

令和3年度政策評価に関する統一研修(講義型研修)

政策効果の把握のための評価・分析手法

令和4年1月20日

早稲田大学政治経済学術院

山田 治徳

1 政策効果の把握はEBPM、政策評価の基本

1-1 EBPMの本質

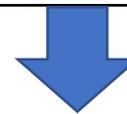
内閣官房行政改革推進本部事務局「EBPMの推進」(平成30年1月12日)

□ 証拠に基づく政策立案(EBPM)とは、

(1) 政策目的を明確化させ、

(2) その**目的のため本当に効果が上がる行政手段は何か**など、

「政策の基本的な枠組み」を証拠に基づいて明確にするための取組。



【政策の基本的な枠組み】

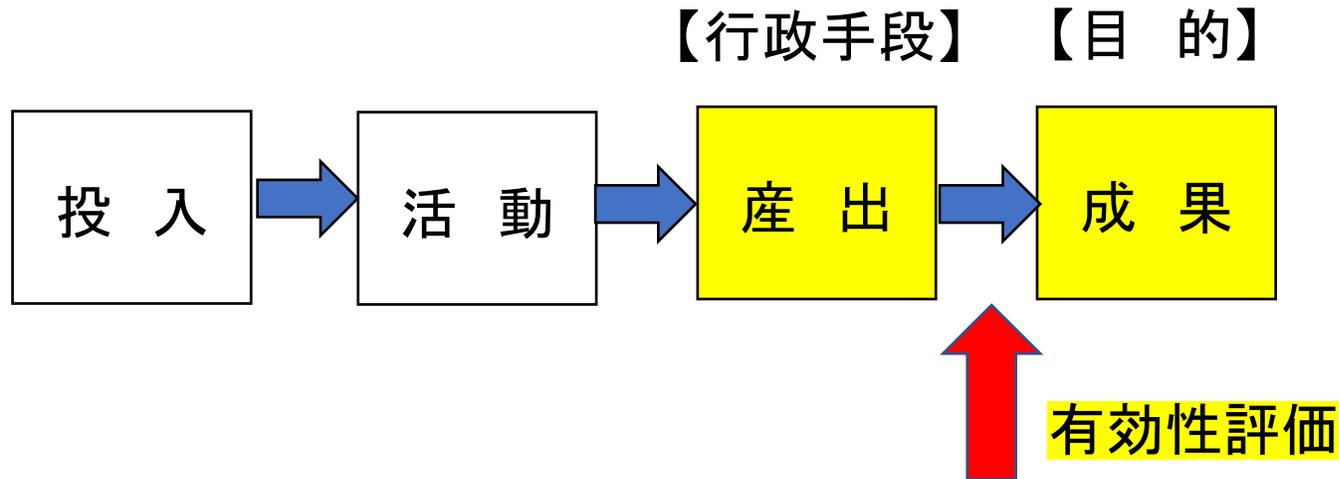


【原因】

【結果】

EBPMの本質は、因果関係の評価

1-2 EBPMの本質は因果関係の評価



目的のため本当に効果が上がる行政手段は何か

1-3 政策効果の把握はEBPM、政策評価の基本

政策評価・独立行政法人評価委員会「提言」(平成27年3月9日)

政策効果を的確に把握することは、政策の有効性を検証する上で不可欠である。

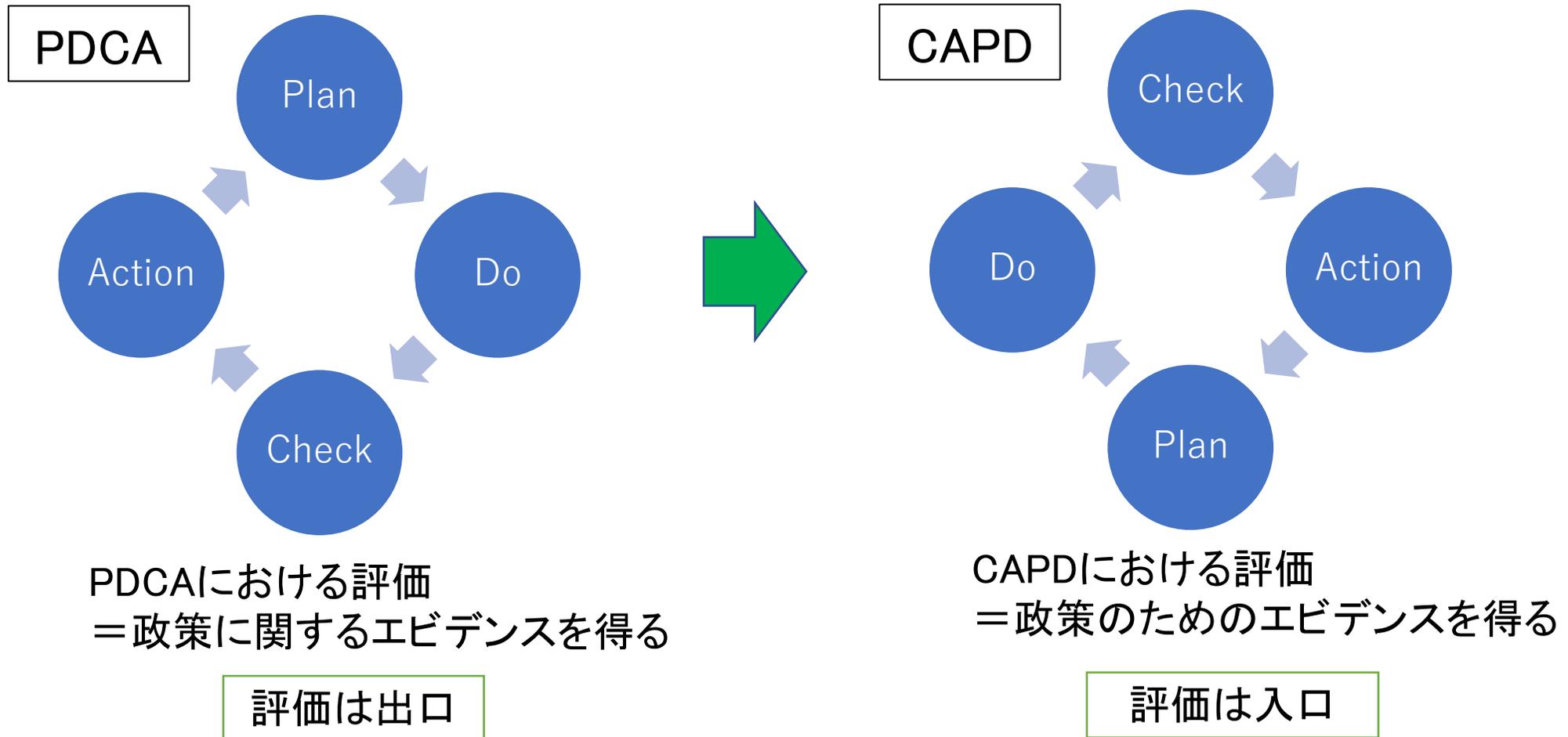
統計改革推進会議「最終取りまとめ」(平成29年5月19日)

EBPMの推進には、(中略)、具体的政策の内容とその効果をつなぐ論理、政策効果とそのコストの関係を明示することが欠かせない。

1-4 PDCAからCAPDへ

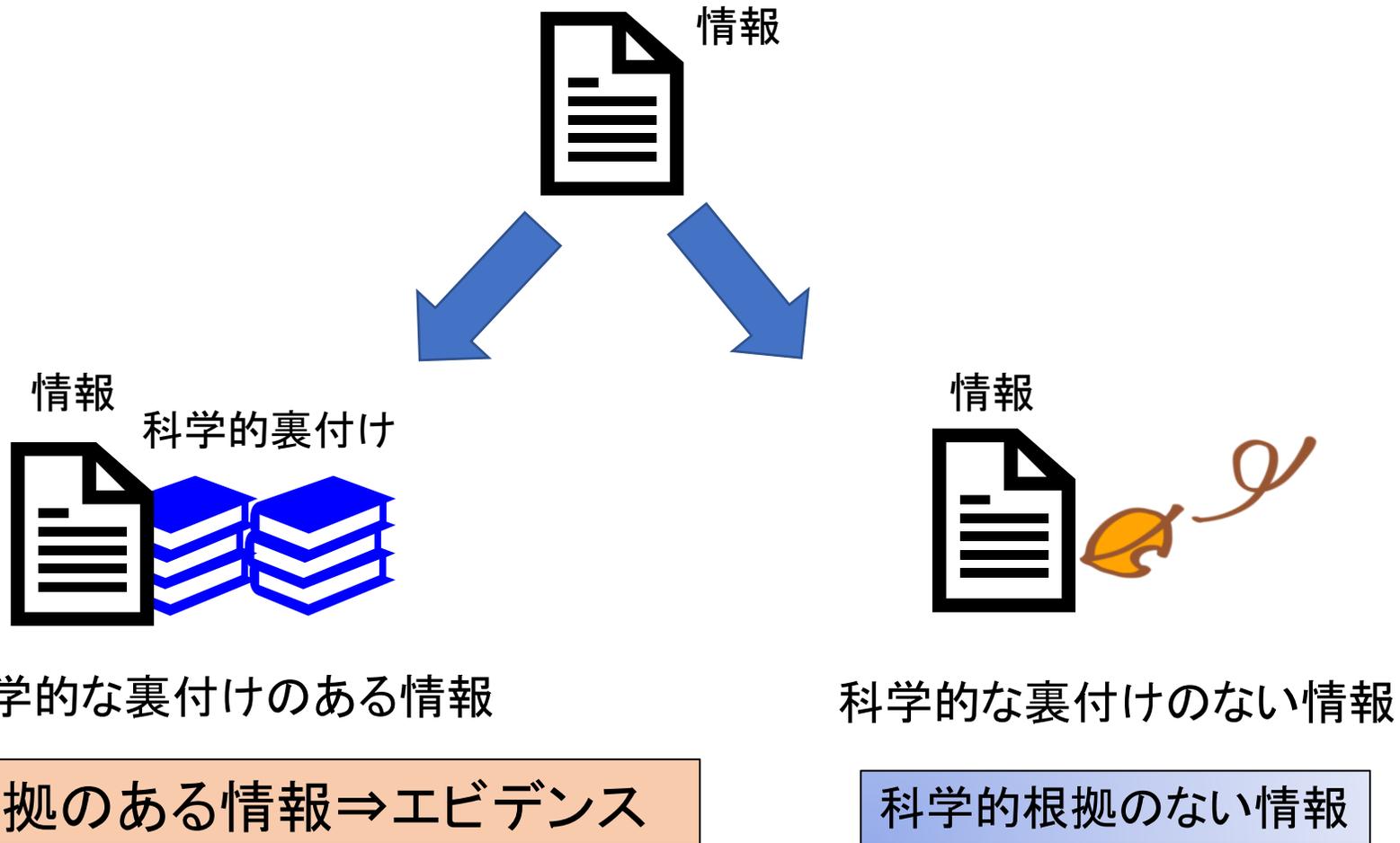
政策評価に関する基本方針(平成17年12月16日閣議決定)

政策評価の結果については、各行政機関において、政策評価の結果が政策の企画立案作業における重要な情報として適時的確に活用され、当該政策に適切に反映されるようにする必要がある。



2 エビデンス (evidence, 証拠) とは何か

2-1 エビデンスに求められるのは科学的であること



2-2 エビデンスと科学の関係

エビデンス＝科学的根拠のある情報

➡現実を重視する政策において科学は現実的と言えるのか？

EBPM先進国・英国は、EBPMにおいて科学をどのように位置付けているか

政策は、健全なエビデンスによって情報が与えられているべきである。社会科学の研究はそのエビデンスの基盤の主要な部分に寄与しているべきである。
(ブレア政権・ブランケット教育雇用相演説、2000年2月)



科学的とは何か？

バイアスのない方法により得られたデータを、
バイアスのない方法で分析して得られた結果であること

2-3 エビデンスに求められる妥当性

- ◆ 結果妥当性 (conclusion validity)

＝「観察された変数間の関係は、正しいデータや方法論に基づくものと言えるのか」

- ◆ 内的妥当性 (internal validity)

＝「観察された変数間の関係は因果関係と言えるのか」

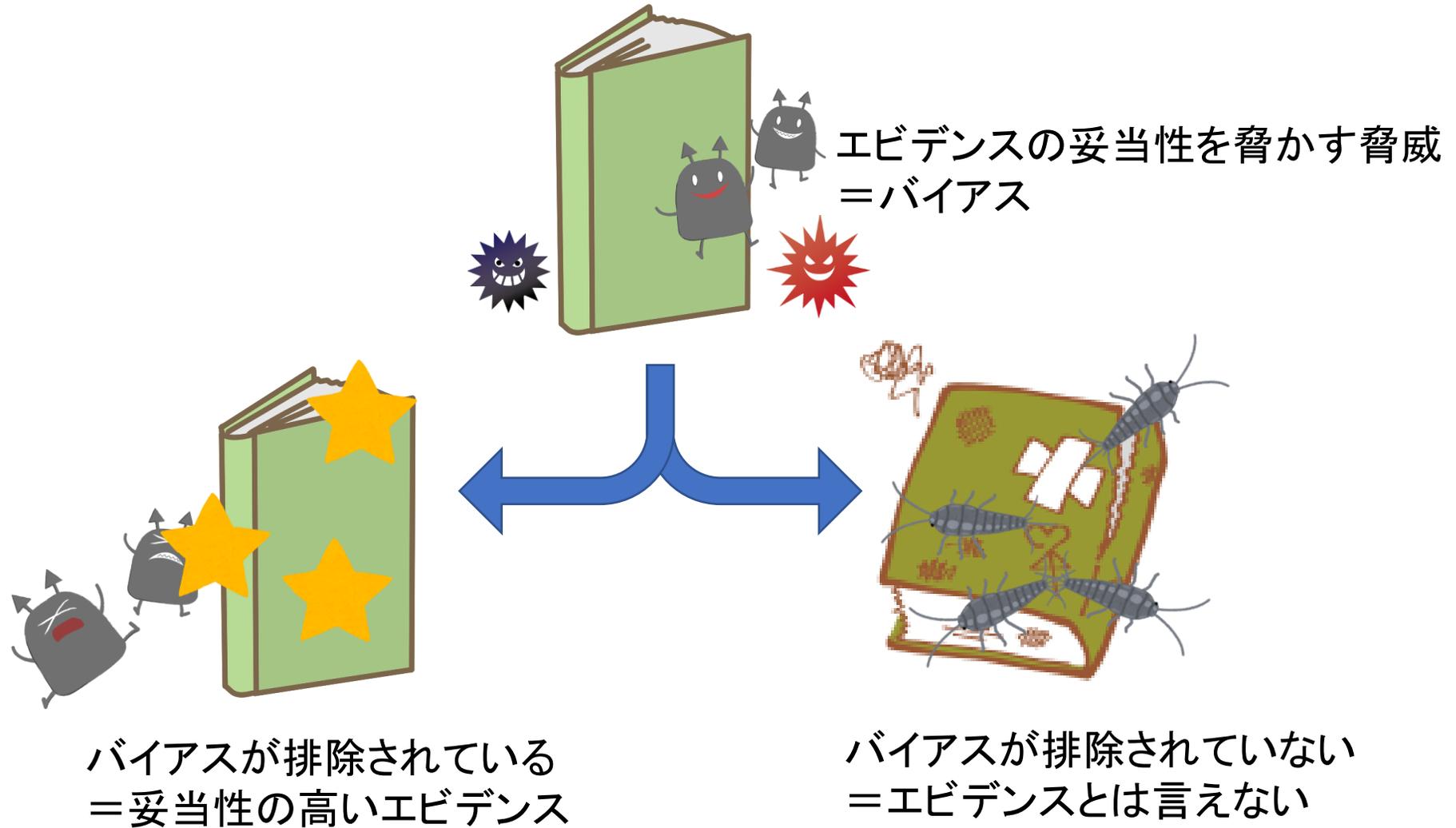
- ◆ 外的妥当性 (external validity)

＝「観察された因果関係はその他の事例に対しても一般化できるのか」

- ◆ 構成概念妥当性 (construct validity)

