



RIETI Policy Discussion Paper Series 26-P-010

EBPM（エビデンスに基づく政策形成）入門 第3話 回帰不連続デザインと差の差分析

関沢 洋一
経済産業研究所



Research Institute of Economy, Trade & Industry, IAA

独立行政法人経済産業研究所

<https://www.rieti.go.jp/jp/>

EBPM(エビデンスに基づく政策形成)入門¹

第3話 回帰不連続デザインと差の差分析

関沢洋一（独立行政法人経済産業研究所）

要 旨

- ・ 実験を伴うことなく政策介入の効果を検証するための主な方法として、回帰不連続デザイン（RDD）、差の差分析（DID）がある。
- ・ RDD は、何らかの指標（ランニング変数）が特定の数値（カットオフ）を上回る時のみ介入が行われる場合（誕生日によるワクチン接種の決定、企業を対象とする補助金の採択、メタボ健診の保健指導など）に、カットオフの周辺にいる介入の対象者と非対象者を比較することでランダム化比較試験（RCT）に近い効果検証を行う。
- ・ DID は、介入が全体の一部のみ特定時点から開始される場合（一部の都道府県のみで介入が行われるなど）において、介入前後のアウトカム変数の変化が、対照群（介入を受けなかったグループ）と比べて介入群において異なったものとなっているかどうかを検証する。

キーワード：回帰不連続デザイン（RDD）、差の差分析（DID）、基準値、自然実験

JEL classification: H11 H43

RIETI ポリシー・ディスカッション・ペーパーは、RIETI の研究に関連して作成され、政策をめぐる議論にタイムリーに貢献することを目的としています。論文に述べられている見解は執筆者個人の責任で発表するものであり、所属する組織及び（独）経済産業研究所としての見解を示すものではありません。

¹本稿の原案は、経済産業研究所（RIETI）のポリシー・ディスカッション・ペーパー検討会で発表を行ったものである。

第3話 回帰不連続デザインと差の差分析

(要約)

- ・実験を伴うことなく政策介入の効果を検証するための主な方法として、回帰不連続デザイン (RDD)、差の差分析 (DID) がある。
- ・RDD は、何らかの指標 (ランニング変数) が特定の数値 (カットオフ) を上回る時にのみ介入が行われる場合 (誕生日によるワクチン接種の決定、企業を対象とする補助金の採択、メタボ健診の保健指導など) に、カットオフの周辺にいる介入の対象者と非対象者を比較することでランダム化比較試験 (RCT) に近い効果検証を行う。
- ・DID は、介入が全体の一部でのみ特定時点から開始される場合 (一部の都道府県のみで介入が行われるなど) において、介入前後のアウトカム変数の変化が、対照群 (介入を受けなかったグループ) と比べて介入群において異なったものとなっているかどうかを検証する。

この第3話では第2話に続いて政策介入の効果検証のための手法を取り上げます。

第2話で取り上げたランダム化比較試験 (RCT) は文字通りの実験です。これに対して、役所の通常業務などで自然に集まってくるデータや、定期的なアンケート調査に答えてもらって収集するデータなど、実験なしに収集したデータ (観察データ) を使っても、RCT のような実験に近い高度な効果検証を行える場合があります。その代表例が回帰不連続デザイン (RDD)、差の差分析 (DID) で、自然実験あるいは準実験と呼ばれることがあります。

第1節 回帰不連続デザイン (RDD, Regression discontinuity design)

