

予防・健康づくりプログラムの統計的評価デザイン

～理解・実践のための手引き～

作成日（第一版）：2025年2月22日

著者：井上 浩輔、池洲 諒、安富 元彦、田近 亜蘭、古川 壽亮、近藤 尚己

＊この手引きは日本医療研究開発機構「予防・健康づくりサービスの平均効果と異質効果の推計デザインとその実装に関する研究（代表：近藤尚己）」の研究の一部として作成されました。

序文

本手引きは、予防・健康づくりプログラムの評価を行う研究者のために、統計的な評価デザインを体系的に整理したものである。

近年、健康行動や生活習慣の改善を目的とした多様なプログラムが開発され、その有効性評価が重要な課題となっている。しかし、評価方法の選択や実施には多くの専門的な判断が求められ、また適切な方法論を採用し、その結果を適切に解釈しなければ因果関係を誤って判断するリスクがある。

本手引きでは、予防・健康づくりプログラムを評価するための統計的アプローチを、主に以下の点に焦点をあてて解説する：

- 各研究デザインの概要
- 各研究デザインを応用する際に必要となる前提条件
- 各研究デザインのメリット・デメリット
- 代表的な実装例
- 実装に向けて詳細を学ぶ際に有用な参考文献

とくに、ランダム化比較試験（RCT）だけでなく、観察データを活用した準実験的デザイン、近年注目を浴びている効果の異質性評価方法についても、必要な仮定や応用上の留意点を明示することで、研究者がテーマや現場の制約に応じた最適なアプローチを選択できることを目的としている。

また、最新の文献に基づく応用例や、評価対象のプログラムに応じた活用シナリオを提示することで、理論と実践を橋渡しする内容となっている。予防・健康づくりプログラムの実務に携わる研究者が、より実践的かつ柔軟に評価研究を設計・実施できる一助となれば幸いである。

目次

序文	2
1 予防・健康づくりプログラム評価を行うための3つのステップ	4
2 具体的な研究デザイン決定プロセス	4
3 各研究デザインのメリット・デメリット	5
4 各研究デザインの詳細	7
4.1 ランダム化比較試験	7
4.2 疑似ランダム化	11
4.3 非ランダム化	17
5 Target Trial Emulation による拡張	26
6 効果の異質性評価	28
6.1 サブグループ解析	28
6.2 機械学習手法による異質性評価	28

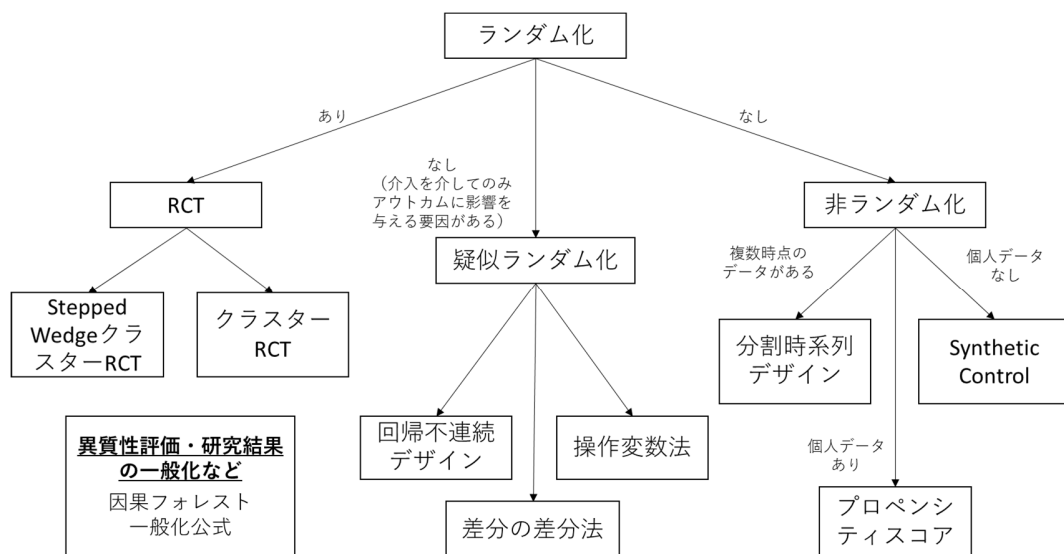
1 予防・健康づくりプログラム評価を行うための3つのステップ

Step 1: アプリ開発

Step 2: 研究デザインの決定

Step 3: 異質性評価

2 具体的な研究デザイン決定プロセス



3 各研究デザインのメリット・デメリット

研究デザイン	解析手法	手法の概要	必要な仮定	メリット	デメリット
RCT	クラスターRCT	集団レベルで介入をランダム化することで介入の効果を推定するランダム化比較試験.	クラスターRCTの実装は、クラスターを用いることが十分に正当化される場合に限られる.	介入を個人レベルでランダム化することが困難な場合に、集団レベルで介入をランダム化することで介入の効果を推定する.	クラスター内の相関によって推定の精度が低下し、必要なサンプルサイズが増加する. 介入の盲検化が困難なことが多い. オープンクラスターの場合、割り付けの隠匿ができない
	Stepped Wedge クラスターRCT	すべてのクラスターが介入を受けるように設計されたクラスターRCT	クラスターRCTと同様に、クラスターを用いることが十分に正当化される場合に限られる	すべてのクラスターに同時に導入することが困難な場合などでも、介入のランダム割り付けを通じて因果効果を推定することができる	介入導入のタイミングがクラスターごとに異なるため、アウトカムの経時的な変化を調整する必要がある
疑似ランダム化	回帰不連続デザイン	あるカットオフ値を境に介入を受ける確率が急激に変化させる連続変数 (running variable) を利用して、因果効果を推定する統計手法	Running variable のカットオフ値の周辺では、(未測定 of 交絡因子も含めて) 介入の有無以外の点で介入群と非介入群は類似している	観察研究のデータを利用して、未測定 of 交絡因子によるバイアスを補正できる	推定値の一般化可能性は制限される可能性がある
	操作変数法	介入を通してのみアウトカムに影響を与える変数 (操作変数) を利用することで、因果効果を推定する	1. 操作変数が介入に関連している 2. 操作変数が介入を通してのみアウトカムに影響を与える 3. 操作変数とアウトカムを引き起こす共通の要因がない	観察研究のデータを利用して、未測定 of 交絡因子によるバイアスを補正できる	理想的な操作変数を見つけることが難しい 操作変数と介入の関連が弱いと、バイアスが生じる
	差分の差分法	介入群と非介入群の経時的データを利用して因果効果を推定するデザインであり、特定の介入の効果を推定するために用いられる	1. 共通トレンドの仮定 2. 共通ショックの仮定	非介入群の経時的なトレンドを利用して、介入群の経時的なトレンドと分離して、介入の因果効果を推定できる	時間によって変化する、介入群と非介入群のアウトカムに異なる影響を与える未測定 of 要因があると、推定値にバイアスが生じる

非ランダム化	分割時系列デザイン	時系列データを用いて、導入時期が明確な集団レベルでの介入の効果を回帰分析で推定する手法	介入導入前のアウトカムの経時的トレンドは、介入がなかった場合の(反実仮定の)介入導入後の経時的トレンドと同じである	経時的なデータを用いることで、(1)アウトカムの経時的トレンドと(2)経時的に変化しない交絡因子を調整できる	経時的に変化して、アウトカムに影響を与える要因を調整する必要がある
	プロペンシティスコア	観察された変数で条件付された、介入を受ける確率を用いて因果効果を推定する手法	未測定 of 交絡因子が存在しない 各個人が介入を受ける確率と介入を受けない確率が共に0より大きい	調整すべき交絡因子の数が多くても利用可能である 介入因子、アウトカム、交絡因子間のモデルの誤特定に頑強である	未測定の交絡因子によるバイアスは対処できない
	Synthetic Control	非介入群(のうちのいくつか)を重み付けすることで仮想的な非介入群(synthetic control)を構築し、介入後時点のアウトカムを介入群と比較することで、因果効果を推定する	Synthetic control をつくるために使用される非介入群が介入群と類似している 介入群への介入が、非介入群に影響を与えない	介入群と非介入群の数が少ない場合でも適応できる 介入群と非介入群の介入開始前のアウトカムについて、経時的なトレンドが平行である必要はない	信頼区間や p 値の計算が困難 アウトカムの測定誤差が大きいデータをSCM で用いるとバイアスを引き起こしうる