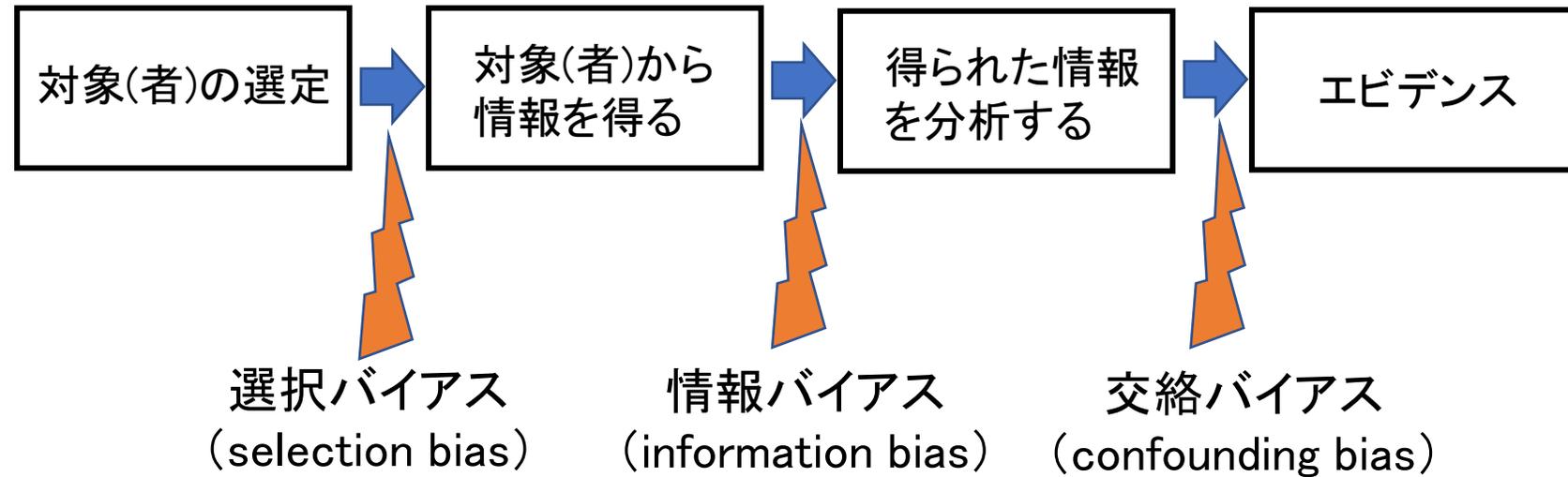
＝「変数は測定しようとしている概念を表していると言えるのか」

2-4 エビデンスの妥当性を脅かすものーバイアスの存在

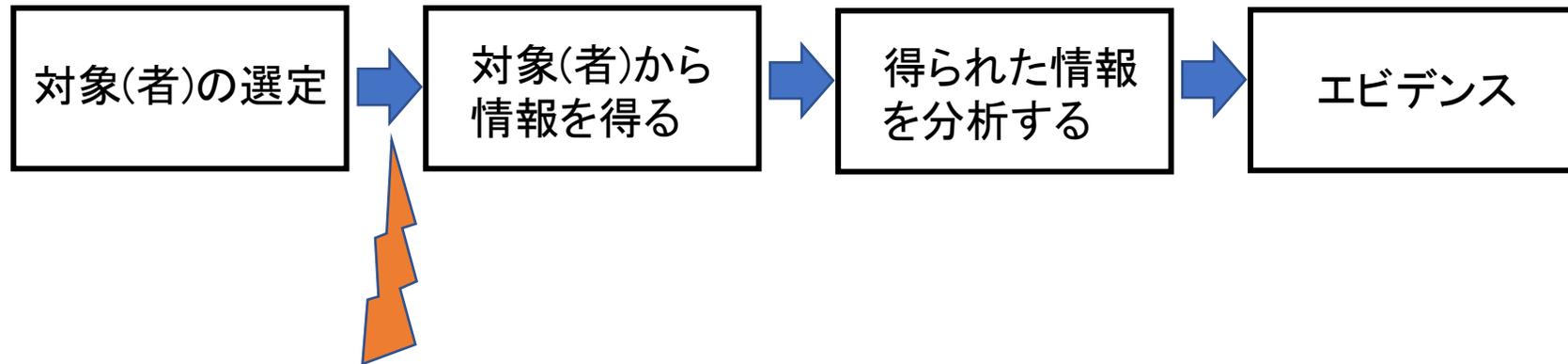


2-5 研究の流れと発生するバイアス

エビデンスを得るための研究の流れ



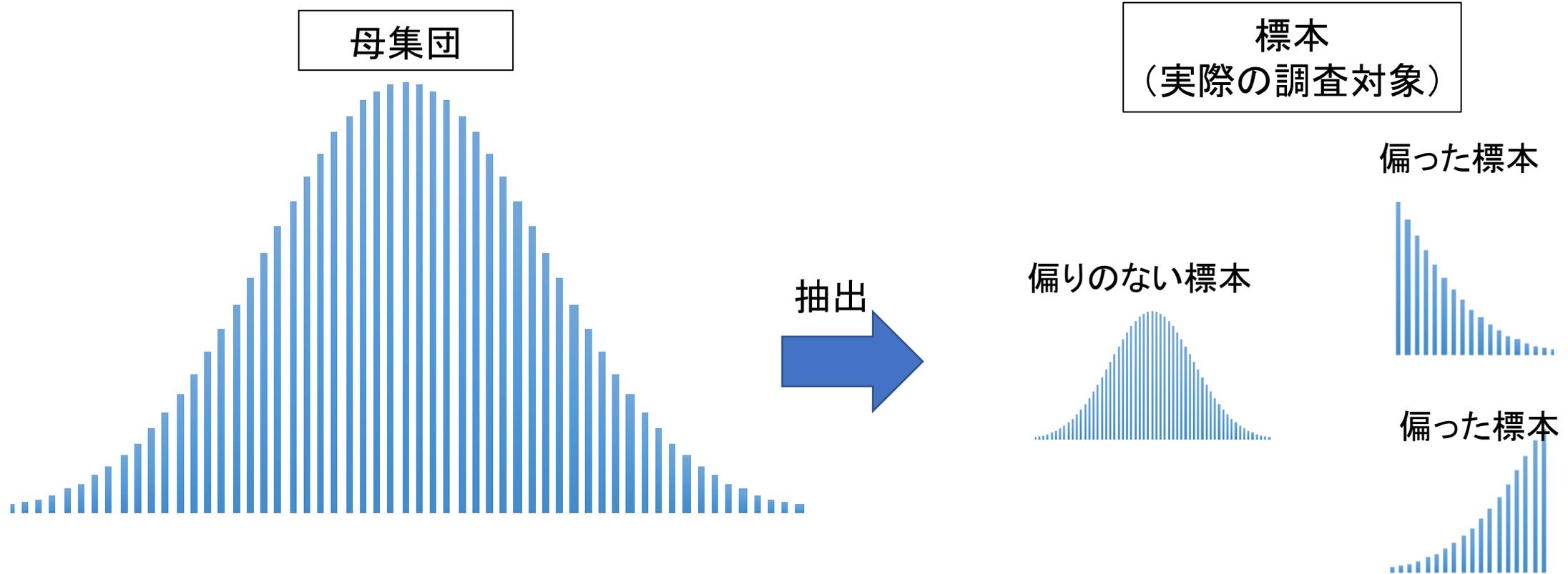
① 選択バイアス (selection bias)



選択バイアス (selection bias)

= 対象(者)を選定する際に発生するバイアス

⇒ 選定された対象に偏りが発生



対象者選定の方法や場所などによって発生

【事例】

立石港立石地区港湾緑地整備事業・来訪者数予測

立石港緑地は地域住民の休息や散策の場として供することから、立石港周辺（半径2km圏内）の住民8,726人を対象として150人にアンケートを配布した結果、回答者数は102人であるとのことでした。

推計された港湾来訪者数741,940人は、周辺住民8,726人の4人に1人が毎日来訪するという推計となっています。また、立石港の近隣に位置する香川県を代表する観光地である「屋島」の平成17年の入込客数571,000人（平成17年度香川県観光客動態調査報告（平成18年4月）による。）を大きく上回っています。

出所：総務省行政評価局（2007）『政策評価の点検結果—評価の実効性の向上に向けて』pp.227-228。

なぜこのような矛盾が起こったのか

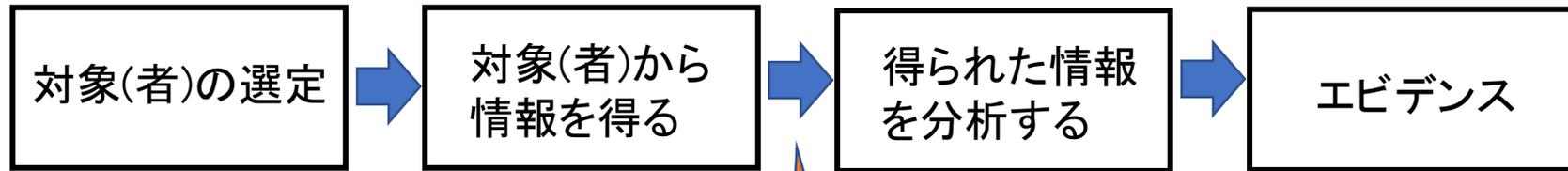
区 分	母集団 半径2km圏内人口	標 本	
		実際の回答者数	母集団構成比の場合
高松市	4,684	99	55
牟礼町	3,093	1	36
庵治町	949	2	11
合 計	8,726	102	102



アンケート回答者の抽出方法について、貴省では、無作為抽出を基本としているのですが、実際には、事業実施箇所において港湾来訪者を対象にアンケートを実施するなど無作為抽出とはなっていません。また、事業実施箇所において港湾来訪者を対象にアンケートを実施することにより、来訪頻度や緑地利用率にバイアスがかかり、緑地の来訪者数が過大に推計されるおそれがあります。

出所：総務省行政評価局（2007）『政策評価の点検結果—評価の実効性の向上に向けて』p.228。

② 情報バイアス (information bias)



情報バイアス (information bias)

= 対象(者)から情報を得る際に発生するバイアス

⇒ 得られた情報に歪みが発生

次の2つの質問に教えてください。

問1 あなたはどちらのタイプの上司が好みですか。
次の選択肢からどちらか一方を選んで、□に✓をしてください。

- ほとんど怒らないが、いつも冷たい上司
- すぐに怒るが、面倒見がよい上司

問2 あなたはどちらのタイプの上司が好みですか。
次の選択肢からどちらか一方を選んで、□に✓をしてください。

- いつも冷たいが、ほとんど怒らない上司
- 面倒見はよいが、すぐに怒る上司

観察や測定の方法が不十分であることによって発生

調査方法によって結果が異なる場合がある

郵送調査と面接調査の比較

(単位:%)

【政治関心度】あなたは、いまの政治にどの程度関心がありますか。	大いにある	少しはある	あまりない	まったくない	その他・ 答えない
郵送調査 (12/1-1/17)	32	50	14	2	2
面接調査 (12/3-4)	31	48	16	4	1
【生活満足度】あなたは、いまの生活にどの程度満足していますか。	満足	まあ満足	やや不満	不満	その他・ 答えない
郵送調査 (12/1-1/17)	4	39	36	20	1
面接調査 (12/3-4)	10	56	24	10	0

出所:松田映二(2008)「面接調査の現状と課題」『行動計量学』第35巻第1号、p.28。

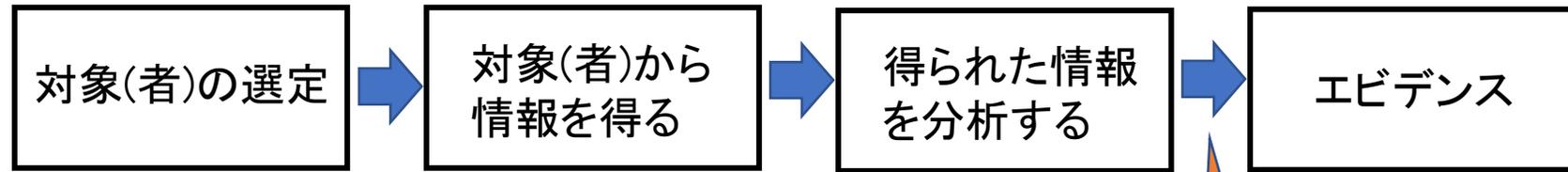
インターネット調査と面接調査の比較

(%)

調査項目	選択肢	訪問面接	インターネットA		インターネットB	
				差		差
Q1 現在の生活に対する満足度	満足	78.2	46.3	-31.9	47.4	-30.8
	不満	20.4	49.7	+29.3	48.7	+28.3
Q2 現在の生活の充実感	充実感を感じている	78.7	50.9	-27.8	53.0	-25.7
	充実感を感じていない	18.6	45.3	+26.7	42.7	+24.1

出所:内閣府政府広報室「訪問面接調査とインターネット調査の比較について」(平成18年8月)

③ 交絡バイアス (confounding bias)



交絡バイアス (confounding bias)

=対象(者)から得られた情報を分析する際に発生するバイアス

⇒得られた情報の分析に誤りが発生

研究の結果、飲酒と肺がんの発症の間には関係がある

区 分		飲 酒		合 計
		あ り	な し	
肺 が ん	あ り	33	27	60
	な し	1,667	2,273	3,940
合 計		1,700	2,300	4,000

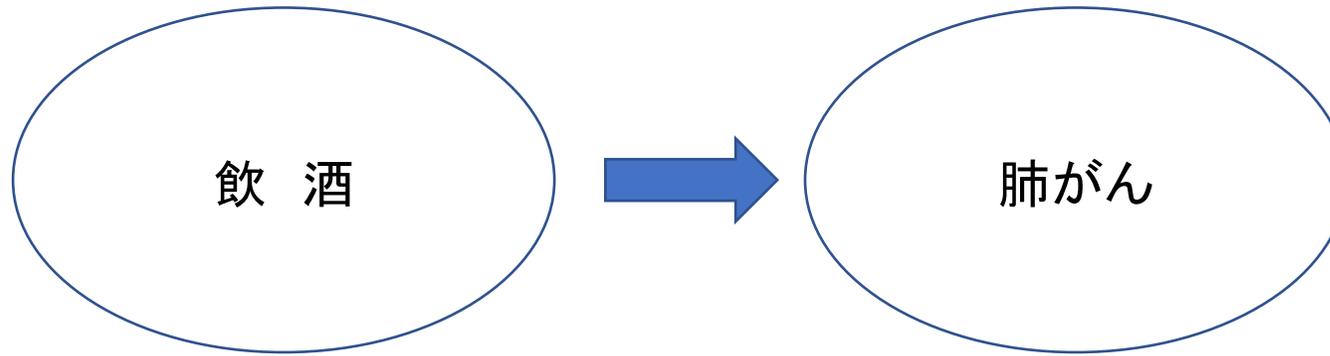
飲酒あり:発症オッズ=33/1667
 飲酒なし:発症オッズ=27/2273



$$\text{オッズ比} = \frac{33/1667}{27/2273} = 1.7$$

飲酒者の肺がん発症確率は、
非飲酒者の1.7倍

研究の結果、飲酒と肺がんの発症の間には関係がある



アルコールには肺がんを引き起こす物質が含まれていないにもかかわらず → なぜだろうか？

【喫煙者】

区 分		飲 酒		合 計
		あ り	な し	
肺 が ん	あ り	24	6	30
	な し	776	194	970
合 計		800	200	1,000

【非喫煙者】

区 分		飲 酒		合 計
		あ り	な し	
肺 が ん	あ り	9	21	30
	な し	891	2,079	2,970
合 計		900	2,100	3,000

飲酒あり:発症オッズ=24/776

飲酒なし:発症オッズ= 6/194

$$\text{オッズ比} = \frac{24/776}{6/194} = 1.0$$

オッズ比は等しい

飲酒者と非飲酒者の
肺がん発症確率は等しい。

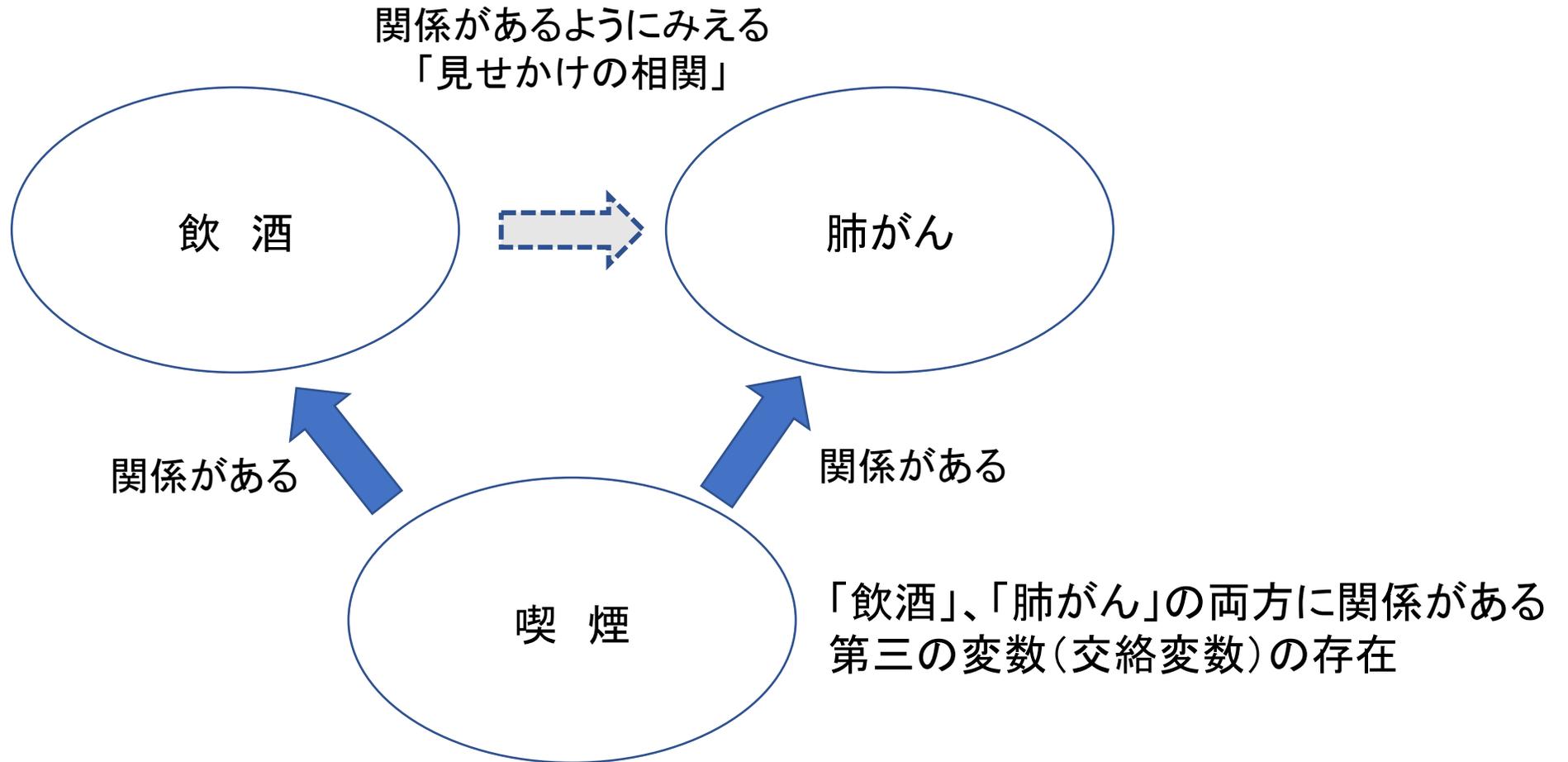
飲酒と肺がん発症の間に関係はない。

飲酒あり:発症オッズ= 9/ 891

飲酒なし:発症オッズ=21/2079

$$\text{オッズ比} = \frac{9/ 891}{21/2079} = 1.0$$

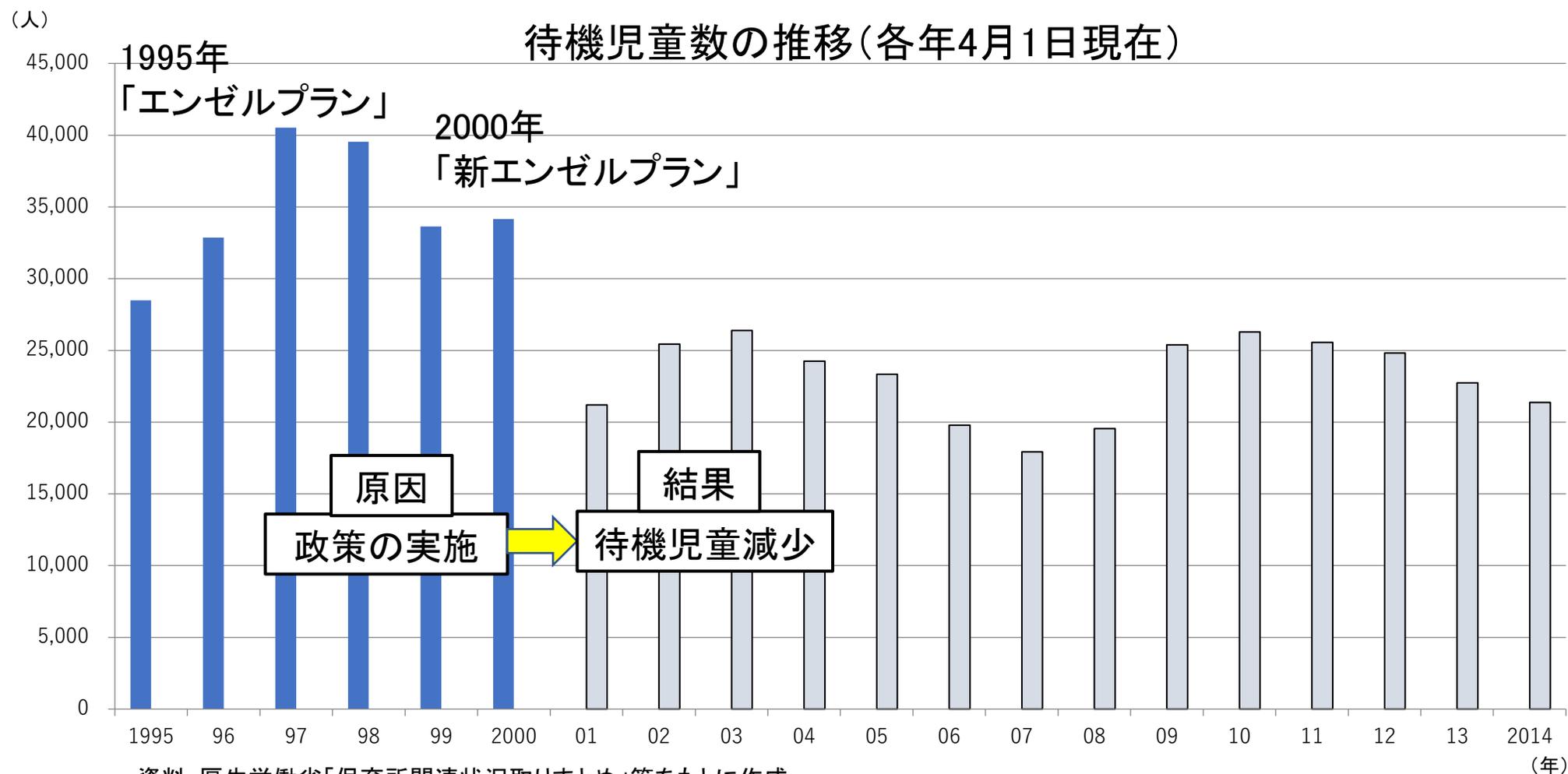
関係があるのは喫煙との間(約3倍)