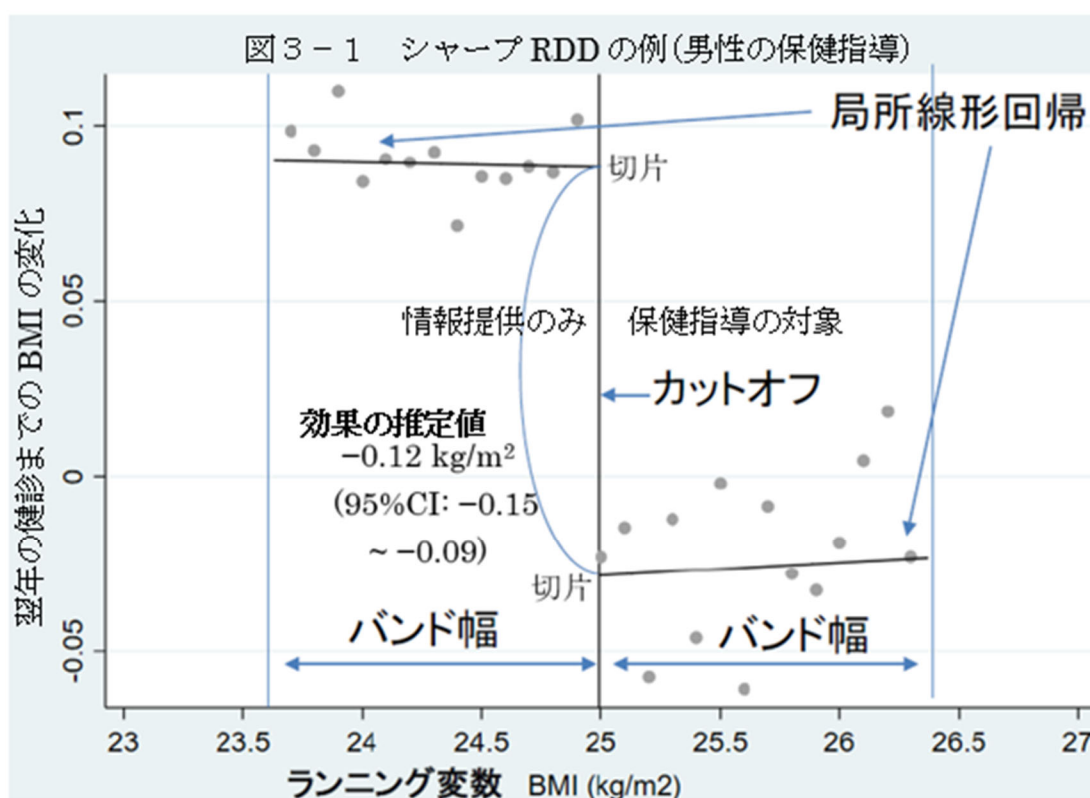
RCT は実験ですが、お金と時間がかかる上に、実験に参加することを嫌がる人々が多いために実行できない場合がしばしばあります。また、既に行われている政策について新たに RCT を行うことは難しいです。そのような場合でも、政策介入を実施するに当たって何らかの数値を基準としていれば、この基準値を利用することによって RCT に近い効果検証を行うことが可能な場合があります。その代表例が以下の RDD です。

たとえば、中小企業を対象とする補助金の多くは申請企業から提出された資料を参考にして企業ごとに得点をつけて、特定の得点に達した場合にのみ採択の対象としています。この特定の得点や数値をランニング変数と呼び、介入の有無が決まる一定の得点や数値をカットオフ (閾値) と呼びます。何らかの変数 (ランニング変数) が特定の点数 (カットオフ) を上回ると介入が行われて

カットオフを下回ると介入が行われない場合に、カットオフの近傍においては介入を受けた人々と受けなかった人々の様々な性質はほとんど違いがなく、ランダムな振り分けに近くなっています。

RDD ではこのことを利用して、介入の効果検証を行います。RDD の強みは実験なしに RCT に近い分析が可能となることですが、ビジュアルに分析結果を示すことができるのも魅力的です。その一方で、カットオフ周辺の研究対象者でしか分析が行えないという限界があります。

図 3-1 で説明します。これはメタボ健診の保健指導（男性）の例で、一定の条件を満たす健診受診者に限定すると（第 5 話参照）、BMI（体重を身長の高さの 2 乗で割ったもの）をランニング変数にして、カットオフ（BMI が 25 kg/m^2 ）の右側では保健指導の対象となり、左側では情報提供のみになります。



(出所) Sekizawa (2023)に基づいて作成した。

縦軸 (Y 軸) のアウトカムは健康診断で計測された BMI の 1 年間の変化です。典型的な RDD ではカットオフからの距離 (バンド幅) をあらかじめ設定して、その範囲内でカットオフの右側と左側で、別々に単回帰分析²を行います (局所線形回帰)。2つの単回帰分析のそれぞれの切片 (カットオフと交わる

² 回帰分析については統計学や計量経済学の入門書を参照してください。

点) の差が保健指導の対象となることの効果になります。

図 3-1 の場合、効果の推定値は -0.12 kg/m^2 (95%信頼区間: $-0.15 \sim -0.09$) で 95%信頼区間の上限が 0 を下回っているため、有意に効果があったこととなります。体重に換算すると身長 170cm の男性で -0.35 kg の減少になります。

RDD にはシャープ RDD とファジー RDD という 2 つのタイプがあります。シャープ RDD は、カットオフを下回れば介入が全く行われずカットオフを上回れば介入が全ての対象者に対して行われる場合に適用されます (0%から 100%にジャンプ)。上記の保健指導の場合は、指導を実際に受けたか否かに関わらずその対象となる場合を介入としたため、シャープ RDD になります。

ファジー RDD は、カットオフを境にして介入を受ける割合が急に増えるものの 0%から 100%へのジャンプになっていない場合に適用されます。

シャープ RDD の例 (研究開発投資を促進するための補助金の効果検証)

研究開発投資を推進するためにチェコ共和国で交付された企業への補助金の効果検証についての研究を紹介します³。この補助金では補助金を申請した企業について得点 (これがランニング変数) が算出され、一定の得点 (これがカットオフ) を超えると補助金が交付されることとなっていました。このことを利用してシャープ RDD による分析が行われました。