4 各研究デザインの詳細

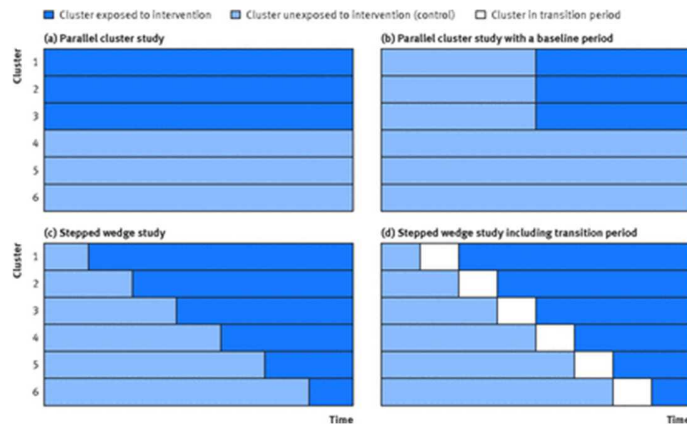
(各手法の概要、必要な仮定、メリット、デメリット、方法論を解説した論文、方法論を応用した論文とそのサマリー)

4.1 ランダム化比較試験

4.1.1 クラスターRCT

4.1.1.1 概要

- 集団レベルで介入をランダム化することで介入の効果を推定するランダム化比較試験
- 個人レベルで介入をランダム化する場合に比べて、バイアスの影響を受けやすく、必要なサンプルサイズも増加する。そのため、適用は慎重に判断する必要がある



(Figure 1 in Hemming et al., BMJ, 2015)

4.1.1.2 必要な仮定

- クラスターRCTの実装は、クラスターを用いることが十分に正当化される場合に限られる。つまり、個人レベルのRCTは実行できないのかを入念に確認する必要がある
- 「メリット」「デメリット(注意すべき点)」を参照

4.1.1.3 メリット

- 介入を個人レベルでランダム化することが困難な場合に、集団レベルで介入をランダム化することで介入の効果を推定する
例：ケアチームのメンバー (例：集中治療室のスタッフ) への教育が、患者アウトカムに与える効果の推定。この場合、同じ集中治療室で治療された患者を個人レベルでランダム化することは不可能
- クラスターRCTは、介入群と非介入群でのそれぞれの介入効果を混同 (contamination) してしまう可能性がある場合に有用である
例：集団におけるワクチン接種は、ワクチン非接種群に対しても感染予防効果をも

たらず。この場合、個人レベルでワクチン接種をランダム化してしまうと、ワクチン非摂取群でのアウトカムは、ワクチン接種群から影響を受けてしまう

4.1.1.4 デメリット

- アウトカムをクラスターと異なる階層で測定する場合、クラスター内でアウトカムが相関する可能性を、統計的に補正する必要がある (generalized estimating equation など)
- クラスター内での相関によって推定の精度が低下し、必要なサンプルサイズが増加する
- 介入の盲検化が困難なことが多く、バイアス (例:アウトカムの評価におけるバイアス、RCT からの脱落に関連したバイアス)を引き起こしうる
- 特にオープンクラスターの場合、allocation concealment (割り付けの隠匿) ができず、Susceptibility bias が免れない (例: クリニックをクラスターに分けたとき、クラスターそのものはランダム化されているにもかかわらず、そのクリニックで介入へ参加する人は割り付けられた介入に向いている人になる可能性)
- 結果を報告する際は、クラスターRCT 向けに拡張された CONSORT statement を参照 (Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012;345:e5661. Epub 2012/09/07. doi: 10.1136/bmj.e5661. PubMed PMID: 22951546.)
- その他の注意すべき点の詳細は「方法論を解説した論文」を参照

4.1.1.5 方法論を解説した論文

- Meurer WJ, Lewis RJ. Cluster randomized trials: evaluating treatments applied to groups. JAMA. 2015;313(20):2068-9. Epub 2015/05/27. doi: 10.1001/jama.2015.5199. PubMed PMID: 26010636.
- Hemming K, Taljaard M. Key considerations for designing, conducting and analysing a cluster randomized trial. Int J Epidemiol. 2023;52(5):1648-58. Epub 2023/05/19. doi: 10.1093/ije/dyad064. PubMed PMID: 37203433; PMCID: PMC10555937.

4.1.1.6 方法論を応用した論文

- Kamada M, Kitayuguchi J, Abe T, Taguri M, Inoue S, Ishikawa Y, Bauman A, Lee IM, Miyachi M, Kawachi I. Community-wide intervention and population-level physical activity: a 5-year cluster randomized trial. Int J Epidemiol. 2018;47(2):642-53. Epub 2017/12/12. doi: 10.1093/ije/dyx248. PubMed PMID: 29228255; PMCID: PMC5913653.
- クラスターRCT デザインを用いて、中高年の身体活動促進のための 5 年間の地域レベルの介入の有効性を評価した論文

日本のある市町村内の 12 地域を介入群 (9 地域) と非介入群 (3 地域) に無作為に割り付けた。さらに、介入群の 9 地域を、目標とする身体活動の種類別に 3 つのサ

ブグループ（有酸素、柔軟、筋力強化）に無作為に割り付けた。無作為に抽出された40～79歳の住民4414人を5年間追跡した。身体活動を促進するためのソーシャル・マーケティングを用いた5年間の介入が身体活動に与える影響を解析したところ、非介入群に比べ、介入群では推奨レベルの運動量達成の割合が増加した。身体活動の種類別の介入はそれぞれ有効であったが、すべての種類を身体活動を同時に促進するアプローチが有効であるという結果は得られなかった。

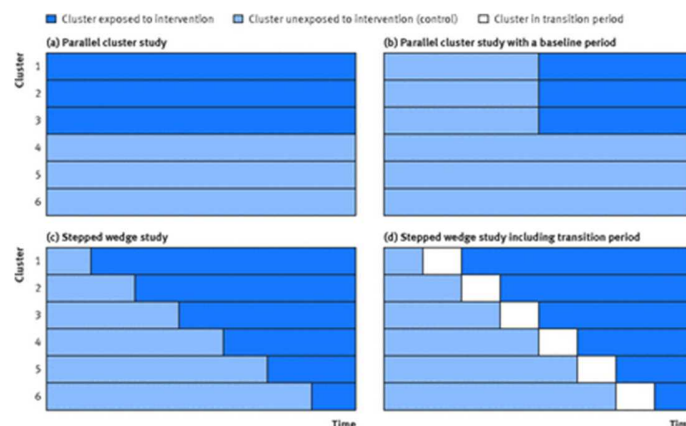
4.1.1.7 予防健康づくりアプリにおける使用例案

- 喫煙歴のある患者への禁煙カウンセリングを促す電子カルテ上のポップアップが禁煙達成に与える効果を推定する。クリニック単位でポップアップを導入し、クラスターRCTを行う

4.1.2 Stepped Wedge クラスターRCT

4.1.2.1 概要

- すべてのクラスターが介入を受けるように設計されたクラスターRCTであり、waiting list designやphased implementationsとも呼ばれる。Stepped wedge クラスターRCTと対比して、従来のクラスターRCTをparallelクラスターRCTと呼ぶことがある
- すべてのクラスターが介入にさらされない初期期間がある。その後、一定の間隔(step)で、1つまたは複数のクラスターが、介入に移行するようにランダムに割り付けられる。
このプロセスは、すべてのクラスターが介入に曝露されるまで続けられる。各クラスターが非介入の期間と介入の期間の両方を経験し、データ収集は研究を通して継



続される

(Figure 1 in Hemming et al., BMJ, 2015)

4.1.2.2 必要な仮定

- クラスターRCTと同様に、クラスターを用いることが十分に正当化される場合に限

られる。つまり、個人レベルの RCT は実行できないのかを入念に確認する必要がある

4.1.2.3 メリット

- クラスターRCTにおいて、介入を受けないクラスターが存在することが受け入れられない場合（例：すでに介入の有効性について一定のエビデンスがある場合）や、介入をすべてのクラスターに同時に導入することが困難な場合でも、介入のランダム割り付けを通じて介入の因果効果を推定することができる
- 同一クラスター内のサンプルが介入群と非介入群の両方として解析に利用でき、クラスター内の相関が強い場合にはクラスターRCTよりも推定の精度が高い

4.1.2.4 デメリット

- 介入導入のタイミングがクラスターごとに異なるため、アウトカムの経時的な変化を調整する必要がある
- ほかの注意点についてはクラスターRCTと重複するため、クラスターRCTの記載を参照。

4.1.2.5 方法論を解説した論文

- Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. 2015;350:h391. Epub 2015/02/11. doi: 10.1136/bmj.h391. PubMed PMID: 25662947.