分析が不十分であることによって発生

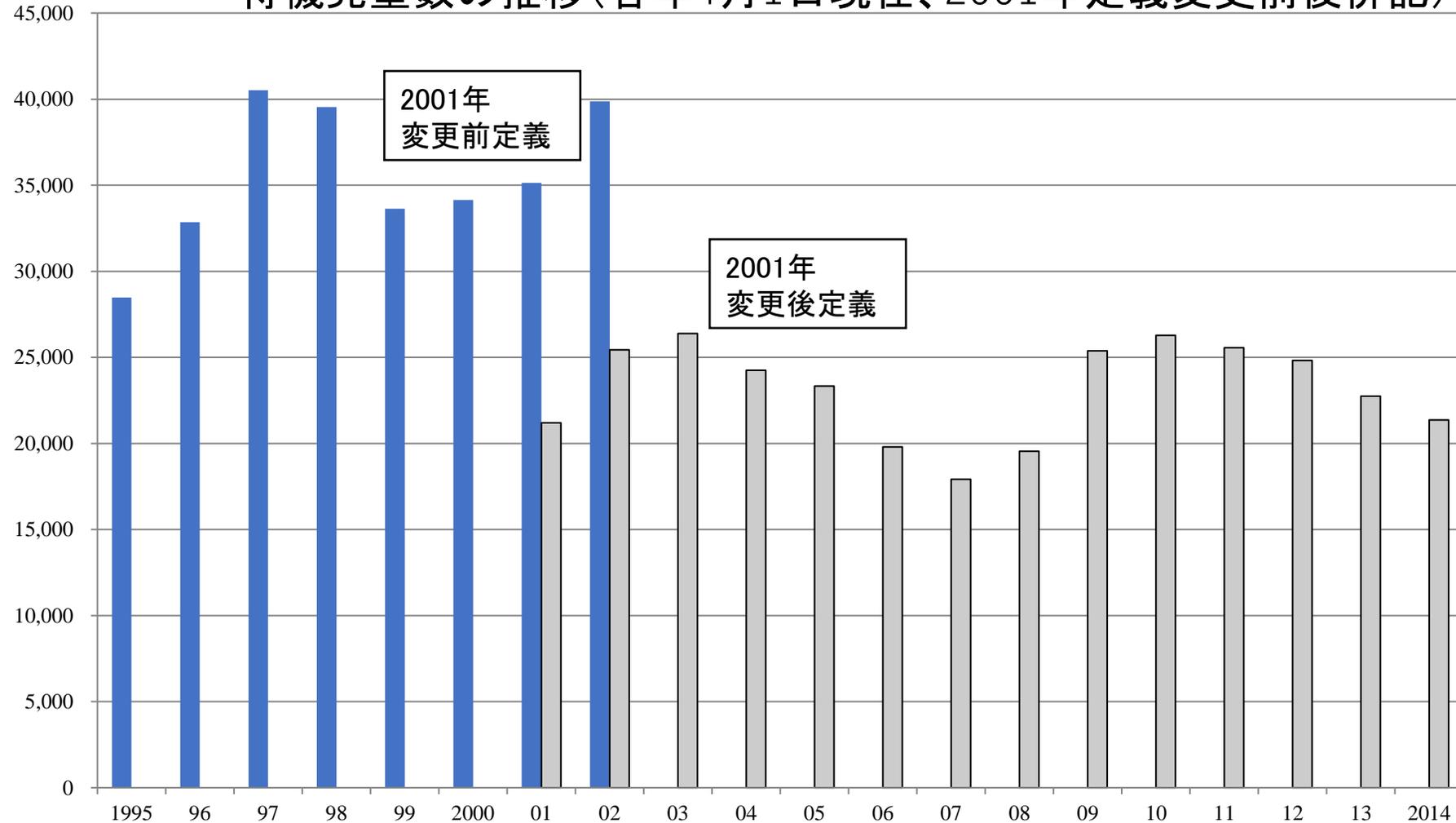
3 因果関係の評価のためのモデル

3-1 EBPMの本質は因果関係の評価—しかし、因果関係の分析は単純ではない



待機児童数の減少は政策効果によるものなのか？

待機児童数の推移(各年4月1日現在、2001年定義変更前後併記)



資料:厚生労働省「保育所関連状況取りまとめ」等をもとに作成。

因果関係＝変数Aの変化によって、一方の変数Bの変化が引き起こされる関係

変数A  変数B

因果関係が成立するための条件

- ① A、Bが共に変化する(共変関係)
- ② Aが先に変化してから、後でBが変化する(時間的先行関係)
- ③ A-B間の関係についての代替的説明が排除される

因果関係と単なる相関関係を混同しない

相関関係 = 変数A、Bの間に共変関係が見られる関係

相関が見られる 変数A \longleftrightarrow 変数B の可能性は... バスケットボール選手は背が高い

① 変数A \longrightarrow 変数B 因果関係 背が高い人(A)がバスケットボールをする(B)

② 変数A \longleftarrow 変数B 因果関係の逆転 バスケットボールをすれば(B)背が高くなる(A)

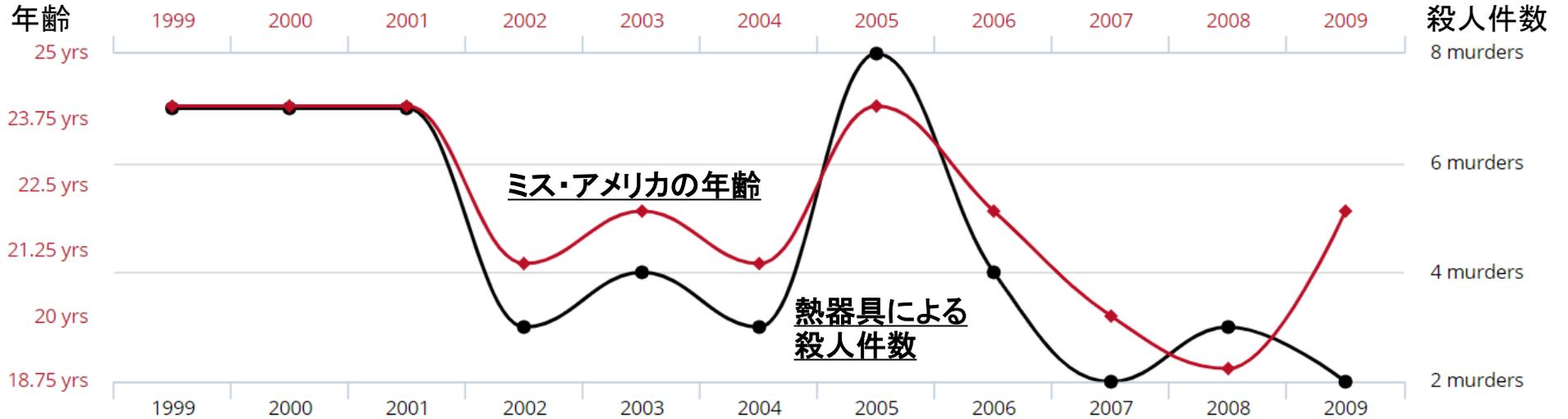
③ 変数A \rightleftharpoons 変数B 比例関係 気体の状態方程式($PV=nRT$)
気体の体積V \leftrightarrow 絶対温度T

④ 変数A \longleftrightarrow 変数B 偶然の一致 ミスアメリカの年齢(A)と熱器具を用いた殺人件数(B)

⑤ 変数X $\begin{matrix} \nearrow & \text{変数A} \\ \searrow & \text{変数B} \end{matrix}$ 見せかけの相関(疑似相関) 飲酒習慣(A)と肺がんの発症(B)

偶然の一致

ミス・アメリカの年齢と熱器具を用いた殺人件数

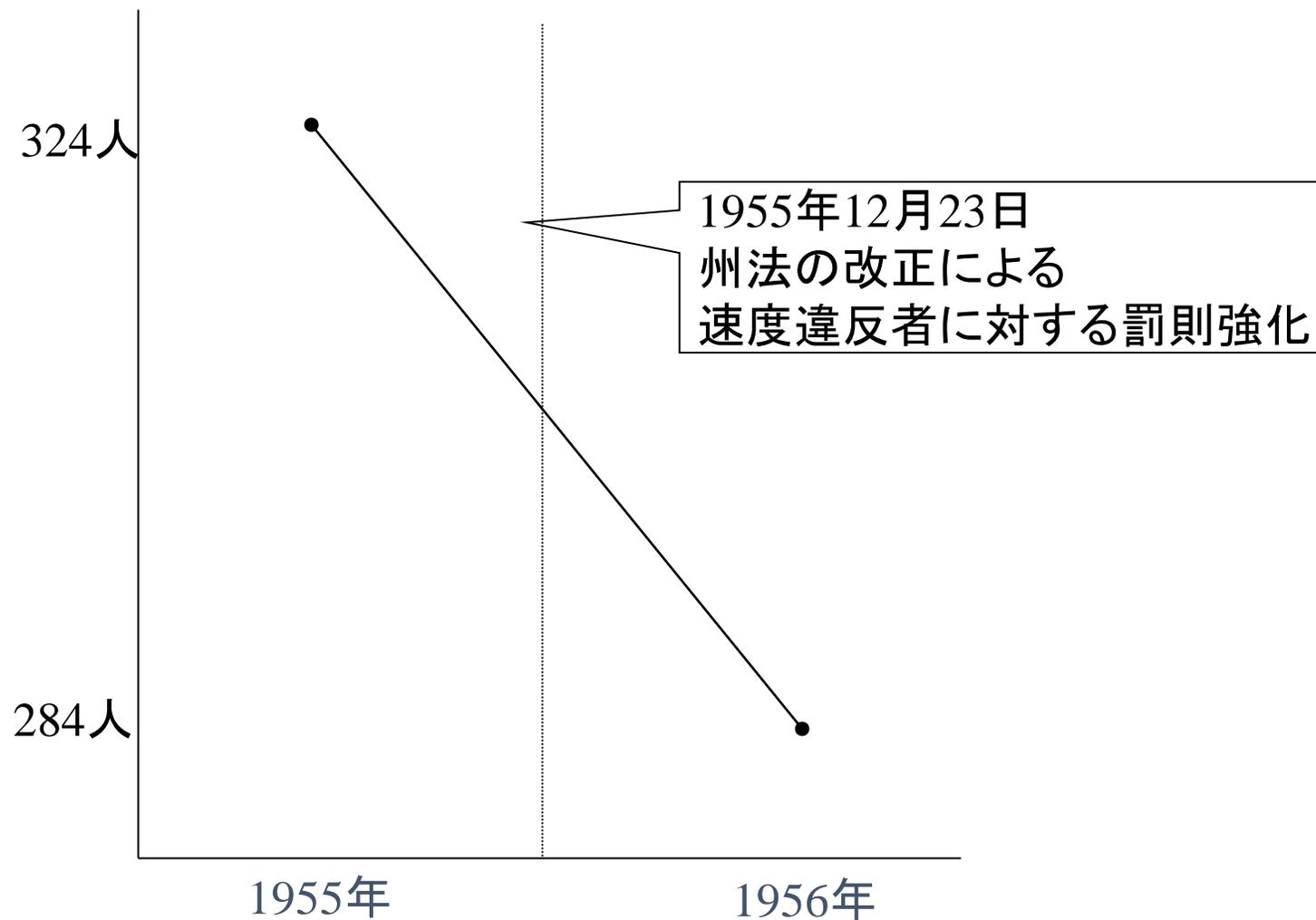


水泳プールでの溺死者数とニコラス・ケイジ出演映画数



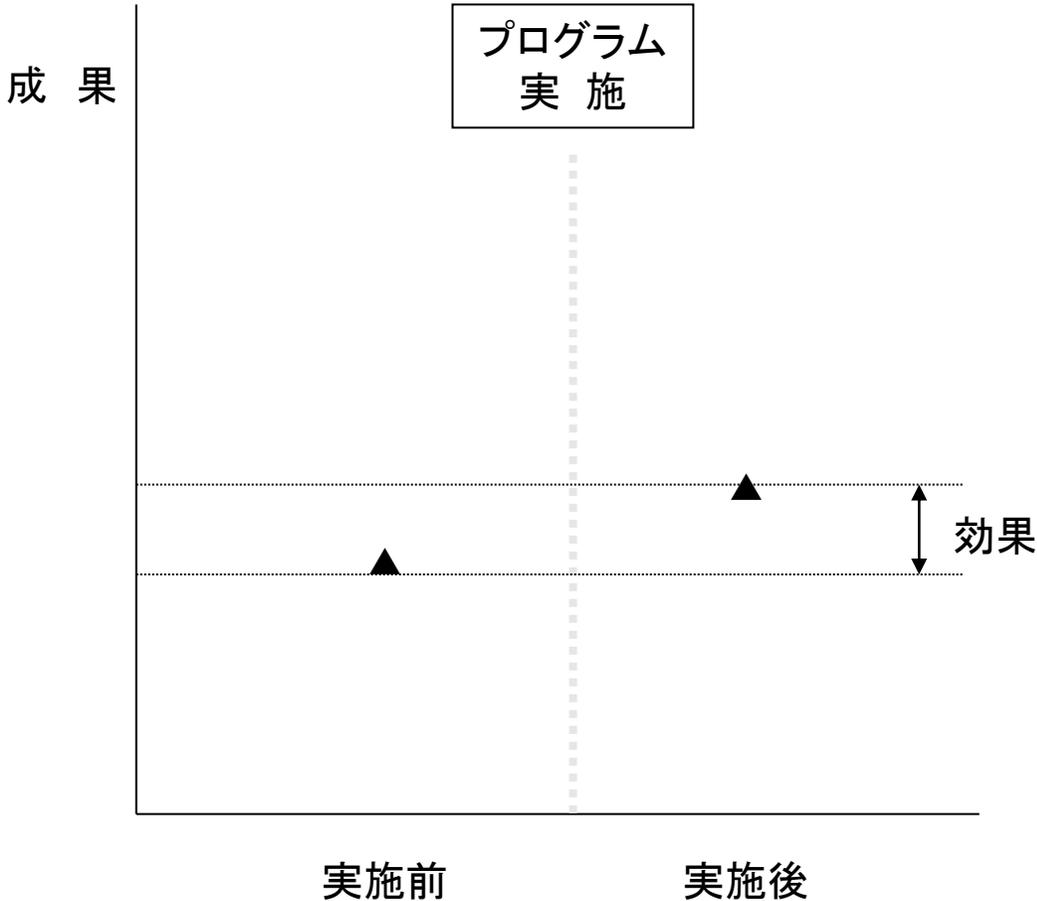
3-2 ビフォーアフター（Before-After）の落とし穴ーよく見かけるモデル

コネティカット州内の交通事故による死者数の推移

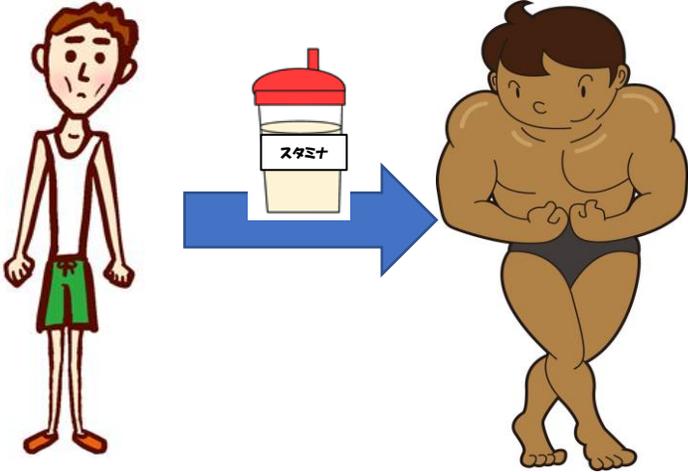


出所: Campbell and Ross (1968) "The Connecticut crackdown on speeding: Time series data in quasi-experimental analysis" *Law and Society Review* Vol.3, p.38.

