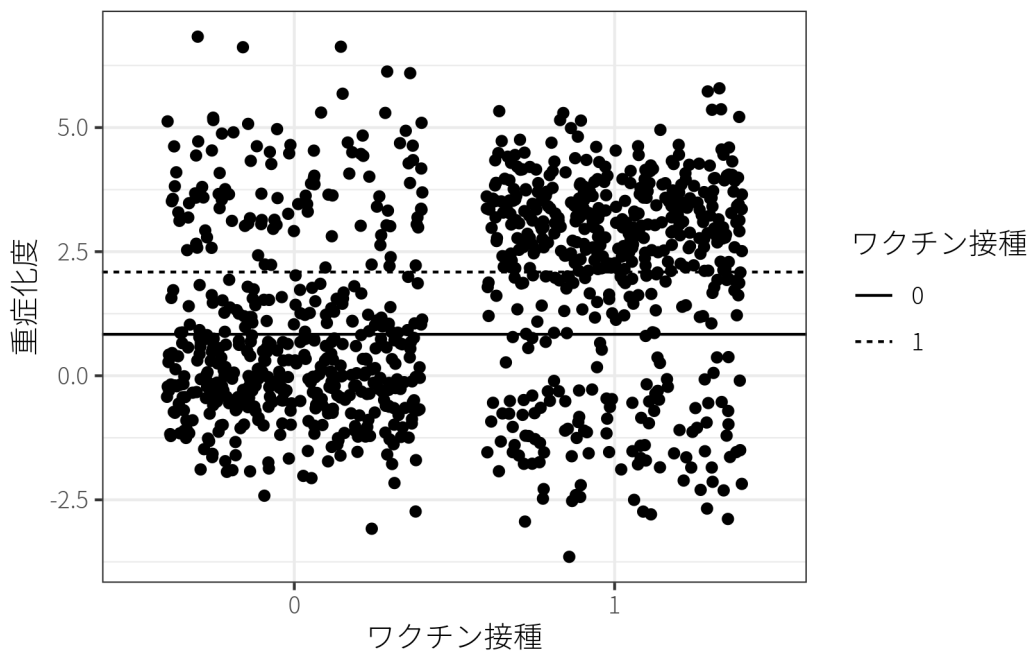


図4: ワクチン接種と重症化の架空の例

## 1.2 交絡

なぜ、単純な比較をするだけでは正しく効果を計算できなかったのか？

交絡 (confounding) : 原因と関係し、結果にも影響するような第三の要因がある状況

- そのような要因を交絡因子 (confounder) や共変量 (covariate) と呼ぶ。

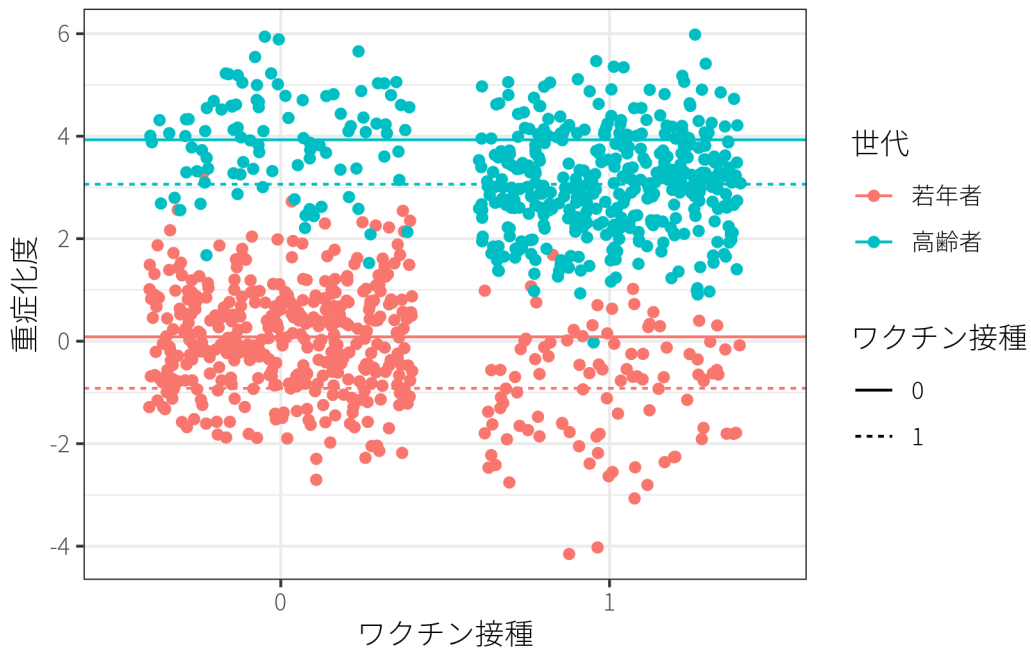


図5: ワクチン接種と重症化の架空の例

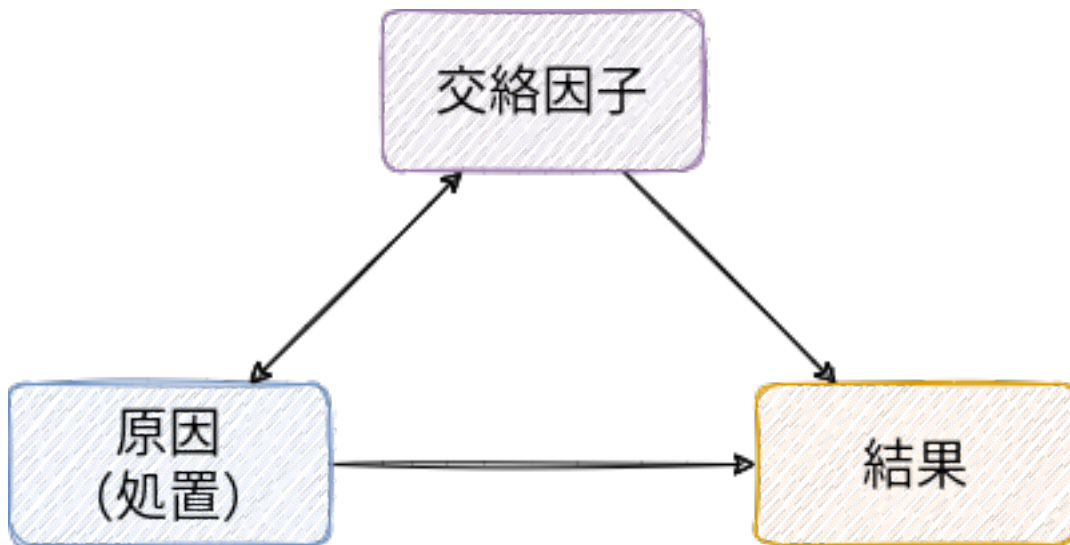


図6: 交絡の可視化

- 原因 → 結果の関係を知りたいけれど、原因 ↔ 交絡因子 → 結果の関係（バックドアパス）があるので、正確に分析できない。
- 因果関係ではないけれど相関関係が生じていることを見かけの相関 (spurious correlation) と呼ぶ。<sup>1</sup>
  - 相関関係は因果関係の前提と言われることがあるが、そうではない点に注意。