ここでは全企業と中小企業、大企業の 3 パターンで分析が行われており、表 3-1 では統計学的な検証の結果を示しています。補助金と民間企業の自己支出を合計した全ての研究開発支出が表の左側に、企業自らが行った民間拠出の研究開発支出については表の右側に掲載されています。

表 3-1 R&D 補助金の研究開発支出 (対数) への影響

	RDD の推計値	
	全ての研究開発支出	民間拠出の研究開発支出
全企業	0.18 (0.19)	0.36 (0.24)
中小企業	0.54** (0.24)	0.94*** (0.29)
大企業	-0.13 (0.17)	-0.48** (0.22)

(注) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$ 。カッコ内は標準誤差。

(出所) Bajgar and Srholec (2025) の Table 4 より作成。

³ Bajgar and Srholec (2025)

表3-1のカッコのない数値は点推定値で、カッコで囲まれた数値は標準誤差です。標準誤差のごく簡単な読み方としては、点推定値を標準誤差で割った数値が概ね2を超えると5%水準で有意差があります。統計的に有意かどうかを示すアスタリスク(*)が書かれていない論文がありますが、割り算をすれば大体わかります。

アウトカム変数が連続変数の場合、点推定値と標準誤差がわかると95%信頼区間は自分で算出できます。

95%信頼区間の下限 = 点推定値 - 2 × 標準誤差

95%信頼区間の上限 = 点推定値 + 2 × 標準誤差

たとえば、表3-1の全企業の全ての研究開発支出の場合、95%信頼区間の下限は $0.18 - 2 \times 0.19 = -0.201$ で、上限は $0.18 + 2 \times 0.19 = 0.56$ で、95%信頼区間は0をまたぐので、5%水準で有意差がないことがわかります。

分析結果によれば、全企業では全ての研究開発支出においても民間拠出の研究開発支出においても有意差はありませんでした。中小企業の場合には、補助金の交付を受けた企業は受けなかった企業よりも、カットオフの近傍において全ての研究開発支出が有意に増えており（有意差を示すアスタリスク(*)が表3-1についています）、民間拠出の研究開発支出も有意に増えていました。大企業においては全ての研究開発支出では有意差がなく、民間拠出の研究開発支出は有意に減少していました。

総括すると、中小企業と大企業では効果に異質性があり（異質性については第4話で取り上げます）、中小企業においてはこの補助金が自己資金による研究開発支出の増加を促したのに対して、大企業ではこの補助金が自己資金による研究開発支出を減らすことを促したことになります。

なお、この補助金の効果検証はチェコ共和国についてのものであり、日本の類似の補助金においても同様の結果になるかどうかは明らかではありません。このように、ある集団や環境において得られた効果検証の結果が、他の集団（国を含む）や異なる状況においても当てはまるかどうかという問題は、外的妥当性と呼ばれます。第4話で改めて取り上げます。

