

政策評価に関する統一研修(地方研修)

政策効果の把握のための評価・分析手法

令和2年1月23日

早稲田大学政治経済学術院・公共経営大学院

山田 治徳

1 政策評価とEBPM

1-1 EBPMの本質は評価

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。



EBPMの本質は、因果関係の評価



【原因】

【結果】

1-2 政策評価はEBPMの中核

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)

「政策をよりよいものとしていくためには、**定量的な根拠に基づいた評価の徹底**が課題である。」

「**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。現状として、政策の目標や測定指標に影響を与える様々な要因の分析が的確になされていない、政策効果が見えにくいといったものもあり、様々な要因について踏み込んだ分析を行うことが求められる。」

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**(予算要求(定員等を含む。)、税制改正要望、法令等による制度の新設・改廃といった作業)**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。**



そもそも評価はEBPMの中核を成すもの

1-3 政策評価の結果＝エビデンス

政策評価の実施に関するガイドライン
(平成17年12月16日、政策評価各府省連絡会議了承)

- ① 政策評価の実施に当たっては、まずは定量的な評価手法の開発を進めるよう努め、可能な限り具体的な指標・数値による定量的な評価手法を用いるよう努める。
- ② 定量的な評価手法の適用が困難である場合又は客観性の確保に結び付かない場合などにおいては、定性的な評価手法を適用するものとし、その際、可能な限り、客観的な情報・データや事実に基づくものとしたり、評価において学識経験者等を活用するなど、評価の客観性の確保に留意する。



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

2 エビデンスの妥当性とバイアス

2-1 エビデンスの要件



エビデンス = 科学的根拠のある情報

科学的根拠のない情報

学問的な裏付けのある客観的な情報

2-2 エビデンスの妥当性

- ◆ 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」

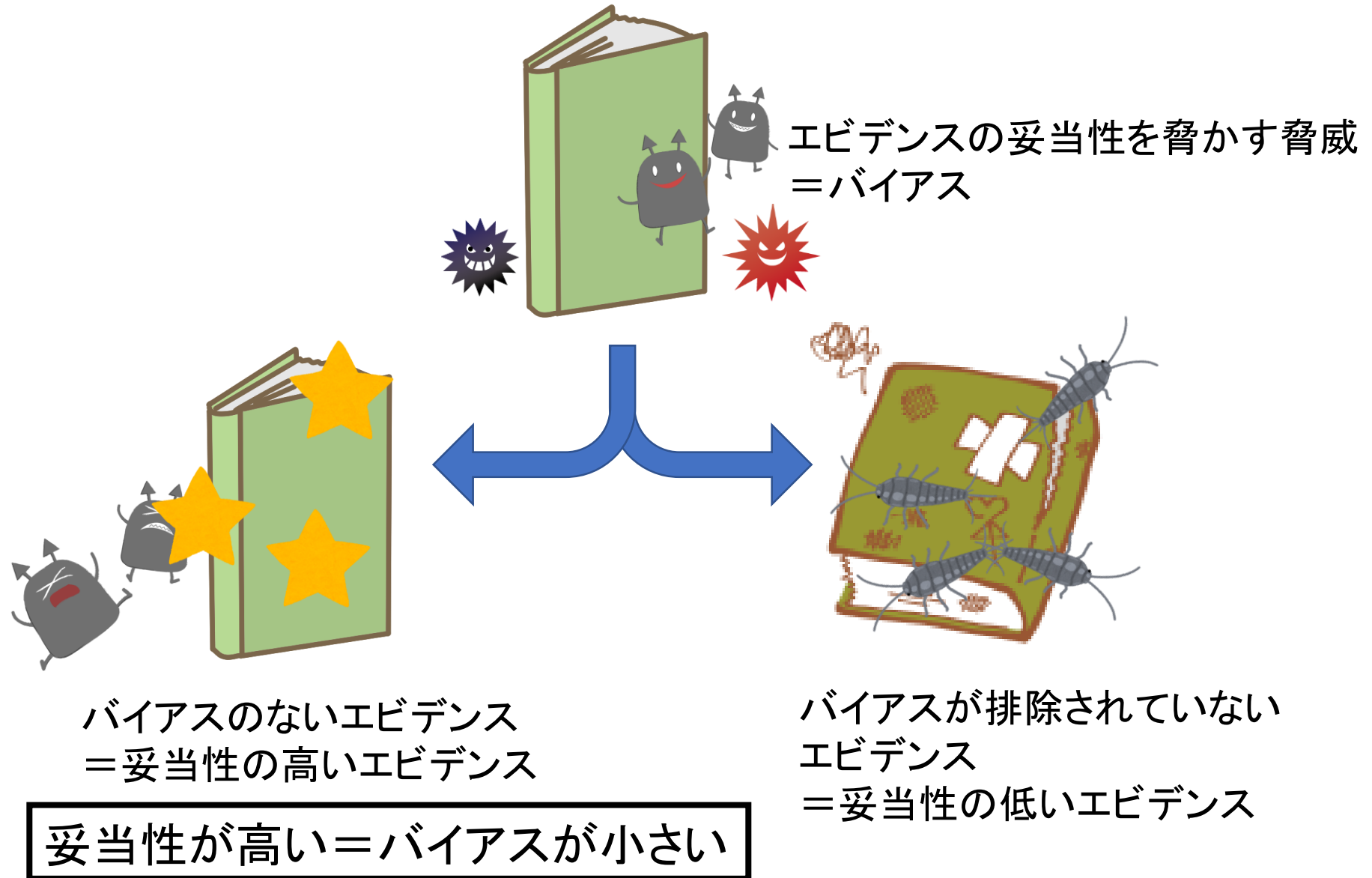
- ◆ 構成概念妥当性 (construct validity)

＝「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」

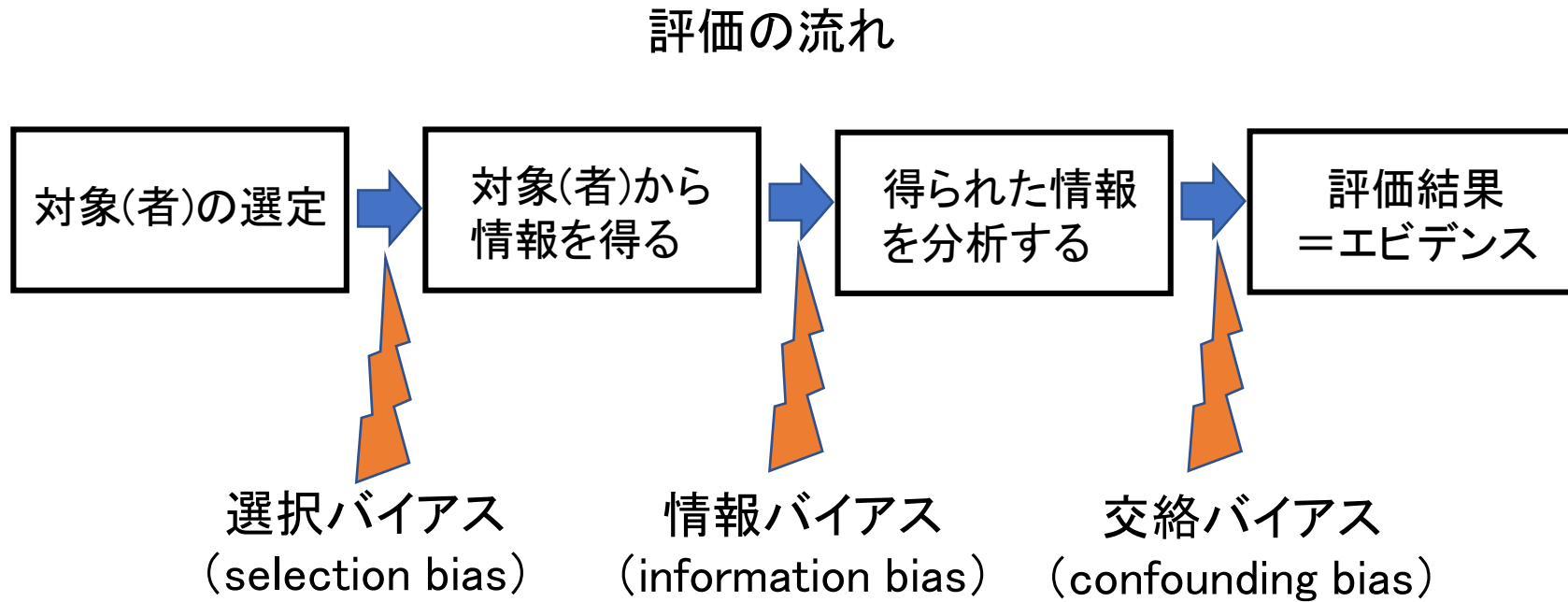
- ◆ 外的妥当性 (external validity)

＝「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」

2-3 エビデンスの妥当性を陥れるもの—バイアスの存在

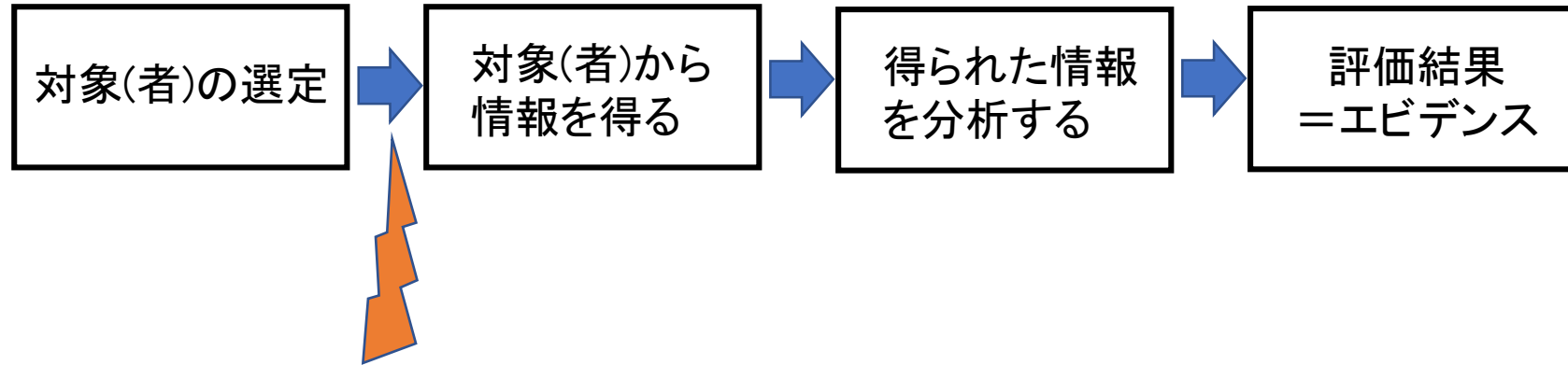


2-4 評価の流れと発生するバイアス



3 政策効果の把握のための評価・分析手法

3-1 選択バイアス (selection bias)

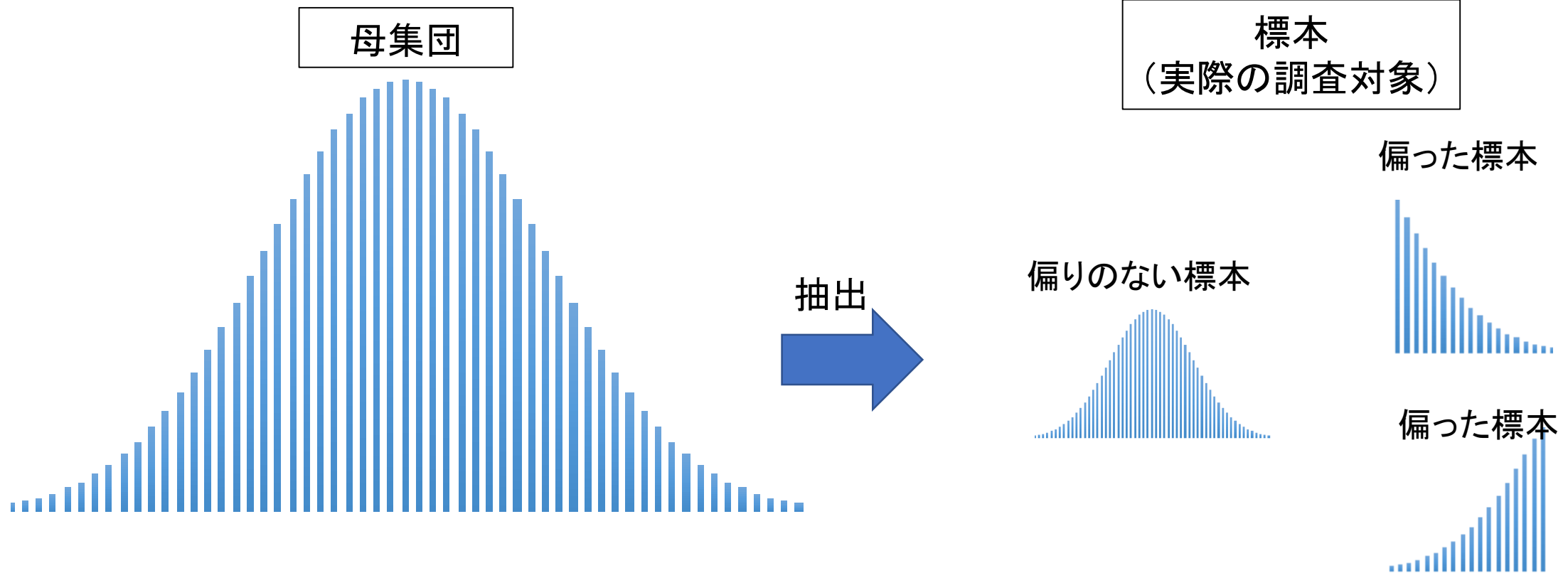


選択バイアス (selection bias)

= 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

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



対象者選定の方法や場所などによって発生

立石港立石地区港湾緑地整備事業・来訪者数予測

【選択バイアスの事例】



出所：Google Map

立石港立石地区港湾緑地整備事業・どのように年間来訪者数を予測したのか

想定利用者：徒歩20分圏内の住民 8,726人（母集団）



8,726人から102人を抽出（標本）



102人の利用頻度を調査する



102人のうち約4人に1人が「毎日来る」と回答



8,726人のうち約4人に1人、2,030人が「毎日来る」



年間来訪者数 = 2,030人 × 365日 = 74万人

分析結果＝「年間来訪者数74万人」

香川県を代表する観光地、屋島 57万人

立石港港湾緑地: 屋島を上回る74万人!?