<sup>1</sup> 本来は無関係なものが相関している状況を指していた。

どのような交絡の例があるだろうか？

↪ 交絡を取り除かない限り、データから効果を示すことはできない。

- 事例分析をする際も同様
  - ある政策を行った自治体とそうではない自治体
  - ある自治体がある政策を行う前と後

## 2 ランダム化比較試験

理想：全く同じ人がワクチンを受けた場合と受けなかった場合に重症化するかどうかを比較する。

↪ 不可能

現実：同じような集団がワクチンを受けた場合と受けなかった場合に重症化するかどうかを比較する。

↪ どのようにして「同じような集団」を作るのか？

シンプルかつ強力な方法としての**ランダム化比較試験** (randomized controlled trial: RCT)

- RCT：対象をランダムに分割して、一方には原因を与え、他方には原因を与えず、集団の結果を比較する。

RCTで交絡（バックドア・パス）を消す！

- ランダムにワクチンを摂取すれば年齢などとは無関係なはず。

### 2.1 フィールド実験

フィールド実験：現実世界にランダムに介入して、**実際の行動**の変化を分析

- A/B テストを初めとするオンラインテスト
- 実際に政策をランダムに試行する。

開発経済学を中心に RCT が活用 ([Banerjee and Duflo, 2012](#); [Leigh, 2020](#))

- 貧困層が移住しないのは資金が足りないからなのか、情報が足りないからなのか？ ([Bryan et al., 2014](#))
- 中等教育は経済的に豊かになるのか？ ([Duflo et al., 2021](#))
- どのようなメッセージだと人々は投票へ行くのか？ ([Gerber et al., 2008](#))

### 2.2 サーベイ実験

サーベイ実験：世論調査（サーベイ）にランダムな項目を入れ、**表明された意見**の変化を分析 ([Song・秦, 2020](#))