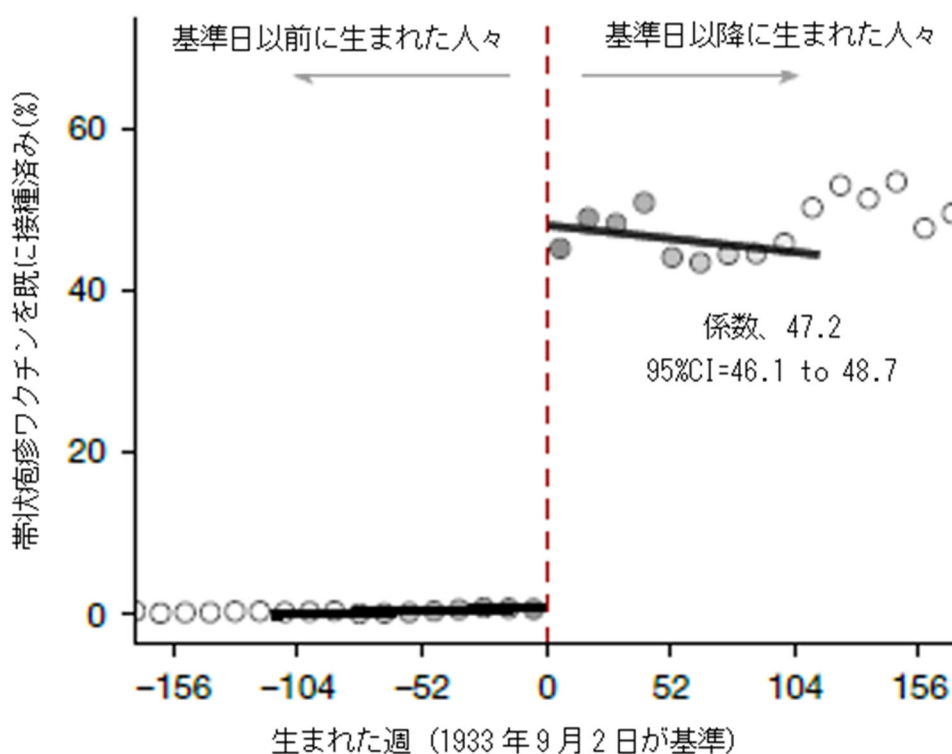
ファジーRDDの例（帯状疱疹ワクチンの認知症予防効果の検証）

ファジーRDDの例も挙げておきます⁴。

⁴ Eyting et al. (2025)

イギリスのウェールズでは、1933年9月2日以前に生まれた人々は帯状疱疹のワクチンを受ける資格がない一方で、この日より後に生まれた人々は受ける資格がありました。このため、この日付以前と比べて、この日付以後でワクチン接種を受ける割合が50%近くジャンプしていました（図3-2）。これを利用してファジーRDDによる分析が行われました。

図3-2 帯状疱疹のワクチン接種における大きな不連続



(出所) Eytting et al. (2025), Fig.1a.

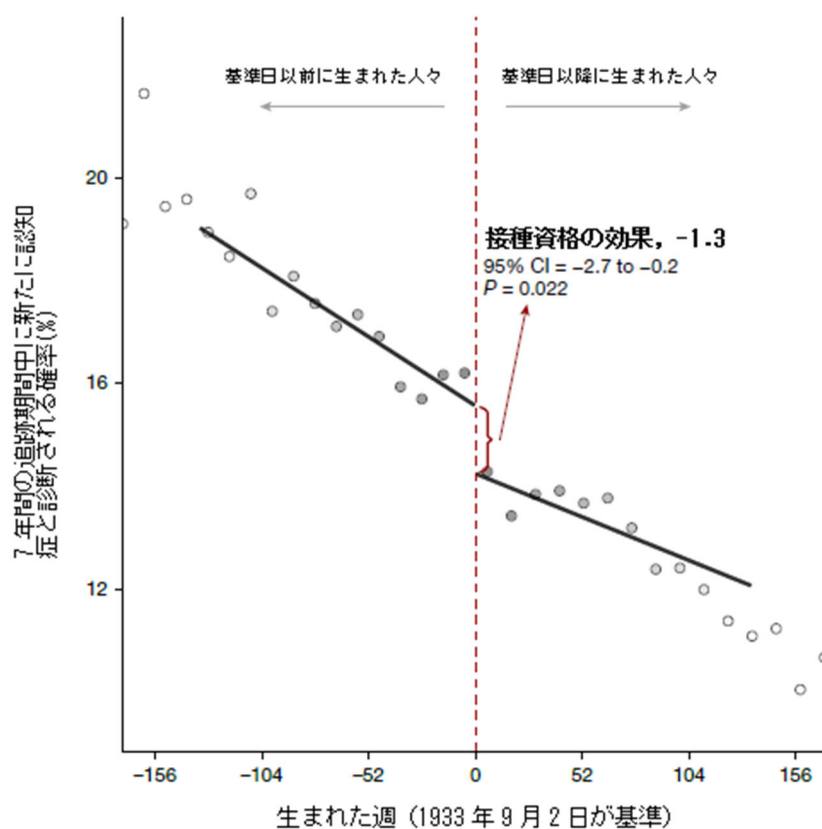
第2話で取り上げたRCTの奨励デザインと同様に、ファジーRDDにおいては主に2つの分析方法があります。1つめは実際に介入に参加したかどうかに関わりなくカットオフを上回った人々と下回った人々の全てを対象として介入効果を推定します。ITT (intention-to-treat) と呼ばれます。もう一つLATE (local average treatment effect、局所平均処置効果)と呼ばれる推計では、カットオフを超えたことによって介入に参加する人々(遵守者)にとって介入の効果がどの程度かを推定します。

まず、当然起きるべきことの確認として、このワクチンが帯状疱疹の発症をどれくらい予防したかを検証しています。それによれば、生年月日の違いによって帯状疱疹のワクチンを受ける資格があった人々は7年間の追跡期間中に帯

状疱疹の診断確率が 1.0%ポイント低くなっていました。LATE だと 2.3%ポイント低くなっており、37.2%の減少があったとされています。

ここから先が重要ですが、生年月日の違いによって带状疱疹のワクチンを受ける資格があった人々は新たに認知症と診断される確率が7年間の追跡期間中に1.3%ポイント低くなっていました（図3-3）。LATE だと 3.5%ポイント低くなっており、20%の減少があったとされています。

図3-3 带状疱疹のワクチンの接種資格の認知症への効果



(出所) Eytting et al. (2025), Fig. 3.