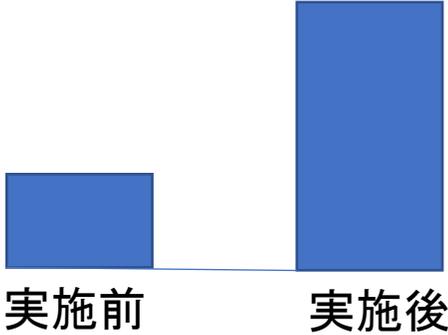
前後比較デザイン (One-group Pretest-Posttest Design, One-case Before-After Design)



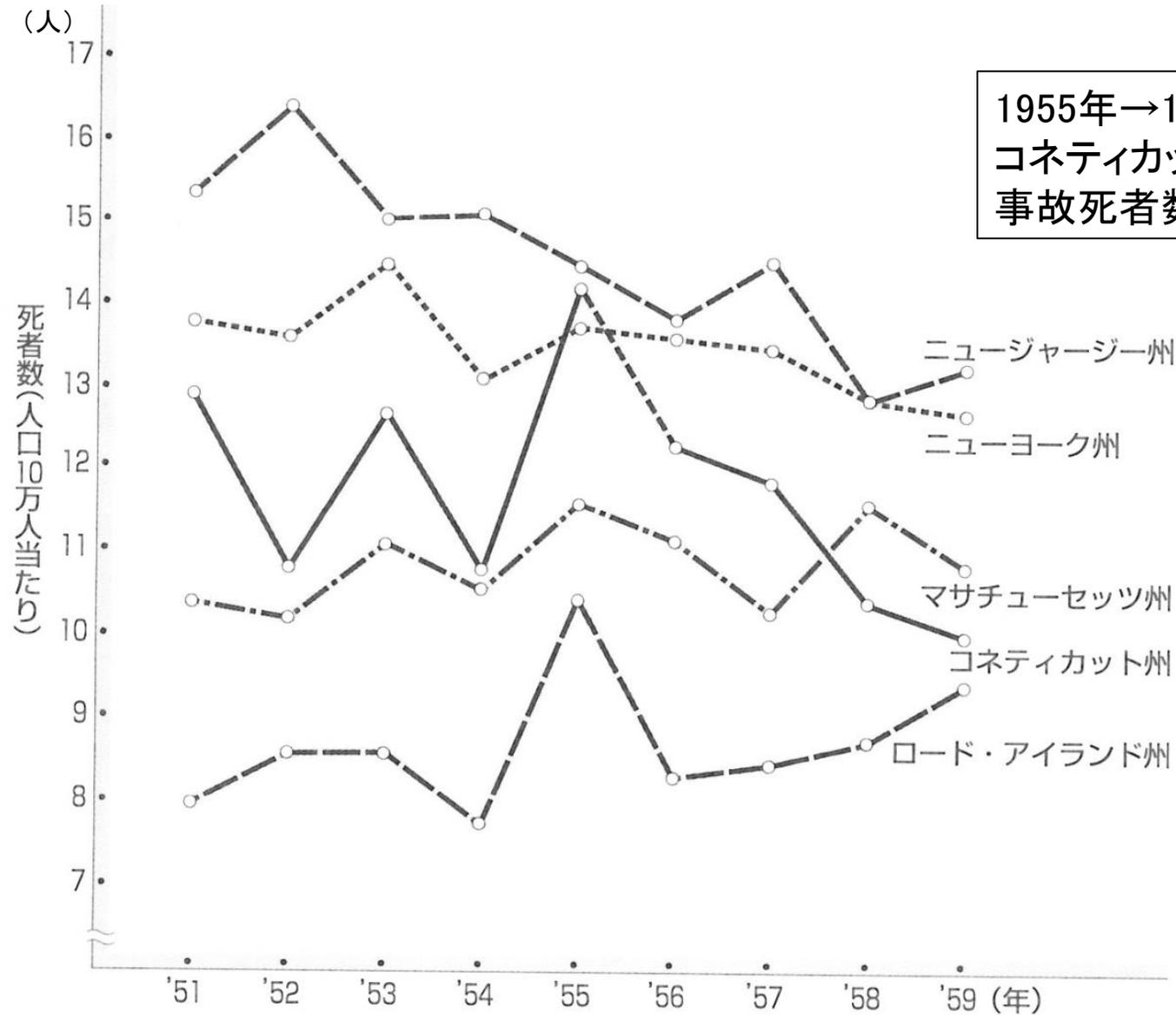
【Before】 【After】



経済対策実施前後の
経済成長率比較



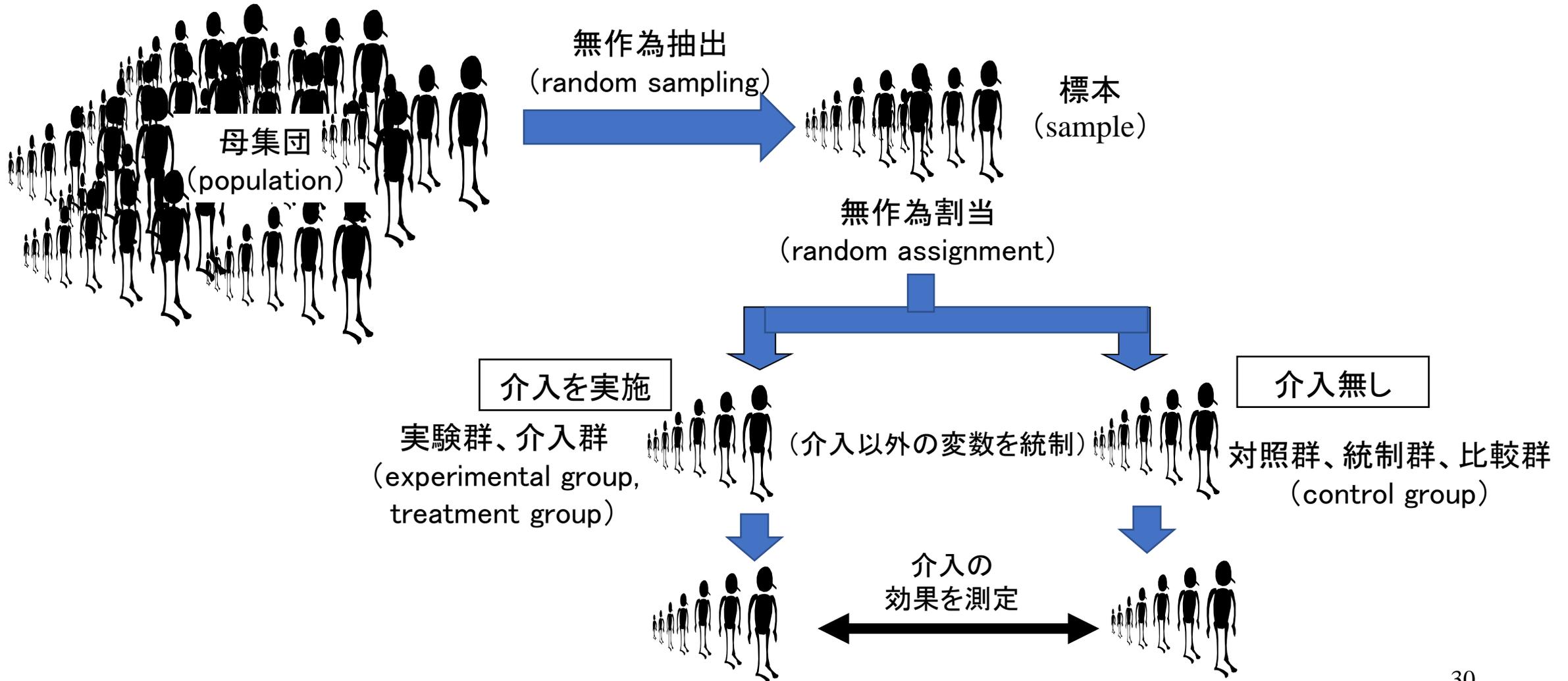
コネティカット州および隣接州の交通事故による死者数の推移 (1951～59年)



出所: Campbell and Ross, p.45.

3-3 ランダム化比較試験 (Randomized Controlled Trial, 略称RCT)

EBPMにおける理想のモデル



無作為抽出 (random sampling) と無作為割当 (random assignment)

- 無作為抽出 (random sampling)

母集団に対する高い外的妥当性 (external validity)

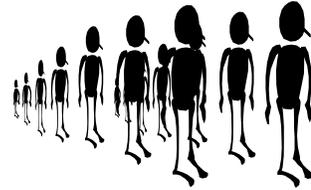
- 無作為割当 (random assignment)

内的妥当性 (internal validity)

ベースライン特性を揃えることで、群間の comparability が確保される
⇒ 交絡要因の影響が排除

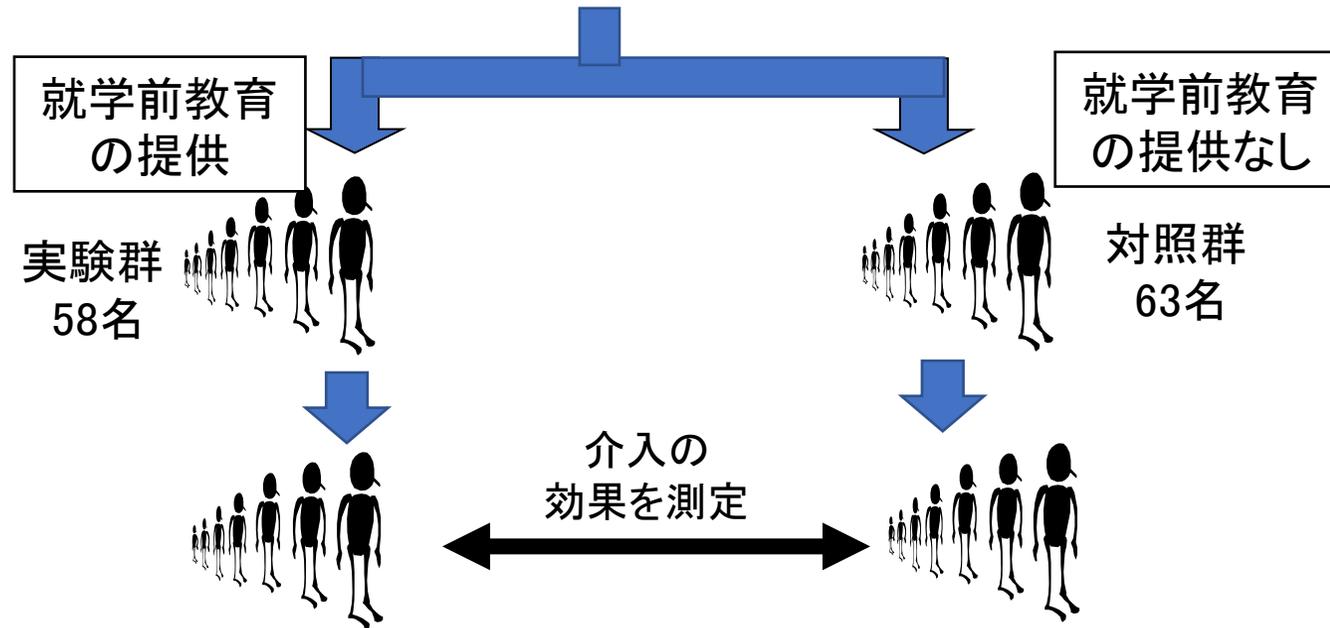
① ペリー就学前計画 (Perry Preschool Project)

1962年：米国ミシガン州



貧困状態にあり、教育上、高リスクと判定された
3-4歳のアフリカ系アメリカ人の子ども達 123名

無作為割当
(random assignment)



ペリー就学前計画 その後の影響

逮捕歴5回以上の割合(40歳時点)
Arrested 5+ times by age 40

Program Group 実験群

36%

Non-program Group 対照群

55%

年収2万ドル以上の割合(40歳時点)
Earned \$20,000+ at age 40

Program Group 実験群

60%

Non-program Group 対照群

40%

高校卒業以上の割合
Graduated high school

Program Group 実験群

77%

Non-program Group 対照群

60%

基礎学力を有する者の割合(14歳時点)
Basic achievement at age 14

Program Group 実験群

49%

Non-program Group 対照群

15%

宿題提出者の割合(15歳時点)
Homework at age 15

Program Group 実験群

61%

Non-program Group 対照群

38%

知能指数90以上の割合(5歳時点)
IQ 90+ at age 5

Program Group 実験群

67%

Non-program Group 対照群

28%

② 次世代エネルギー・社会システム実証事業

次世代エネルギー社会システム協議会の概要

- 次世代エネルギー・社会システム事業については、平成 21 年度より検討を開始し、平成 22 年 1 月に次世代エネルギー・社会システム協議会中間報告をとりまとめ、これに基づき同年 4 月に実証地域の選定をし、平成 23 年度より実証事業を開始した。
- 平成 26 年度 4 月に全体の中間評価が、5 月に各地域からの中間報告が行われ、実証事業は平成 27 年 3 月に完了した。
- 今回は、実証事業が完了して約 1 年間が経過したところ、実証事業の総括と今後の方向性について議論を行う。

- 平成 21 年 1 月 「次世代エネルギー・社会システム協議会」の設置
- 平成 22 年 1 月 「次世代エネルギー・社会システム協議会」中間取りまとめ
- 平成 22 年 4 月 次世代エネルギー・社会システム実証地域（横浜市、豊田市、けいはんな学研都市（京都府）、北九州市）の選定
- 平成 22 年 8 月 実証事業のマスタープランの策定
- ※平成 23 年 3 月 東日本大震災
- 平成 23 年 4 月 実証事業の開始
- ※平成 24 年 7 月 固定価格買取制度(FIT)開始
- 平成 26 年 4 月、5 月 実証事業の中間評価、中間報告
- 平成 27 年 3 月 実証事業の完了

【参考】四地域実証の概要

- 平成23年度から平成26年度まで、様々なパターンの代表例を構成する全国4つの地域（横浜市、豊田市、けいはんな市、北九州市）で、大規模なスマートコミュニティ実証事業を展開。

住宅団地型

住宅約700戸等を対象とし、系統の状況に応じて需要サイドで追従を行う実証を実施。また、家庭部門のより一層の省エネに向けた電力会社による省エネコンサルを実施。（関西電力・三菱電機・三菱重工）

けいはんな学研都市

広域大都市型

住宅約4000戸、大規模ビル等約10棟を対象とした大規模な実証。また、大型蓄電池等を統合的に管理することで、仮想的に大規模発電所と見立てる実証を実施。（東芝・東京電力）

横浜市

北九州市

地方中核都市型

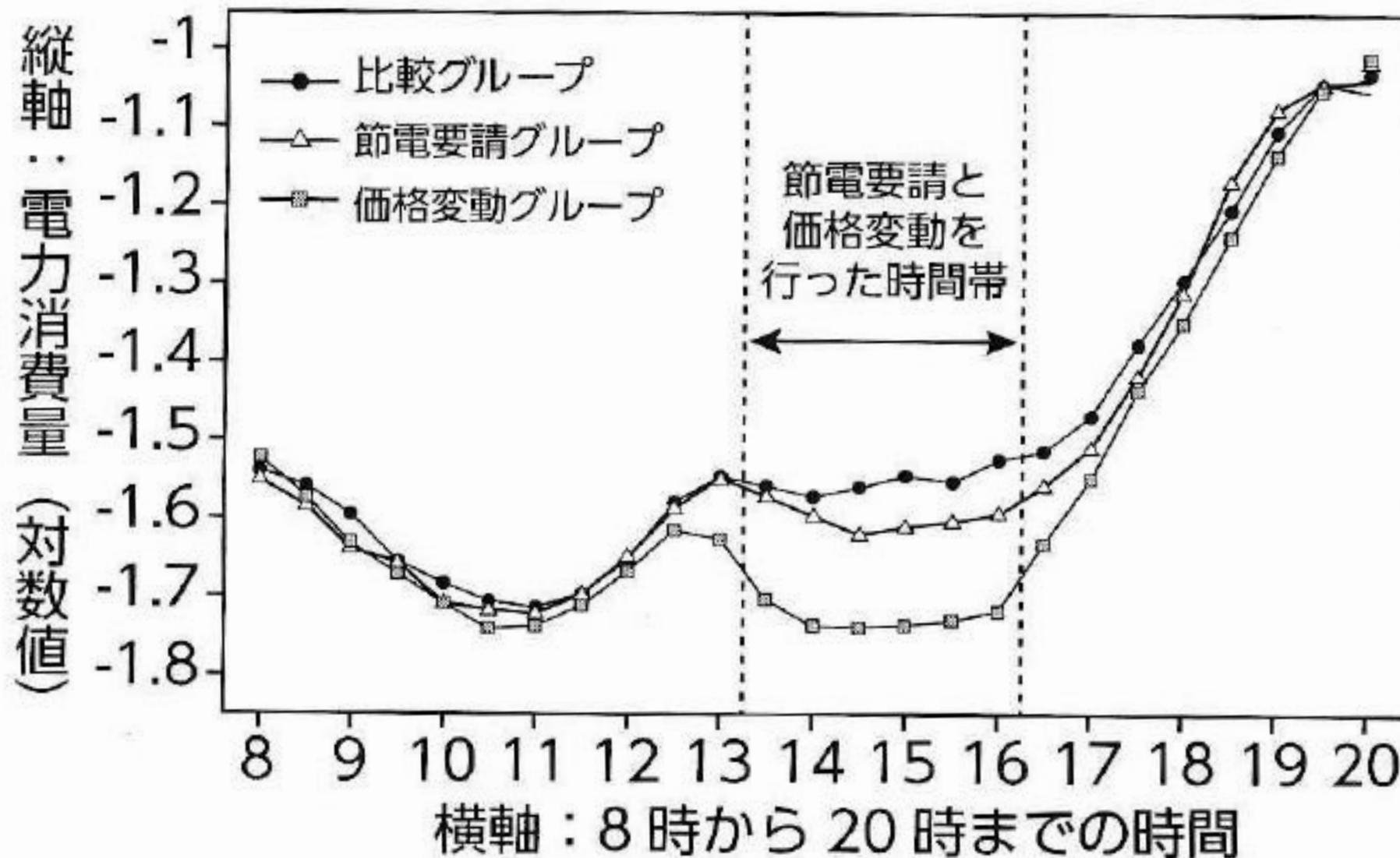
新日鐵住金の特定供給エリアで実証。コジェネをベースロード電源と見立て、需要家180戸において、需給状況に応じて電力料金を変動させるダイナミックプライシングの実証を実施。（富士電機・新日鐵住金）