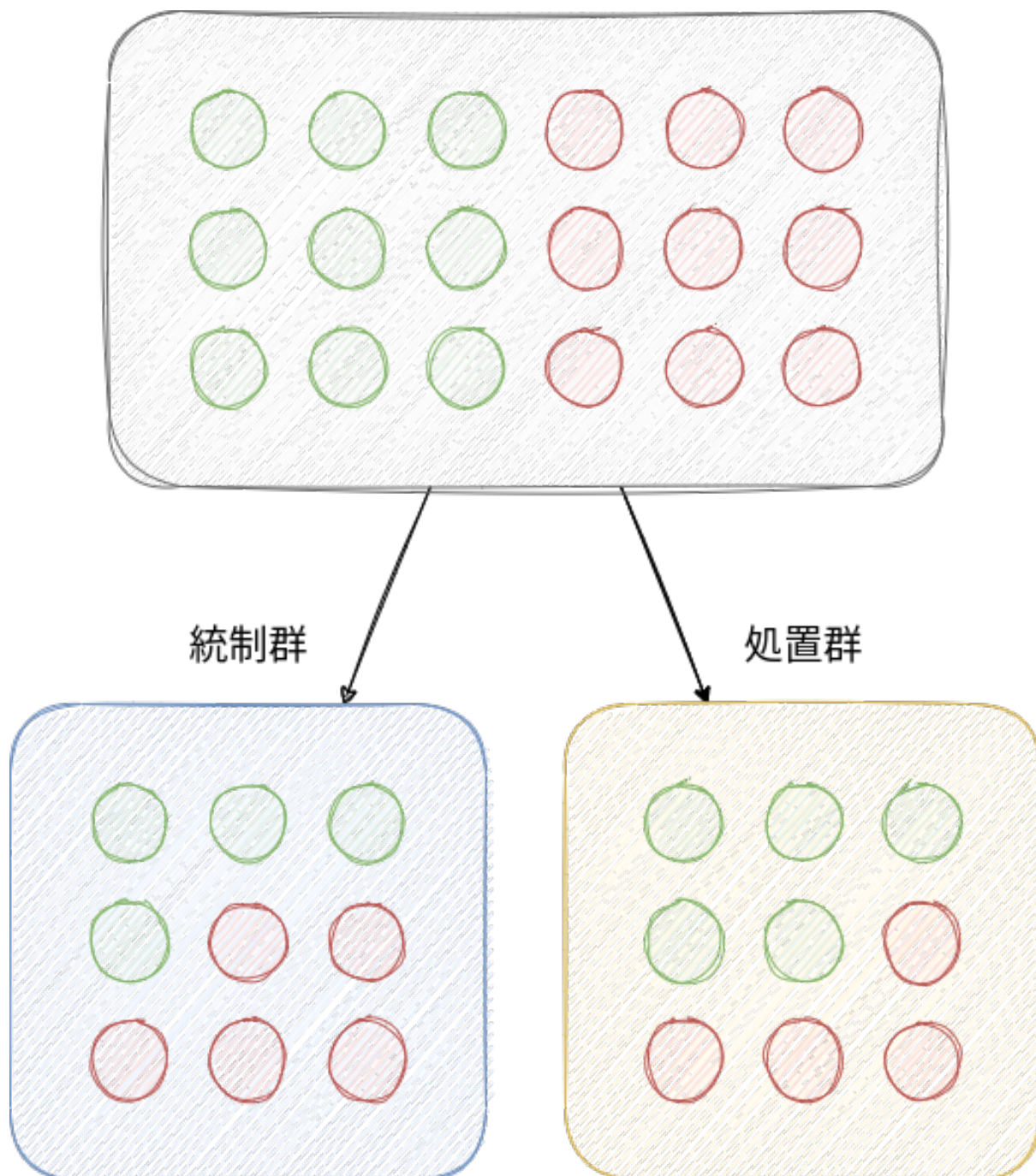


図7: RCT のイメージ

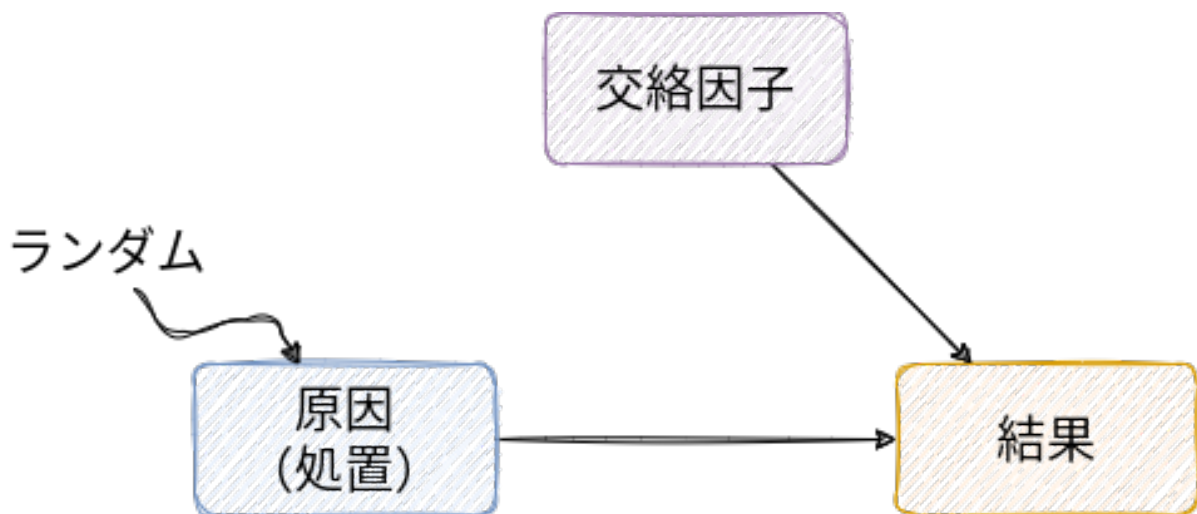


図8: RCT の可視化

TABLE II  
PROGRAM TAKE-UP RATES<sup>a</sup>

	<i>Incentivized</i>	Cash	Credit	<i>Not Incentivized</i>	Info	Control	<i>Diff. (I – NI)</i>
Migration rate in 2008	58.0% (1.4)	59.0% (1.9)	56.8% (2.1)	36.0% (2.0)	35.9% (2.8)	36.0% (2.8)	22.0*** (2.4)
Migration rate in 2009	46.7% (1.4)	44.6% (1.9)	49.1% (2.1)	37.5% (2.0)	34.4% (2.8)	40.5% (2.9)	9.2*** (2.5)
Migration rate in 2011 <sup>b</sup>	39% (2.1)			32% (2.5)			7.0** (3.3)

図9: Bryan et al. (2014)

サーベイ実験は政治学や社会学を中心に利用

- 人々は移民に関する事実を知ると寛容になるのか? (Alesina et al., 2023; Barrera et al., 2020)
  - 人々は移民の割合などを過大に評価している。
- 移民の事実に関する質問と再配分政策への意見に関する質問の順番をランダムにする。
  - 移民の事実に関する誤解に気づいた人は再配分に寛容になる?
- 移民に関する情報を以下のうちから 1 つだけランダムに提示し、マリーヌ・ル・ペンへの支持を調査

1. なんにも示さない
2. マリーヌ・ル・ペンの主張 (事実ではない)
3. 事実
4. 2 と 3 の両方

Table 6: Labor Market Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Worked for pay in past 6 months (2019)	Has wage contract with employer (2019)	Job with benefits (2019)	Public sector employee (2019)	Lives in urban area (2019)	Self- employed (2019)	Total earnings in past 6 months (2019)	Could not cope with 200 GHX emergency (2019)
<b>Panel A: All</b>								
Treatment	0.011 (0.021)	0.039 (0.015)	0.030 (0.015)	0.019 (0.013)	-0.015 (0.015)	-0.029 (0.020)	37.123 (93.450)	-0.027 (0.017)
P-value	0.589	0.008	0.052	0.157	0.330	0.153	0.691	0.117
Comparison mean	0.730	0.084	0.099	0.077	0.123	0.245	1456.217	0.161
N	1952	1951	1951	1952	1921	1952	1915	1951
<b>Panel B: Female</b>								
Treatment	0.033 (0.033)	0.041 (0.019)	0.020 (0.019)	0.041 (0.019)	-0.029 (0.021)	-0.012 (0.031)	35.794 (108.464)	-0.044 (0.024)
P-value	0.314	0.032	0.283	0.031	0.152	0.683	0.741	0.070
Comparison mean	0.602	0.063	0.075	0.063	0.119	0.287	951.456	0.176
N	986	986	986	986	973	986	972	986
<b>Panel C: Male</b>								
Treatment	-0.020 (0.024)	0.035 (0.023)	0.037 (0.024)	-0.003 (0.019)	-0.001 (0.023)	-0.042 (0.026)	-12.740 (145.790)	-0.009 (0.024)
P-value	0.405	0.119	0.126	0.874	0.959	0.106	0.930	0.718
Comparison mean	0.864	0.106	0.125	0.092	0.128	0.201	1993.862	0.146
N	966	965	965	966	948	966	943	965
P-val male=fem	0.207	0.856	0.572	0.092	0.421	0.536	0.838	0.319

Notes: See Table 2 notes. Total earnings in last 6 months winsorized at the 99% level. Labor market outcomes for 2017 shown in Table A8, those for 2020 shown in Table 7.

Figure 10: Duflo et al. (2021)

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

Figure 11: Gerber et al. (2008)



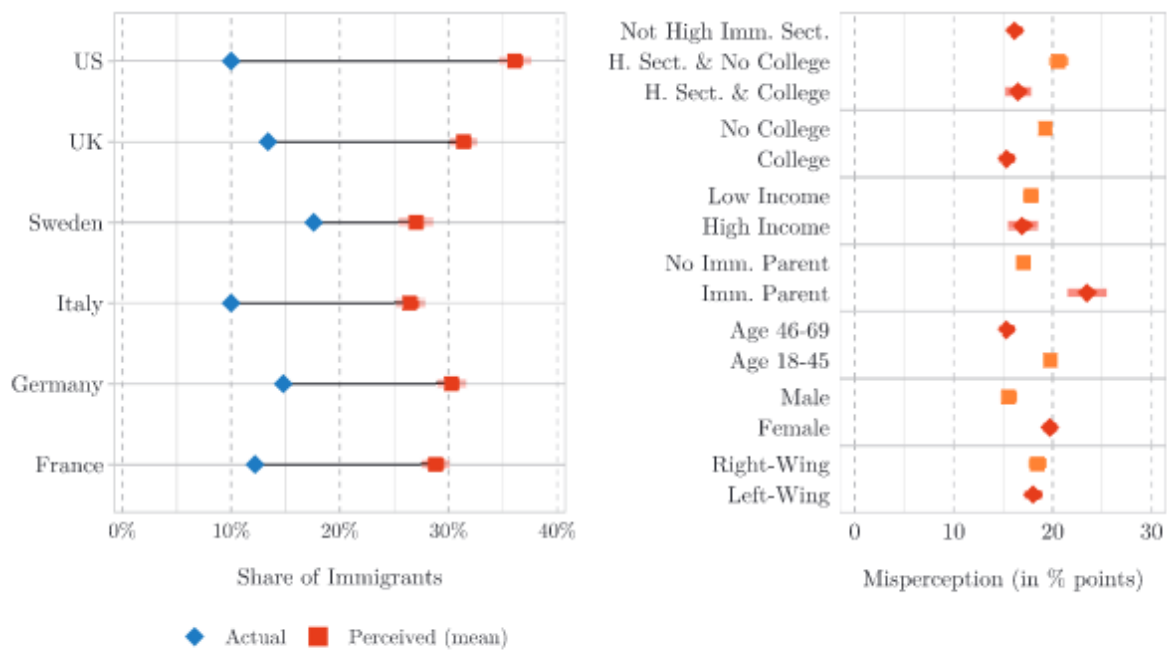


FIGURE 2  
Perceived vs. actual share of immigrants

Figure 12: Alesina et al. (2023)

TABLE 4  
Treatment effects on support for redistribution

	Tax top 1 (1)	Tax bottom 50 (2)	Social budget (3)	Education budget (4)	Inequality serious problem (5)	Donation above median (6)
Order/salience T	-1.948*** (0.416)	0.914*** (0.276)	-0.543** (0.238)	0.439** (0.175)	-0.0280** (0.0132)	-0.0479*** (0.0138)
T: share of immigrants	-0.627 (0.419)	0.0449 (0.278)	-0.479** (0.233)	0.188 (0.172)	-0.00590 (0.0133)	-0.0165 (0.0140)
T: origin of immigrants	-0.0662 (0.425)	0.0322 (0.284)	-0.465* (0.239)	0.164 (0.173)	0.00626 (0.0132)	0.00208 (0.0140)
T: hard work	0.0772 (0.422)	-0.212 (0.279)	-0.0944 (0.235)	0.333** (0.170)	0.0158 (0.0132)	0.00910 (0.0139)
Observations	19,765	19,765	19,765	19,765	19,763	19,765
Control mean	37.12	10.94	29.53	16.00	0.59	0.47