4.1.2.6 方法論を応用した論文

- Leontjevas R, Gerritsen DL, Smalbrugge M, Teerenstra S, Vernooij-Dassen MJ, Koopmans RT. A structural multidisciplinary approach to depression management in nursing-home residents: a multicentre, stepped-wedge cluster-randomised trial. *Lancet*. 2013;381(9885):2255-64. Epub 2013/05/07. doi: 10.1016/S0140-6736(13)60590-5. PubMed PMID: 23643110.

2009 年 5 月から 2011 年 4 月にかけて、オランダの 17 の介護施設を対象とした stepped wedge クラスターRCT を実施し、多職種によるケアプログラム（Act in Case of Depression[AiD]）が、介護施設入所者におけるうつ病の有病率に与える影響を評価した。研究開始時には研究対象の介護施設はいずれも AiD を実施していなかった。約 4 ヶ月の間隔で AiD を順次介護施設に導入した。線形混合モデルを用いて、介護施設内と同一個人でのアウトカムの反復測定による相関を調整した。さらに経時的なトレンドを調整した。その結果、AiD 導入後のうつ病有病率は 7% 減少した、と推定された。

4.1.2.7 予防健康づくりアプリにおける使用例案

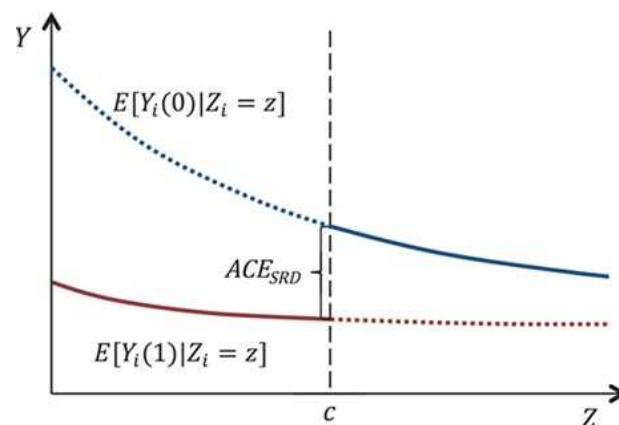
- 喫煙歴のある患者への禁煙カウンセリングを促す電子カルテ上のポップアップが禁煙達成に与える効果を推定する。すべてのクリニックにポップアップを導入するが、クリニック単位で導入時期をずらし、stepped wedge クラスタ-RCT を行う

4.2 疑似ランダム化

4.2.1 回帰不連続デザイン

4.2.1.1 概要

- 回帰不連続デザインは、あるカットオフ値を境に介入を受ける確率を急激に変化させる連続変数 (running variable と呼ぶ) を利用して、介入の因果効果を推定する統計手法である
- 臨床診療または公衆衛生プログラムが、それぞれ治療またはプログラムの適格性を割り当てる決定規則として連続変数のカットオフ値を使用する場合に適用できる
Running variable の例：出生体重 (1500 g 未満 vs. 1500 g 以上)、日付 (ある日付を境に政策が施行された場合など)
- Running variable のカットオフ値の周辺で介入の有無以外の点で介入群と非介入群は類似しているとみなすことで、介入の因果推論が可能となる。未測定の変絡因子に関しても類似しているとみなす
- Running variable のカットオフ値の前後でのアウトカムの変化を、カットオフ値の前後での介入の有無の変化による効果とみなす
- シャープな回帰不連続デザイン：Running variable のカットオフ値の一方の側にいるサンプルは全員介入を受け、他方の側にいるサンプルは全員介入を受けない
- ファジーな回帰不連続デザイン：Running variable のカットオフ値の一方の側では他方の側よりも介入を受ける確率が高い
- 回帰不連続デザインは、操作変数法の一例と捉えることもできる
詳細は、[方法論を解説した論文]を参照



(Figure 1 in Bor et al., Epidemiology, 2014)

4.2.1.2 必要な仮定

1. Running variable のカットオフ値の周辺では、(未測定の交絡因子も含めて)介入の有無以外の点で介入群と非介入群は類似している
測定されている要因が介入群と非介入群で類似していることを確認する

厳密には、カットオフ値の前後で「介入と潜在アウトカムが独立である」という仮定である。詳しくは[方法論を解説した論文]を参照

2. (1. の仮定に関連して) Running variable の値の故意的な操作がない
故意的な操作の例：患者の血圧が 139 mmHg である場合に、医師がカルテに（高血圧症診療のカットオフ値の 1 つである）140 mmHg と記載し、降圧薬の処方を開始する。

この仮定が満たされない場合は、統計手法 (McCray test) で検知できる (ただし、この仮定が満たされていることを厳密に示すことはできない)

4.2.1.3 メリット

- 観察研究のデータを利用して、未測定 of 交絡因子 (例：介入群と非介入群での測定しきれていない重症度の違い) によるバイアスを補正できる

4.2.1.4 デメリット

- Running variable のカットオフ値に近い集団に対してのみ有効な因果効果の推定値が得られるため、推定値の一般化可能性は制限される可能性がある
- 複数の関数形で推定を行い (感度分析)、推定結果が関数形の仮定に依存しないことを確認する

4.2.1.5 方法論を解説した論文

- Maciejewski ML, Basu A. Regression Discontinuity Design. JAMA. 2020;324(4):381-2. Epub 2020/07/03. doi: 10.1001/jama.2020.3822. PubMed PMID: 32614409.
- Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. BMJ. 2016;352:i1216. Epub 2016/03/16. doi: 10.1136/bmj.i1216. PubMed PMID: 26977086; PMCID: PMC6884311.
- Moscoe E, Bor J, Barnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. J Clin Epidemiol. 2015;68(2):122-33. Epub 2015/01/13. doi: 10.1016/j.jclinepi.2014.06.021. PubMed PMID: 25579639.
- Bor J, Moscoe E, Mutevedzi P, Newell ML, Barnighausen T. Regression

discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology. 2014;25(5):729-37. Epub 2014/07/26. doi: 10.1097/EDE.0000000000000138. PubMed PMID: 25061922; PMCID: PMC4162343.

4.2.1.6 方法論を応用した論文

- Fukuma S, Iizuka T, Ikenoue T, Tsugawa Y. Association of the National Health Guidance Intervention for Obesity and Cardiovascular Risks With Health Outcomes Among Japanese Men. JAMA Intern Med. 2020;180(12):1630-7. Epub 2020/10/09. doi: 10.1001/jamainternmed.2020.4334.

特定健診に参加した 40～74 歳の男性を対象として、回帰不連続デザインを用いて特定保健指導が健康アウトカムに与える影響を検討した。

特定保健指導：特定健診で腹囲 85cm 以上で心血管危険因子が 1 つ以上見つかった人に対する健康的な生活習慣に関するカウンセリングとフォローアップ

Running variable: 腹囲

40～74 歳の男性において、特定保健指導は臨床的に意義のある体重減少や他の心血管危険因子の減少とは関連しなかった。

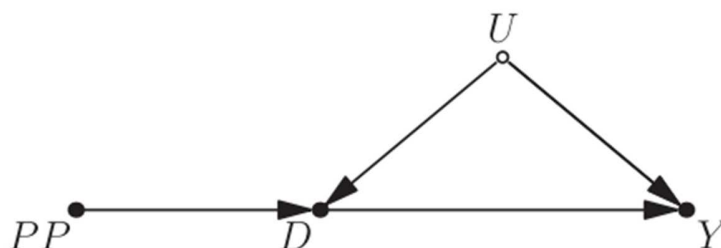
4.2.1.7 予防健康づくりアプリにおける使用例案

- 禁煙アプリ使用の適応基準 (例：年齢) を利用した禁煙アプリの禁煙効果の検証

4.2.2 操作変数法

4.2.2.1 概要

- 介入の有無を規定しているが、介入を通してのみアウトカムに影響を与える変数 (操作変数) を利用することで、観察研究のデータを用いながら未測定交絡因子によるバイアスを補正して介入効果を推定する



PP, 操作変数 ; D, 曝露 ; Y, アウトカム ; U, 未測定交絡因子

(Figure 1 in Widding-Havneraas et al., J Clin Epidemiol., 2014)