ランニング変数が操作されないことの重要性

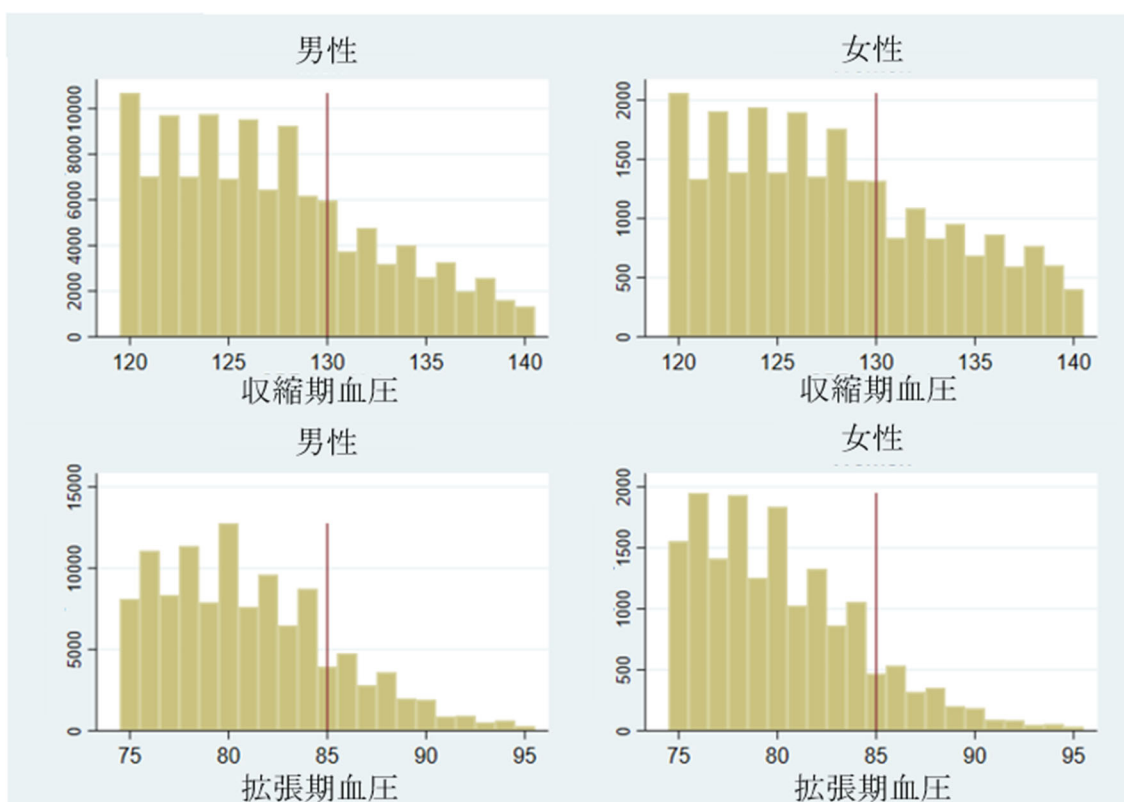
RDD では、介入の対象になろうとしたり外れようとしていたりしてカットオフ周辺でランニング変数の操作が行われると正確な分析が行えなくなります。以下は実際に私が出くわしたメタボ健診の例です⁵。

メタボ健診では収縮期血圧（上の血圧）が 130 以上（または拡張期血圧が 85 以上）の場合に高血圧のリスクがあると認定され、130 未満だと認定されな

⁵ Sekizawa (2024)

いという運用がされています。図3-4がこれによって引き起こされていると思われる操作について示しています。血圧計は2刻みのものと1刻みのものがあるため、偶数として計測される場合が多くなるためわかりにくいのですが、たとえば、130と128（あるいは131と129）で比べると、低い数字の人々が急に増えていることがわかります。このような場合には操作が疑われます。

図3-4 メタボ健診における血圧の操作



(注) 対象者は Sekizawa (2024)における血圧をランニング変数とする RDD の分析対象者。縦線は高血圧のリスクの有無を示すカットオフ値。

(出所) Sekizawa (2024)

このような操作は意図的に血圧の値を変えたのではなく、現場における血圧の測り方にあると思われる。メタボ健診では厚生労働省の指導によれば血圧は2回計測するのが基本です。ところが、現場では血圧を1回だけ測ることが多く、高血圧のリスクがあるとされる収縮期血圧130以上（または拡張期血圧85以上）だと念のためもう1回測ることが多いようです⁶。血圧測定では1回目よりも2回目の方が低い数値となる場合が多いので、記録される血圧は130

⁶ 渡邊ら(2020)