Figure 13: Alesina et al. (2023)

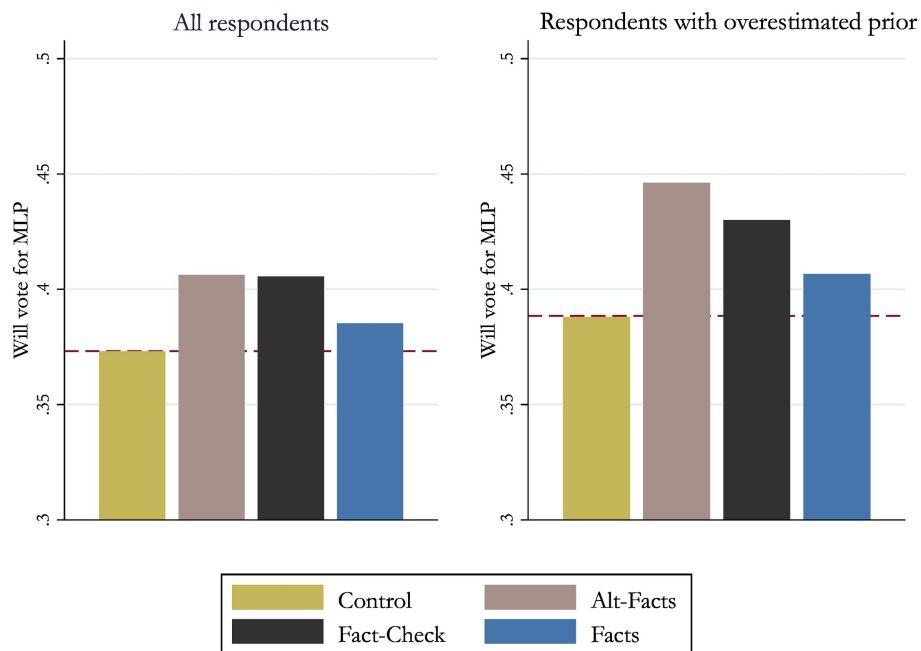


図14: Barrera et al. (2020)

### 2.2.1 コンジョイント実験

コンジョイント実験：2つ（以上）の選択肢を提示し、その要因をランダムに変化させ、どの要因が選択に影響を与えているのかを分析

- 人々はどのような政策を重視して投票するのか？
  - 衆院総選挙、緊急解析！ データが明かした有権者の本音
  - マニフェスト選挙を疑え：2021年総選挙の計量政治学

### 2.2.2 リスト実験

社会的望ましきバイアス (social desirability bias: SDB)：回答者は社会的に望ましい答えをしようとして本音を話さない傾向

リスト実験：該当する項目の数を尋ねることでSDBを回避する実験手法

- 人々はどの程度、人種差別をしているのか？(Kuklinski et al., 1997)
- 知りたい項目が入っているものと、そうでないものをランダムに表示させ、該当数を尋ねる。

## 3 自然実験

自然実験 (natural experiment)：RCTではないがRCTと同じような状況

仮に、次のような公約を掲げた2つの政党が今回の総選挙で候補者を擁立していると想定してください。あなたは、どちらの政党を支持しますか。もし、どちらを支持するかはつきりとは言えない場合でも、どちらか一方、あえていえば支持する方を選んでください。

	政党1	政党2
雇用政策	年功序列を撤廃して 労働市場の流動化を促す	多様な働き方を認め、正規・ 非正規を問わず雇用を拡大
金融財政政策	大胆な金融緩和と機動的な 財政出動によりデフレ脱却	過度の金融緩和や円安、 公共事業のパラマキを是正
TPP	TPPへの参加反対	参加して積極的に自由化を推進
議員定数削減	議員定数削減を実現する	選挙制度調査会の答申を尊重し、 よりよい選挙制度改革に取り組む
集団的自衛権	閣議決定のみに基づく行使の 容認には反対	閣議決定のみに基づく行使の 容認には反対
成長戦略	地方産業・中小企業の 活性化による成長実現	農業・医療など 岩盤規制を打破
憲法改正	現行憲法条文のいかなる変更にも 反対。平和憲法を守る	現行憲法の基本原理を維持した上で 必要な条文を追加
消費増税	期限を決めずに延期	2017年4月に10%にし、 軽減税率を導入
原発再稼働	安全基準に合格すれば 認める	責任ある逃避計画など 厳しい条件で容認