4.2.2.2 必要な仮定

1. 操作変数が介入に関連している
 2. 操作変数が介入を通してのみアウトカムに影響を与える
 3. 操作変数とアウトカムを引き起こす共通の要因がない
 - 4a. 介入効果の異質性がない（介入効果が集団内のすべての個体で同一）
 - 4b. 操作変数による介入の有無の変化は一方向のみ（操作変数が変化した時に、介入が無
→有に変化する個体と介入が有→無に変化する個体はいずれか一方しか存在しない）
- (1-3 + 4a または 4b が必要な仮定)

4.2.2.3 メリット

- 観察研究のデータを利用して、未測定の変因(例：介入群と非介入群での測定しきれていない重症度の違い)によるバイアスを補正できる

4.2.2.4 デメリット

- (ランダム割り付け以外では)理想的な操作変数をみつけることが難しい
- 操作変数と介入の関連が弱いと、バイアスが生じる(さらに、推定が不精確になる)
- 操作変数法のための必要な仮定が満たされていることをデータから証明できない(1の仮定はデータで確認可能)
- (4bの仮定をおいた場合に)推定値の解釈が難しい。4bの仮定のもとでは、操作変数の変化によって介入の有無が変化した集団における介入効果を推定することになる。そうした集団がどういった集団なのか、同定することができない(一般集団においては推定値が必ずしも当てはまらない)

4.2.2.5 方法論を解説した論文

- Maciejewski ML, Brookhart MA. Using Instrumental Variables to Address Bias From Unobserved Confounders. JAMA. 2019;321(21):2124-5. Epub 2019/05/03. doi: 10.1001/jama.2019.5646.
- Maciejewski ML, Dowd BE, Norton EC. Instrumental Variables and Heterogeneous Treatment Effects. JAMA. 2022;327(12):1177-8. Epub 2022/03/23. doi: 10.1001/jama.2022.2505.
- Widding-Havneraas T, Zachrisson HD. A Gentle Introduction to Instrumental Variables. J Clin Epidemiol. 2022;149:203-5. Epub 2022/07/11. doi: 10.1016/j.jclinepi.2022.06.022.

4.2.2.6 方法論を応用した論文

- Werner RM, Coe NB, Qi M, Konetzka RT. Patient Outcomes After Hospital Discharge to Home With Home Health Care vs to a Skilled Nursing Facility. JAMA Intern Med. 2019;179(5):617-23. Epub 2019/03/12. doi: 10.1001/jamainternmed.2018.7998.

Medicare（高齢者を主な対象とした、アメリカの公的医療保険）受給者の退院後ケアの選択が患者の転帰と Medicare の支出に与える影響を、操作変数法を用いて評価した。2010 年から 2016 年にかけての 1700 万以上の入院データを分析し、在宅医療と介護施設でのケアを比較した。患者の自宅から最寄りの在宅医療機関までの距離と、最寄りの介護施設までの距離を比較し、どちらの施設がより近いか、を操作変数として利用した。介護施設でのケアと比較して、在宅医療は 30 日以内の再入院率が 5.6%ポイント高かった。一方で、30 日以内の死亡率や身体機能の改善には有意な差は認めなかった。介護施設でのケアと比較して、在宅医療のほうが支出は低かった

4.2.2.7 予防健康づくりアプリにおける使用例案

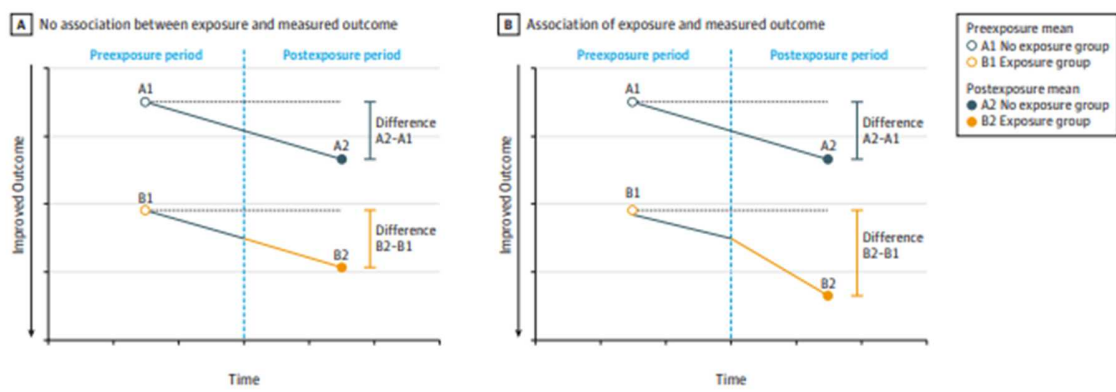
- 禁煙アプリの保険適応を操作変数とした、禁煙アプリの禁煙達成効果の検証

4.2.3 差分の差分法

4.2.3.1 概要

- 差分の差分法（Difference-in-Differences: DID）は、介入群と非介入群の経時的データを利用して因果効果を推定するデザインであり、特定の介入（法律の成立、政策の制定、大規模なプログラムの実施など）の効果を推定するために用いられる
- 非介入群の経時的な変化を用いて、介入がなかった場合の介入群の経時的な変化の影響を補正する
- DID は、介入後ダミー変数と介入群ダミー変数の交互作用項を含んだ回帰モデルで実装することができる

Figure. Conceptual Illustration of a Difference-in-Differences Analysis for 2 Scenarios



(Figure in Dimick and Ryan, JAMA, 2014)

4. 2. 3. 2 必要な仮定

- 共通トレンドの仮定
 介入がなければ、介入群と非介入群のアウトカムの経時的推移の違いは一定である。介入群と非介入群の介入前の経時的データを比較することでこの仮定の確からしめを確認できる
- 共通ショックの仮定
 介入と同時期もしくは介入以後に、介入群と非介入群のアウトカムに異なる影響を与える要因が存在しない

4. 2. 3. 3 メリット

- 非介入群の経時的なトレンドを利用して、介入群の経時的なトレンドと分離して、介入の因果効果を推定できる
- 介入群と非介入群の間に未測定の交絡因子がある場合でも、「必要な仮定」のもとで介入群への介入の効果を推定できる（ただし、許容される未測定の交絡因子は、時間によって変化しない変数に限られる）

4. 2. 3. 4 デメリット

- 時間によって変化する、介入群と非介入群のアウトカムに異なる影響を与える未測定の要因（未測定の時間依存性の交絡因子）があると、推定値にバイアスが生じる。そのため、時間依存性の交絡因子は回帰モデルで調整する必要がある
- 共通トレンドの仮定と共通ショックの仮定を満たす非介入群を見つけることが難しい

4.2.3.5 方法論を解説した論文

- Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. JAMA. 2014;312(22):2401-2. Epub 2014/12/10. doi: 10.1001/jama.2014.16153. PubMed PMID: 25490331.
- Wing C, Dreyer M. Making Sense of the Difference-in-Difference Design. JAMA Intern Med. 2024;184(10):1250-1. Epub 2024/08/26. doi: 10.1001/jamainternmed.2024.4135. PubMed PMID: 39186264.
- Columbia University Mailman School of Public Health. Difference-in-Difference Estimation [cited 2024 October 24].

4.2.3.6 方法論を応用した論文

- Miyawaki A, Noguchi H, Kobayashi Y. Impact of medical subsidy disqualification on children's healthcare utilization: A difference-in-differences analysis from Japan. Soc Sci Med. 2017;191:89-98. Epub 2017/09/17. doi: 10.1016/j.socscimed.2017.09.001. PubMed PMID: 28917140.