未満の場合が増えることとなります。この運用は2回測る人を減らすという意味では効率的ですが、RDDの適用には妨げになります。

操作されているかどうかはこの図のようにビジュアルにチェックする場合がありますが、データを使って統計的に検証する方法もあります⁷。検証によって不正行為が起きていることが明らかになる可能性もあるので、不正行為を未然に防ぐ注意が行政側に求められます。

第2節 差の差分析 (DID, Difference-in-Differences)

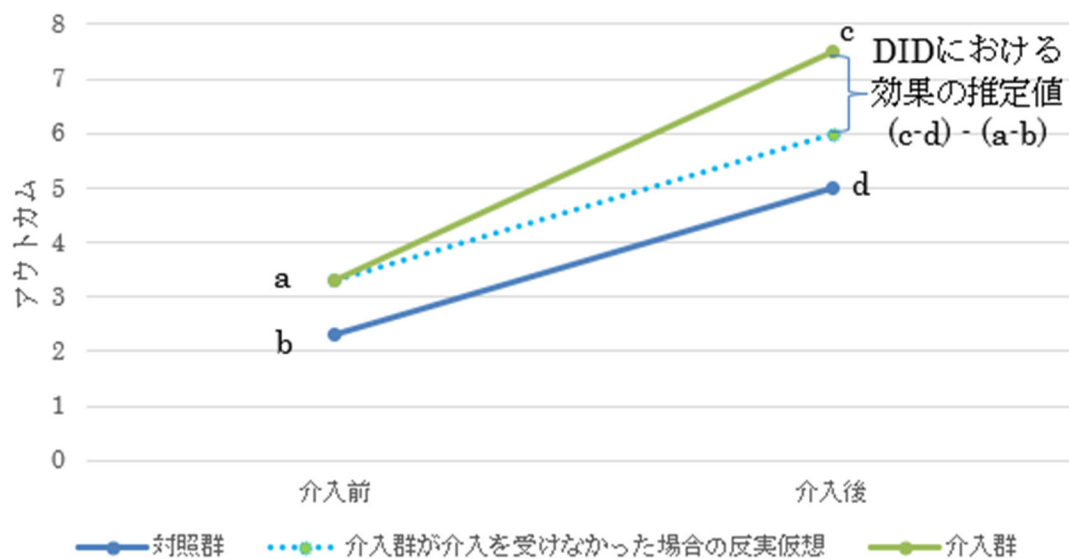
DIDでは、ある政策が特定時点までは行われず、その後一部に一部の集団（地域、人々、企業）に対してのみ行われた場合に、その政策が適用された集団を介入群、適用されなかった集団を対照群として扱って分析するものです。表3-2のそれぞれの数値は、たとえばaは介入群のアウトカムの介入前の平均値です。

表3-2 各群・各時点のアウトカムの平均値

	介入前	介入後
介入群	a	c
対照群	b	d

ここで、DIDにおける効果の推定値は $(c - a) - (d - b) = (c - d) - (a - b)$ となり、差から差を引いたもの、つまり、差の差ということになります（図3-5）。

図3-5 DIDのイメージ



⁷ Frandsen (2017); McCrary (2008)

規模の大きな RCT であれば介入前の a と b は概ね似たような数値になりますが、DID では必ずしもそうなりません。仮に介入群が介入を受けなければ、対照群の介入前から介入後への変化 (d-b) と同じ変化が介入群にも見られるという仮定を満たすことが必要になります。つまり、介入がなければ、介入群と対照群のアウトカム変数は平行に動くという仮定があります (平行トレンド)。平行トレンドは図 3-5 の点線で示しています。

介入群が介入を受けなかったというのは実際には起きていないことですので (反実仮想)、平行トレンドが満たされているかどうかはデータからはわかりません。これを補うために実際の DID では事前トレンド(pre trend)といって、介入開始の一時点前を基準として (ここの両群のアウトカムの差をゼロに置き換えることが多い)、それ以前の各時点において、介入群と対照群の間の差がゼロに近い状態であり続けたかを確認することが多いです。この場合、介入前と介入後の双方において複数時点のアウトカムの計測が行われます。このような設計をイベントスタディと呼びます。

DID の例 1 (Facebook が大学生のメンタルヘルスに及ぼす影響)

Facebook の大学キャンパスでの利用は最初にハーバード大学、次はスタンフォード大学というように、大学毎に順次利用可能になりました。米国大学健康協会 (ADHA) は年に 2 回の頻度で大学生に対して健康についてのアンケート調査を行っており、その中にはメンタルヘルスについて尋ねる質問が含まれていました。