歴史が動いた源平合戦の大舞台、屋島。
那須与一が射抜いた扇の的のエピソードなど、
数々の歴史ロマンが誕生しました。

げんべいがっせんこせんじょう
源平合戦古戦場



やしま
屋島スカイウェイ



山上にある全国でも珍しい水族館

しんやしますいぞくかん
新屋島水族館



出所：Google Map

「どこか変だ。まず74万人。山頂へのドライブウェイもありホテルもあり水族館もあり、8歳で海のもくずと消えた安徳天皇のゆかりの寺社など、香川県を代表する観光地の屋島でさえ観光客は57万人でしかない。屋島の1000分の1の大きさの埋立地の客のほうが多い？」

出所：猪瀬直樹(2007)「借金列島のつくり方、教えます」週刊文春H19.3.29号連載より一部抜粋。

出所：屋島関係情報ポータルサイト「屋島ナビ」 (<http://www.yashima-navi.jp/>)

調査対象者の選定方法



なぜこのような矛盾が起こったのか

区 分	母集団 徒歩20分圏内人口	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

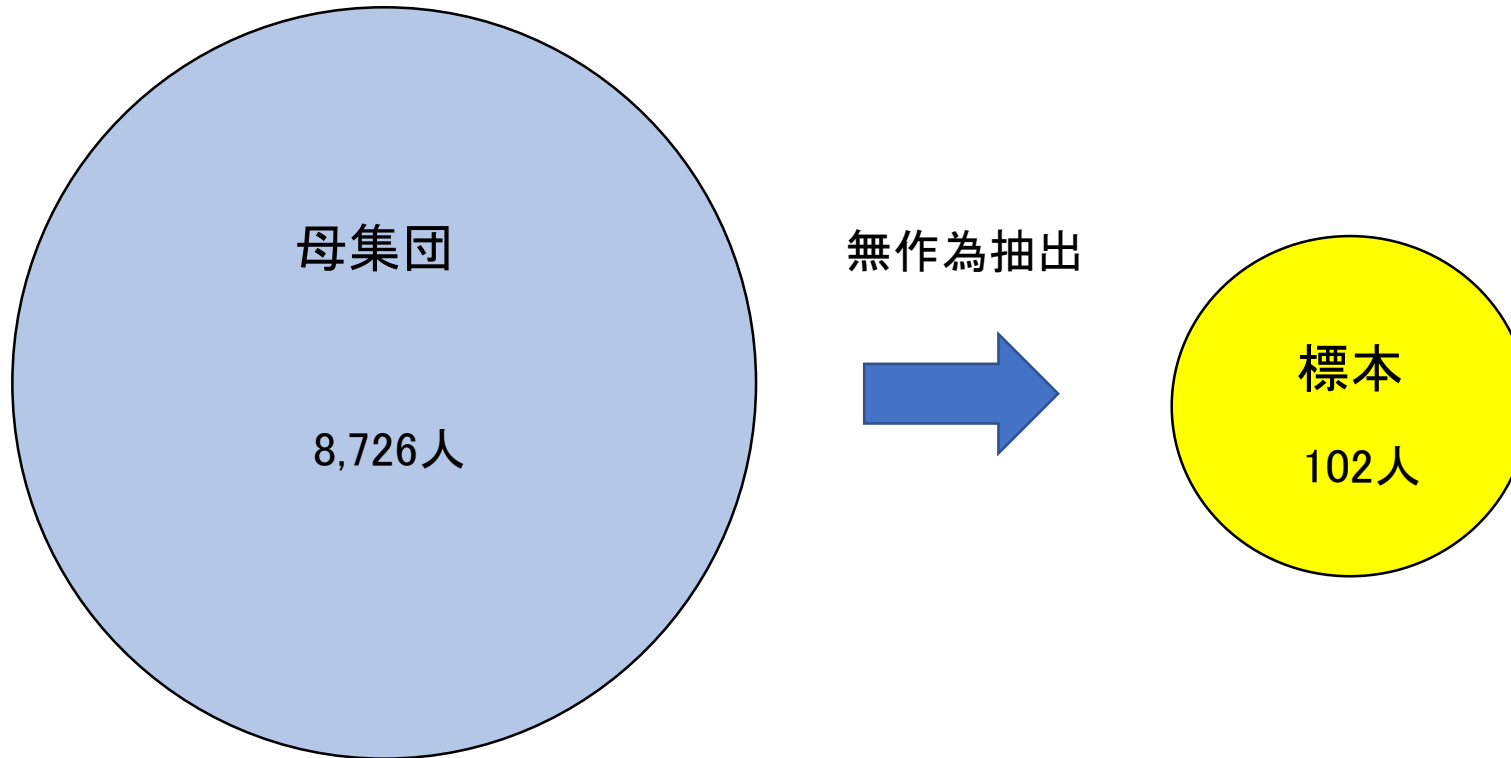


「半径2キロはタテマエでごく近所の人たち、半径100メートルの人たちにだけアンケートをとれば、4人に1人が毎日来訪する、という結果が出るに違いない。そうすると母集団8000人以上ならば2000人余とサギみたいな計算をした。2000人余が365日通うと74万人になる。実際には工事現場の前を通る人に声をかけ、アンケート調査した。近所の人なら、犬の散歩にでも毎日行く、と答えるかもしれない。そういう102人に訊ねたのだ。」

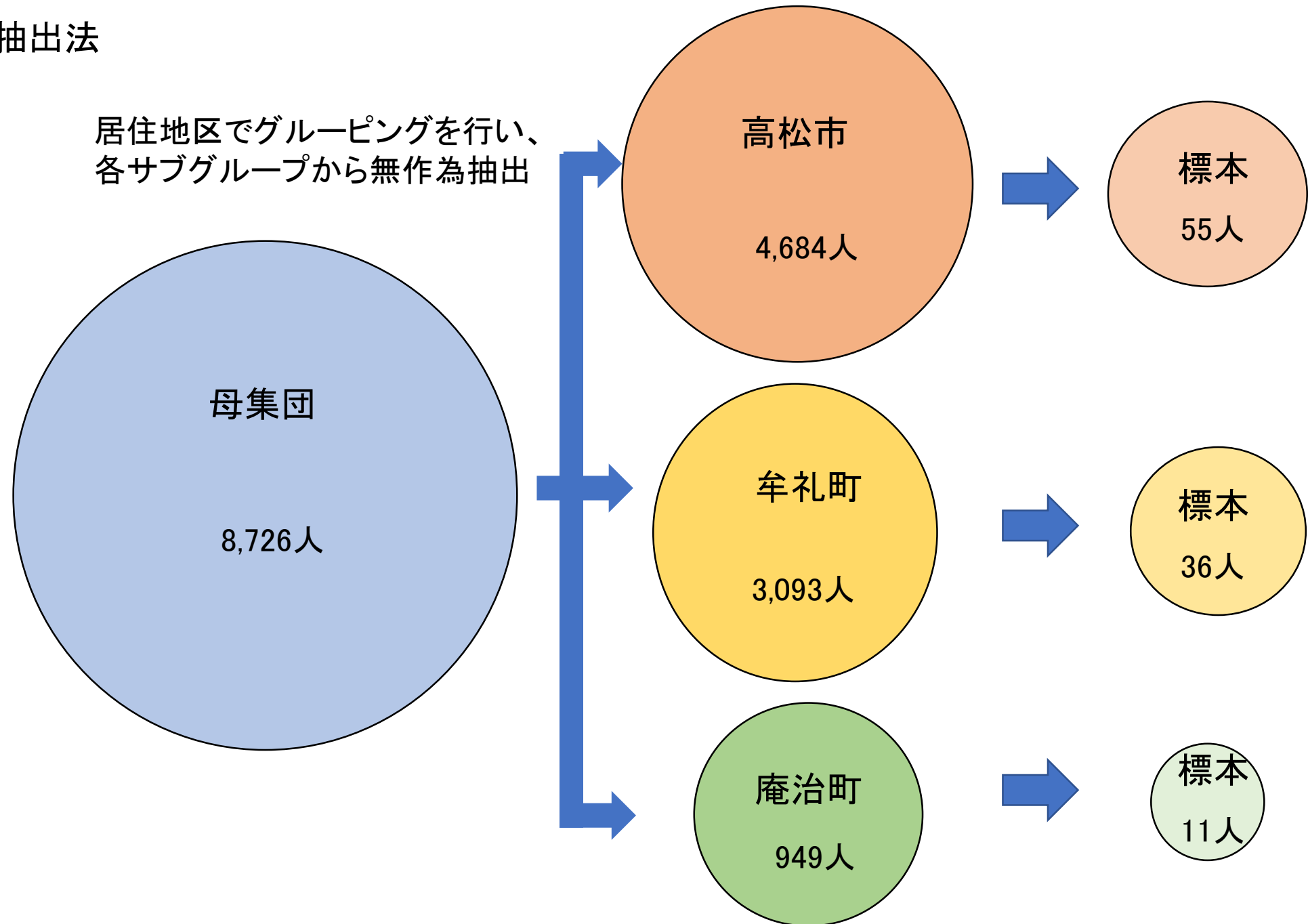
出所：猪瀬直樹(2007)「借金列島のつくり方、教えます」週刊文春H19.3.29号連載より一部抜粋。

標本(調査対象)の基本は無作為抽出(random sampling)

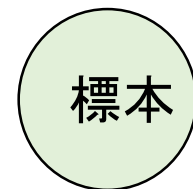
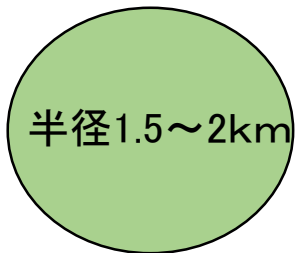
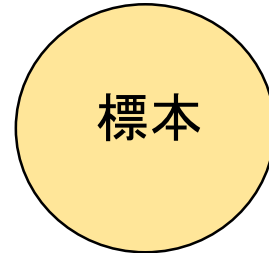
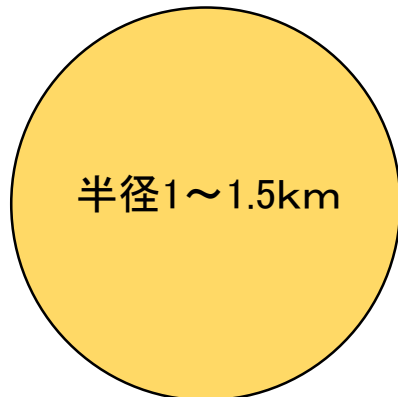
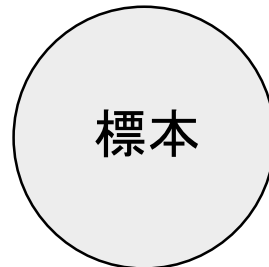
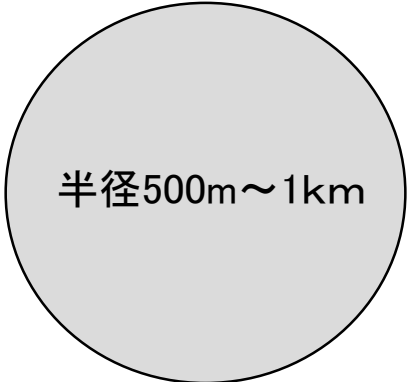
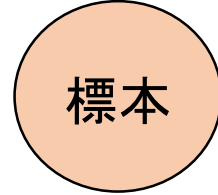
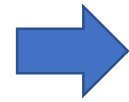
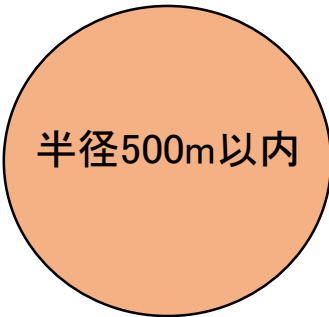
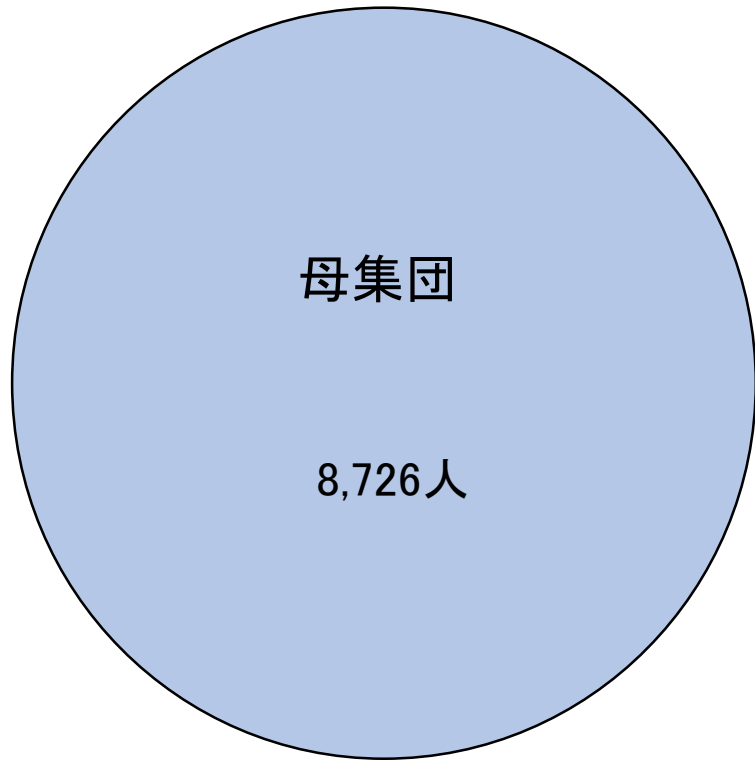
① 単純無作為抽出法



② 層化抽出法



緑地帯からの距離でグルーピングを行い、
各サブグループから無作為抽出



必要な標本数

- ①母集団の規模、②信頼係数、
③標本誤差で決まる

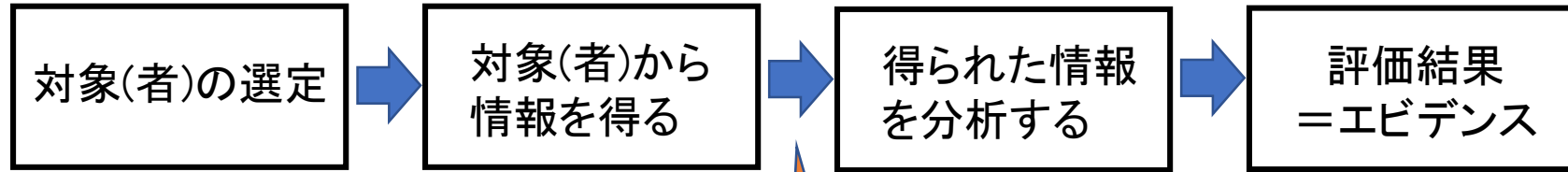
本事例の母集団8,726の場合、
標本数 102では標本誤差は±10%。
少なくとも369以上の標本数が望ましい。