どちらを支持するか

政党1

政党2

図15: コンジョイント分析の例

## 争点ごとに示される各政党の政策に対する支持態度

●コンジョイント分析が示した「政策パッケージ」選択確率

● 政党名が表示されない場合 ● 自民党の政策として表示された場合

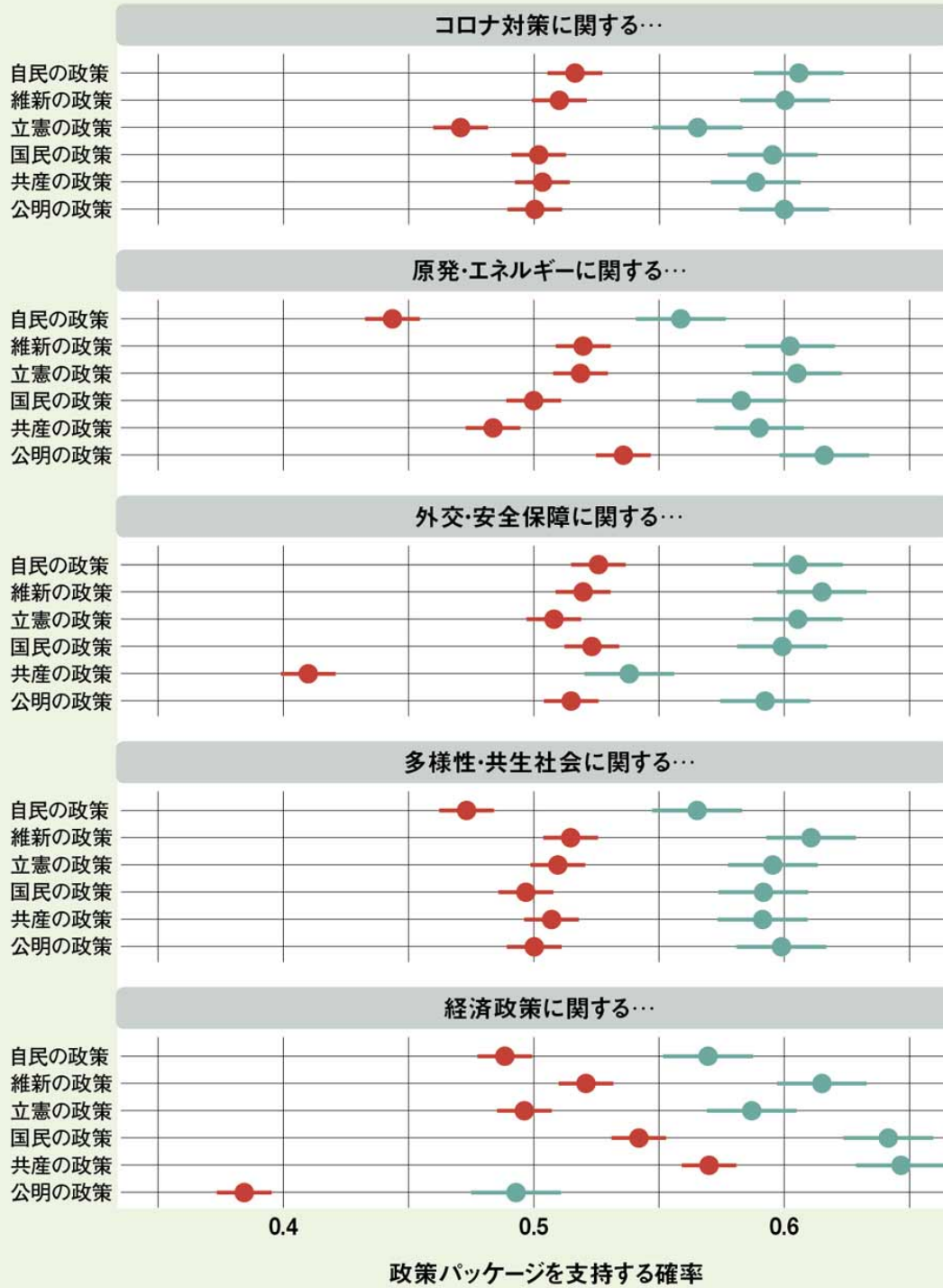


図16: コンジョイント分析の結果

Now I am going to read you three things that sometimes make people angry or upset. After I read all three, just tell me HOW MANY of them upset you. I don't want to know which ones, just HOW MANY.

With these ground rules established, the interviewer then reads a list of three items:

- (1) the federal government increasing the tax on gasoline;
- (2) professional athletes getting million-dollar contracts;
- (3) large corporations polluting the environment.<sup>3</sup>

Some randomly assigned respondents receive the baseline version. Others, in the test condition, receive the three baseline items plus a fourth, in this case “a black family moving in next door.”

図17: Kuklinski et al. (1997)

- ナショナリズムの高揚は武力紛争に繋がるのか? (Bertoli, 2017)

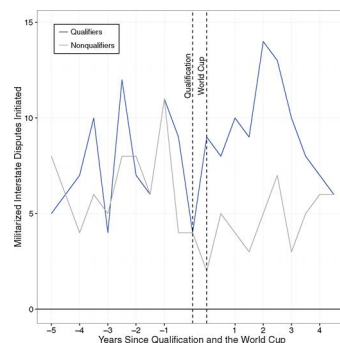


図18: Bertoli (2017)

- 政治的指導者の交代は民主化や平和に繋がるのか? (Jones and Olken, 2009)
- 女性医師による治療は死亡率に影響するのか? (Tsugawa et al., 2017)

### 3.1 ランダムな公共政策

行政がランダムな施策を行う場合、それは RCT とみなせる。

- 女性の政治参加は重要なのか?
- 汚職の監査には効果があるのか?

TABLE 5—ASSASSINATIONS AND INSTITUTIONAL CHANGE

	Absolute change in POLITY2 dummy (1)	Directional change in POLITY2 dummy (2)	Percentage of “regular” leader transitions in next 20 years (3)
<i>Panel A: Average effects</i>			
Success	0.091 (0.047)	0.079 (0.051)	0.111 (0.057)
Parm. <i>p</i> -value	0.06*	0.12	0.06*
Nonparm. <i>p</i> -value	0.03**	0.02**	0.18
Observations	221	221	138
Data source	Polity IV	Polity IV	Archigos
<i>Panel B: Split by regime type in year before attempt</i>			
Success × autocracy		0.131 (0.055)	0.191 (0.085)
Success × democracy		−0.012 (0.083)	0.034 (0.043)
Autocracy—parm. <i>p</i>		0.02**	0.03**
Autocracy—nonparm. <i>p</i>		0.01***	0.05**
Democracy—parm. <i>p</i>		0.89	0.43
Democracy—nonparm. <i>p</i>		0.13	0.96
Observations		221	133
Data source	Polity IV	Polity IV	Archigos

図19: Jones and Olken (2009)

## 4 統計的仮説検定

### 警告

統計的仮説検定は非常に難しいので、分からなくても構わない。講師を含めてちゃんと理解できているか怪しい。

RCT や自然実験であれば、因果効果の大きさは明らかにできる？

↪ 偶然、(本来は効果がないはずなのに) 2つのグループで差が出てしまった可能性

**統計的仮説検定**：効果が現れたのが偶然ではないかどうかを判別する方法

1. 仮に本当は効果がないとする (帰無仮説: null hypothesis)
2. 本当は効果がないのに、効果があるように見える実際のデータが生じる確率 (p 値: p-value) を求める。
  - p 値を求めるときには推定結果の不確か性を表す標準誤差 (standard error: SE) を用いる。
3. p 値が予め設定しておいた値 (例えば 5%) を下回っている場合、**統計的に有意**であると呼ぶ。
  - 本当は効果がないのに 5% の確率で生じる結果が出たのだとしたら、もはや「効果がない」という

TABLE 7—ASSASSINATIONS AND CONFLICT: CHANGE ONE YEAR AFTER ATTEMPT

	Gleditsch-COW dataset 1875–2002 (1)	Gleditsch-COW dataset 1946–2002 (2)	PRIO/Uppsala dataset 1946–2002 (3)
<i>Panel A: Average effects</i>			
Success	−0.072 (0.068)	0.041 (0.093)	0.162 (0.071)
Parm. <i>p</i> -value	0.29	0.66	0.02**
Nonparm. <i>p</i> -value	0.57	0.83	0.03**
Observations	223	116	116
Data source	Gleditsch	Gleditsch	PRIO
<i>Panel B: Split by war status in year before attempt</i>			
Success × intense war	−0.255 (0.144)	−0.103 (0.257)	−0.110 (0.294)
Success × moderate war			0.334 (0.163)
Success × not at war	−0.024 (0.068)	0.020 (0.086)	0.070 (0.057)
Intense war—parm. <i>p</i> -value	0.08*	0.69	0.71
Intense war—nonparm. <i>p</i> -value	0.13	1.00	0.69
Moderate war—parm. <i>p</i> -value	N/A	N/A	0.05**
Moderate war—nonparm. <i>p</i> -value	N/A	N/A	0.13
Not at war—parm. <i>p</i> -value	0.73	0.82	0.22
Not at war—nonparm. <i>p</i> -value	0.62	0.71	0.21
Observations	222	116	116
Data source	Gleditsch	Gleditsch	PRIO

Figure 20: Jones and Olken (2009)

Table 2. Association Between Physician Sex and 30-Day Patient Mortality and Readmissions