豊田市

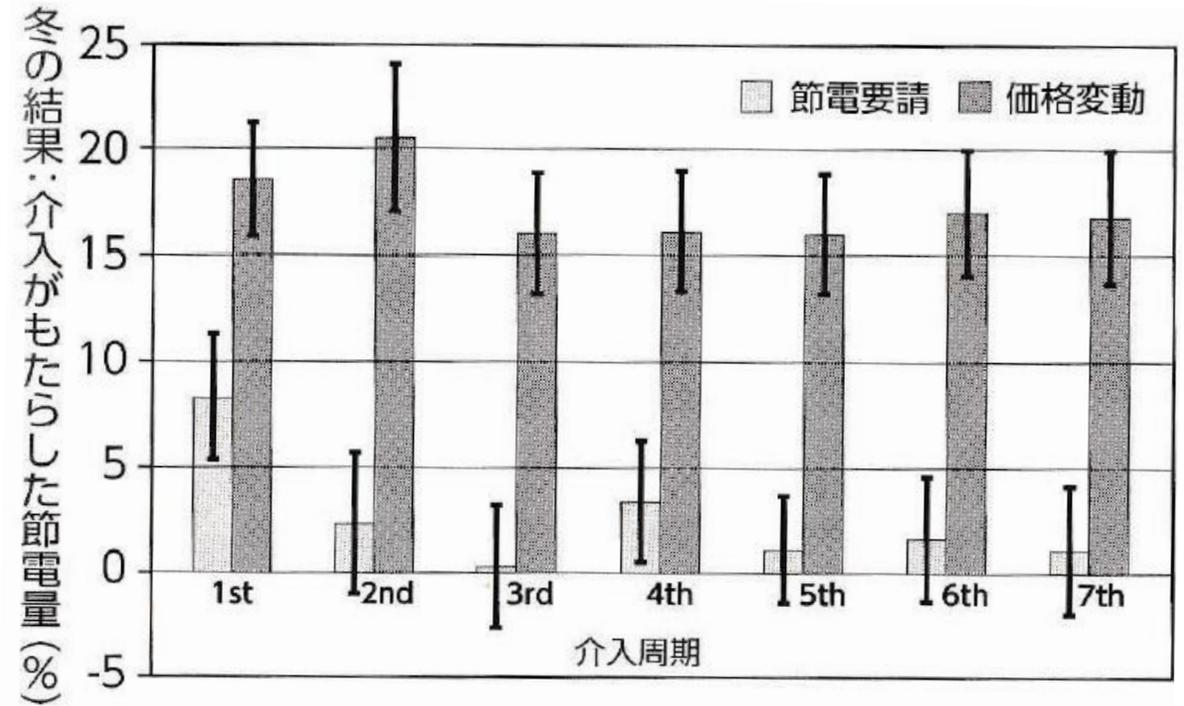
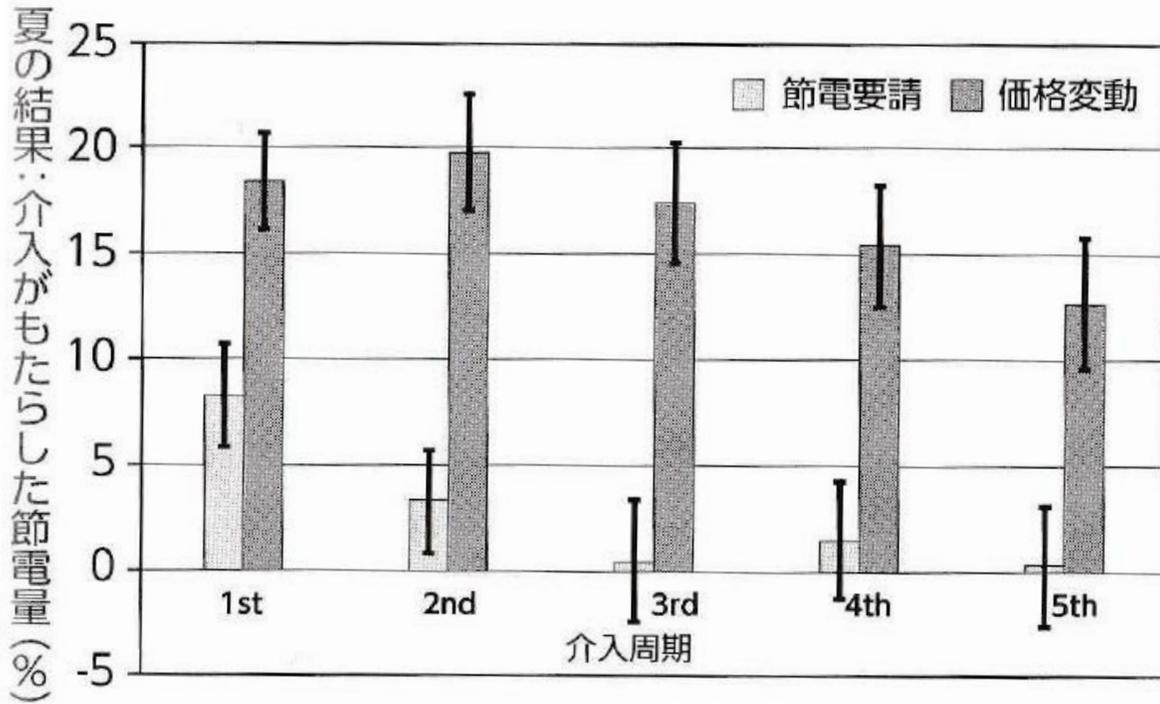
戸別住宅型

創エネ、蓄エネ機器を導入した67戸の新築住宅を中心とし、地産地消を行う実証を実施。また、暮らしの中における次世代自動車を含む次世代交通システムを実証。（トヨタ自動車・中部電力）

京都けいはんな地区で行われたRCTの結果



節電要請・価格変動 節電効果の持続性(京都けいはんな地区)



出所：伊藤公一朗(2017)『データ分析の力ー因果関係に迫る思考法』光文社新書、p.106。

ランダム化比較試験(RCT)の問題点

- ◆ 時間や費用がかかる(各方面の協力が不可欠)
- ◆ 失敗した場合の社会的コストが大きい
- ◆ 倫理的問題
- ◆ ホーソン効果(Hawthorne effect)
- ◆ 人為的に設定された条件下での話であり、どこまで現実に適用可能か？(外的妥当性)



人為的に介入群と対照群を作り出すモデル＝「介入研究」は理想的なモデルだが大変！

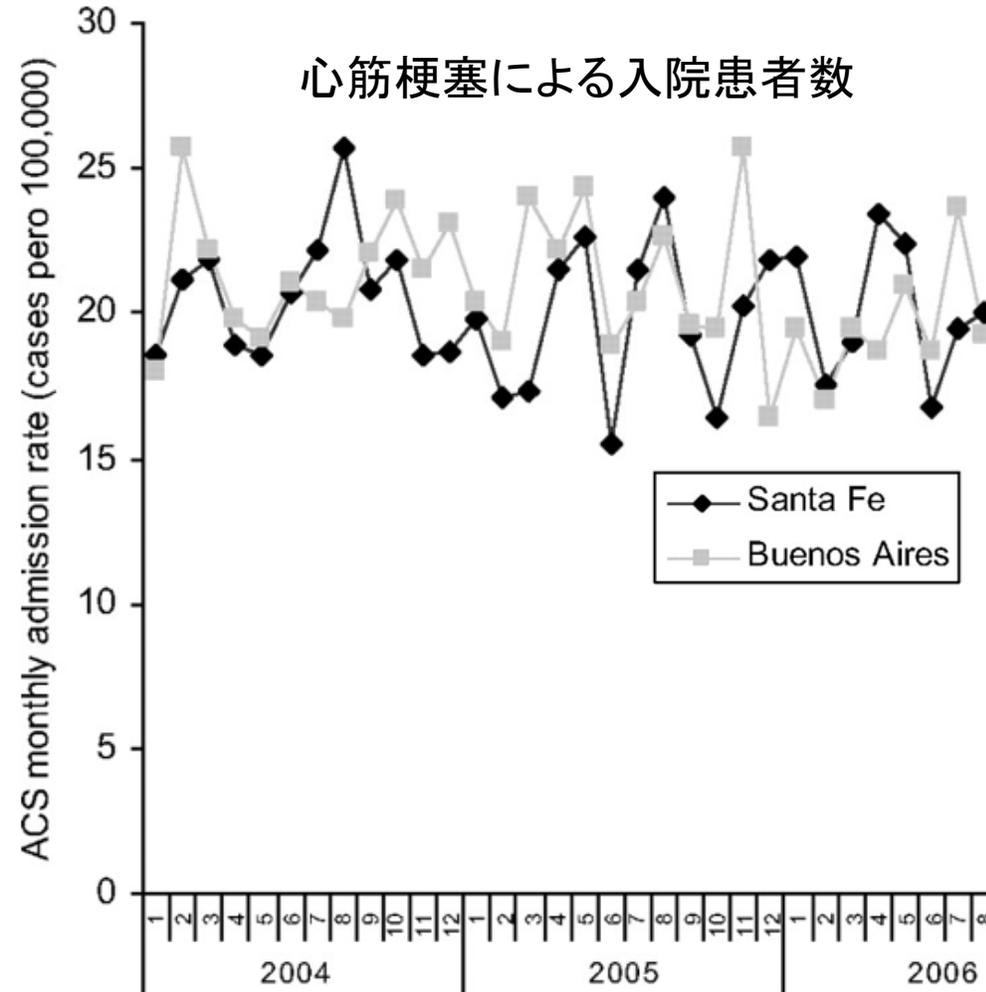


人為的に介入を行うことなく、対象者の活動や行動を観察することで得られたデータを用いる研究
＝「観察研究(非介入研究)」

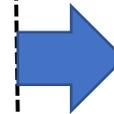
3-4 自然実験 (natural experiment)

政策の変更のような制度の見直しや自然災害の発生などの外生的要因によって、自然に実験群と対照群に分かれてしまった状況を利用するモデル

① 受動喫煙防止法制化 (サンタフェ)

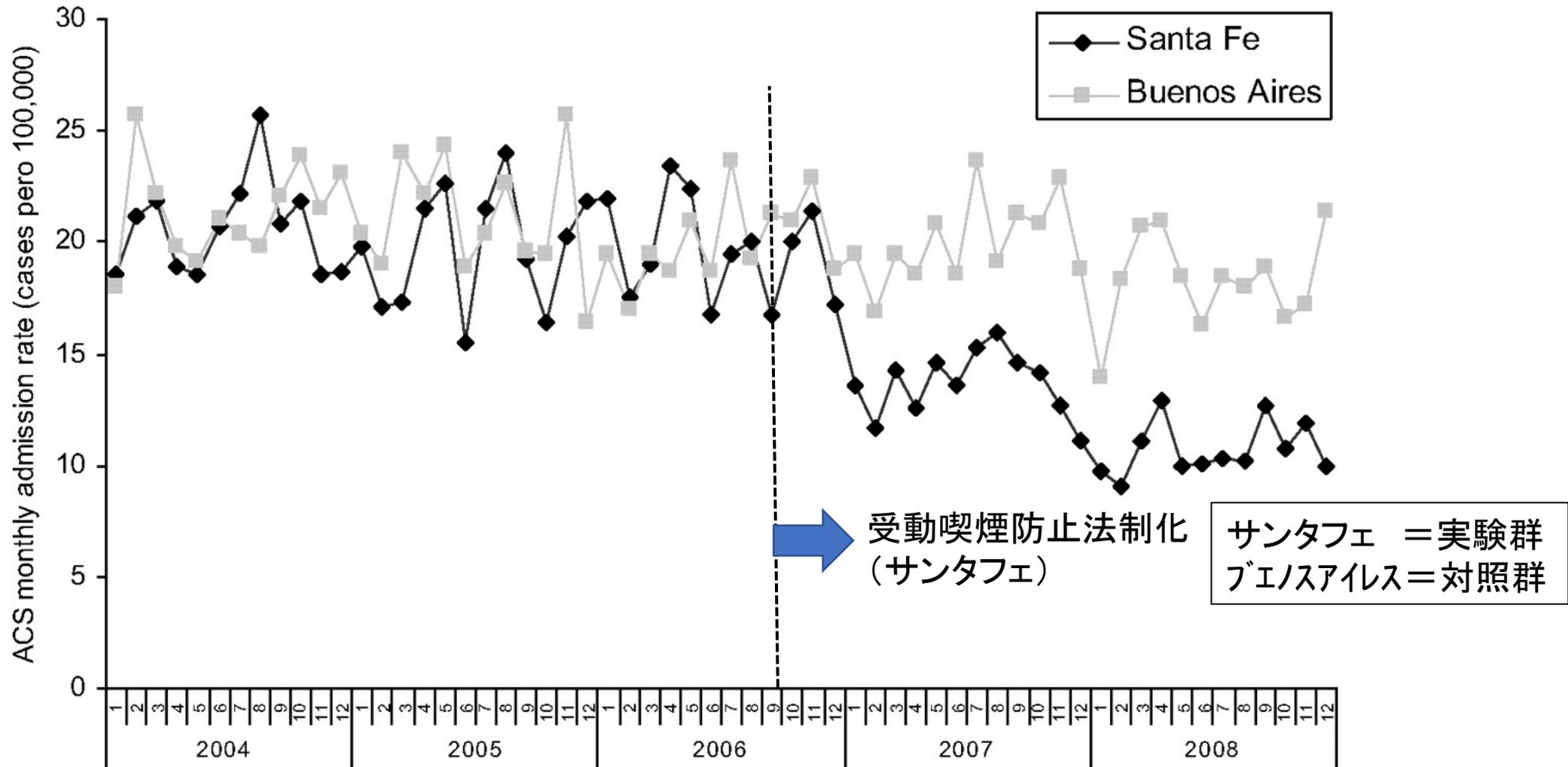


受動喫煙防止法制化 (サンタフェ)



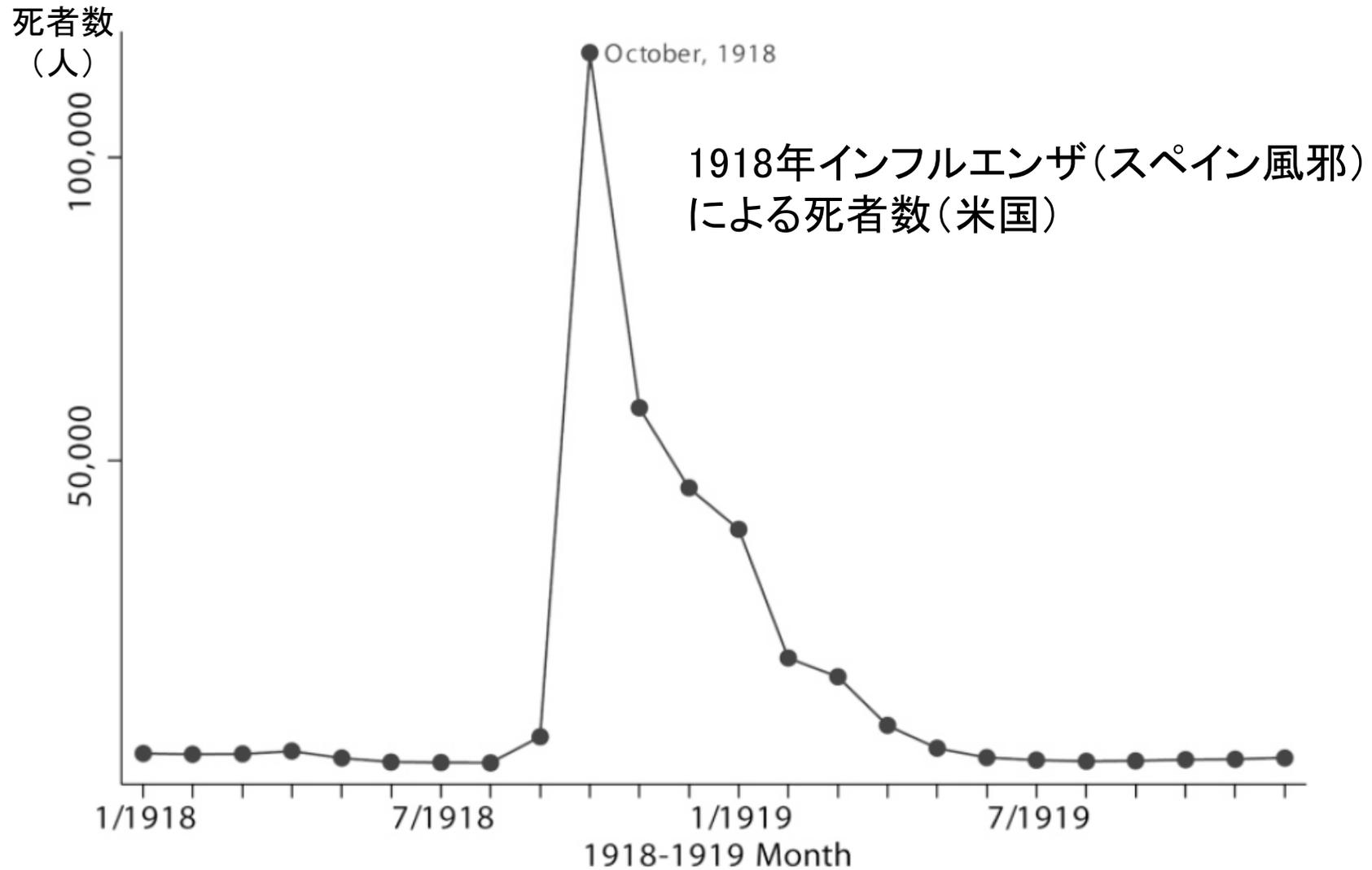
出所: D Ferrante, B Linetzky, M Virgolini, V Schoj, B Apelberg(2012) "Reduction in hospital admissions for acute coronary syndrome after the successful implementation of 100% smoke-free legislation in Argentina: a comparison with partial smoking restrictions" Tobacco Control, 21(4),402-406を加工。

心筋梗塞による入院患者数



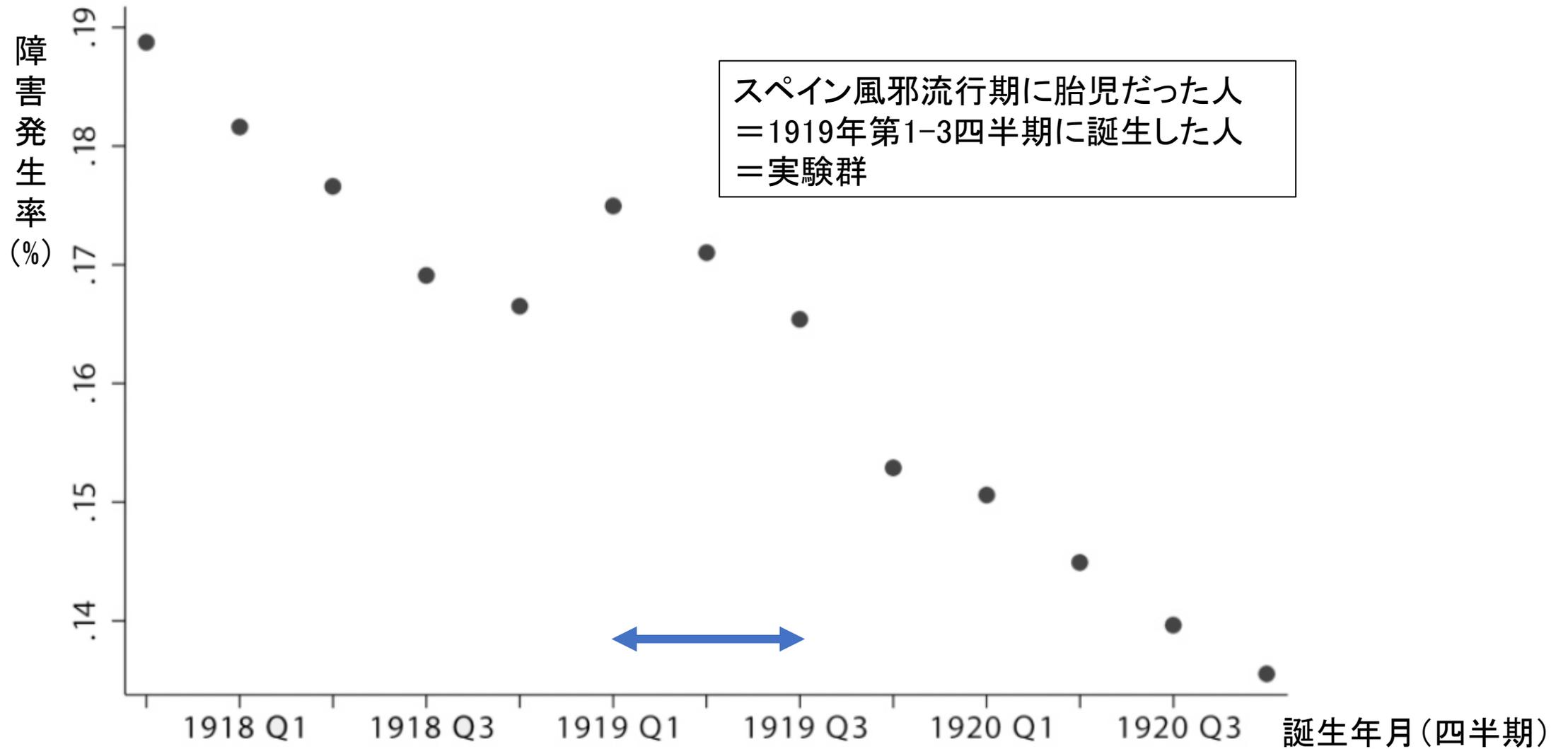
出所: D Ferrante, B Linetzky, M Virgolini, V Schoj, B Apelberg(2012) "Reduction in hospital admissions for acute coronary syndrome after the successful implementation of 100% smoke-free legislation in Argentina: a comparison with partial smoking restrictions" Tobacco Control, 21(4),402-406

② 1918年インフルエンザ(スペイン風邪)流行による健康への長期的影響



出所 : Almond,D (2006) Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population, *Journal of Political Economy*, 114(4),p.674.