仮に標本調査で25%となった場合、
母集団における比率は
95%の信頼係数(信頼度)で、
標本誤差±10%では15~35%
標本誤差± 5%では20~30%
標本誤差± 3%では22~28%

標本数の早見表(信頼係数95%)

母集団の規模	標本誤差					
	±1%	±2%	±3%	±4%	±5%	±10%
1,000	906	706	516	375	278	88
2,000	1,655	1,091	696	462	322	92
3,000	2,286	1,334	787	500	341	94
4,000	2,824	1,500	842	522	350	94
5,000	3,288	1,622	879	536	357	95
8,726	4,573	1,884	951	562	369	96
10,000	4,899	1,936	964	566	370	96
20,000	6,489	2,144	1,013	583	377	96
50,000	8,057	2,291	1,045	593	381	96
100,000	8,763	2,345	1,056	597	383	96
500,000	9,423	2,390	1,065	600	384	97
1,000,000	9,513	2,395	1,066	600	384	97
10,000,000	9,595	2,401	1,067	600	384	97
100,000,000	9,604	2,401	1,068	600	384	97

3-2 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

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

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)

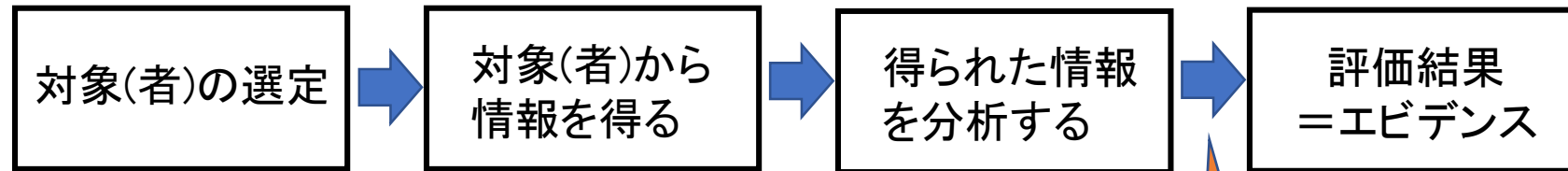
前の質問に対する回答が後の質問に対する回答に影響を及ぼす場合がある

＝キャリーオーバー効果 (carry-over effect)

都政モニターアンケート「防災対策について」(平成13年12月11日実施)

- ① ここ数年、都内の犯罪は増加傾向にあります。あなたはこうした状況を知っていましたか。
- ② あなたは、最近の東京の防犯や治安の面について不安を感じていますか。
- ③ 各警察署には、犯罪容疑者をとどめておくための留置場が併設されています。一方、国の施設として被告人等を収容する拘置所や受刑者を収容する刑務所がありますが、あなたは、留置場、拘置所、刑務所のそれぞれの違いについて知っていましたか。
- ④ 犯罪の増加等により、留置場など都内にあるこうした施設の収容定員の絶対数が不足しています。あなたはこうした状況を知っていましたか。
- ⑤ 各警察署には留置場が併設されていますが、「今後、警察署の建て替えを行う際には、留置定員を増やすため、留置場を拡充する必要がある」という意見があります。こうした意見に対し、あなたはどのように考えますか。

3-3 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

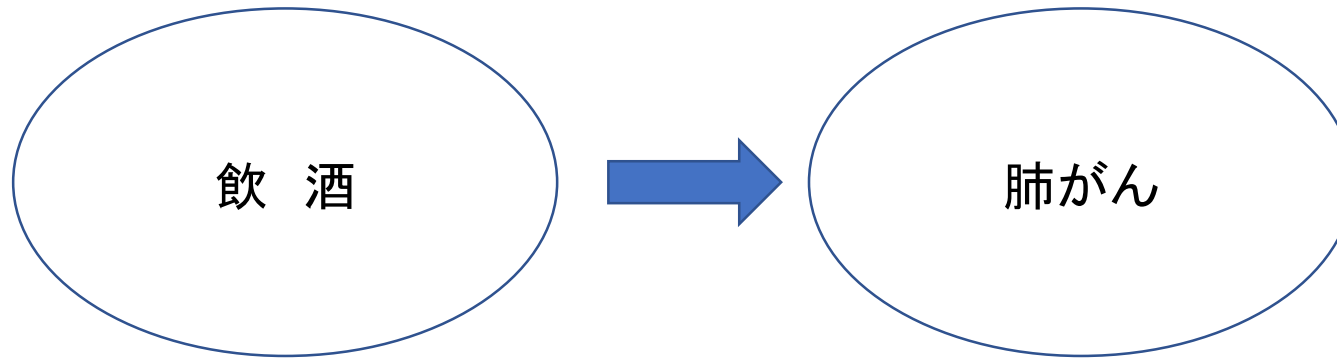
区 分		飲 酒		合 計
		あ り	な し	
肺 が ん	あ り	33	27	60
	な し	1,667	2,273	3,940
合 計		1,700	2,300	4,000