Characteristic	No. of Hospitalizations (No. of Physicians)	Patient Outcomes, % (95% CI)		Adjusted Risk Difference, % (95% CI) <sup>a</sup>	P Value
		Female Physicians	Male Physicians		
<b>30-d Mortality rate</b>					
Model 1: risk-adjusted 30-d mortality rate <sup>b</sup>	1 583 028 (57 896)	10.82 (10.71 to 10.93)	11.49 (11.42 to 11.56)	−0.67 (−0.80 to −0.54)	<.001
Model 2: Model 1 + hospital fixed effects	1 583 024 (57 896)	10.91 (10.81 to 11.01)	11.46 (11.40 to 11.52)	−0.55 (−0.67 to −0.42)	<.001
Model 3: Model 2 + physician characteristics	1 283 621 (46 201)	11.07 (10.95 to 11.19)	11.49 (11.43 to 11.56)	−0.43 (−0.57 to −0.28)	<.001
<b>30-d Readmission rate</b>					
Model 1: risk-adjusted 30-d readmission rate <sup>b</sup>	1 540 797 (57 876)	15.01 (14.89 to 15.14)	15.57 (15.49 to 15.65)	−0.55 (−0.71 to −0.41)	<.001
Model 2: Model 1 + hospital fixed effects	1 540 797 (57 876)	15.00 (14.89 to 15.13)	15.57 (15.50 to 15.64)	−0.56 (−0.70 to −0.42)	<.001
Model 3: Model 2 + physician characteristics	1 249 210 (46 205)	15.02 (14.88 to 15.15)	15.57 (15.49 to 15.64)	−0.55 (−0.71 to −0.39)	<.001

<sup>a</sup> Adjusted risk differences between female and male physicians. SEs were clustered at the physician level.

<sup>b</sup> Risk adjustment using patients' age, sex, race, primary diagnosis, coexisting conditions (Elixhauser comorbidity index), median household income, Medicaid status, and year indicators.

Figure 21: Tsugawa et al. (2017)

Dependent Variables	West Bengal		
	Mean, Reserved GP (1)	Mean, Unreserved GP (2)	Difference (3)
<i>A. Village Level</i>			
Number of Drinking Water Facilities Newly Built or Repaired	23.83 (5.00)	14.74 (1.44)	9.09 (4.02)
Condition of Roads (1 if in good condition)	.41 (.05)	.23 (.03)	.18 (.06)
Number of Panchayat Run Education Centers	.06 (.02)	.12 (.03)	-.06 (.04)
Number of Irrigation Facilities Newly Built or Repaired	3.01 (.79)	3.39 (.8)	-.38 (1.26)
Other Public Goods (ponds, biogas, sanitation, community buildings)	1.66 (.49)	1.34 (.23)	.32 (.48)
Test Statistics: Difference Jointly Significant ( <i>p</i> -value)			4.15 (.001)

Figure 22: Chattopadhyay and Duflo (2004)

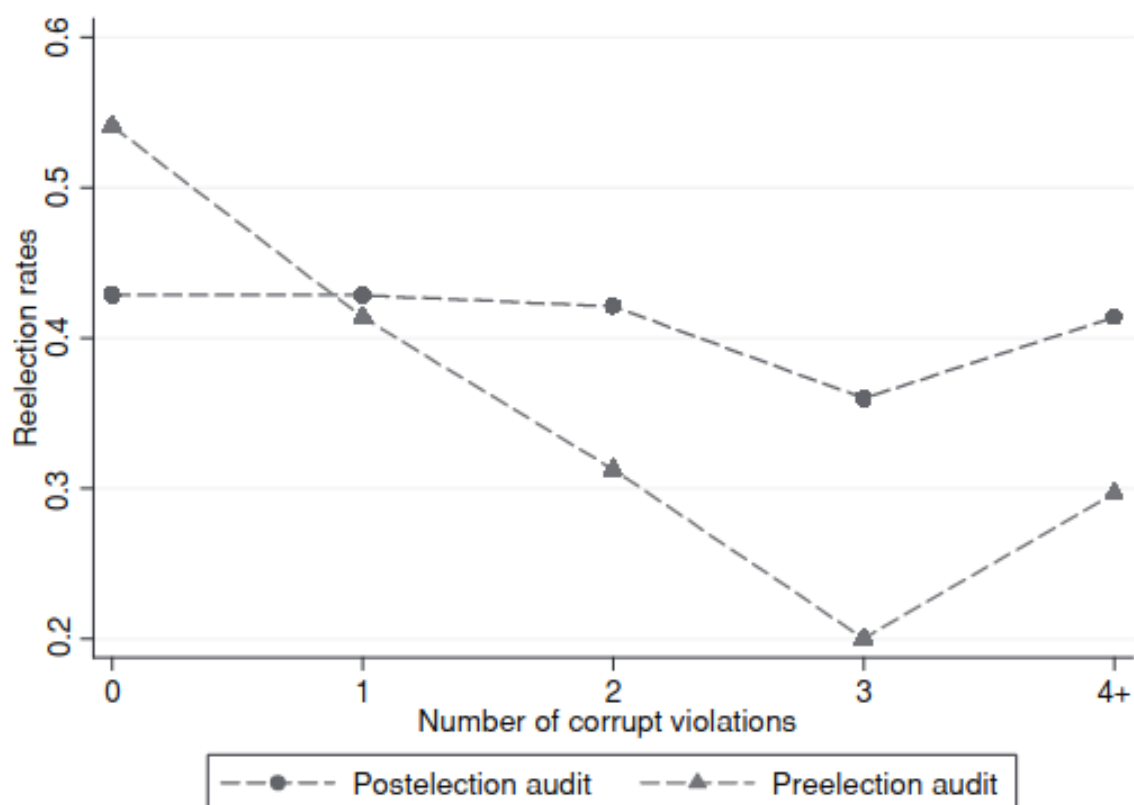


FIGURE III  
Relationship between Reelection Rates and Corruption Levels

Figure 23: Ferraz and Finan (2008)