誕生年による障害発生率(男性、1980年)



出所：Almond,D (2006) Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population, *Journal of Political Economy*, 114(4),p.675.

【自然実験の長所】

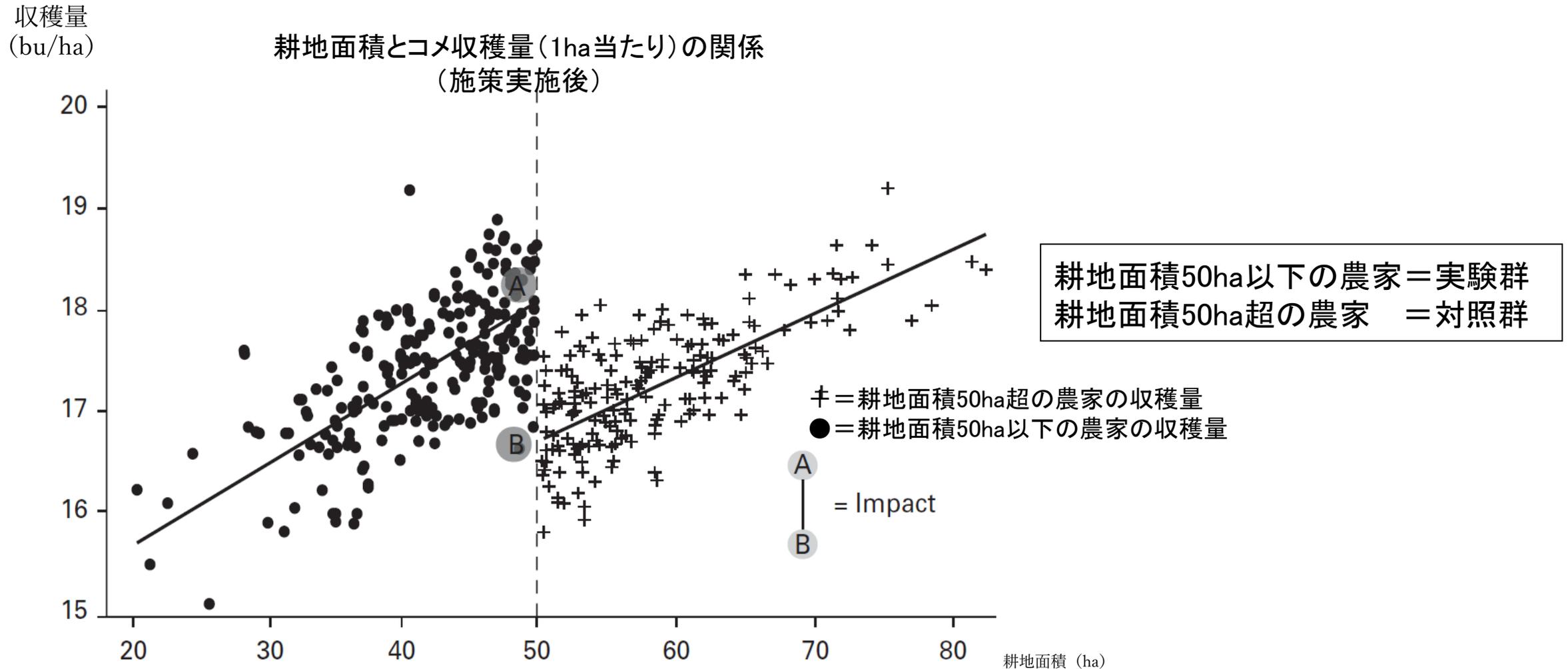
- 外的妥当性が高い—現実への適合性が高い
- 無作為割当に伴う倫理的問題が少ない
- 費用が少なくて済む

【自然実験の短所】

- 内的妥当性の点で劣る—観察されていない交絡要因を完全には排除できない

回帰分断法

= 実験群と対照群への割当を無作為割当ではなく、一定の基準値(閾値、cut-off point)を設け、この基準値を境にそれぞれの群に割当を行う方法

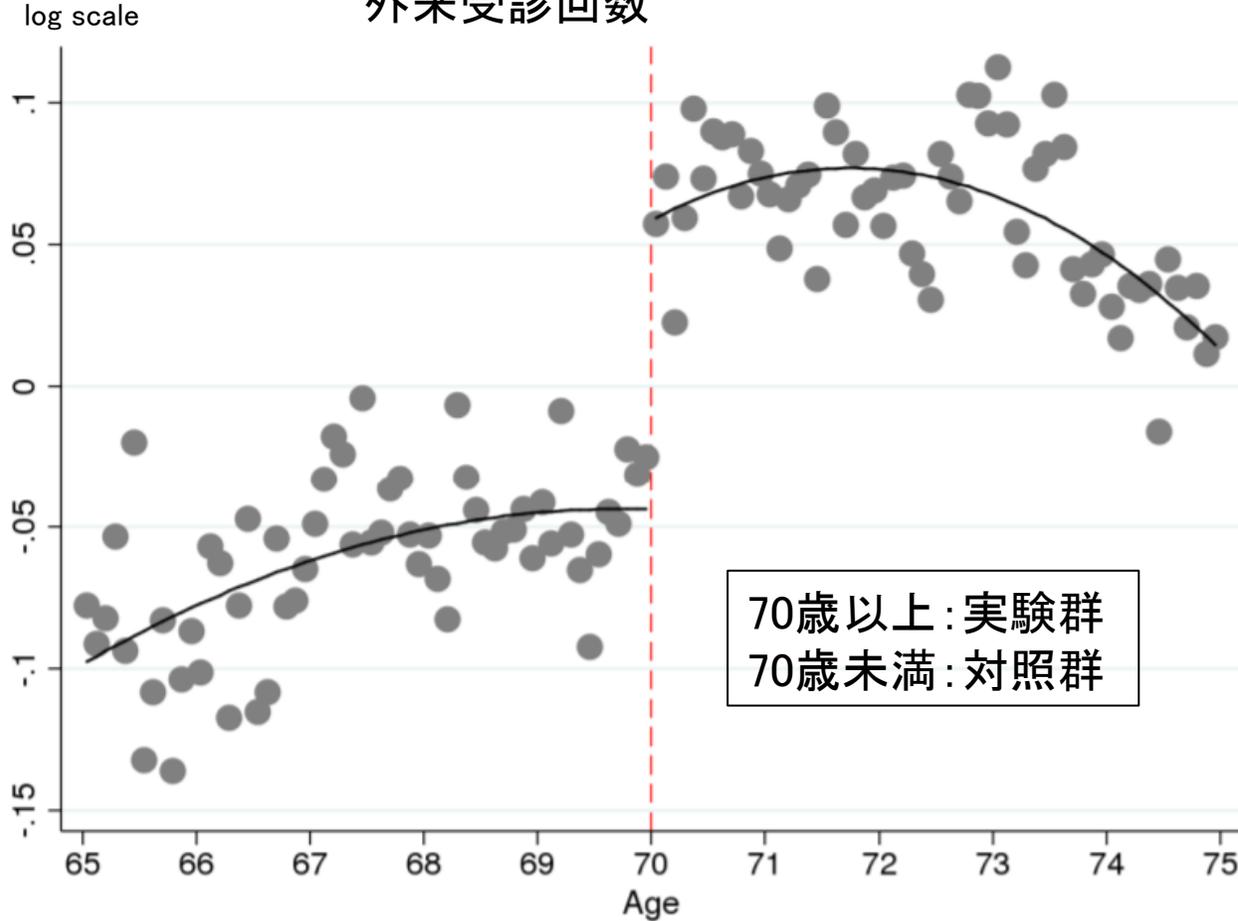


出所: Paul J. Gertler, Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch (2016) *Impact Evaluation in Practice (Second Edition)*, World Bank Group, 116-117.

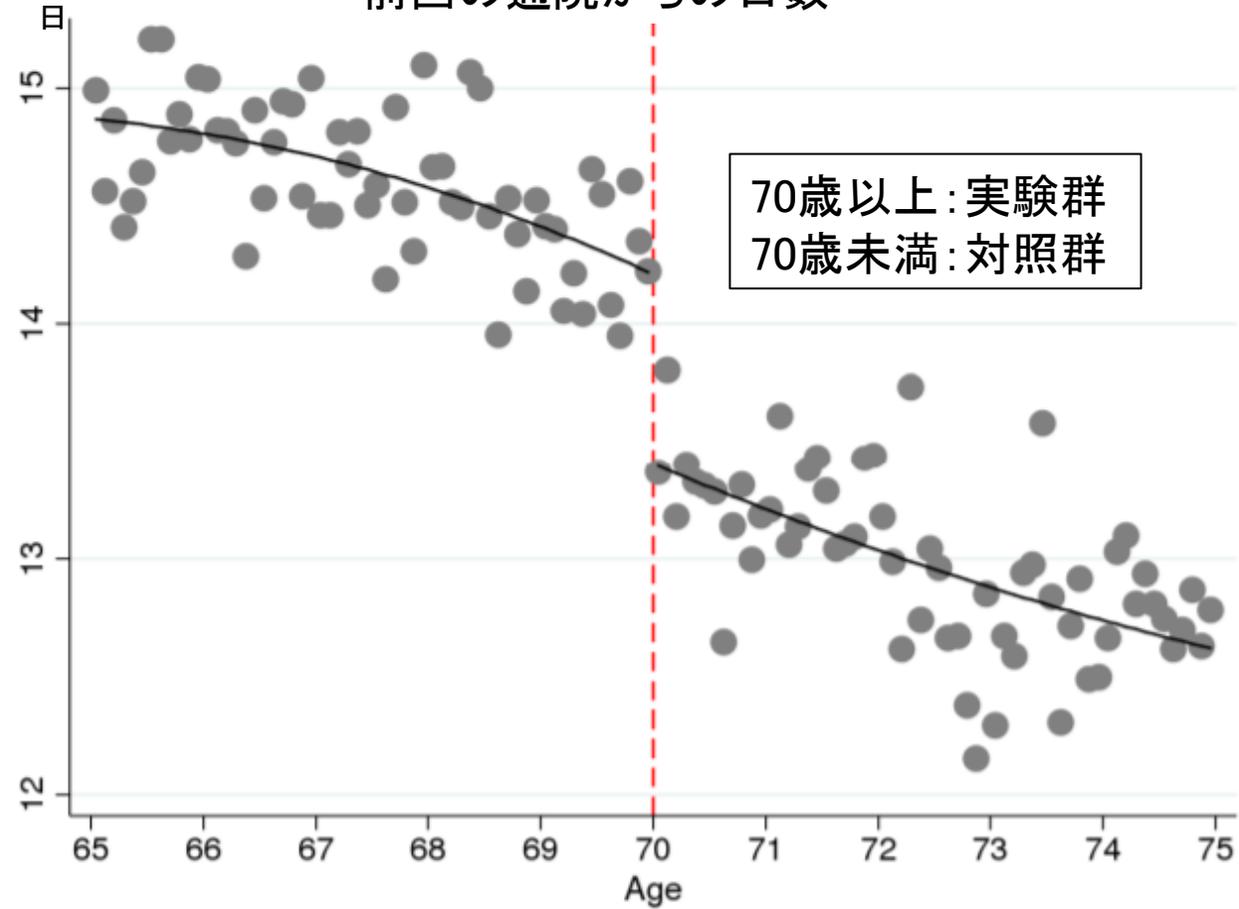
高齢者医療費・自己負担割合低下の効果

(70歳未満:3割 → 70歳以上:1割)

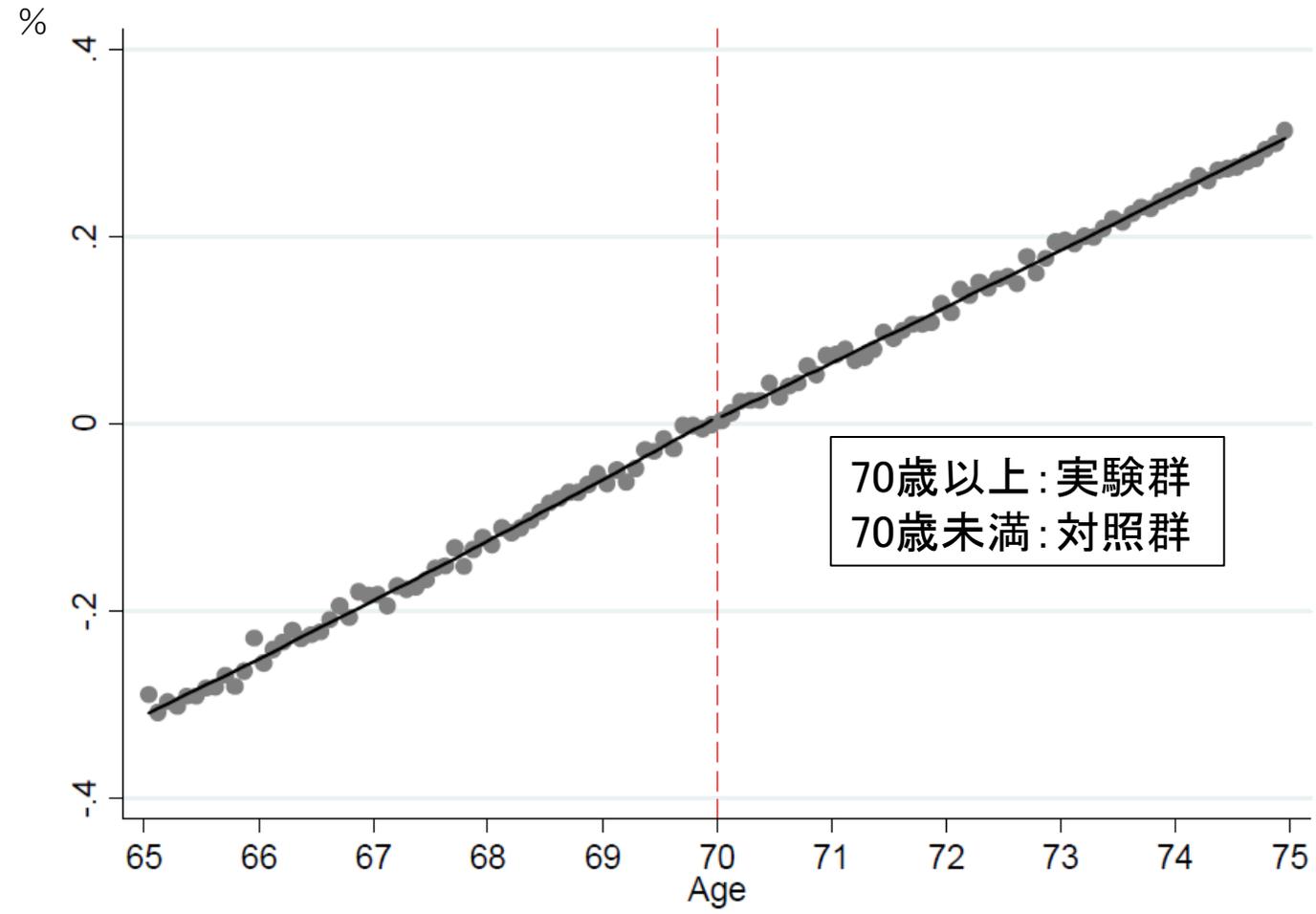
外来受診回数



前回の通院からの日数



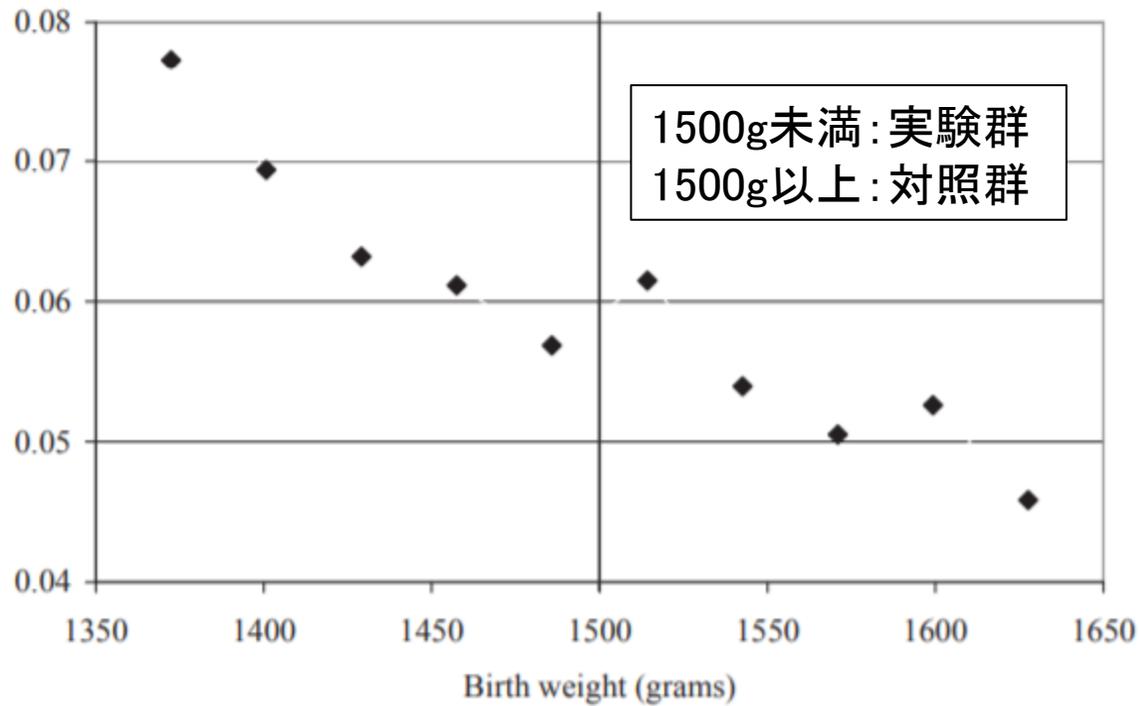
死亡率



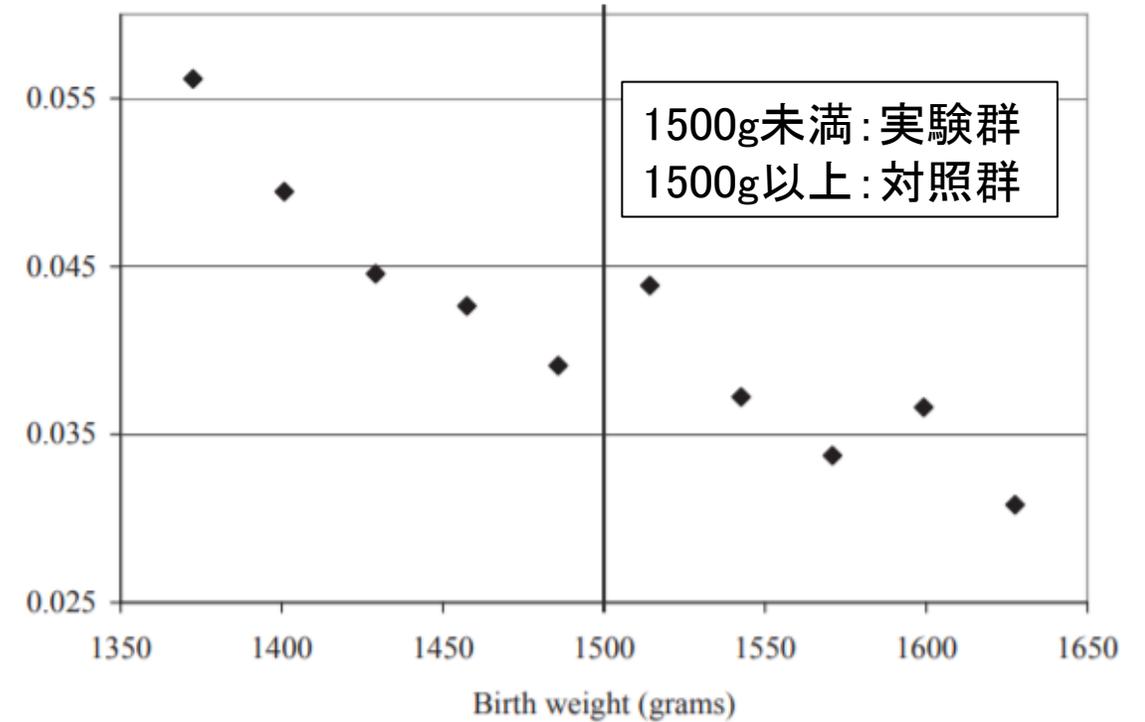
出所: Hitoshi Shigeoka, The Effect of Patient Cost Sharing on Utilization, Health, and Risk Protection, NBER Working Paper No. 19726, Issued in December 2013, p.32.