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

\Rightarrow 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

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

飲酒者と非飲酒者の
肺がん発症確率は等しい。

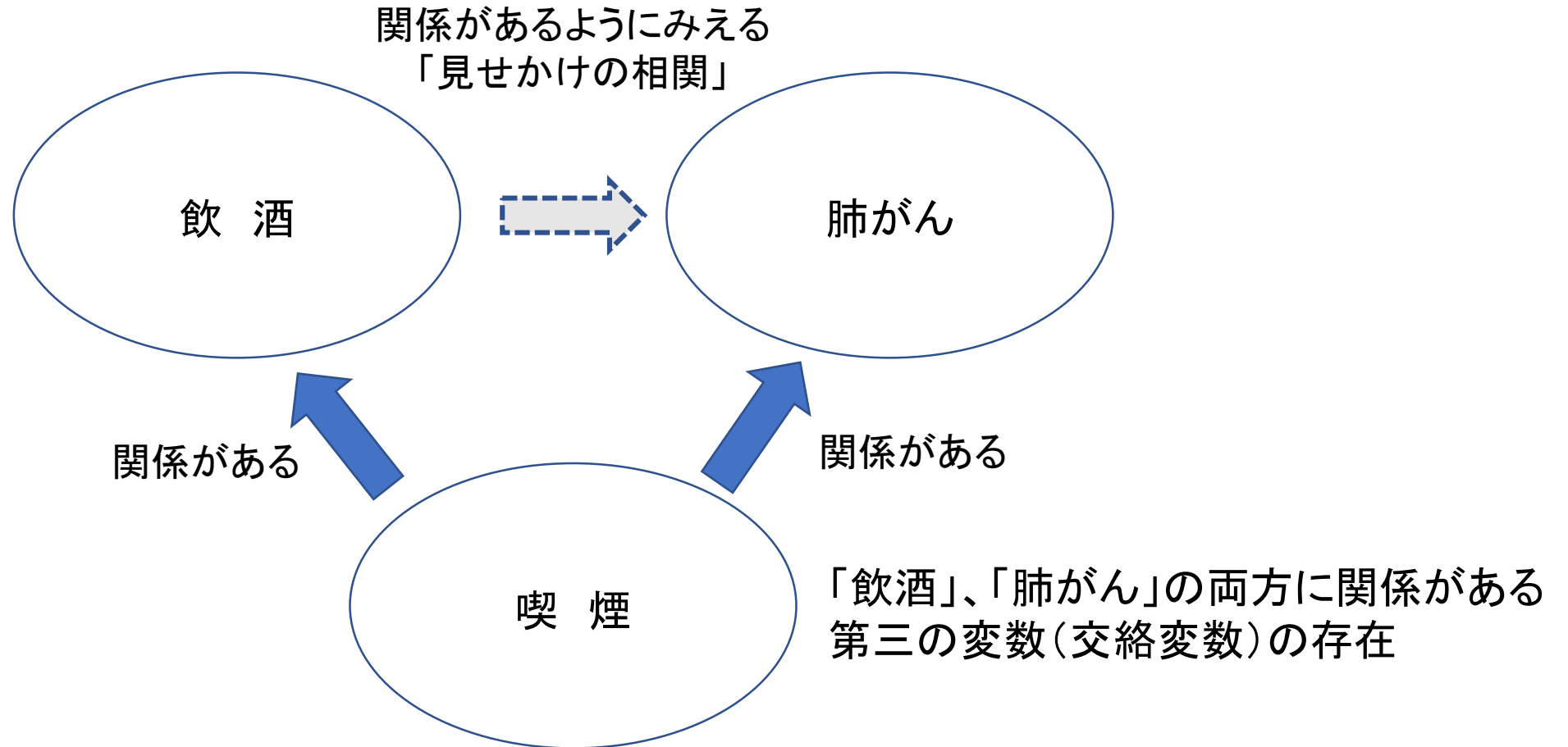
飲酒あり:発症オッズ= 9/ 891

飲酒なし:発症オッズ=21/2079

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

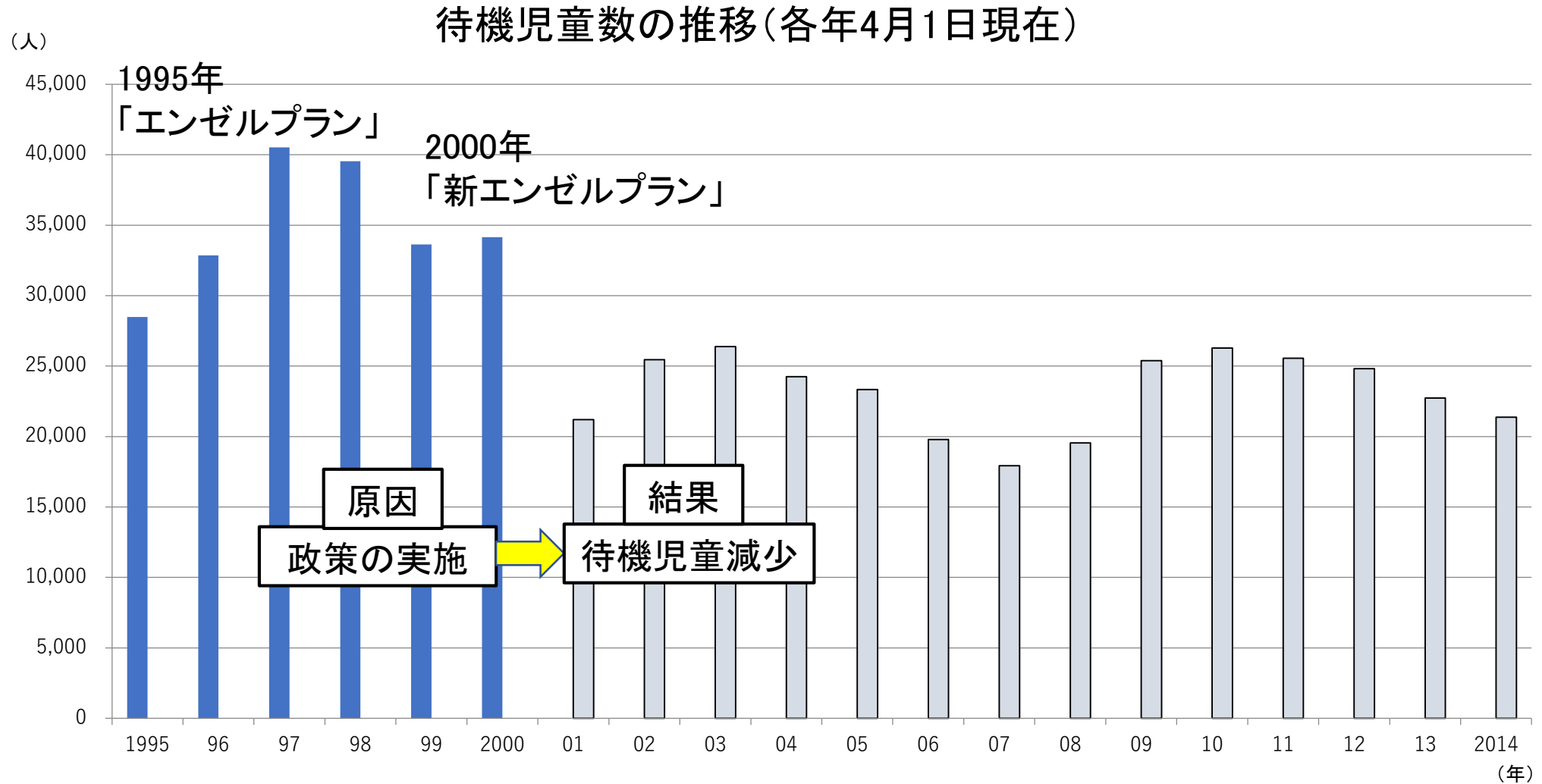
飲酒と肺がん発症の間には関係はない。関係があるのは喫煙習慣と肺がん発症の間。

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



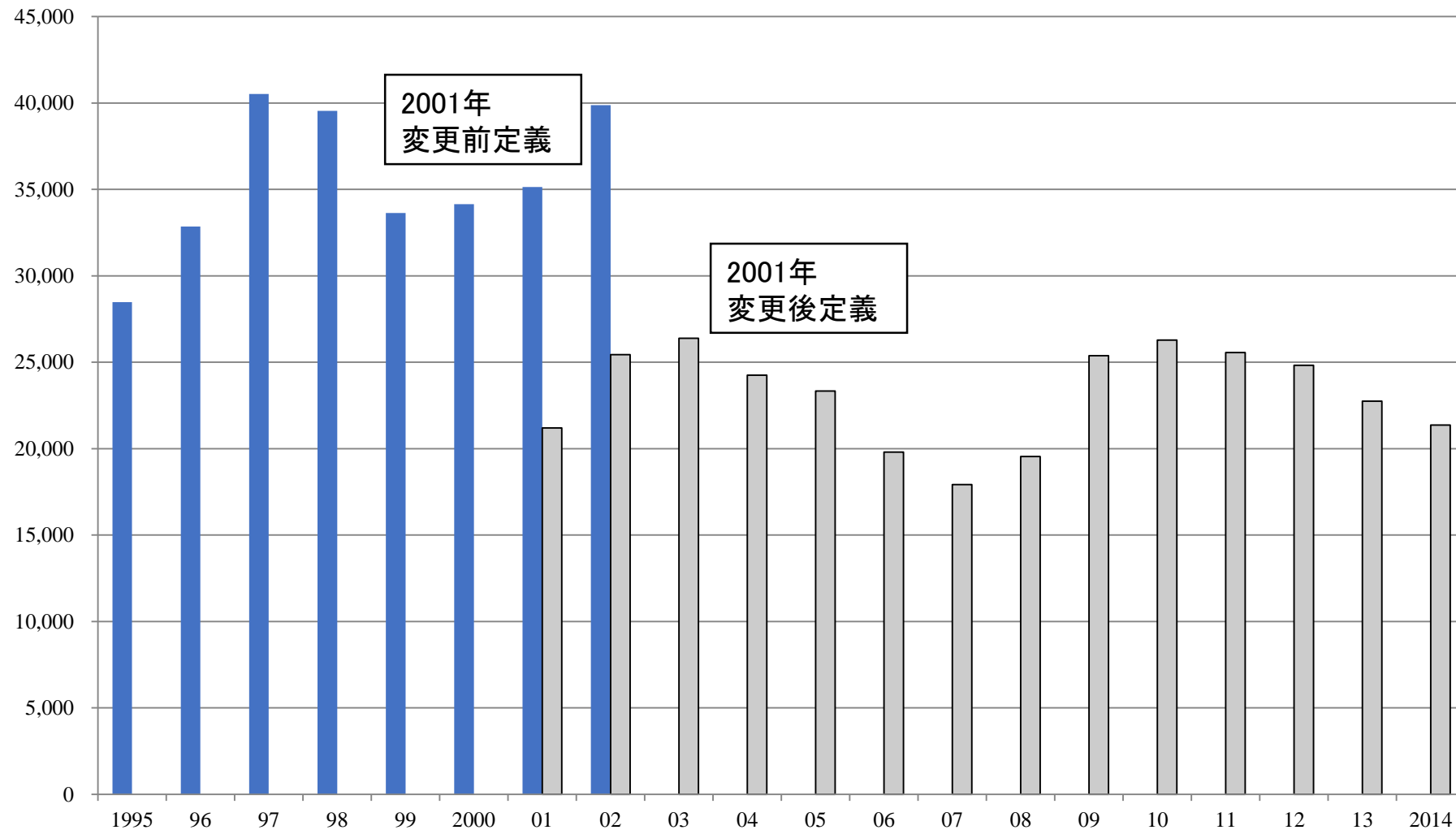
分析が不十分であることによって発生

① 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？



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。

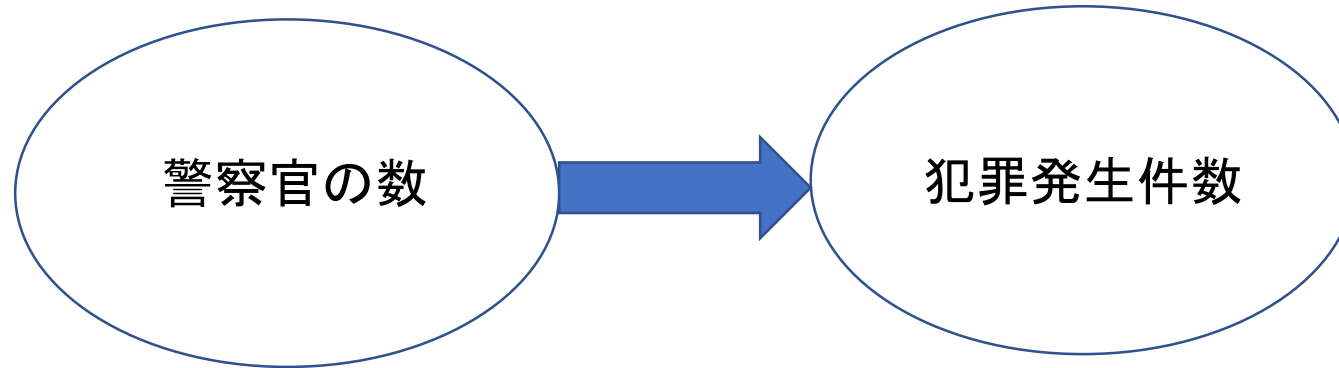
待機児童数の推移(各年4月1日現在、2001年定義変更前後併記)