日本のある市町村における国民健康保険の診療報酬請求書（レセプトデータ）を用いて、医療費負担率の引き上げ（医療費助成の対象から外れること）が小児の医療サービス利用に与える影響を、差分の差分法(differences-in-differences approach)で検討した。小学生が4年生に進級すると、自治体の医療費助成の対象から外れることを利用して、介入群を小学4年生に進級した児童とした。非介入群は、小学2年生または3年生に進級した児童とした。医療費負担率の引き上げは、外来診療の利用を減少させることがわかった。また、入院診療の増加が観察されたが、これは健康状態が悪化したためではなく、むしろ外来診療から入院診療への置き換えが起きたためであった。外来診療の減少の程度は病態によって不均一であり、軽症または慢性的な病態ほど減少が顕著であった

4.2.3.7 予防健康づくりアプリにおける使用例案

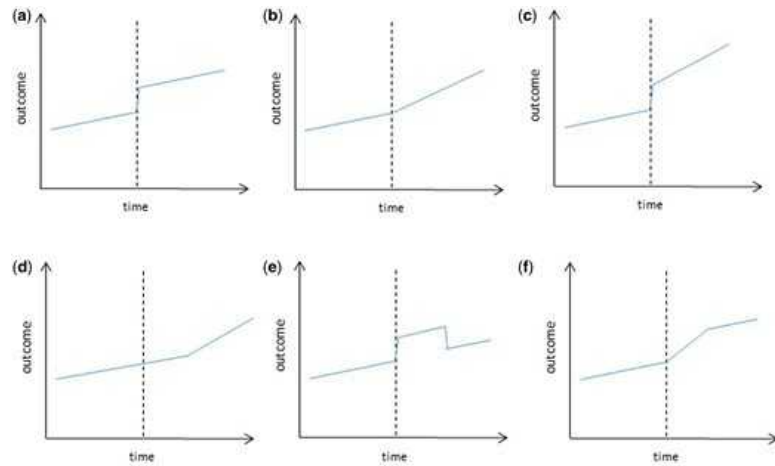
- 禁煙アプリをある特定の時点で導入したクリニックとそうでないクリニックで、患者の禁煙率への効果を DID で推定する

4.3 非ランダム化

4.3.1 分割時系列デザイン

4.3.1.1 概要

- 時系列データを用いて、導入時期が明確な集団レベルでの介入の効果を回帰分析で推定する手法



(Figure 2 in Bernal et al., Int J Epidemiol., 2017)

4.3.1.2 必要な仮定

- 介入導入前のアウトカムの経時的トレンドは、介入がなかった場合の(反実仮定の)介入導入後の経時的トレンドと同じである(同時期にアウトカムに影響を与えるほかの介入が存在しない)
- (上記の仮定と関連して)時間依存的に変化する未測定の交絡因子が存在しない
- 介入の効果が早期に、もしくは決まったタイムラグを持ってアウトカムに反映される。介入の効果が発生するタイミング(文献などに基づく)に基づいて回帰モデルを設定する

4.3.1.3 メリット

- 経時的なデータを用いることで、導入時期が明確な集団レベルでの介入に関して、(1)アウトカムの経時的トレンドと(2)経時的に変化しない交絡因子を調整できる

4.3.1.4 デメリット

- 経時的に変化して、アウトカムに影響を与える要因(例:季節性)を調整する必要がある
- 介入前後のデータが複数の時点で必要(時点数が多いほうが検出力が高まる)
- 経時的データの誤差項の相関(系列相関)が起きていないか確認する必要がある
- 感度分析として以下のような推定を行う必要がある
 - ・異なる回帰モデルでの推定
 - ・介入の効果を受けない変数をアウトカムにして推定(ネガティブ・コントロール)

4.3.1.5 方法論を解説した論文

- Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 2017;46(1):348-55. Epub 2016/06/11. doi: 10.1093/ije/dyw098. PubMed PMID: 27283160; PMCID: PMC5407170.
- Lopez Bernal J, Soumerai S, Gasparrini A. A methodological framework for model selection in interrupted time series studies. *J Clin Epidemiol*. 2018;103:82-91. Epub 2018/06/10. doi: 10.1016/j.jclinepi.2018.05.026. PubMed PMID: 29885427.
- Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther*. 2002;27(4):299-309. Epub 2002/08/14. doi: 10.1046/j.1365-2710.2002.00430.x. PubMed PMID: 12174032.
- Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prev Sci*. 2000;1(1):31-49. Epub 2001/08/18. doi: 10.1023/a:1010024016308. PubMed PMID: 11507793; PMCID: PMC4553062.
- Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *BMJ*. 2015;350:h2750. Epub 2015/06/11. doi: 10.1136/bmj.h2750. PubMed PMID: 26058820; PMCID: PMC4460815.

4.3.1.6 方法論を応用した論文

- Barone-Adesi F, Gasparrini A, Vizzini L, Merletti F, Richiardi L. Effects of Italian smoking regulation on rates of hospital admission for acute coronary events: a country-wide study. *PLoS One*. 2011;6(3):e17419. Epub 2011/03/15. doi: 10.1371/journal.pone.0017419. PubMed PMID: 21399685; PMCID: PMC3047543.

イタリアで2005年1月に施行された全国規模の禁煙規制（公共の場における禁煙）が急性冠症候群による入院に与える影響を、分断時系列デザインを用いて検討した。

70歳未満の急性冠症候群による入院は、禁煙規制の導入後に4%減少した。70歳以上では効果は認められなかった。

4.3.1.7 予防健康づくりアプリにおける使用例案

- 禁煙アプリの保険適応前後での、禁煙アプリの禁煙達成効果の検証

4.3.2 プロペンシティスコア

4.3.2.1 概要

- 観察された変数で条件付された、介入を受ける確率（プロペンシティスコア）を用いて因果効果を推定する統計手法
- 未測定の変数がないという仮定のもとで、プロペンシティスコアを利用して観察研究のデータを用いながら介入効果を推定することが可能である（ほかに必要な仮定については、[必要な仮定]を参照）
- プロペンシティスコアを利用した解析にはいくつか種類がある（マッチング、重み付け、など。詳細は、[方法論を解説した論文]を参照）
- プロペンシティスコアマッチング

Step 1: プロペンシティスコアを推定する

交絡因子とアウトカムの予測因子で介入の確率をモデルする

Step 2: 推定したプロペンシティスコアを利用して、介入群と非介入群をマッチングする

Step 3: マッチした介入群と非介入群の間で交絡因子の分布が類似していることを確認する

交絡因子の分布が類似しない場合は、Step 1 に戻ってプロペンシティスコアの推定モデルを変更する

Step 4: マッチした介入群と非介入群の間でアウトカムを比較する

詳細は、Zhao et al., Ann Transl Med. 2021 を参照

- 逆確率重み付け

Step 1: プロペンシティスコアを推定する

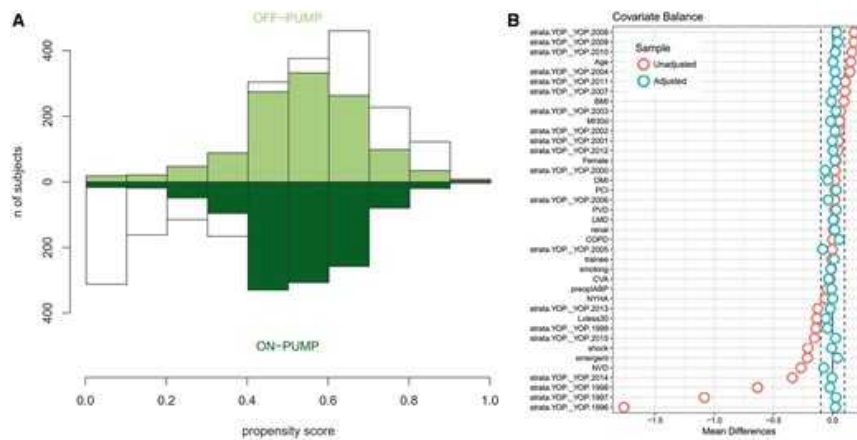
交絡因子とアウトカムの予測因子で介入の確率をモデルする

Step 2: 推定したプロペンシティスコアを利用して、介入群と非介入群にそれぞれ以下の重みを掛ける

介入群: $1/(\text{推定したプロペンシティスコア})$

非介入群: $1/(1-\text{推定したプロペンシティスコア})$

Step 3: 重みづけした介入群と非介入群の間でアウトカムを比較する



(Figure 2 in Benedetto et al., Eur J Cardiothorac Surg., 2018)

4.3.2.2 必要な仮定

1. 未測定の交絡因子が存在しない（共変量で調整すると潜在アウトカムと介入が独立）
2. 各個人が介入を受ける確率と介入を受けない確率が共に 0 より大きい（この仮定は、プロペンシティスコアを定義するために必要な共変量で条件付しても成立する必要がある）
3. ある個人への介入は、その個人のみに効果をもたらす（ほかの個人への介入によって影響を受けない）

4.3.2.3 メリット

- （層別化やマッチングと比較して）調整すべき交絡因子の数が多くても利用可能である。プロペンシティスコアは複数の交絡因子を一次元の数値としてまとめて処理する
- （回帰分析による交絡因子の調整と比較して）介入因子、アウトカム、交絡因子間のモデルの誤特定に頑強である（特に、プロペンシティスコアによるマッチングまたは層別化解析）

4.3.2.4 デメリット

- 未測定の交絡因子によるバイアス是对処できない
- プロペンシティスコアの推定モデルに以下の変数を含めしまうとバイアスを引き起こす
 - ・介入のアウトカムへの効果を媒介する変数