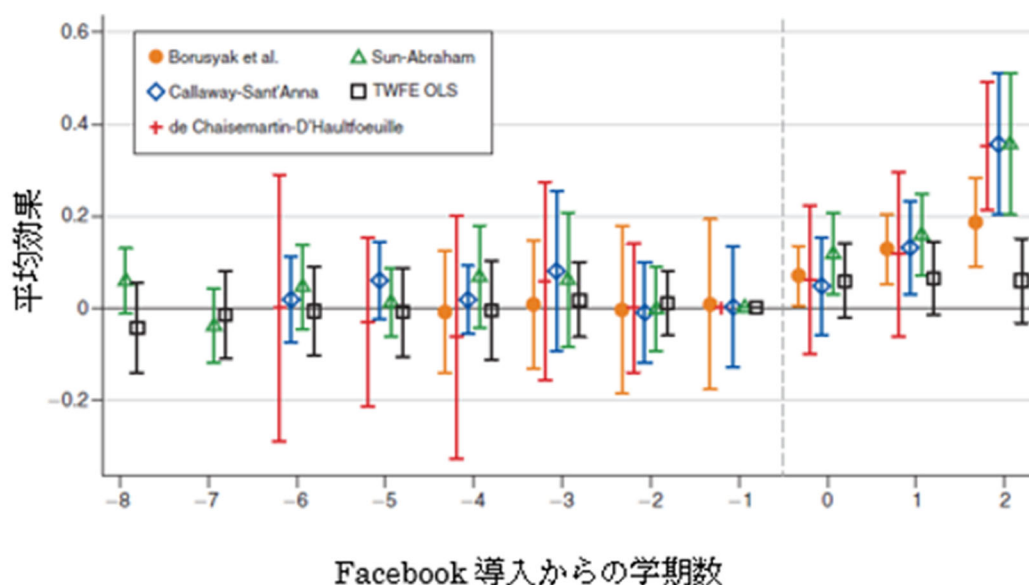
Braghieri et al. (2022) では、各大学毎の Facebook 利用開始時期の情報と、ADHA の調査によって得られた大学毎の大学生のメンタルヘルスのスコアを使って、Staggered DID (介入開始時点が分析対象[この場合は大学]によって異なる場合の DID) によって Facebook の利用が大学生のメンタルヘルスにもたらした影響を明らかにしました。アウトカムとなるメンタルヘルスの指標はこの研究の著者が標準化したものが使われており、数値が高いほどメンタルヘルスが悪化することを示します。

イベントスタディ型の Staggered DID については様々な分析手法が提案されていて、ここでは 5 つの分析手法による分析結果が示されています (図 3-6)。介入開始時点 (この場合は Facebook の利用が始まった学期) が 0 でその一時点前の -1 において介入群と対照群の差がゼロに置き換えられています。この場合、介入時点より前 (-2 以下) では 95%信頼区間がゼロをまたいでおり、事前トレンドは満たされていると判断されます。介入開始後、特に 2

時点後においては二方向固定効果回帰 (TWFE, two-way fixed effect) ⁸という分析方法を除いて有意にプラスの効果が出ています (ここではプラスはメンタルヘルスの悪化を意味します)。

つまり、この研究は大学生における Facebook の利用がメンタルヘルスの悪化につながることを強く示唆しています。

図 3-6 精神不調の指標に対する Facebook の影響 (イベントスタディ)



(出所) Reproduced from Figure 2 of Braghieri et al. (2022). © American Economic Association, used with permission.

DID の例 2 (産科病棟の閉鎖が出産関連行動に及ぼす影響)

Fischer et al. (2024)はアメリカの郡において産科病棟が閉鎖されて集約化されることの効果を staggered DID によって検証しています (図 3-7)。

TWFE と dCDH⁹という 2つの分析方法を用いてその両方の結果を報告しています。上述した Facebook の場合には TWFE と他の分析手法の分析結果が乖離していましたが、こちらではほぼ同じになっています。

妊婦が居住する郡における出産は明らかに減少していますが (パネル A)、事前トレンドは満たしていません。このことについてはこの論文の著者は事

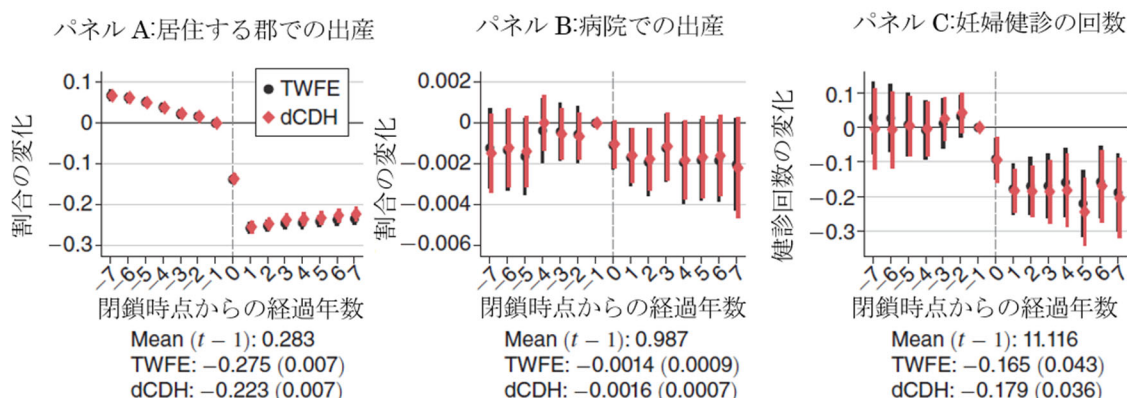
⁸ TWFE は伝統的に DID で使われてきた分析方法ですが、特に介入開始時点が複数になる場合に問題があることが指摘されています(川口・澤田, 2024)。