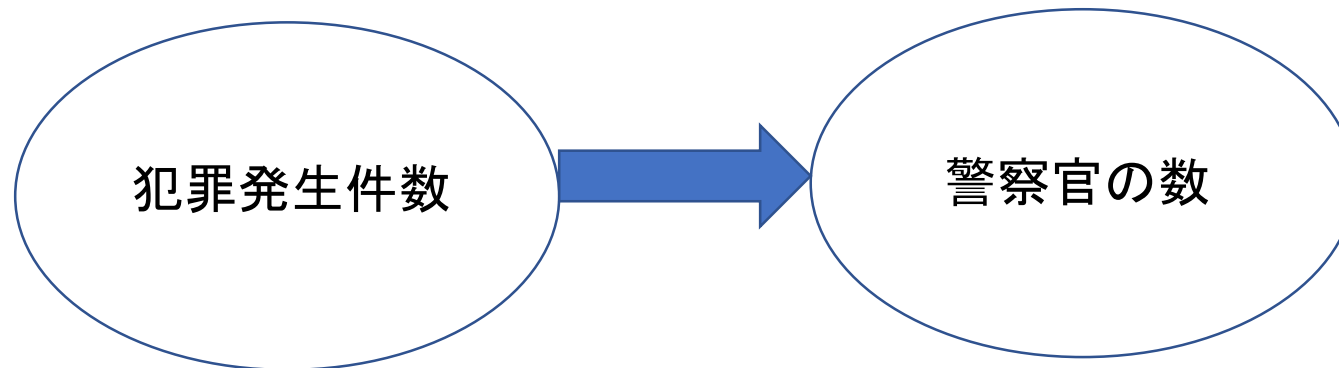
資料:厚生労働省「保育所関連状況取りまとめ」等をもとに作成。

因果関係の逆転かもしれない

「お巡りさんが多い地域ほど犯罪が多い」 = 因果関係？

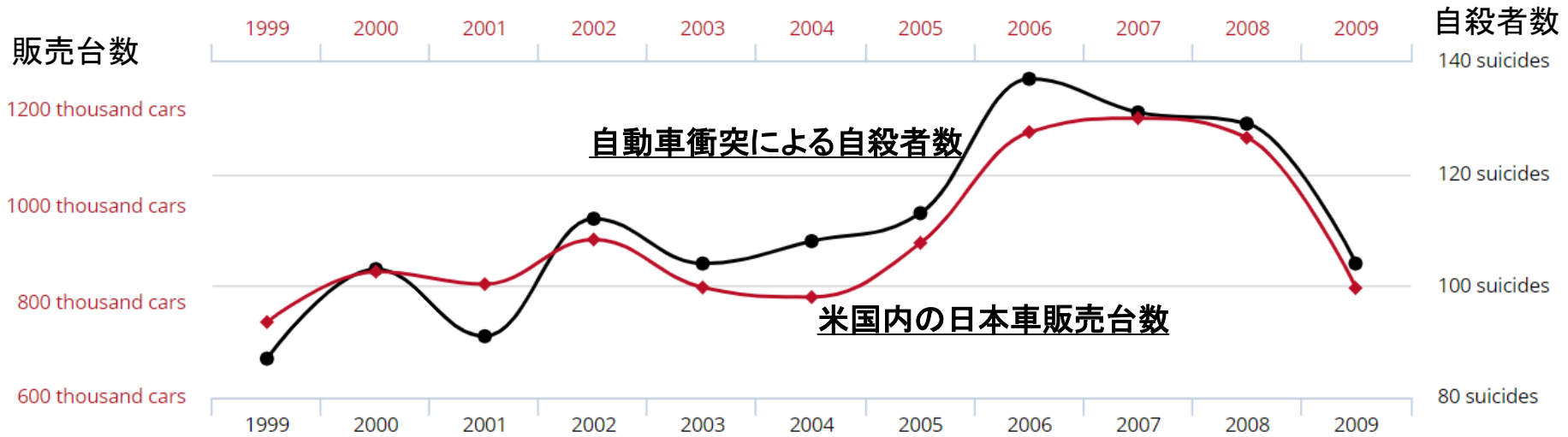


「犯罪が多いからお巡りさんが多く配置されている」 = 因果関係

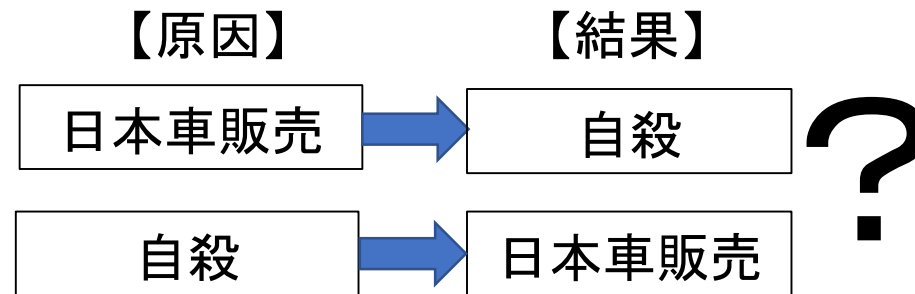


偶然の一致の場合もある

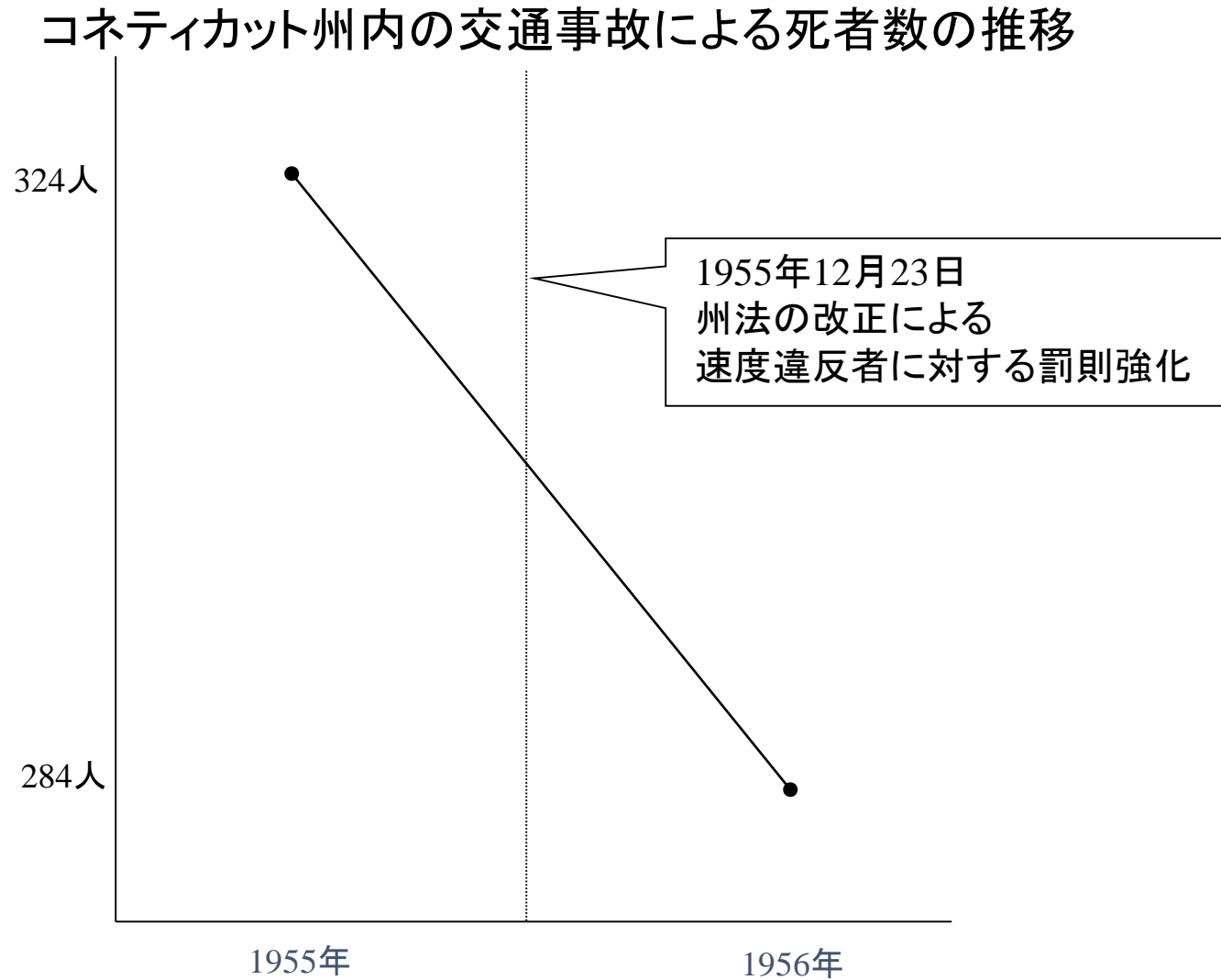
米国内の日本車販売台数と自動車衝突による自殺者数の関係



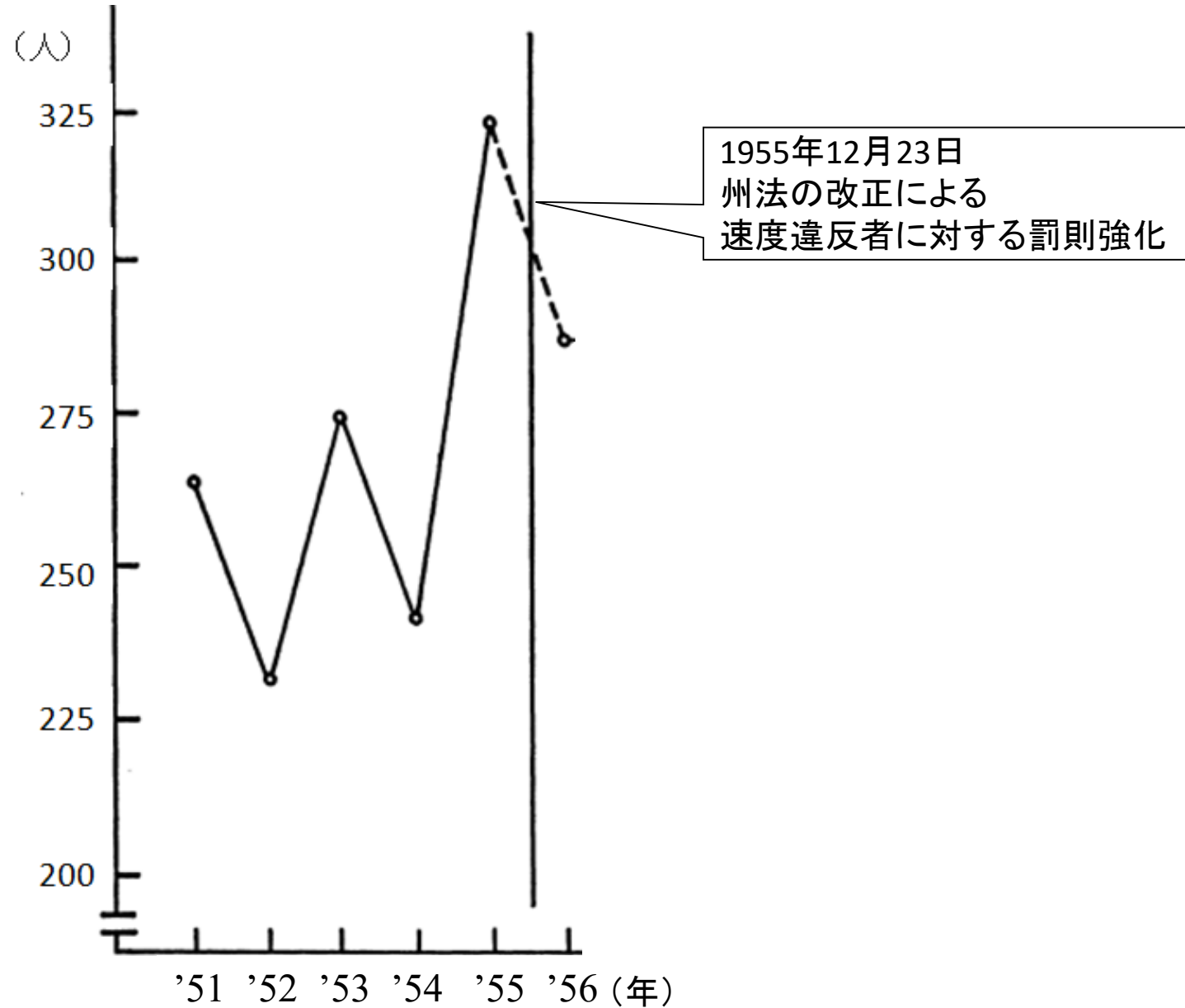
出所：Spurious correlations(<http://tylervigen.com/spurious-correlations>)



② ビフォーアフター（Before-After）の落とし穴

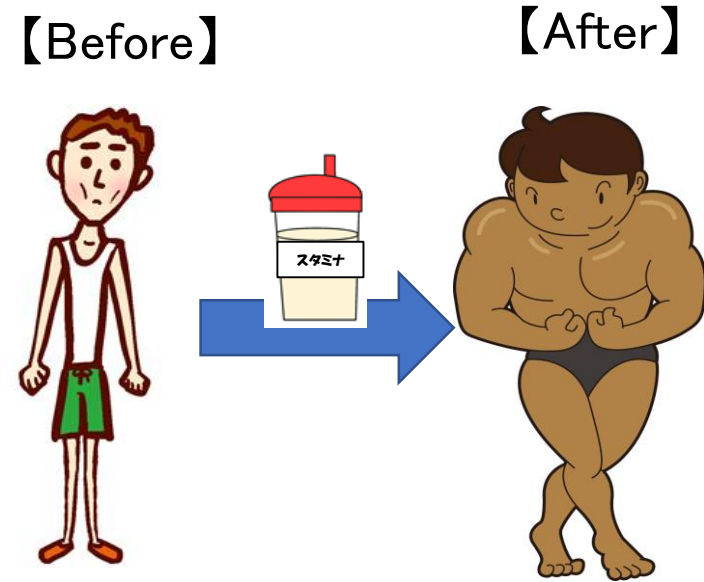
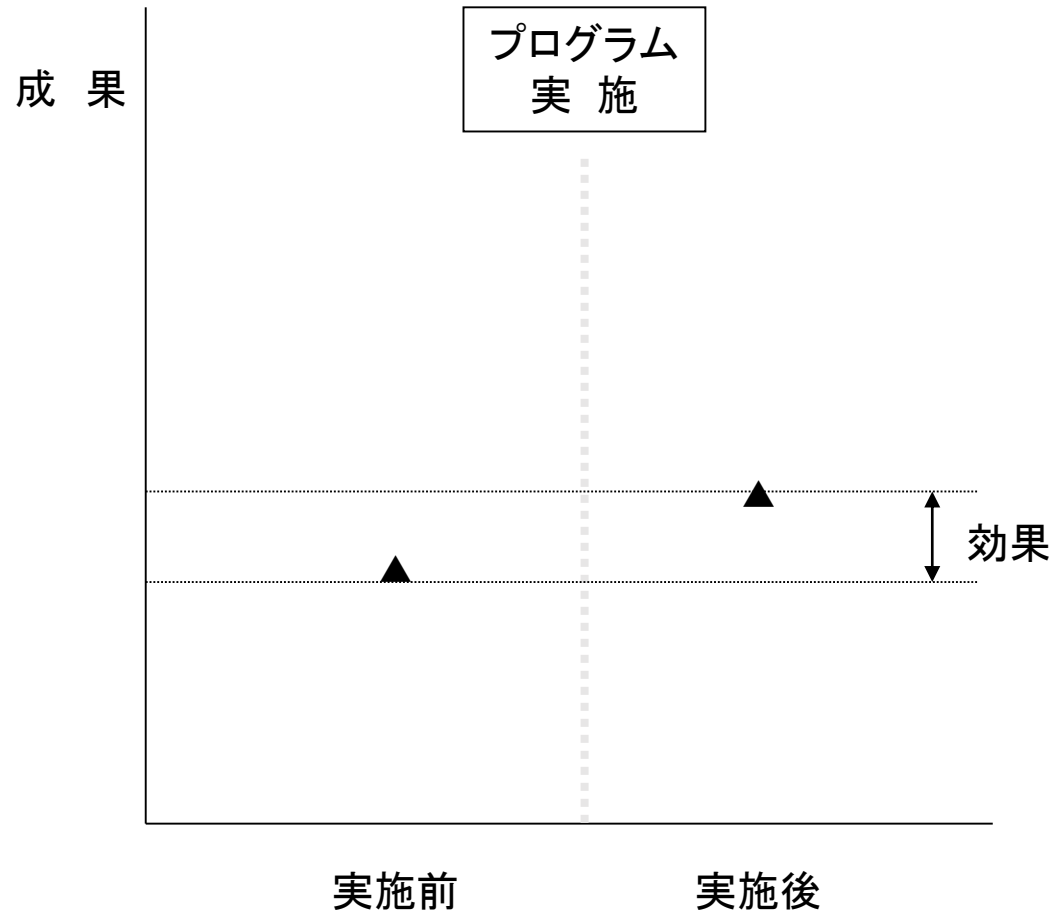


コネティカット州内の交通事故による死者数の推移(1951~56年)

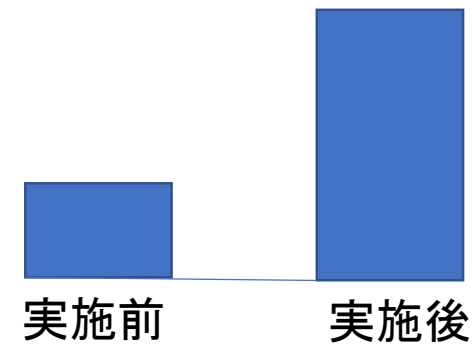


出所: Campbell and Ross, p.42を一部修正。

1事例事前・事後比較デザイン (One-group Pretest-Posttest Design, One-case Before-After Design)



経済対策実施前後の
経済成長率比較



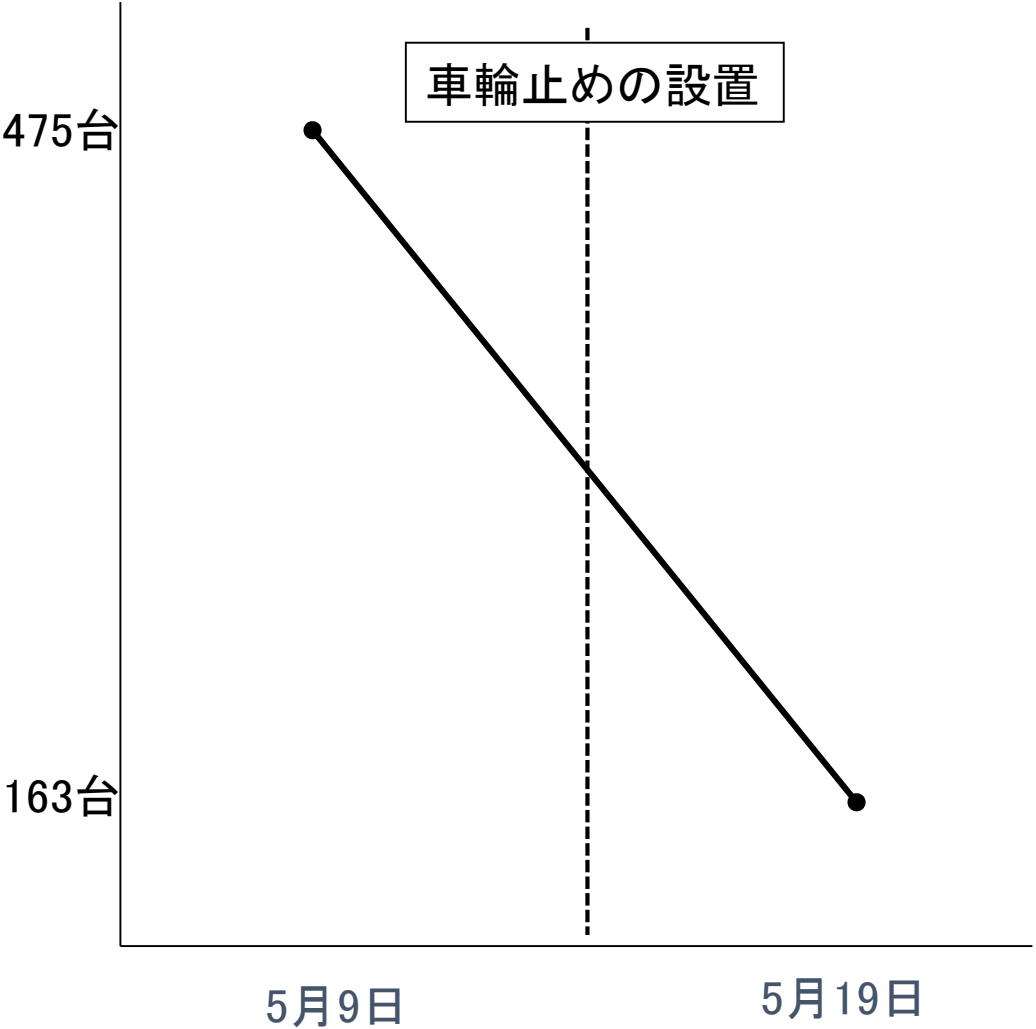
車輪止めの設置前後における駐車違反台数の変化

「駐車違反 6 割減 / 『車輪止め』 にまいった」

〇〇県警は、改正道交法の施行で5月10日から使用を始めた車輪止め装置「クランプ」による駐車違反取り締まり結果をまとめた。県警は施行前の9日と19日の午後3-4時に指定路線の〇〇通りでの駐車違反台数を調べた結果、9日の475台が19日には163台になり、65.6%も減少していた。

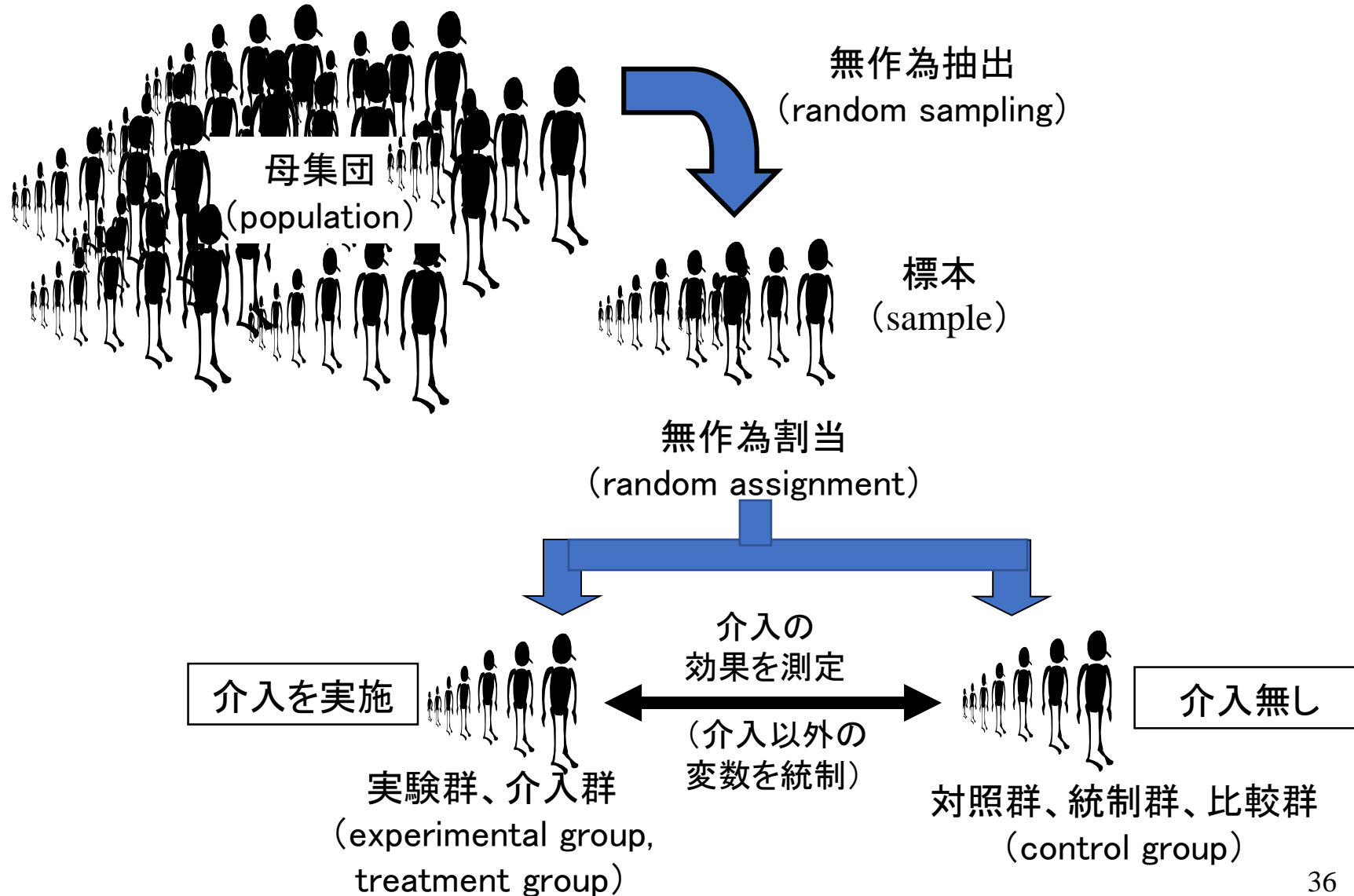
取り締まりにあった市民は「これを取り付けられるととても恥ずかしい」「ここまでされたら逃げられん」などと話していたという。県警では「〇〇通りだけでなく、徹底した取り締まりを展開したい」としている。