- ・ 介入（もしくは介入の原因）とアウトカム（もしくはアウトカムの原因）の両方の結果となる変数

- ・ 操作変数

- プロペンシティスコアの推定にはアウトカムの予測因子を含めたほうがよい（理由：推定の統計的な効率性を高める）
- 推定する介入効果が、どの集団における効果なのか、使用するプロペンシティスコアの手法によって異なる
 - ・ たとえば、プロペンシティスコアによって介入群を非介入群にマッチングした場合、推定される治療効果は、「実際に治療を受けた集団」における介入効果であり、「実際に治療を受けていない集団」についての介入効果については推論できない

詳細は、[方法論を解説した論文]を参照

4.3.2.5 方法論を解説した論文

- Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: from naive enthusiasm to intuitive understanding. *Stat Methods Med Res.* 2012;21(3):273-93. Epub 2011/01/26. doi: 10.1177/0962280210394483. PubMed PMID: 21262780.
- Haukoos JS, Lewis RJ. The Propensity Score. *JAMA.* 2015;314(15):1637-8. Epub 2015/10/27. doi: 10.1001/jama.2015.13480. PubMed PMID: 26501539; PMCID: PMC4866501.
- Thomas L, Li F, Pencina M. Using Propensity Score Methods to Create Target Populations in Observational Clinical Research. *JAMA.* 2020;323(5):466-7. Epub 2020/01/11. doi: 10.1001/jama.2019.21558. PubMed PMID: 31922529.
- Zhao QY, Luo JC, Su Y, Zhang YJ, Tu GW, Luo Z. Propensity score matching with R: conventional methods and new features. *Ann Transl Med.* 2021;9(9):812. Epub 2021/07/17. doi: 10.21037/atm-20-3998. PubMed PMID: 34268425; PMCID: PMC8246231.

4.3.2.6 方法論を応用した論文

- Nofuji Y, Seino S, Abe T, Yokoyama Y, Narita M, Murayama H, Shinkai S, Kitamura A, Fujiwara Y. Effects of community-based frailty-preventing intervention on all-cause and cause-specific functional disability in older adults living in rural Japan: A propensity score analysis. *Prev Med.* 2023; 169:107449. Epub 2023/02/16. doi: 10.1016/j.ypmed.2023.107449. PubMed PMID: 36791973.

日本の自治体において、地域在住の高齢者を対象とした虚弱予防プログラムの効果をプロペンシティスコアマッチングと逆確率重み付けの手法で推定した。虚弱予防プログラムの参加者を含む高齢者 3350 人のデータを分析し、プロペンシティスコアマッチングと逆確率重み付けの解析においてプログラム参加者の機能障害の発生（ハザード比）が有意に少なかった。

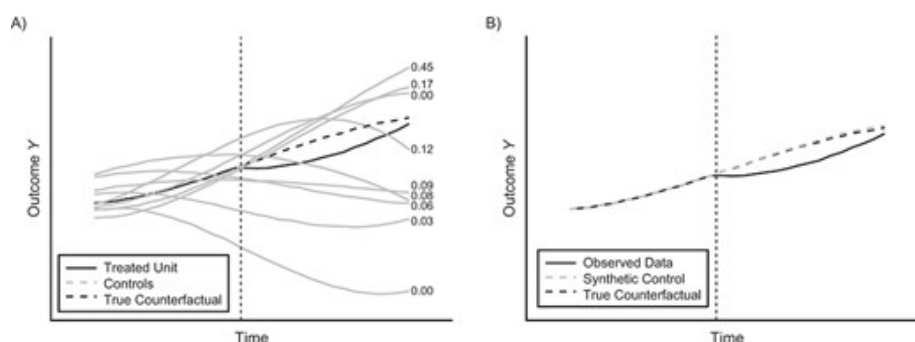
4.3.2.7 予防健康づくりアプリにおける使用例案

- (実験データではない)リアルワールドデータを用いた体重減少アプリの効果検証

4.3.3 Synthetic Control

4.3.3.1 概要

- Synthetic control method (SCM) は、明確に定義された集団に対して、導入時点が明確な介入の効果を評価するために開発された因果推論の手法
- 経時的に測定されたデータを用いて、介入を受けていない非介入群(のうちのいくつか)の重み付け平均を、介入群が介入を受けていない場合の反実仮想的なアウトカムとする。こうした作成された仮想的な非介入群を synthetic control (SC) と呼ぶ
- SC における介入後時点のアウトカムを、介入群における実際の介入後のアウトカムと比較することで、介入の効果を推定する
- SC は、単一の介入群と複数の非介入群を含む経時的なデータ(パネルデータ)に最適化アルゴリズム*適応することで作成される
(*介入前のアウトカムや共変量に関して、介入群と SC の類似性を最適化するアルゴリズム)



(Figure 1 in Bonander et al., AJE, 2021)

4.3.3.2 必要な仮定

- SC をつくるために使用される非介入群が介入群と類似している (例: アメリカのある州に導入された介入の効果を推定するために、アメリカの他の州すべてを使用して SC をつくることは適切か?)

- 介入群への介入が、非介入群に影響を与えない (spillover effects がない)
- 介入前または介入後のいずれの期間においても、介入群または非介入群に、アウトカムに差別的な影響を及ぼす可能性のある他のイベントがない

4.3.3.3 メリット

- 介入群と非介入群の数が少ない場合でも適応できる (例: 都道府県レベルで介入の有無が異なる政策の評価、など)
- 介入群と非介入群の介入開始前のアウトカムについて、経時的なトレンドが平行である必要がない (差分の差分法との比較)
- 時間によって変化しない未測定の変因のみでなく、時間によって変化する未測定の変因があっても因果効果を推定できる

4.3.3.4 デメリット

- 信頼区間や p 値の計算が困難: 非介入群を介入群と見立てた placebo effect を推定し、SCM との推定値を比較することで、SCM の推定値がバイアスによるものかどうかを考察できる
- SC 作成に用いる変数に関して、介入群における値は SCM に用いられる非介入群での分布の「内部」にある必要がある (Figure 4 in Bonander et al., AJE)
- アウトカムの測定誤差が大きいデータを SCM で用いるとバイアスを引き起こしうる

4.3.3.5 方法論を解説した論文

- Bouttelle J, Craig P, Lewsey J, Robinson M, Popham F. Synthetic control methodology as a tool for evaluating population-level health interventions. *J Epidemiol Community Health*. 2018;72(8):673-8. Epub 2018/04/15. doi: 10.1136/jech-2017-210106. PubMed PMID: 29653993; PMCID: PMC6204967.
- Bonander C, Humphreys D, Degli Esposti M. Synthetic Control Methods for the Evaluation of Single-Unit Interventions in Epidemiology: A Tutorial. *Am J Epidemiol*. 2021;190(12):2700-11. Epub 2021/08/04. doi: 10.1093/aje/kwab211. PubMed PMID: 34343240; PMCID: PMC8634614.

4.3.3.6 方法論を応用した論文

- Rado MK, van Lenthe FJ, Sheikh A, Been JV. Investigating the effects of comprehensive smoke-free legislation on neonatal and infant mortality in Thailand using the synthetic control method. *EClinicalMedicine*. 2020;27:100560. Epub 2020/10/10. doi: 10.1016/j.eclinm.2020.100560. PubMed PMID: 33033797; PMCID: PMC7533363.

2010 年にタイで施行された禁煙法が新生児死亡率および乳児死亡率に及ぼす影響を SCM を用いて推定した。禁煙法制化を実施していない中所得国 (世界銀行と Penn World の 2001-2017 年のパネルデータから抽出) を用いて SC を作成した。SCM によ

って、新生児死亡率は 2.9 %/年、乳児死亡率は 2.8%/年減少し、8 年間で推定 7463 人の乳児死亡（新生児死亡 4623 人を含む）が回避された、と推定された。

4.3.3.7 予防健康づくりアプリにおける使用例案

- 禁煙アプリを導入した国における禁煙アプリが喫煙率に与える影響を推定するために、導入していない国を非介入群として用いて SCM を行う

5 Target Trial Emulation による拡張

5.1 概要

- Target Trial Emulation は観察データを用いて因果効果の推定を行う際に用いることを提唱されたフレームワークである
- 手順としては、1) まず研究疑問に対して仮想的な RCT（標的試験：Target Trial）の研究計画を提示し、2) 次に観察データを用い、Target Trial を明示的に模倣する (Emulation)
- 標的試験の研究計画とその模倣は、それぞれ 7 つの要素に分解される

標的試験の研究計画	観察データを用いた標的試験の模倣
1. 資格基準	1. 資格基準
2. 治療戦略	2. 治療戦略
3. 治療割り付け	3. 治療割り付け
4. 追跡期間	4. 追跡期間
5. アウトカム	5. アウトカム
6. 因果効果の推定対象	6. 因果効果の推定対象
7. 統計解析	7. 統計解析

- この 2 段階の手順を踏むことで、標的試験を模倣する際に生じうる研究デザイン上の誤りを発見し、観察研究で頻繁にみられるバイアス（例：既存使用者バイアスや不死時間バイアスなど）を回避することができる

5.2 方法論を解説した論文

- Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. JAMA. 2022 Dec 27;328(24):2446-2447. doi: 10.1001/jama.2022.21383.
- Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016 Nov;79:70-75. doi: 10.1016/j.jclinepi.2016.04.014.
- Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr 15;183(8):758-64. doi: 10.1093/aje/kwv254.