極低出生体重児(1500g未満)に対する付加的治療の効果

新生児死亡率
(出生後1年以内)



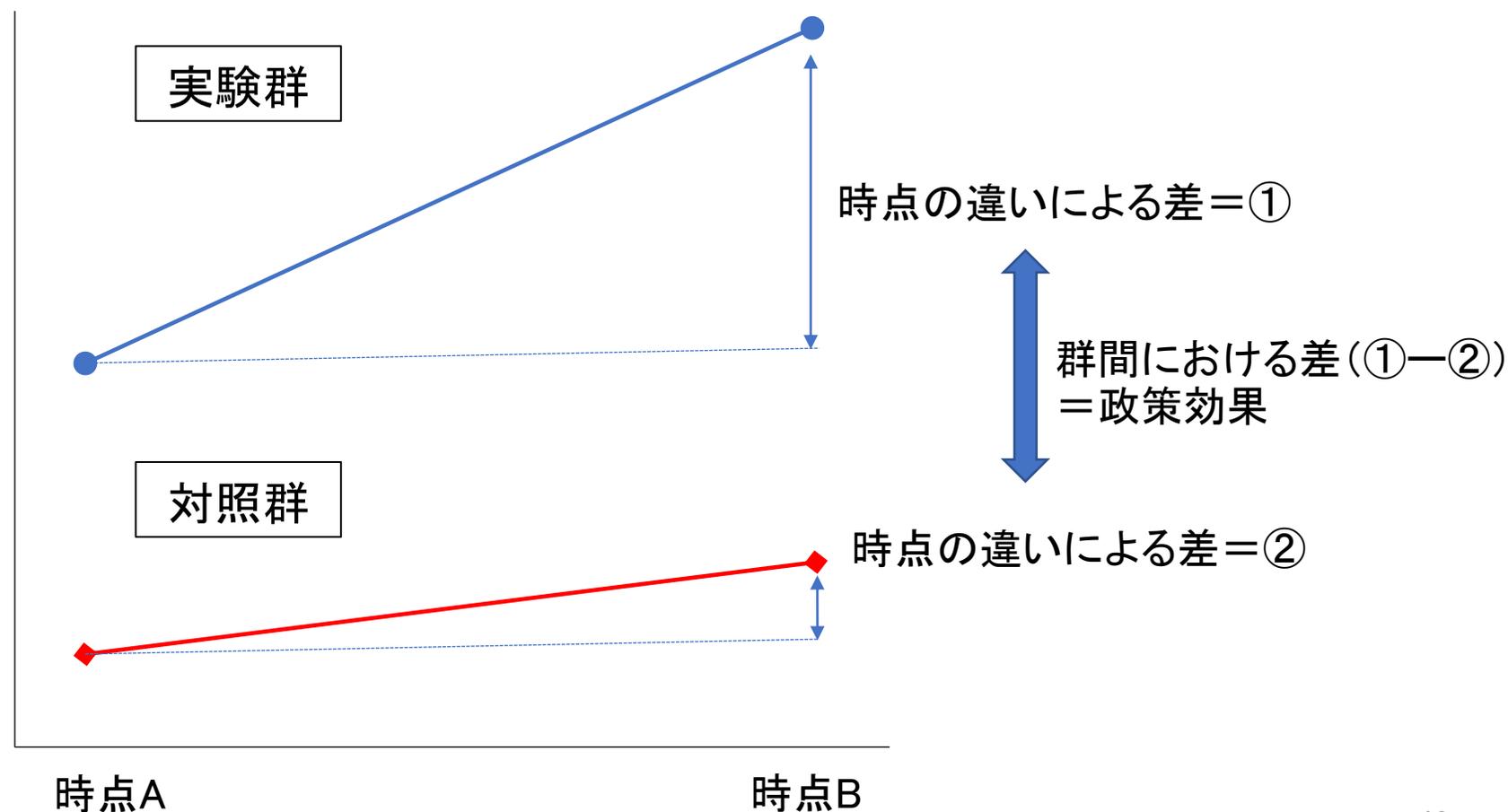
新生児死亡率
(出生後28日以内)



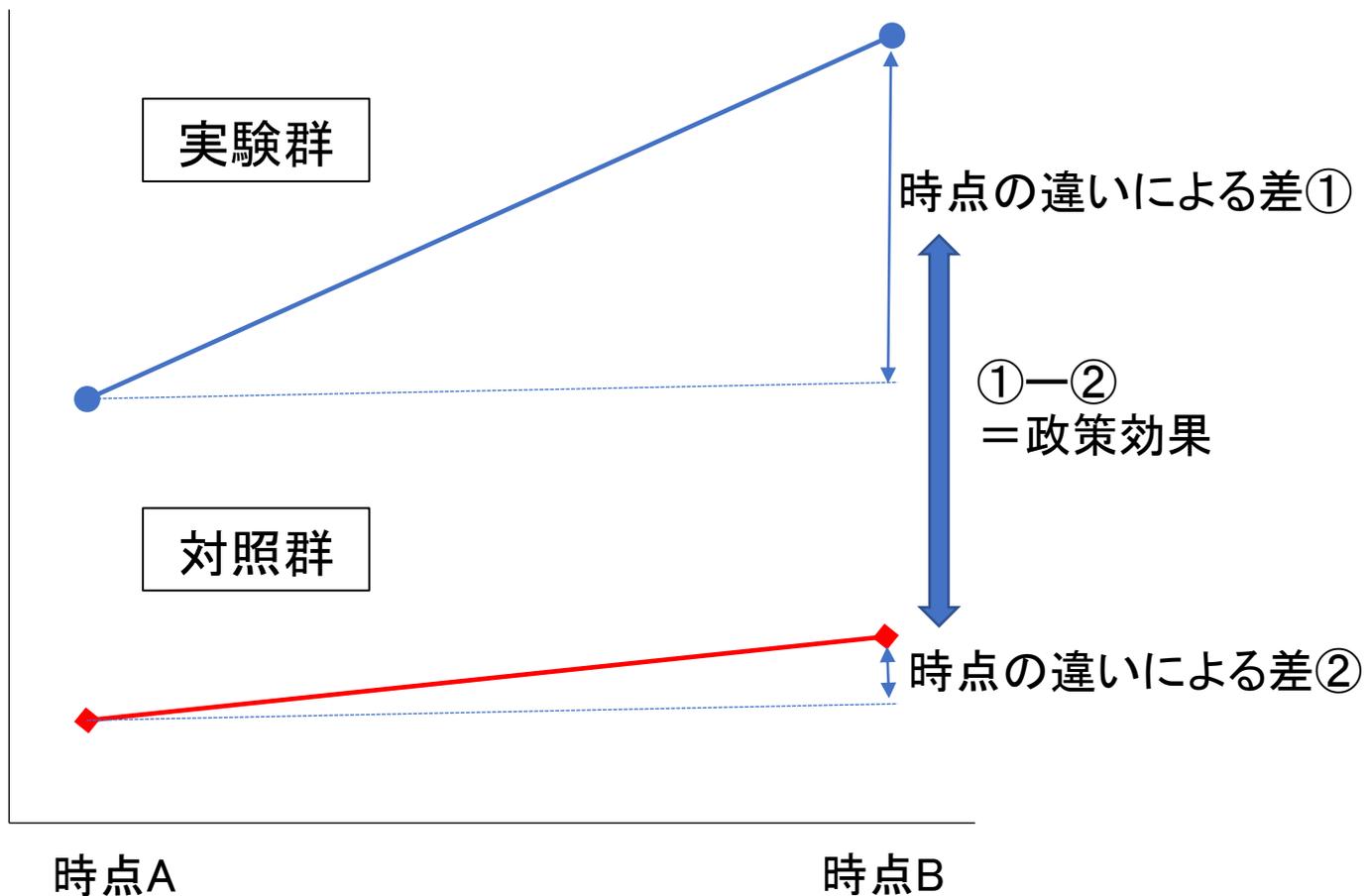
Source: D Almond, J J. Doyle, Jr., A E. Kowalski, H Williams (2010) Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns *The Quarterly Journal of Economics*, 125(2), 591-634.

3-6 DID (difference-in-difference) 分析

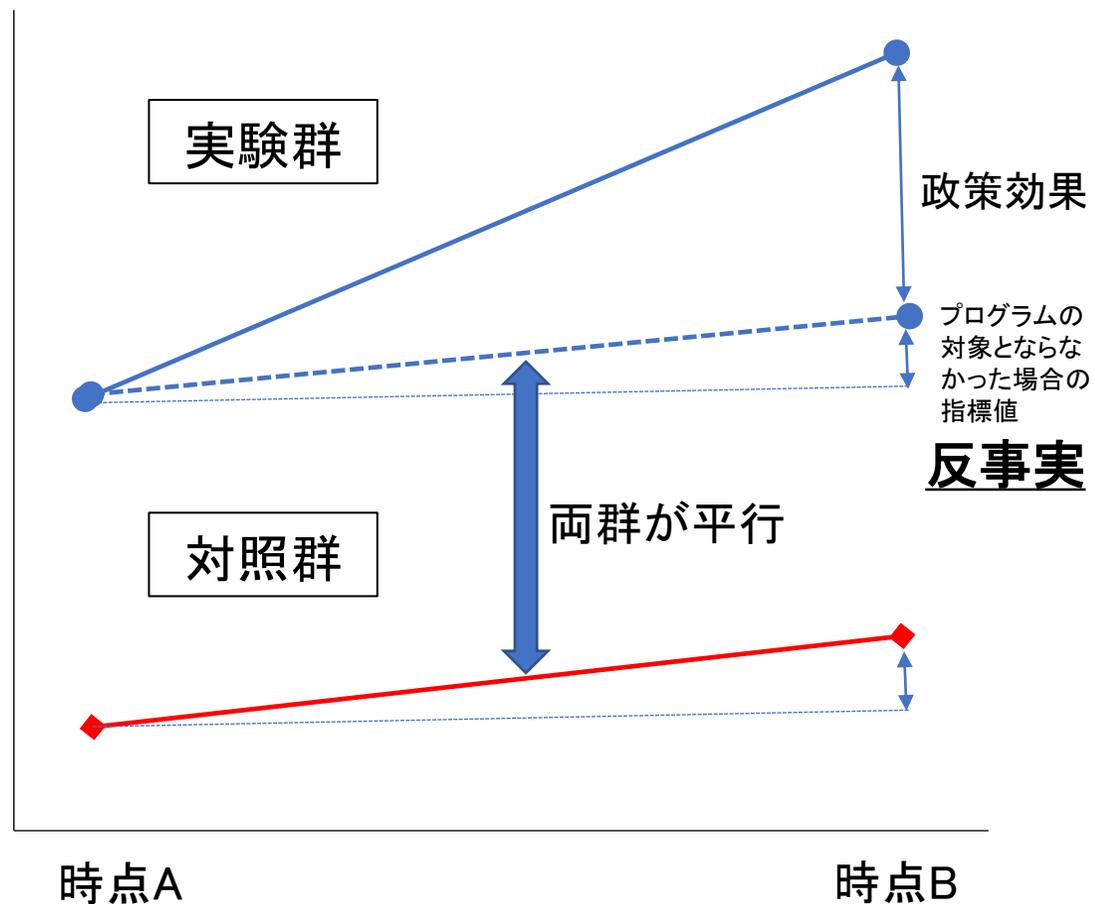
実験群と対照群を設定し、両群について時点の違いによる差と、群間における差を測定することで、政策効果を検証するモデル



実験群の差分①－対照群の差分②＝政策効果（インパクト）



平行トレンドの仮定 (parallel trends assumption)



消費喚起—現金給付か？商品券配布か？

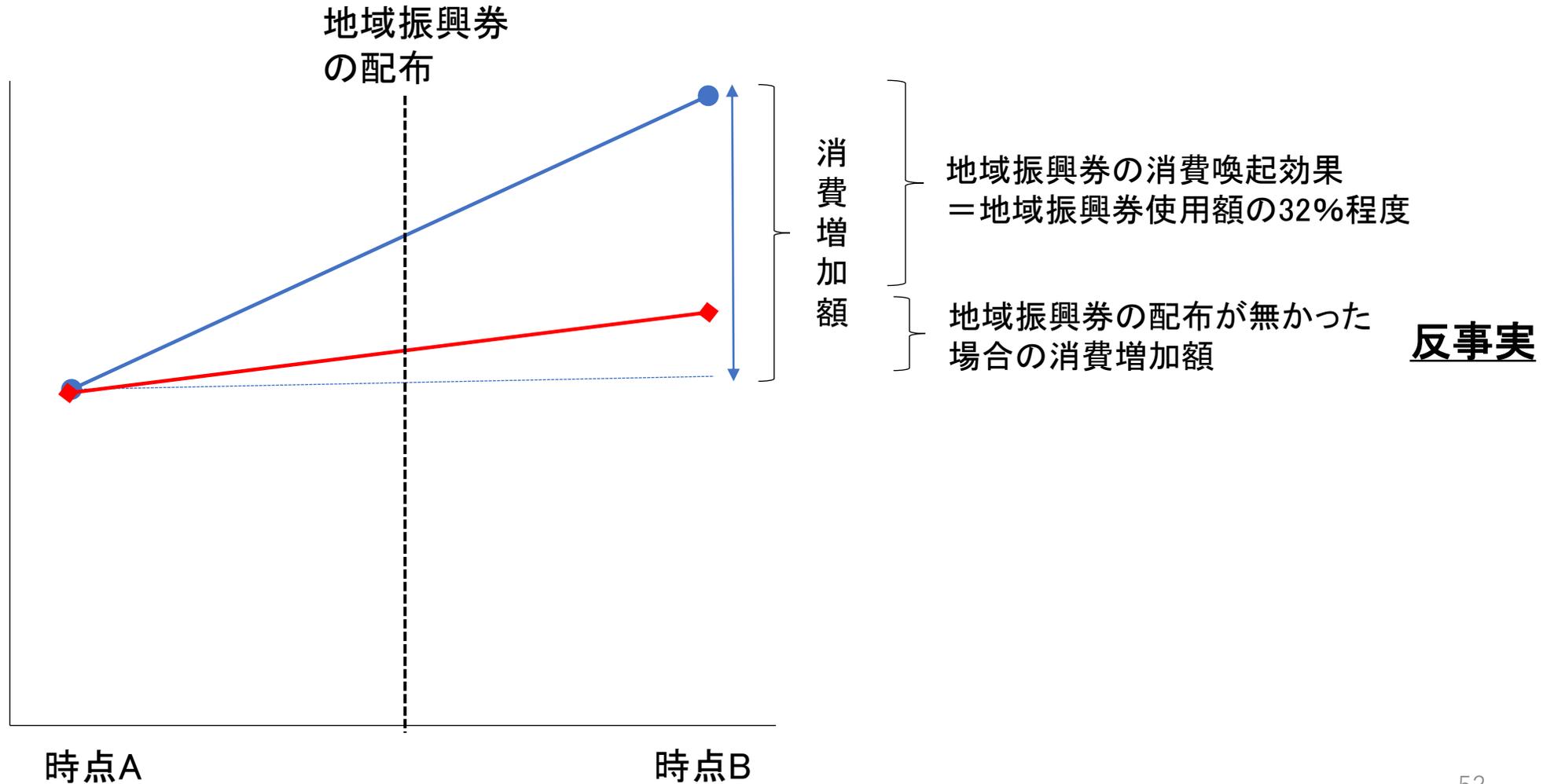
「現金給付は貯蓄に回る。現金給付より商品券(クーポン)が望ましい。」

地域振興券、定額給付金の比較

区 分	地域振興券	定額給付金
実施年	1999年	2009年
事業規模	6,194億円	1兆9,367億円
給付対象者	<ul style="list-style-type: none">• 15歳以下の児童が属する世帯の世帯主• 老齢福祉年金の受給者等	全世帯主