出所：読売新聞1994.5.24記事を参考に作成



③ どのようなモデルが望ましいのか

1) ランダム化比較試験 (Randomized Controlled Trial, 略称RCT) — 最も頑健なモデル



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11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

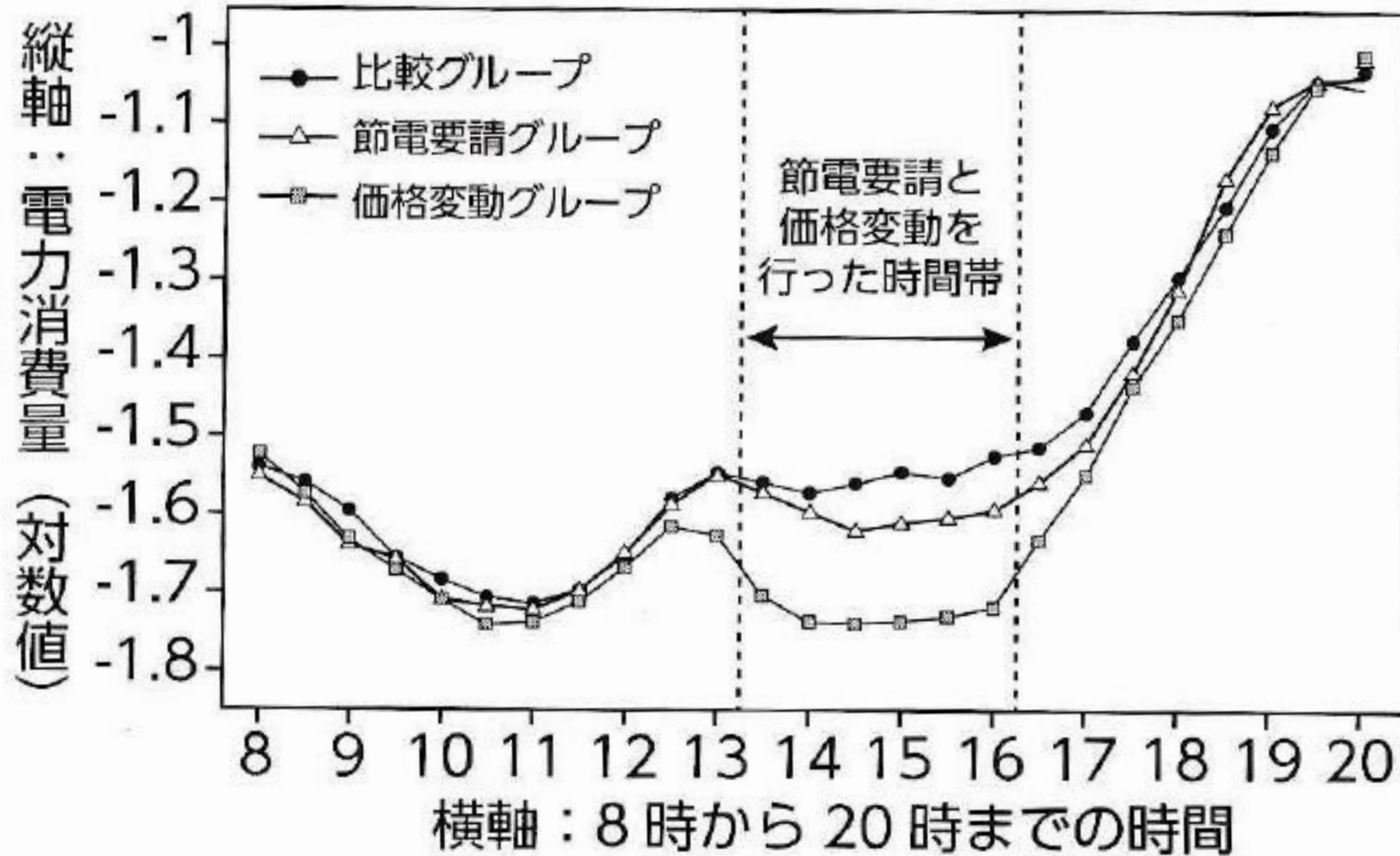
新日鐵住金の特定供給エリアで実証。コジェネをベースロード電源と見立て、需要家180戸において、需給状況に応じて電力料金を変動させるダイナミックプライシングの実証を実施。（富士電機・新日鐵住金）

豊田市

戸別住宅型

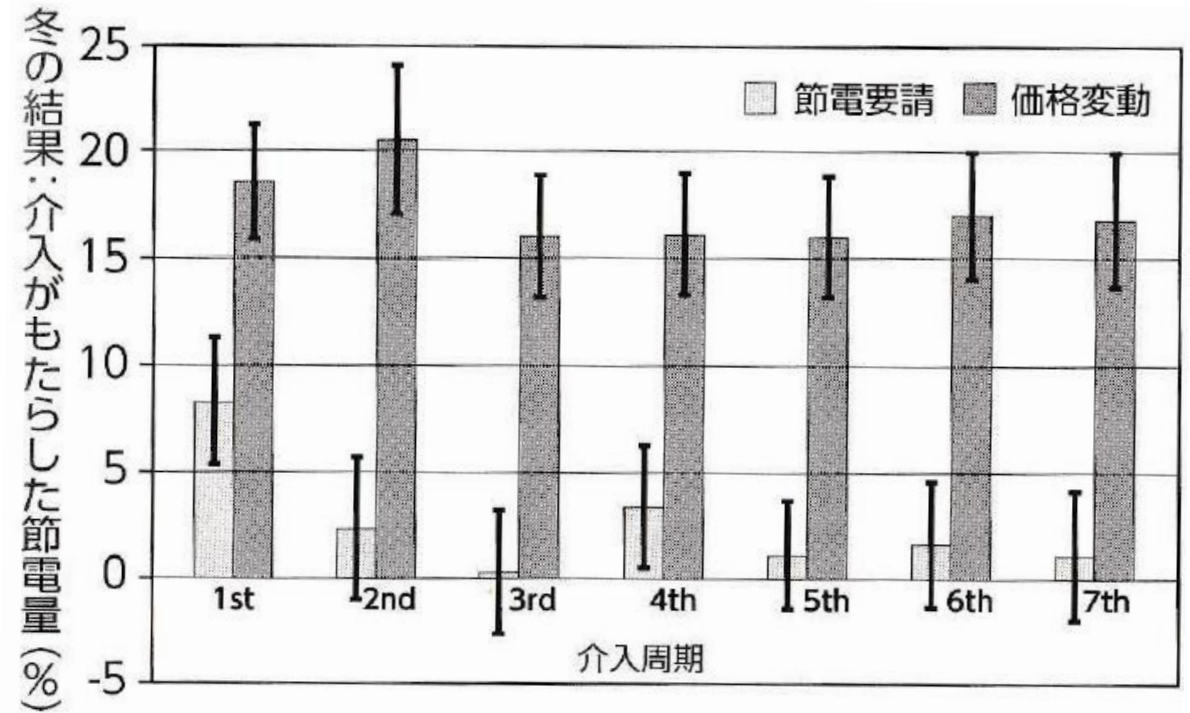
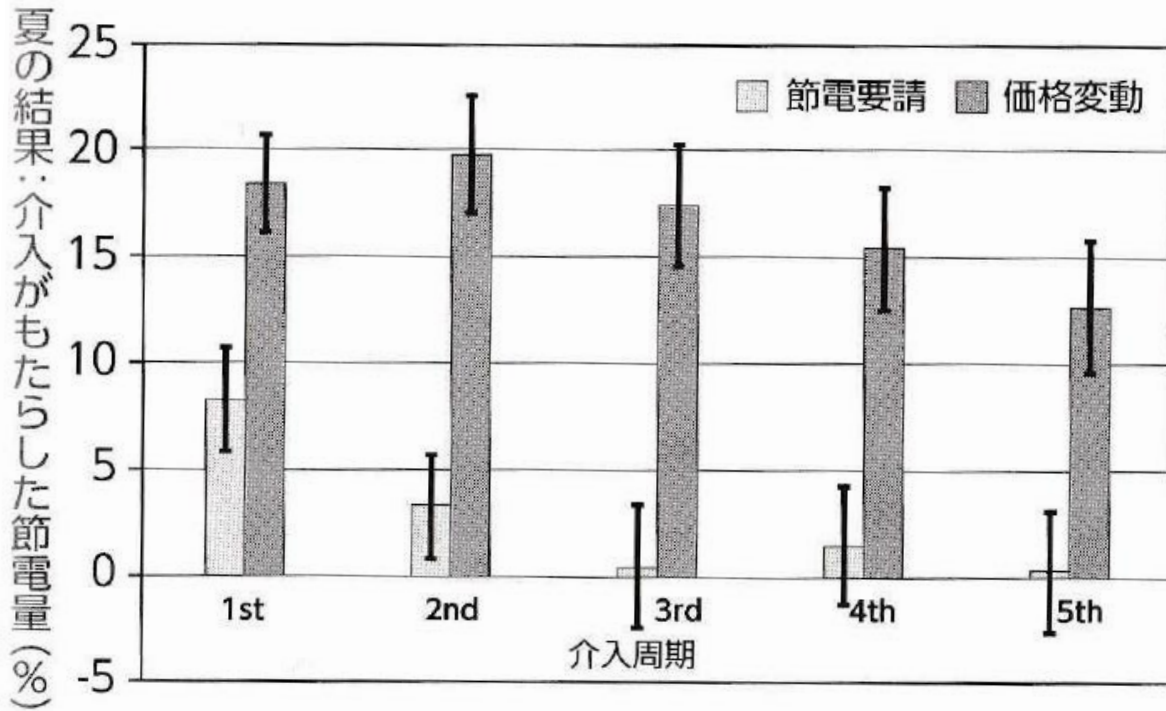
創エネ、蓄エネ機器を導入した67戸の新築住宅を中心とし、地産地消を行う実証を実施。また、暮らしの中における次世代自動車を含む次世代交通システムを実証。（トヨタ自動車・中部電力）

京都けいはんな地区で行われたRCTの結果



出所：伊藤公一朗(2017)『データ分析のカー因果関係に迫る思考法』光文社新書、p.104。

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)
- ◆ 失敗した場合のコストが膨大
- ◆ 無作為割当に伴う倫理的問題
- ◆ あくまで理想的な条件下での話であり、どこまで現実に適用可能か？



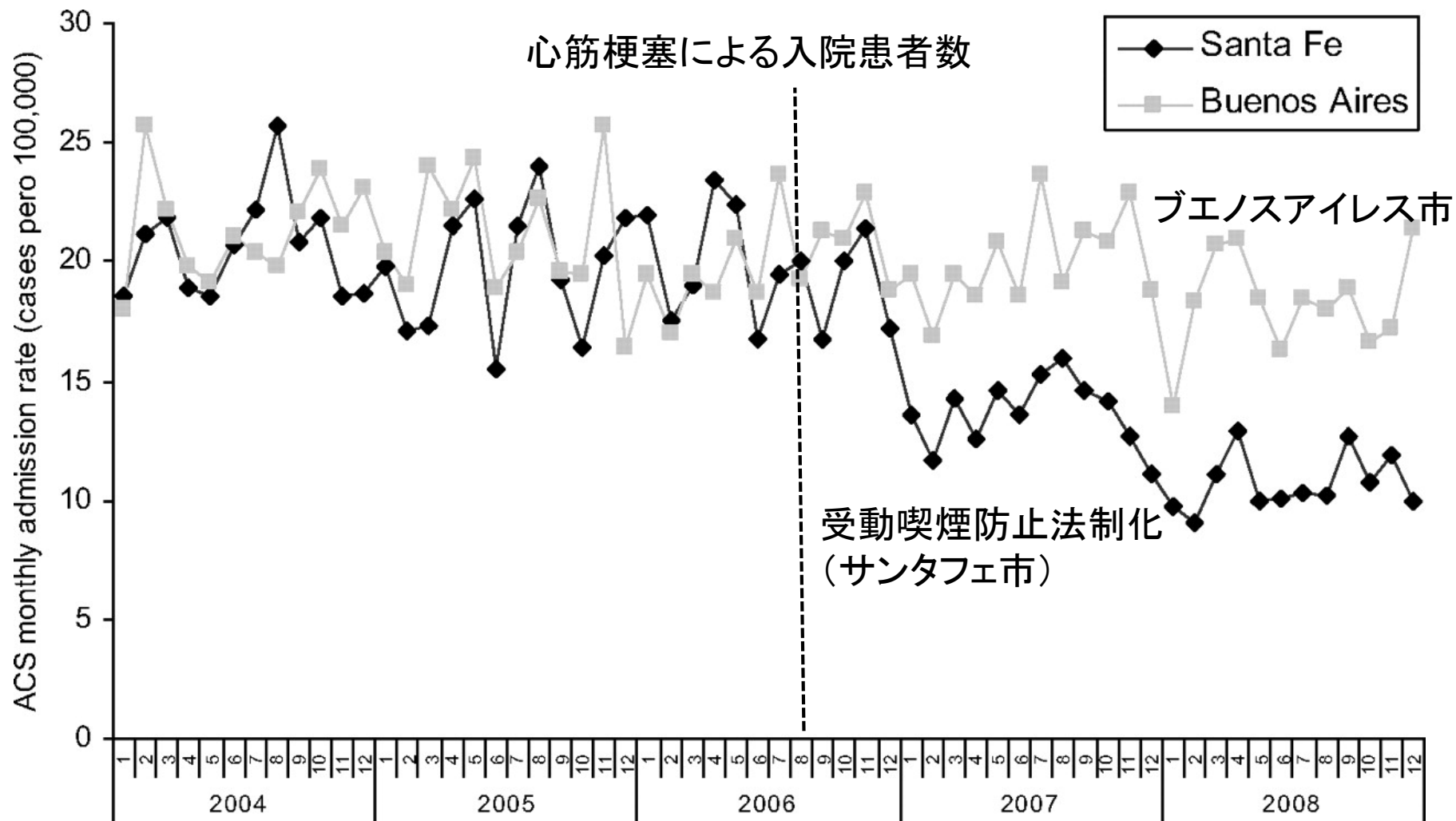
人為的に介入群と対照群を作り出すモデル＝「介入研究」は理想的なモデルだが大変！



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」

2) 自然実験(natural experiment)