5.3 方法論を応用した論文

- Mori Y, Komura T, Adomi M, Yagi R, Fukuma S, Kondo N, Yanagita M, Duru OK, Tuttle KR, Inoue K. Sodium-glucose cotransporter 2 inhibitors and cardiovascular events among patients with type 2 diabetes and low-to-normal body mass index: a nationwide cohort study. Cardiovasc Diabetol. 2024 Oct 22;23(1):372. doi: 10.1186/s12933-024-02478-7.

SGLT2 阻害薬開始群、DPP4 阻害薬継続群の 2 群について、心血管病予防効果を比較した。まず、標的試験の研究計画を提示し、次に、日本の全国健康保険協会のデータベース（2015 年—2022 年）を使用し、この標的試験を模倣した。通常の RCT では薬剤の新規開始者を比較する研究デザインが用いられるが、実臨床では糖尿病治療薬の変更は頻繁に行われるため、SGLT2 阻害薬新規開始者と DPP4 阻害薬継続者を比較する RCT を標的試験として設定した。この Trial を模倣するために、Prevalent new-user design という研究デザインを使用した

傾向スコアマッチングにより 14 万人が各群に組み入れられ、追跡期間中に 8 千人に心血管疾患の発症を認めた。この結果は、過去の大規模 RCT で報告されているものと合致していた。さらに、過去の RCT で対象となっていなかった BMI25 未満の患者について効果を検証したところ、BMI25 未満の患者群では、明らかな心血管病予防効果を認めなかった

5.4 予防健康づくりアプリにおける使用例案

- 禁煙アプリの導入による因果効果を推定するために、禁煙アプリを導入したクリニックとそうでないクリニックの比較を行う。この際、**Target Trial Emulation** フレームワークを用い、まず仮想的な標的試験を提示し、次に観察データを用いてこの標的試験を模倣する

6 効果の異質性評価

6.1 サブグループ解析

6.1.1 概要

- 研究対象集団全体における因果効果を推定した後、因果効果の異質性を検討するために、ある下位集団について因果効果を推定する解析。RCT で頻繁に用いられ、例として、性別や年齢群によって区分された集団ごとに因果効果を推定することがある

6.1.2 メリット

- 単一の属性で区分した集団ごとに因果効果を推定するため、解釈が容易である（例：男性と女性で因果効果が異なる場合、性別による効果の異質性があるという解釈になる）

6.1.3 デメリット

- 複数の属性で区分した場合の因果効果が検討されることは非常に稀である（例：年齢群と性別、既存疾患の有無を組み合わせた、効果の異質性の検討など）ため、複雑な異質性はほとんど評価されない
- RCT では予定しているサブグループ解析を事前にプロトコルで明記する必要があるため、効果の異質性が見逃される可能性がある

6.2 機械学習手法による異質性評価

手法	概要	利点	限界
罰則付き回帰（LASSO, ridge, elastic net）	・ LASSO は、回帰係数を縮小する正則化回帰モデルであり、新しいサンプルでの予測性能を最大化することを目的とする - また、いくつかの予測変数をモデルから完全に除外することで、変数選択も行う ・ Ridge は LASSO に似ているが、変数選択を行わない ・ Elastic net は、LASSO と Ridge を組み合わせたものである 上記 3 つの正則化回帰モデルでは、条件付き平均因果効果（CATE）を推定するために、治療変数と共変量の交互作用を含めることができる。	・ 変数選択を行う ・ 手法がシンプル	・ 共変量との交互作用を事前に指定していない場合、考慮できない ・ 高次元の共変量の組み合わせにおいて、交互作用の検出が困難
Causal	・ Causal forest は、フォ	・ ノンパラメトリ	・ 計算負荷が大きい

forest	レストベースのアルゴリズム ・個人の特性によって定義されるリーフ間の因果効果推定値の分散を最大化するようにサンプルを分割する ・R-learner フレームワークを用い損失関数を最小化する ・「誠実な (honest)」推定アプローチを用いて、決定木作成と CATE 推定に異なるサンプルを使用することで、決定木の構造と効果推定の独立性を保証する	ック（非線形で複雑な交互作用の誤特定に対して影響を受けにくい） ・不確実性の推定プロセスがアルゴリズムに内包されている	い ・比のスケール（リスク比など）で評価ができない
Bayesian causal forest	・ Bayesian causal forest は 2 つのベイジアン加法回帰木（BART）関数を用いて効果の異質性を評価する ・基礎木の合計を計算してアウトカムを予測し、残差を最小化するため、MCMC により木を反復的に更新する ・過学習を回避し、特に観察研究において発生しやすい「正則化による交絡」を低減する	・ノンパラメトリック（非線形で複雑な交互作用の誤特定に対して影響を受けにくい） ・不確実性の推定プロセスがアルゴリズムに内包されている	・計算負荷が大きい ・比のスケール（リスク比など）で評価ができない
Meta-learner Framework			
S-learner	・治療変数と共変量の交互作用をモデル化する基礎学習器を用い、治療群および対照群のアウトカムを予測する	・手法がシンプル ・治療効果が単純またはゼロの場合、T-learner よりも優れた性能を発揮する	・処置効果の推定を直接最適化しない ・変数選択を行う基礎学習器を用いる場合、治療変数がモデルから除外される危険性がある ・治療群と比較群のサンプルサイズが大きく異なるとき推定が不安定
T-learner	・基礎学習器を用いて治療	・手法がシンプル	・処置効果の推定

	群と対照群のアウトカムを別々に予測し、その後、両群の期待アウトカム値の差を計算する	・治療群と対照群を別々に明示的にモデリングする ・治療効果が強く異質性が大きい場合、S-learner よりも優れた性能を発揮する	を直接最適化しない ・治療群と比較群のサンプルサイズが大きく異なるとき推定が不安定
X-learner/DR-learner	・観測されたアウトカムと各群の潜在アウトカムの差を用いて、治療群および対照群における治療効果を推定する ・共変量分布の不均衡に対処するため、傾向スコア重み付けを組み込んでいる ・DR-learner は、傾向スコア重み付けの代わりに二重ロバスト推定量を使用する X-learner の類似形	・異質性のある因果効果を直接推定する ・観察研究で頻繁に発生する不均衡なデザインにおいて効果的	・手法が複雑 ・極端な傾向スコアが存在する場合、不安定になる ・傾向スコアモデルの誤特定に対して脆弱である
R-learner	・曝露および周辺アウトカムの傾向スコアを計算し、治療とアウトカムの残差を算出し、これらの残差で定義される損失関数を最小化する ・R-learner は、損失関数を最小化するために、正則化を組み込んだ機械学習を必要とする	・異なるサブサンプルを使用して、局外パラメータ (nuisance parameter) を推定し、直接的に損失関数を構築することで CATE を予測する	・手法が複雑 ・極端な傾向スコアが存在する場合、不安定になる ・傾向スコアモデルの誤特定に対して脆弱である

(Table 3, 4 in Inoue et al., J Clin Epidemiol, 2024 より改変)

6.3 各機械学習手法の詳細（各手法の概要、必要な仮定、メリット、デメリット、方法論を解説した論文）

6.3.1 罰則付き回帰

6.3.1.1 概要

- 大きく LASSO, ridge, elastic net の 3 つに分かれる
- LASSO は L1 正則化項を用いることで、回帰係数を 0 に縮め、予測変数自体をモデルから除く変数選択を行う
- Ridge は LASSO と異なり、回帰係数を 0 まで縮小させることはなく、変数選択を行わない
- Elastic net は LASSO と Ridge を組み合わせたものである

6.3.1.2 メリット

- 手法がシンプルである

6.3.1.3 デメリット

- モデル中に明示しない限り、共変量を含む交互作用項を考慮しない
- 高次元の交互作用を考慮することは難しい

6.3.1.4 方法論を解説した論文

- Imai K, Ratkovic M. Estimating treatment effect heterogeneity in randomized program evaluation. Ann Appl Stat 2013;7(1):443e70. <https://doi.org/10.1214/12-AOAS593>.