前提がおかしいのではないか。

→ とりあえず、統計的に有意でなければ効果があると強く主張できない。

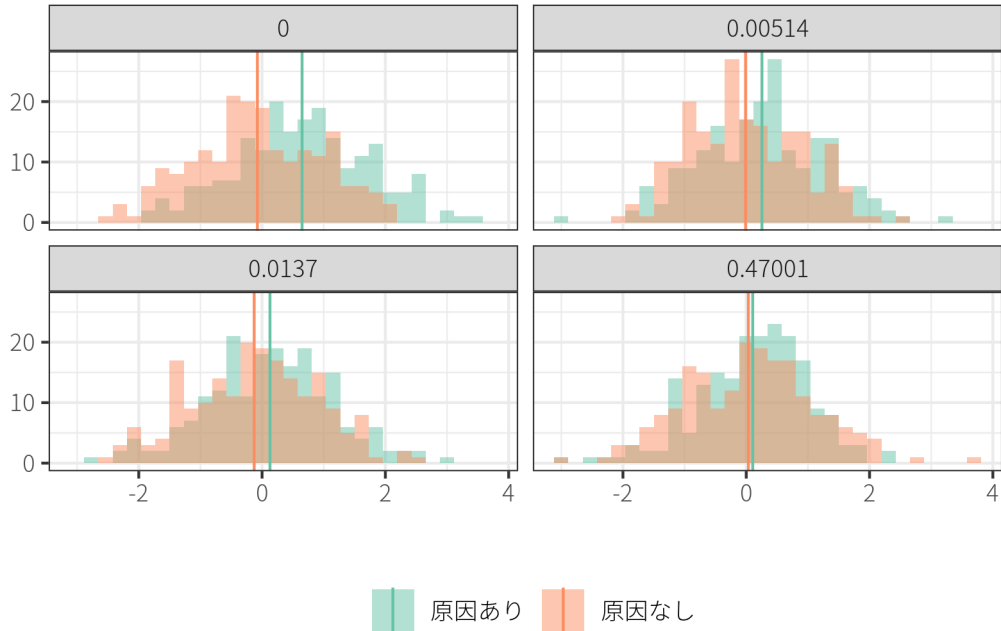


図24: 効果の大きさと統計的有意性

1. 上記の代わりに、効果の値を標準誤差で割ったもの（ $z$  値や  $t$  値）が、およそ 2 以上であれば 5% 有意水準で統計的に有意である。
2. 信頼区間を求めて、信頼区間が 0（などの基準点）を含まなかったら統計的に有意であると判断する方法もある。

5% 水準で統計的に有意  $\Leftrightarrow$   $p$  値が 0.05 以下  $\Leftrightarrow$   $z$  値や  $t$  値が約 2 以上  $\Leftrightarrow$  95% 信頼区間が 0 を含まない

#### 4.1 誤解・注意事項

1.  $p$  値が低ければ効果が大きい
2. 統計的に有意ではないから関係ない
3.  $p$  値は「効果がない」確率ではない

#### 4.2 問題点

$p$  値が有意水準以下であるかどうかで二者択一の判断をすることが問題視

- $p$ -hacking：データや分析手法を変えて、統計的に有意になるようにする
- 出版バイアス：統計的に有意ではない結果 (null result) は出版されにくい。

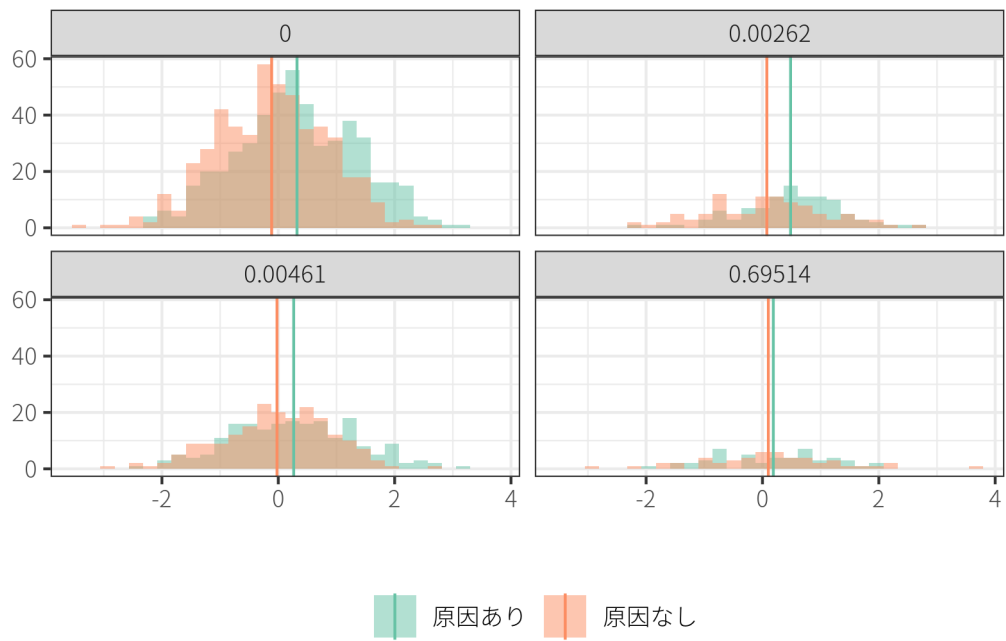
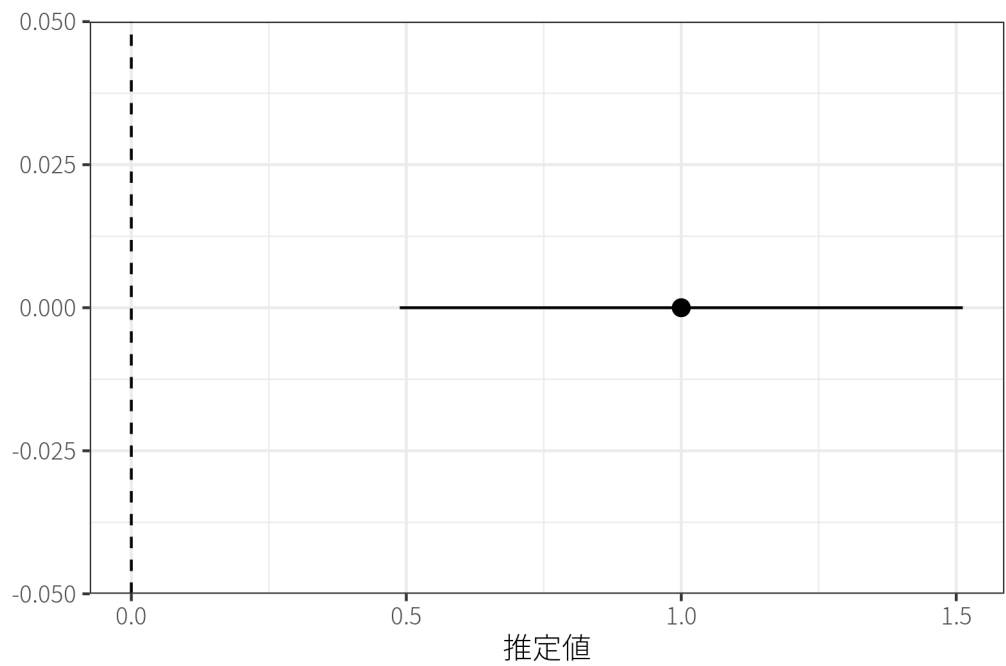


図25: サンプルの大きさと統計的有意性



- HARKing (hypothesizing after the results are known) : データ分析を行い、統計的に有意な結果から仮説を後付けする。