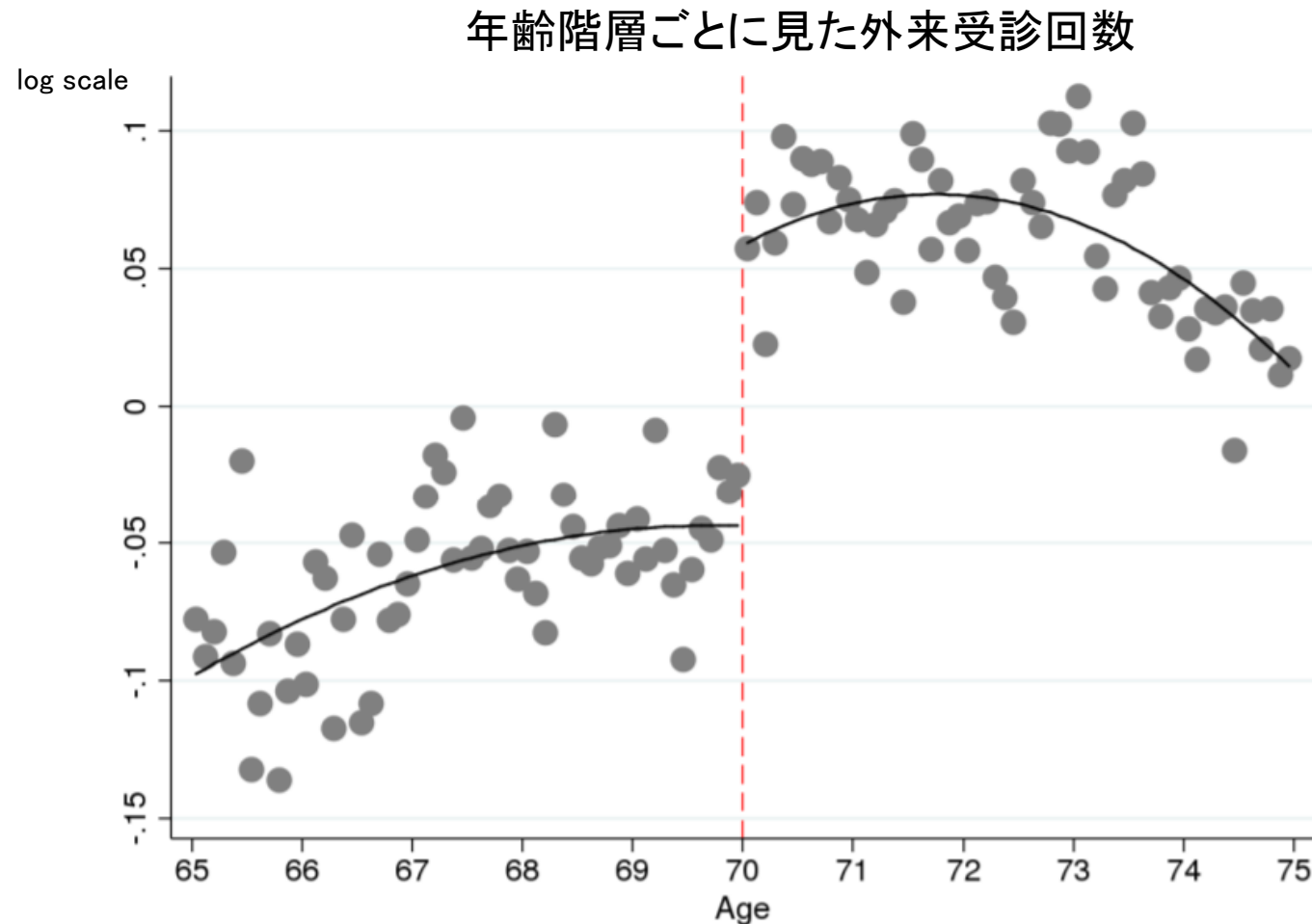
政策の変更のような制度の見直しや自然災害の発生などの外生的要因によって、自然に実験群と対照群に分かれてしまった状況を利用するモデル



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

3) 回帰分断法 (regression-discontinuity design)

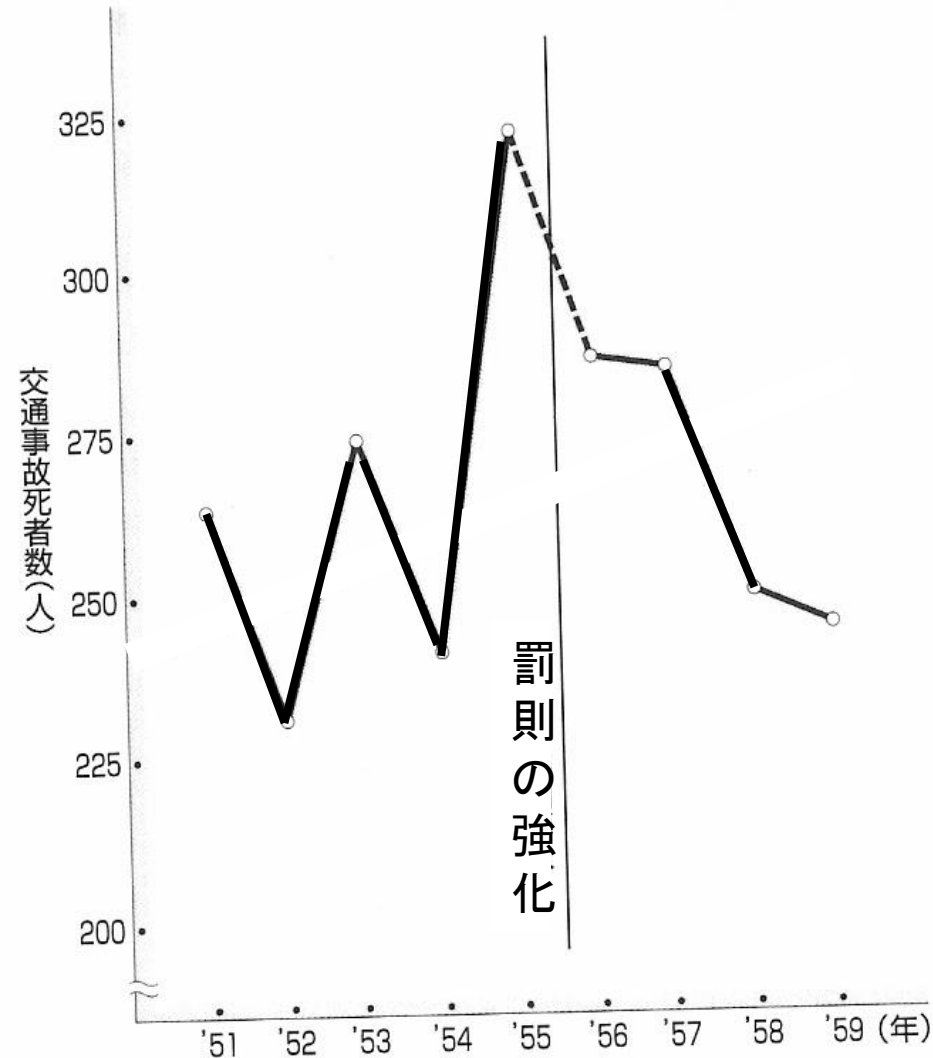
実験群と対照群への割当を無作為割当ではなく、一定の基準値 (cut-off point) を設け、この基準値を境にそれぞれの群に割当を行う方法



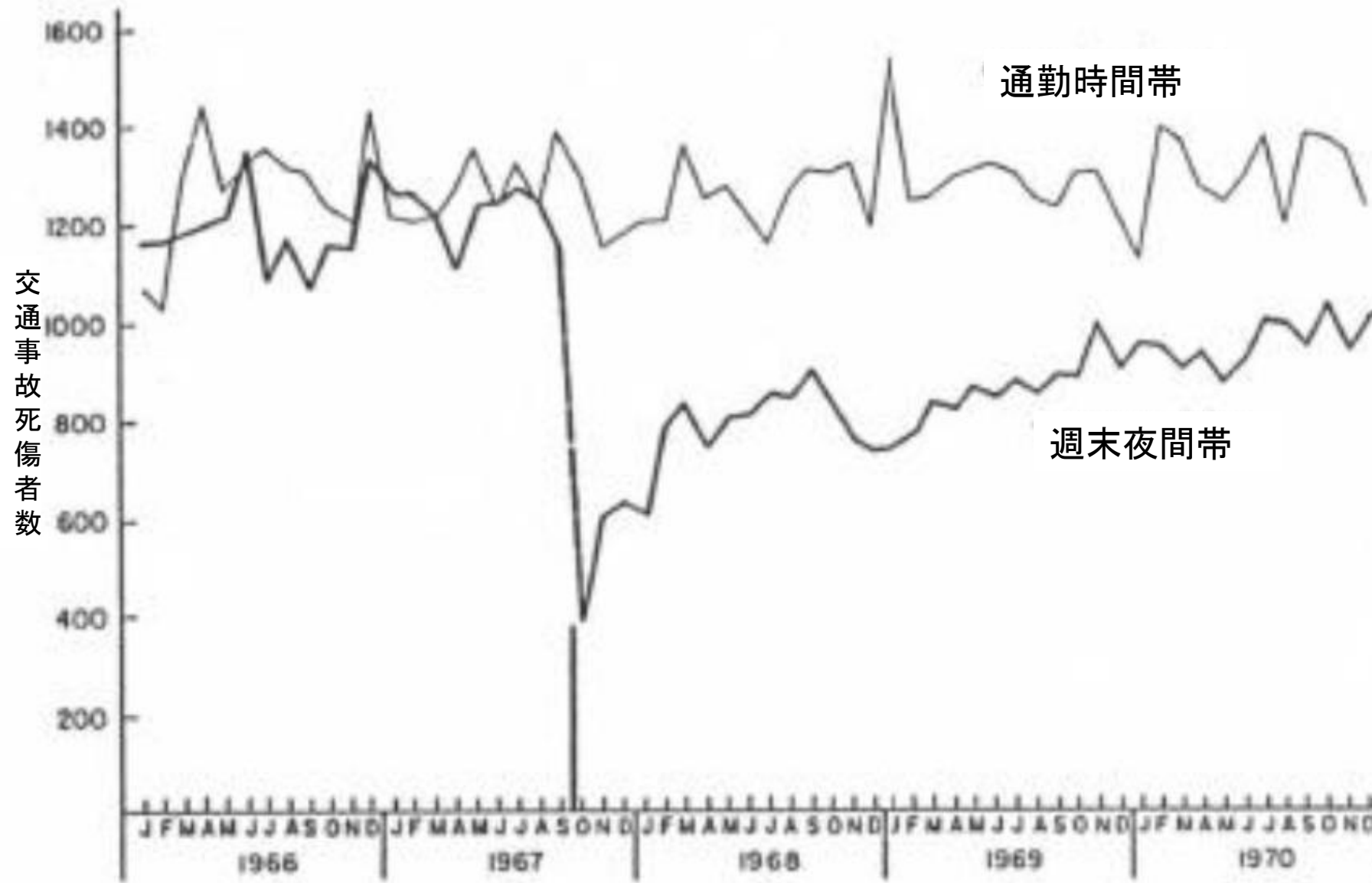
4) 断続時系列法 (Interrupted Time Series Design)