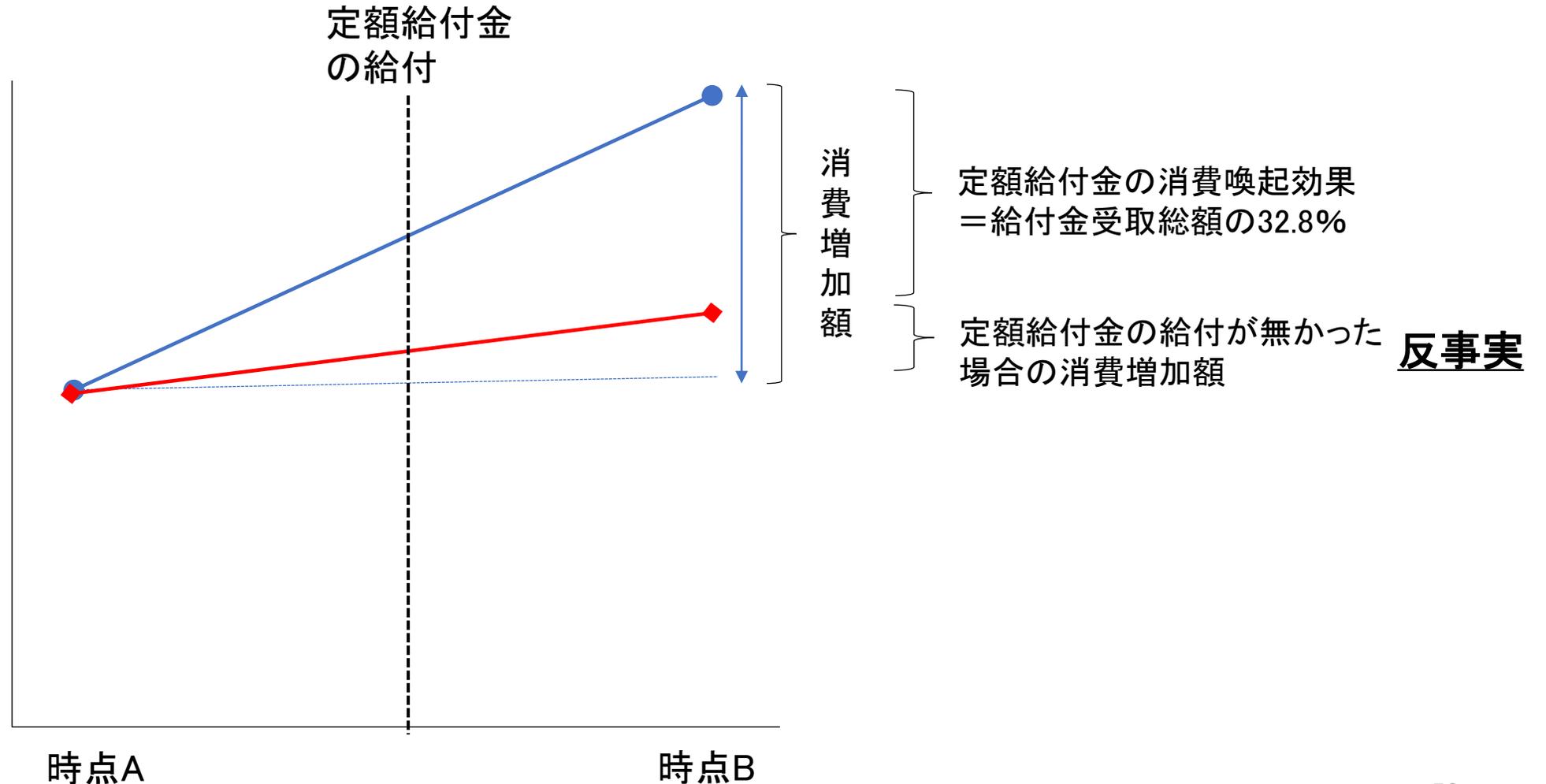
① 地域振興券(世帯に商品券配布)の効果

資料: 経済企画庁(1999)「地域振興券の消費喚起効果等について」



② 定額給付金(世帯に現金給付)の効果

資料:内閣府(2010)「『定額給付金に関連した消費等に関する調査』の結果について」



「現金給付は貯蓄に回る。現金給付より商品券(クーポン)」と言えますか？

地域振興券(世帯に商品券配布)



定額給付金(世帯に現金給付)



消費喚起効果

32%程度

32.8%

4 むすびにかえて

4-1 統計はEBPMを支える基礎

統計改革推進会議「最終取りまとめ」(平成29年5月19日)

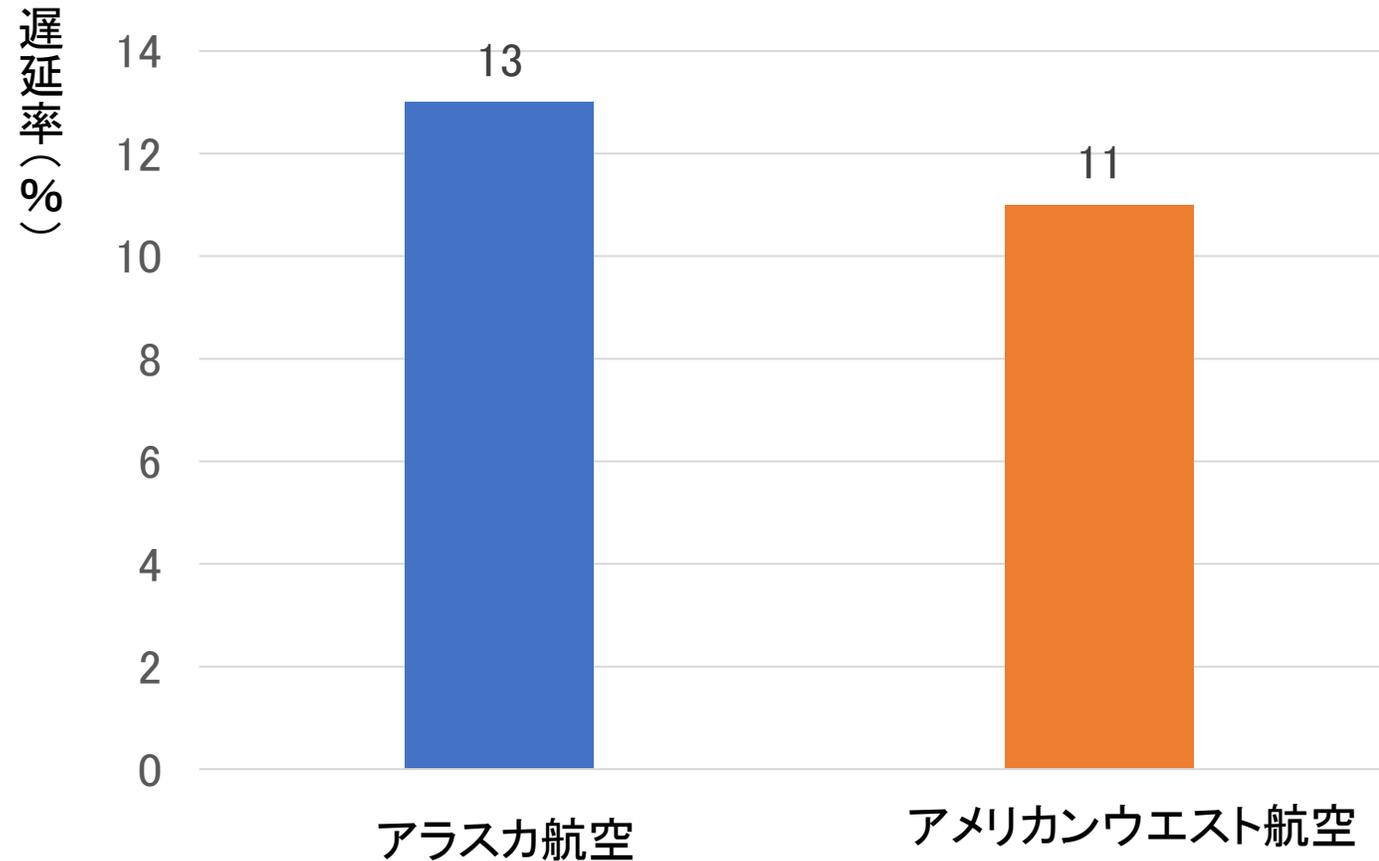
EBPMの基盤をなすのが、統計等データを始めとする各種データなどの客観的な証拠であり、政策課題の把握、政策効果の予測・測定・評価による政策の改善と統計等データの整備・改善が有機的に連動するサイクル(EBPMサイクル)を構築することが必要である。



EBPMと統計は「車の両輪」であり、統計はEBPMを支える基礎。

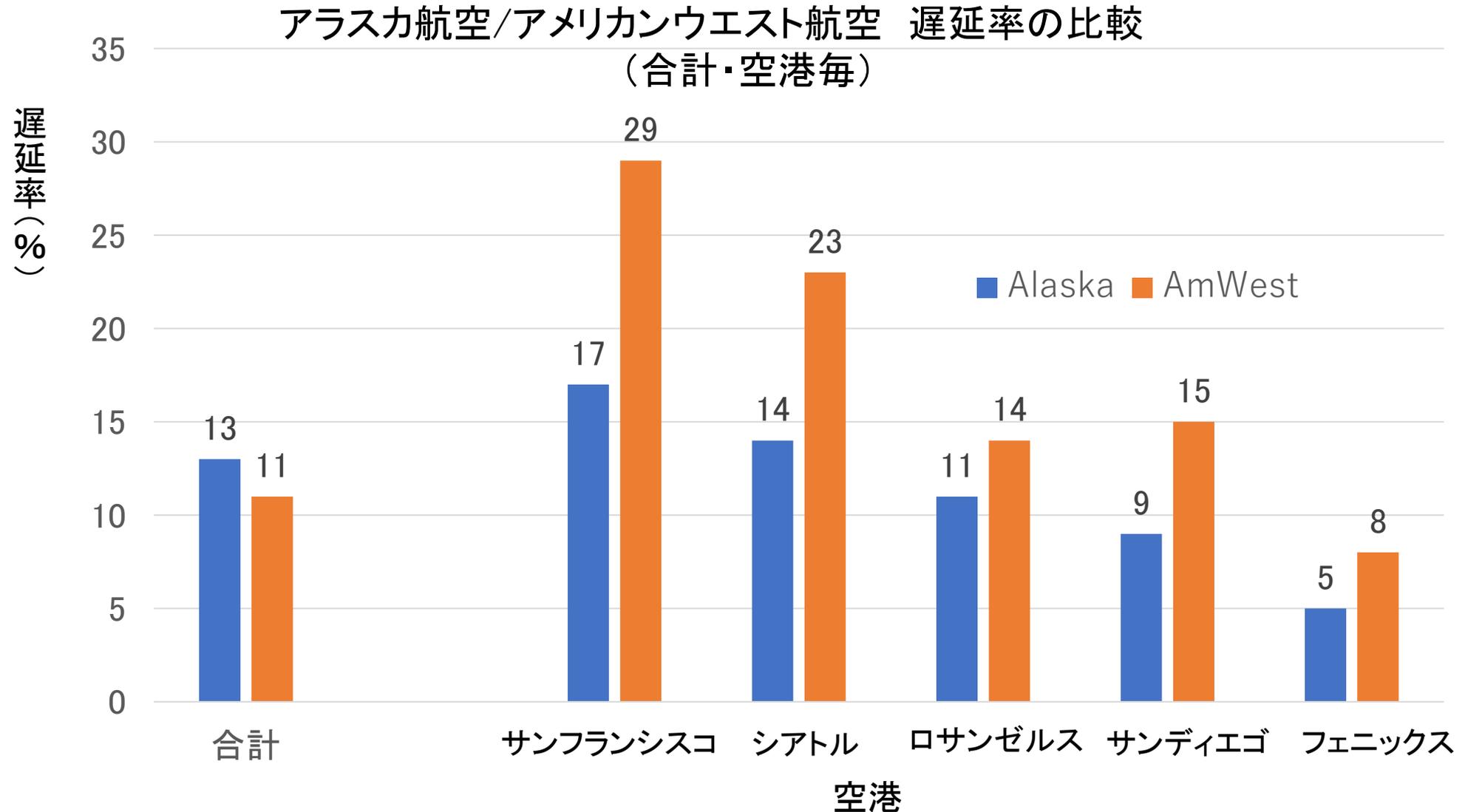
4-2 「世の中には3種類の嘘がある。嘘、真っ赤な嘘、そして…」

アラスカ航空/アメリカンウエスト航空 遅延率の比較



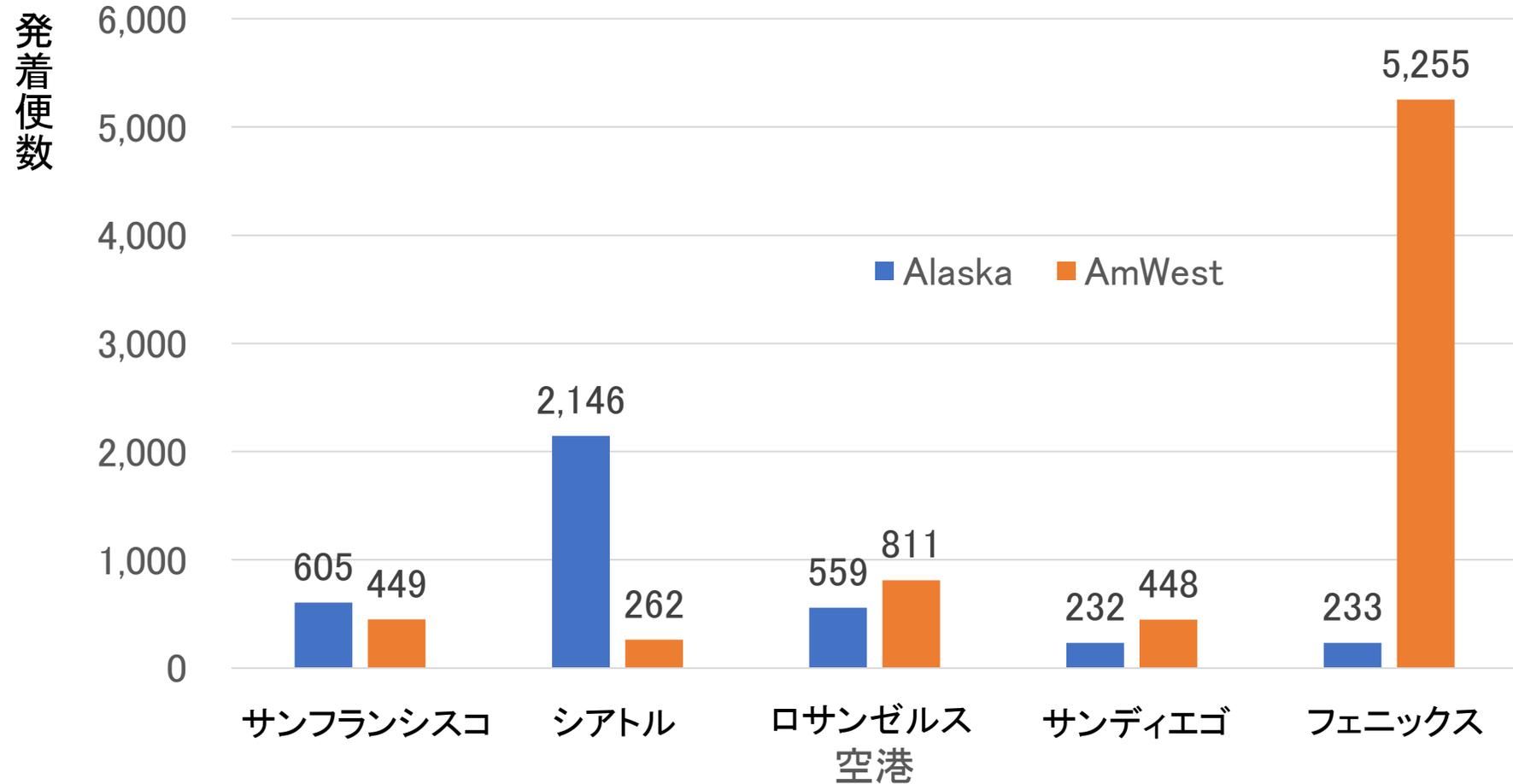
出所：Fung,Kaiser(2013)“Numbersense:How to Use Big Data to Your Advantage”,McGraw-Hill,p.14より作成.

シンプソンのパラドックス (Simpson's Paradox)



シンプソンのパラドックスの原因

空港ごとの発着便数



出所：Fung,Kaiser(2013)“Numbersense:How to Use Big Data to Your Advantage”,McGraw-Hill,p.14より作成.

統計改革推進会議「最終取りまとめ」(平成29年5月19日)

EBPMと統計は「車の両輪」であり、統計はEBPMを支える基礎。

しかし、…

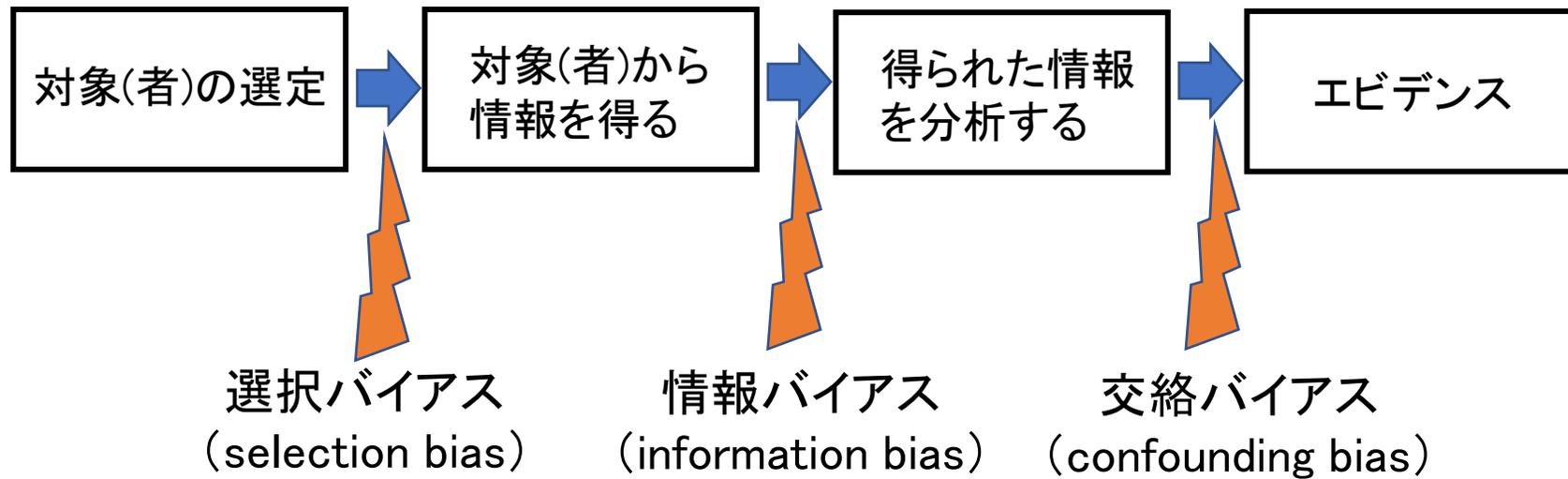
"There are three kinds of lies:
lies, damned lies, and statistics."

「世の中には3種類の嘘がある。
嘘、真っ赤な嘘、そして統計。」

出所：Twain, M. (1907). Chapters from My Autobiography: XX. North American Review, 185(618):465-576.

4-3 改めての確認

エビデンスを得るためのプロセスはあらゆる段階においてバイアスに晒されている



評価は客観的な情報、データ、事実、手法に基づき行われることで、
その結果は政策の企画立案に活用される重要な情報【エビデンス】となる