6.3.2 Causal Forest

6.3.2.1 概要

- Causal Forest は、ランダムフォレストと類似したツリーベースのアルゴリズム
- 複数の決定木を組み合わせて予測を行うアンサンブル法を用いる
- サブグループ解析と異なり、異質性を示す候補となる数十の属性を予測モデルに組み込み、因果効果の差が最大化されるように決定木が作成される
- 作成された複数の決定木を組み合わせ、条件付き平均因果効果（conditional average treatment effect: CATE）を推定する
- 決定木の作成と CATE 推定に異なるデータを用いることで、データの過学習を避けることができる

6.3.2.2 メリット

- ノンパラメトリック手法であるため、非線形の複雑な交互作用などによるモデル誤特定に対して頑健である
- 推定値のばらつきを定量することが可能である
- サブグループ解析と異なり、効果の異質性を探索的に評価することが可能である

6.3.2.3 デメリット

- 解析の計算機負荷が大きい
- Estimand は差のスケール（リスク差）であるため、比のスケール（リスク比など）の異質性評価はできない。この点は一般的にリスク差がリスク比と比較して別の集団に外挿することが難しい点（Control event rate が異なるため）と合わせて、注意深く考慮する必要がある。

6.3.2.4 方法論を解説した論文

- Athey S, Wager S. Policy Learning with Observational Data. Published online September 4, 2020. doi:10.48550/arXiv.1702.02896
- Yadlowsky S, Fleming S, Shah N, Brunskill E, Wager S. Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects. Published online November 28, 2023. doi:10.48550/arXiv.2111.07966

6.3.3 Bayesian Causal Forest

6.3.3.1 概要

- Bayesian Causal Forest は、Bayesian Additive Regression Trees (ベイズ加法回帰木) と呼ばれるノンパラメトリックなベイズ手法を応用した手法である
- 作成した回帰木を足し合わせてアウトカムを予測し、実際に観測されたアウトカムと比較する。この残差が最小化されるように回帰木の作成ルールを最適化し、マルコフ連鎖モンテカルロを用いて事後分布を推定する

6.3.3.2 メリット

- ノンパラメトリック手法であるため、非線形の複雑な交互作用などによるモデル誤特定に対して頑健である
- 推定値のばらつきを定量することが可能である
- サブグループ解析と異なり、効果の異質性を探索的に評価することが可能である

6.3.3.3 デメリット

- 解析の計算機負荷が大きい
- Estimand は差のスケール（リスク差）であるため、比のスケール（リスク比など）の異質性評価はできない

6.3.3.4 方法論を解説した論文

- Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. *Ann Appl Stat.* 2010;4(1). doi:10.1214/09-AOAS285
- Hill J, Linero A, Murray J. Bayesian Additive Regression Trees: A Review and Look Forward. *Annual Review of Statistics and Its Application.* 2020;7(1):251-278. doi:10.1146/annurev-statistics-031219-041110

6.3.4 方法論を応用した論文

- Inoue K, Athey S, Baicker K, Tsugawa Y. Heterogeneous effects of Medicaid coverage on cardiovascular risk factors: secondary analysis of randomized controlled trial. *BMJ.* 2024; 386: e079377.

オレゴン医療保険実験において、医療保険（メディケア）への加入資格を無作為に割り付けられた RCT データに因果フォレストを応用することで、因果効果の異質性を検討した。結果として、集団全体では医療保険加入による血圧改善効果は認められなかったものの、その効果は個人ごとのばらつき（異質性）があることが明らかとなった。特に介入前の医療費が高かった集団においては血圧改善効果を認めなかった一方で、医療費が低い・もしくはかかっていた集団においては臨床的に意義のある血圧改善効果を認めた。本エビデンスは、医療保険が健康に与える影響について因果フォレストの応用により今までわかっていなかった知見を提供するものであり、White House が発表した 2025 年の President Economic Report にも引用された。

6.3.5 Meta-learner

- 1 つまたは複数の機械学習モデルを組み合わせることで、CATE の推定、因果効果の異質性の評価を行う手法の総称
- S-learner, T-learner, X-learner/DR-learner, R-learner に大別される

6.3.5.1 S-learner

6.3.5.1.1 概要

- 因果効果を直接予測する手法
- 治療群と比較群の両群を含めたデータセットに対して治療変数、共変量とその交互作用項を含めた単一のモデルを用いる

6.3.5.1.2 メリット

- 手法がシンプルで計算機負荷が低い
- 因果効果の異質性が複雑でない場合や因果効果がない場合、T-learner より良いパフォーマンスを発揮する

6.3.5.1.3 デメリット

- 因果効果を推定するために最適化された手法ではない
- 治療群と比較群のサンプルサイズが大きく異なる時推定が不安定

6.3.5.1.4 方法論を解説した論文

- Hill J, Linero A, Murray J. Bayesian additive regression trees: a review and look forward. Annu Rev Stat Its Appl 2020;7(1):251e78.

6.3.5.2 T-learner

6.3.5.2.1 概要

- 因果効果を直接予測する手法
- S-learner と異なり、治療群と比較群を別々にモデリングする

6.3.5.2.2 メリット

- 手法がシンプルで計算機負荷が低い
- 効果の異質性が複雑であるとき、S-learner より良いパフォーマンスを発揮する

6.3.5.2.3 デメリット

- 因果効果を推定するために最適化された手法ではない
- 治療群と比較群のサンプルサイズが大きく異なる時推定が不安定

6.3.5.2.4 方法論を解説した論文

- Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. Proc Natl Acad Sci 2016;113(27):7353e60.

6.3.5.3 X-learner/DR-learner

6.3.5.3.1 概要

- 直接因果効果を予測しない手法
- 潜在アウトカムアプローチを用い、実際に観察されたアウトカムと反実仮想アウトカムの差をとることで因果効果を推定する
- **X-learner** のフレームワーク内で因果効果の推定に二重頑健推定量を用いるものを **DR-learner** と呼称する

6.3.5.3.2 メリット

- 傾向スコアによる重みづけを用いることで、治療群と比較群のサンプルサイズが大きく異なる場合でも推定が安定する

6.3.5.3.3 デメリット

- 手法が複雑
- 極端な傾向スコアの存在下では推定が不安定になる
- 傾向スコアモデルの誤特定に影響される

6.3.5.3.4 方法論を解説した論文

- Kunzel SR, Sekhon JS, Bickel PJ, Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc Natl Acad Sci* 2019;116(10):4156e65.
- Kennedy EH. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electron J Stat* 2023;17(2):3008e49.

6.3.5.4 R-learner

6.3.5.4.1 概要

- 直接因果効果を予測しない手法
- 正則化を用い、**CATE** 推定を最適化する

6.3.5.4.2 メリット

- 最適化に用いるデータセットと **CATE** 推定に用いるデータセットを分けることで過学習を回避する

6.3.5.4.3 デメリット

- 手法が複雑
- 極端な傾向スコアの存在下では推定が不安定になる
- 傾向スコアモデルの誤特定に影響される

6.3.5.4.4 方法論を解説した論文

- Nie X, Wager S. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* 2021;108(2):299e319.

6.3.5.5 方法論を応用した論文

- Sinha P, Spicer A, Delucchi KL, McAuley DF, Calfee CS, Churpek MM. Comparison of machine learning clustering algorithms for detecting heterogeneity of treatment effect in acute respiratory distress syndrome: A secondary analysis of three randomised controlled trials. EBioMedicine. 2021 Dec;74:103697. doi: 10.1016/j.ebiom.2021.103697.

急性呼吸窮迫症候群に対する治療効果を検討した 3 つの RCT データを用い、90 日死亡をアウトカムとした効果の異質性を検討した。ランダムフォレストを base learner とした X-learner、ベイズ加法回帰木を base learner とした X-learner を用い、治療効果の異なるクラスターが特定された

6.4 予防健康づくりアプリにおける使用例案

- 対象集団全体での禁煙アプリの導入に関する因果効果を推定した後に、上記機械学習モデルを用い、効果の異質性を検討する。実際に用いる手法は、ランダム割り付けの有無、治療群・比較群のサンプルサイズを参考に決定する