実験群と対照群への割当のための変数に時間を用いる回帰分断法

コネティカット州における交通事故死者数の推移



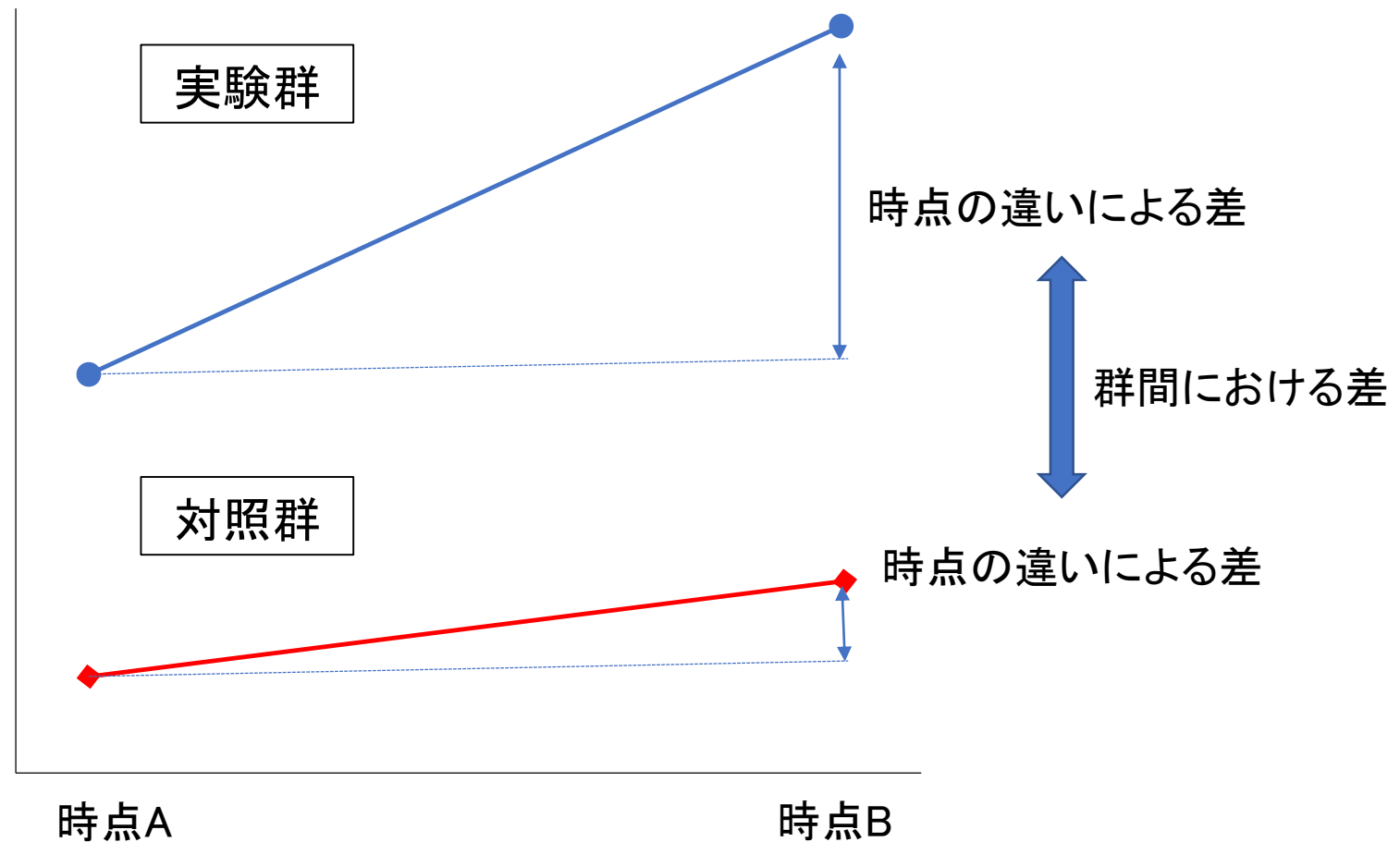
飲酒運転に対する取締り強化と交通事故死傷者数の状況



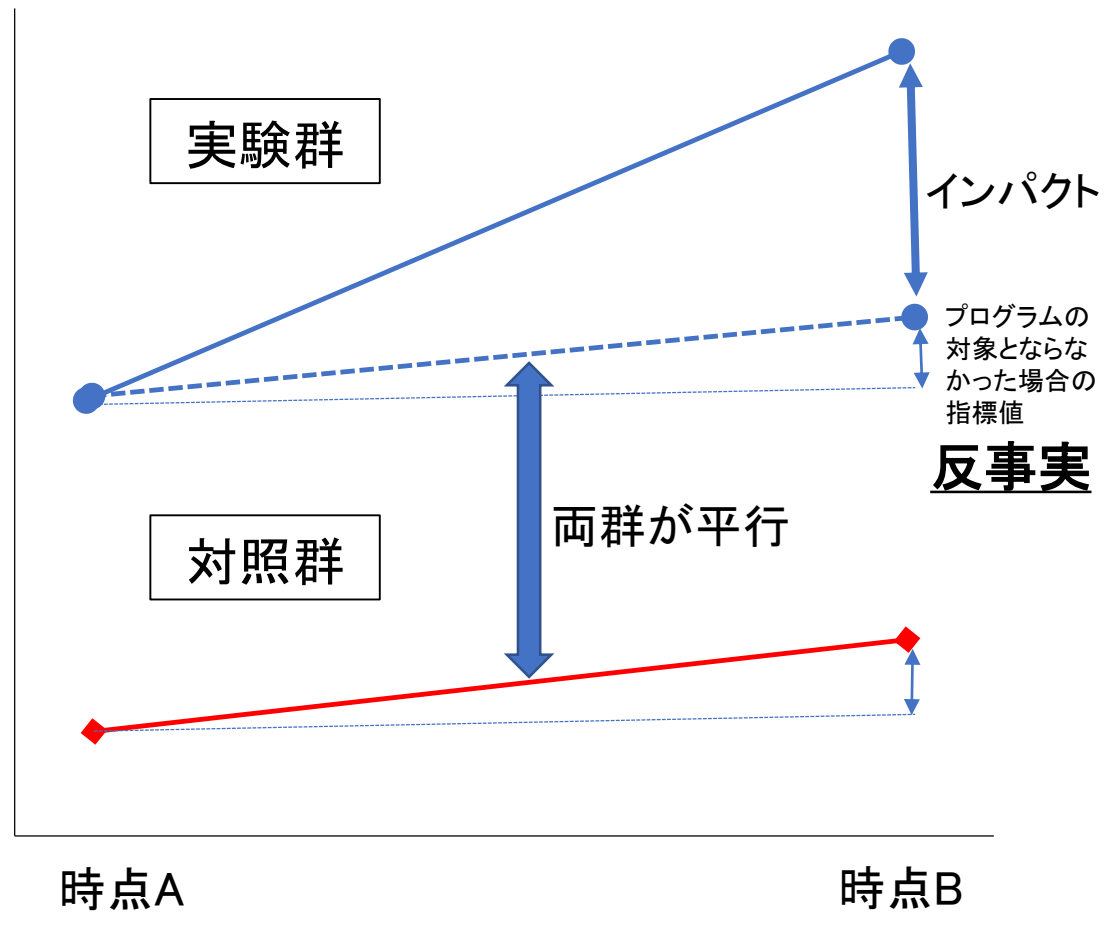
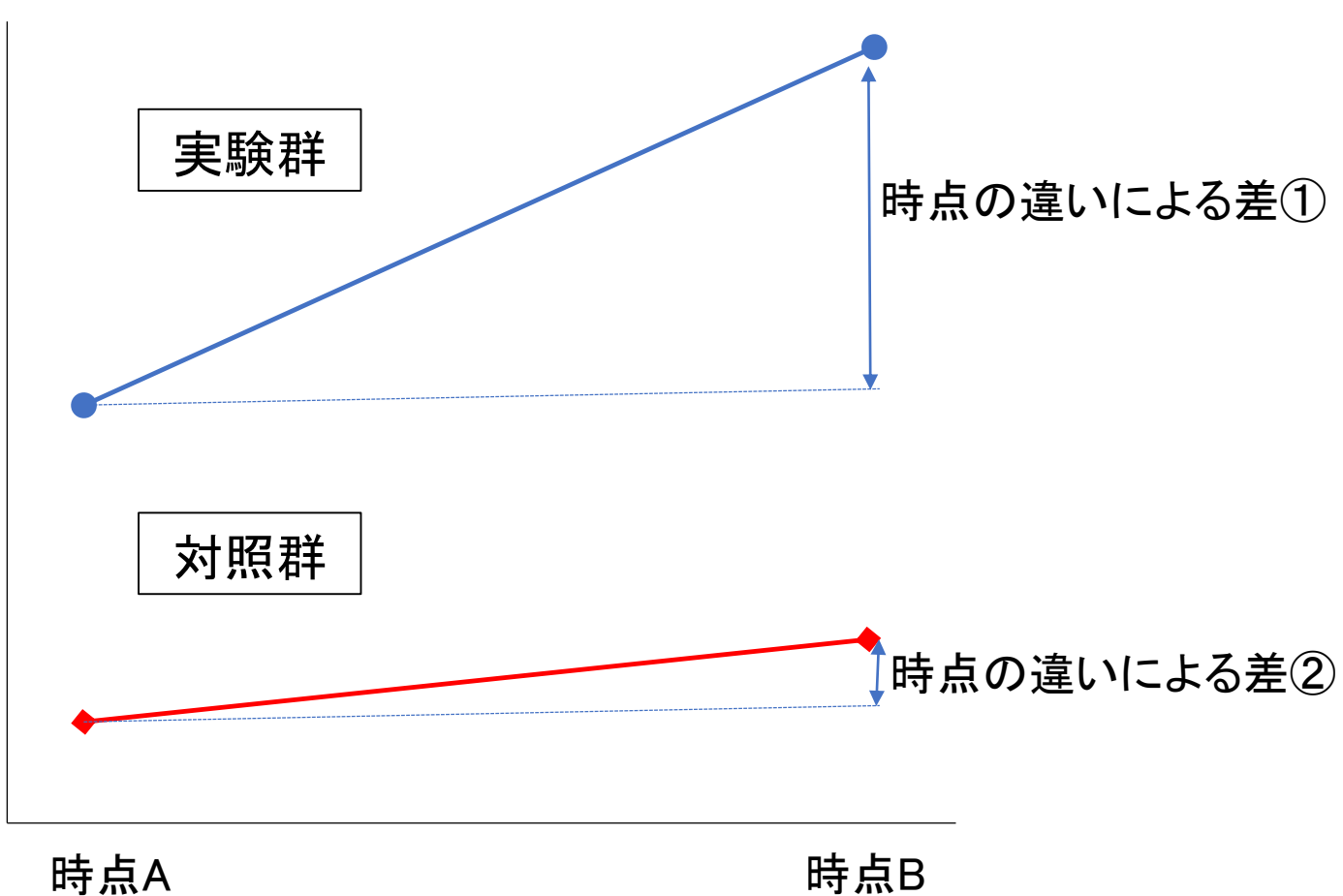
Source: Donald T. Campbell and Beatrice J. Krauss (2012) Speculations on Quasi-Experimental Design in HIV/AIDS Prevention Research, Journal of Methods and Measurement in the Social Sciences, Vol.3, No. 1, 56.

5) 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、インパクトを検証するモデル



実験群の差分①－対照群の差分②＝インパクト

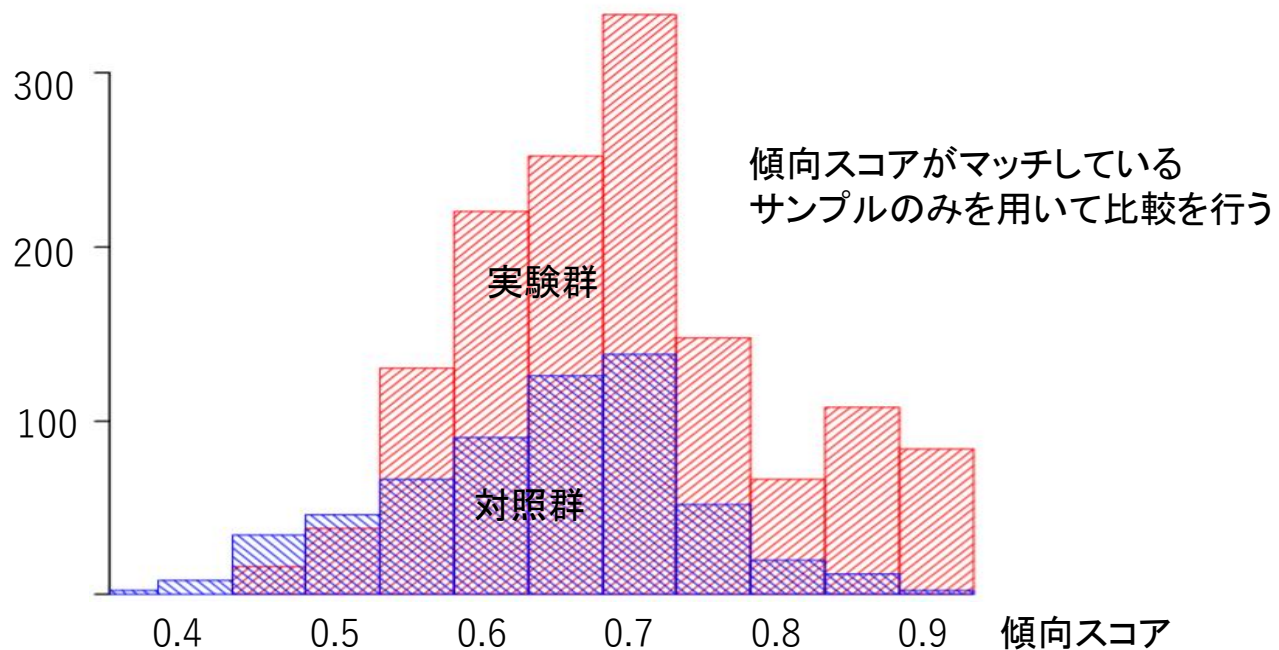


6) 操作変数法

見せかけの相関(疑似相関)や因果関係の逆転など内生性の問題が予想される場合に、操作変数を用いることで、因果関係を検証する手法

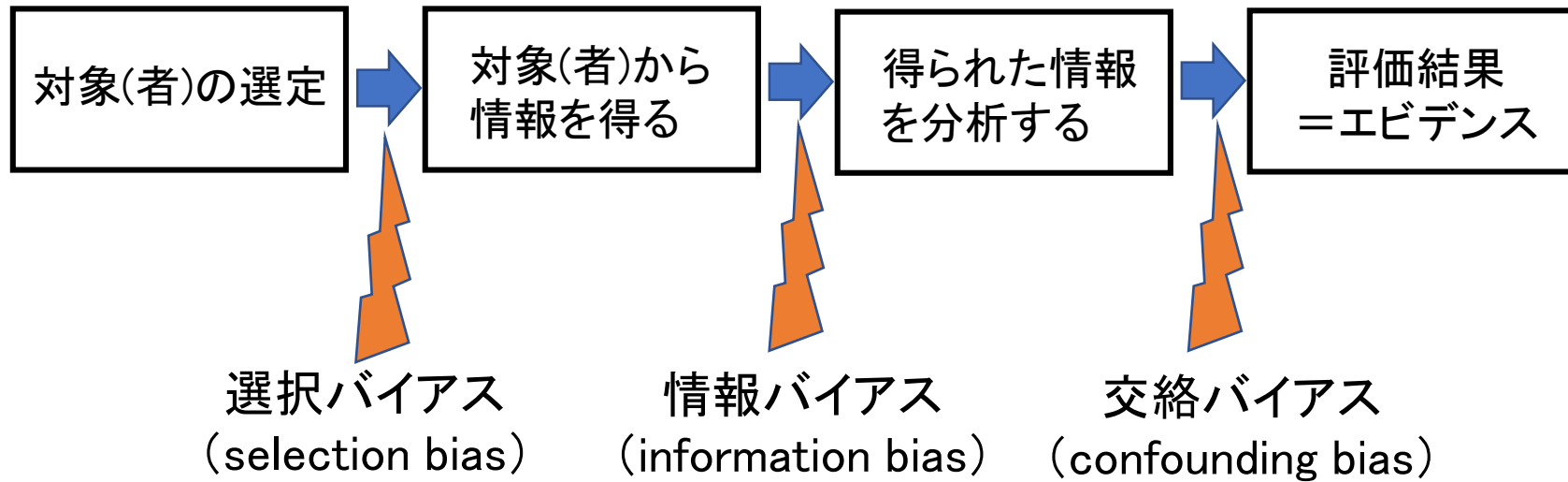
7) マッチング法(傾向スコアマッチング)

傾向スコア(実験群に入る確率)を計算し、このスコアを用いてマッチングを行う方法



4 まとめとしての再確認

政策評価、EBPM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