⁹ dCDH は de Chaisemartin and D'Haultfoeuille (2020)で提案された DID の推計法の略語です。

前トレンドにおける傾向よりも大きな変化が介入後に見られたとして、効果があったとしています。病院における出産は減っているように見えますが（産科病棟が閉鎖された群の閉鎖の1年前の病院での出産割合が98.7%で、閉鎖後に0.14%ポイント [TWFEの場合] ないし0.16%ポイント [dCDHの場合]の減少）、ごくわずかです（パネルB）¹⁰。出産前の妊婦健診の回数は有意に減っています（パネルC）。

これら以外の結果として、産科病棟が閉鎖された郡において帝王切開が減っていること、妊娠37～39週目における陣痛誘発が増加していることなどが発見されていて、母子共々健康に対する問題は見つけられなかった（わずかながら望ましくなる傾向があった）と報告されています。

図3-7 産科病棟の閉鎖が出産や妊婦健診回数に及ぼす平均効果



(出所) Reproduced from Figure 2 of Fischer et al. (2024). © American Economic Association, used with permission.

終わりに

RCTもRDDもDIDもテクニカルな面を突き詰めればきりがありませんが、基本的な発想はシンプルで、因果推論の専門家でない政策現場にいる人にとっても十分に理解可能です。このような基本的な理解を行政官や政治家が頭に入れておくと、信頼できる効果検証を政策介入の実施後に行うための事前の制度設計が円滑に実施できるようになります（第4話参照）。

¹⁰ 図3-7のパネルBの下の方に TWFE: -0.0014 (0.0009)とあり、介入効果を示します。カッコ内は標準誤差で、点推定値 (-0.0014) を標準誤差 (0.0009) で割った数値は2を下回っているので5%水準では有意ではありません（第1節を参照）。dCDHでは有意差があります。

また、第2話とこの第3話で述べた程度の理解があれば、これらの分析手法についての細かい知識を行政官などの政策立案者が持つ必要はありません。分析は信頼に値する専門的知識と中立性を持った研究者に委ねることが望ましいです。

なお、ここで取り上げなかった因果推論の重要な手法として合成コントロール法(Abadie, 2021)がありますが、別の機会に取り上げたいと思います。

引用文献

- Abadie, A. (2021). "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*, 59(2), 391-425.
- Bajgar, M., & Srholec, M. (2025). "Crowding in or crowding out? Evidence from discontinuity in the assignment of business R&D subsidies," *Journal of Public Economics*, 245, 105357.
- Braghieri, L., Levy, R. e., & Makarin, A. (2022). "Social Media and Mental Health," *American Economic Review*, 112(11), 3660–3693.
- de Chaisemartin, C., & D'Haultfœuille, X. (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110(9), 2964-2996.
- Eyting, M., Xie, M., Michalik, F., Heß, S., Chung, S., & Geldsetzer, P. (2025). "A natural experiment on the effect of herpes zoster vaccination on dementia," *Nature*, 641(8062), 438-446.
- Fischer, S., Royer, H., & White, C. (2024). "Health Care Centralization: The Health Impacts of Obstetric Unit Closures in the United States," *American Economic Journal: Applied Economics*, 16(3), 113–141.
- Frandsen, B. R. (2017). Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design when the Running Variable is Discrete *Regression Discontinuity Designs* (Vol. 38, pp. 281-315): Emerald Publishing Limited.
- McCrary, J. (2008). "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142(2), 698-714.
- Sekizawa, Y. (2023). "Effects of being eligible for specific health guidance on health outcomes: A regression discontinuity analysis using Japan's data on specific health checkups," *Prev Med*, 172, 107520.

Sekizawa, Y. (2024). "Japan's intensive health guidance program has limited effects on cardiovascular risk factors: a regression discontinuity analysis," *Public Health*, 232, 108-113.

川口康平・澤田昌行. (2024). 因果推論の計量経済学: 日本評論社.

渡邊倫子・佐藤ひとみ・菅原由紀江・附田順子・後藤敏和・菊地惇. (2020). "健診における血圧測定法の現状～アンケート調査より～," *総合健診*, 47(3), 463-468.