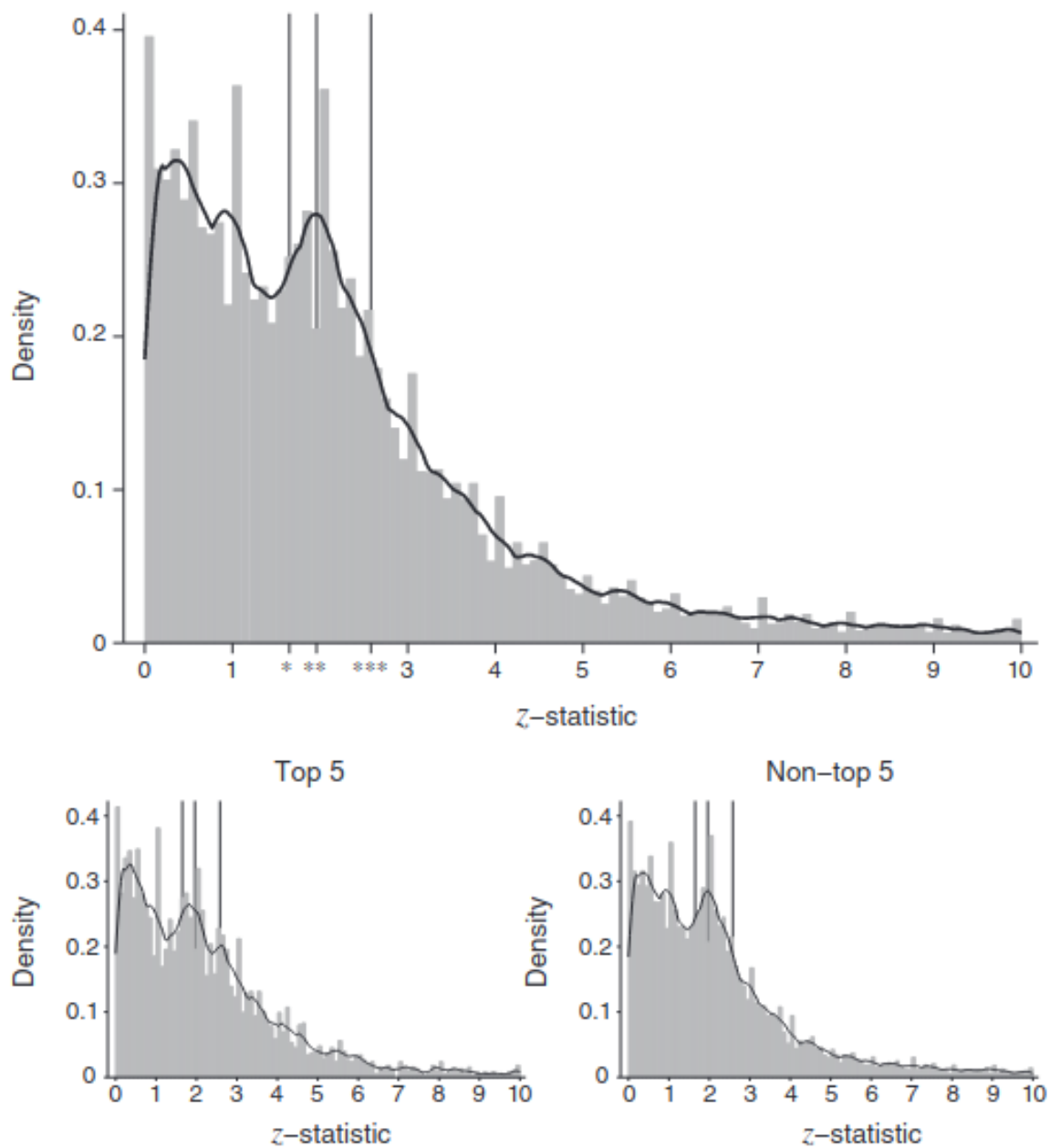


FIGURE 1. z-STATISTICS IN 25 TOP ECONOMICS JOURNALS

図26: Brodeur et al. (2020)

## 5 RCT の限界

交絡がある限り、単純な比較では効果は分からない！

→RCT や自然実験のように、同じようなグループを作り出す工夫

RCT や自然実験は因果推論における強力な手法だが、実行可能？

- 多くの場合、個人を対象とするマイクロな分析
  - 投票行動や消費者行動とは親和性が高い
  - 国家の行動や状態を分析することは困難
- 処置が倫理的に問題がある可能性<sup>2</sup>
  - 嘘の情報や心理的に負担となる情報を与える。
  - 資金やトレーニングの提供など一部の人に有利（不利）なものかも
- 高額な資金が必要かも
  - サーベイ実験の場合、オンラインのクラウドソーシングのサービス<sup>3</sup>を利用すれば比較的安価に行える。
  - 調査会社のサンプルプールを利用する場合は高額
  - フィールド実験の場合は運営費用&現地のパートナーを確保
- (特にサーベイ実験の場合) 表明選好に過ぎず、顕示選好ではないかも？
  - 質問への回答 ≠ 現実の政治的行動
- 都合の良い自然実験はなかなか起こらない。

→RCT や自然実験以外に政策効果を検証できないか？

### 参考文献

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva (2023) “Immigration and redistribution,” *The Review of Economic Studies*, Vol. 90, No. 1, pp. 1–39.

Banerjee, Abhijit V. and Esther Duflo (2012) 『貧乏人の経済学: もういちど貧困問題を根っこから考える』, みすず書房.

Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya (2020) “Facts, alternative facts, and fact checking in times of post-truth politics,” *Journal of public economics*, Vol. 182, p. 104123.

Bertoli, Andrew D (2017) “Nationalism and conflict: Lessons from international sports,” *International Studies Quarterly*, Vol. 61, No. 4, pp. 835–849.

<sup>2</sup> 実験を行う場合は大学の倫理審査委員会で審査を受け、認可される必要がある。

<sup>3</sup> Yahoo! クラウドソーシング、Amazon Mechanical Turk、LUCID Marketplace など。

- Brodeur, Abel, Nikolai Cook, and Anthony Heyes (2020) “Methods matter: P-hacking and publication bias in causal analysis in economics,” *American Economic Review*, Vol. 110, No. 11, pp. 3634–3660.
- Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak (2014) “Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh,” *Econometrica*, Vol. 82, No. 5, pp. 1671–1748.
- Chattopadhyay, Raghavendra and Esther Duflo (2004) “Women as policy makers: Evidence from a randomized policy experiment in India,” *Econometrica*, Vol. 72, No. 5, pp. 1409–1443.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer (2021) “The impact of free secondary education: Experimental evidence from Ghana,” Technical report, National Bureau of Economic Research.
- Ferraz, Claudio and Frederico Finan (2008) “Exposing corrupt politicians: the effects of Brazil’s publicly released audits on electoral outcomes,” *The Quarterly journal of economics*, Vol. 123, No. 2, pp. 703–745.
- Gerber, Alan S, Donald P Green, and Christopher W Larimer (2008) “Social pressure and voter turnout: Evidence from a large-scale field experiment,” *American political Science review*, Vol. 102, No. 1, pp. 33–48.
- Jones, Benjamin F and Benjamin A Olken (2009) “Hit or miss? The effect of assassinations on institutions and war,” *American Economic Journal: Macroeconomics*, Vol. 1, No. 2, pp. 55–87.
- Kuklinski, James H, Michael D Cobb, and Martin Gilens (1997) “Racial attitudes and the” New South,”” *The Journal of Politics*, Vol. 59, No. 2, pp. 323–349.
- Leigh, Andrew (2020) 『RCT 大全: ランダム化比較試験は世界をどう変えたのか』, みすず書房.
- Song, Jaehyun・秦正樹 (2020) 「オンライン・サーベイ実験の方法」, 『理論と方法』, 第 35 卷, 第 1 号, 92–108 頁, DOI: [10.11218/ojjams.35.92](https://doi.org/10.11218/ojjams.35.92).
- Tsugawa, Yusuke, Anupam B Jena, Jose F Figueroa, E John Orav, Daniel M Blumenthal, and Ashish K Jha (2017) “Comparison of hospital mortality and readmission rates for Medicare patients treated by male vs female physicians,” *JAMA internal medicine*, Vol. 177, No. 2, pp. 206–213.