

非ランダム化比較試験データを用いて治療効果を
推定するための統計的手法

日本製薬工業協会

医薬品評価委員会 データサイエンス部会

2019 年度継続タスクフォース 5

「医療経済評価に関連する統計的手法の整理と紹介」

Ver 1.0

2020 年 2 月

本報告書の中で **Technical Support Document 17 (Decision Support Unit, National Institute for Health and Care Excellence)** を引用／参考にした記載は、本タスクフォースの解釈であり、NICE には一切の責任はない。また TSD 17 のすべてを紹介したわけではなく、一部を抜粋し紹介している。それらの紹介にあたり、タスクフォースによる解釈や追加の説明を加えている。これらは本タスクが独自にまとめたものであり、NICE から見解を得たものではない。

目次

はじめに	3
1. ランダム化比較試験と非ランダム化比較試験	4
2. なぜ因果推論が必要か?	5
3. 因果推論の考え方	6
3.1 RUBIN の因果モデルと潜在アウトカム	6
3.2 交絡因子	6
3.3 治療の選択メカニズム	7
3.4 平均因果効果	8
3.5 一般的事項	9
4. 因果効果の推定	9
4.1 観察可能な変数による治療の選択	9
4.1.1 回帰モデルに基づく方法	11
4.1.2 <i>Inverse Probability Weighting</i> 法	11
4.1.3 2重ロバスト法	12
4.1.4 傾向スコアを説明変数に含めた回帰分析	13
4.1.5 マッチング法	13
4.1.6 マッチングした対象集団に対する回帰分析	14
4.1.7 実装	15
4.2 観察不可能な変数による治療の選択	18
4.2.1 操作変数法	18
4.2.2 実装	20
4.3 自然実験	21
4.3.1 差の差分析	21
4.3.2 回帰不連続デザイン	22
5. 適切な解析手法の選択	24
おわりに	25
引用文献	26
APPENDIX	29

はじめに

近年、医療情報データベースやリアルワールドデータを活用してエビデンスを構築する環境が整いつつあり、National Institute for Health and Care Excellence (NICE) の総合的評価（アプレイザル）においても、興味のある治療効果の推定値を得るために、ランダム化比較試験を唯一の情報源とするのではなく、非ランダム化比較試験ⁱの結果を利用するケースが増えつつある。

2013年に公表されたNICEのMethod Guide¹では、非ランダム化比較試験で生じる潜在的なバイアスとして、交絡、非盲検、不完全な追跡及び事前規定されていない評価項目などが原因で生じるバイアスを挙げており、これらの潜在的なバイアスを特定し、可能であれば定量化した上で調整することを推奨している。しかしながら、このMethod Guideでは、非ランダム化比較試験データを用いた治療効果の推定方法や潜在的なバイアスを適切に対処するための方法に関する説明はなされていなかった。このように、医療経済評価における非ランダム化比較試験データの活用方法に関する明確な指針の欠如は、解析手法の不適切でかつ一貫性のない使用を招く可能性があり、結果としてバイアスを伴った治療効果を推定することで、医療技術の有用性や費用対効果に基づく決定に悪影響を及ぼす可能性があった。

そうした中、2015年に公表されたテクニカルサポート文書（Technical Support Document : TSD）17²において、Method Guideでは説明がなされていなかった、非ランダム化比較試験の個別患者データ（Individual Patient Data : IPD）を用いて因果効果を推定する方法の説明や、非ランダム化比較試験データを用いて因果効果の推定を行ったアプレイザルの紹介、並びに非ランダム化比較試験のデータ解析の質を評価するためのチェックリストの紹介などがなされた。更に、TSD 17では、今後の技術評価の質と透明性を改善するための提言を行っている。

更に、2016年にはリアルワールドデータの活用に関するガイダンス³が公表された。このガイダンスはTSD 17を基にして作成されており、NICEの意思決定に用いる治療効果を推定するため、どのようにリアルワールドデータを活用するかについて、事例を含めた紹介がなされている。また、これまでも、NICEでは幅広く科学的エビデンスを利用してきたが、今後は、病院の手術記録やレジストリなどを含めたりリアルワールドデータの活用を更に拡大する意向であることが2019年に発表された。

このように、各種ガイダンスが順次公表され、NICEの医療経済評価に非ランダム化比較試験データを用いる環境が整いつつある。本邦では2019年より医薬品及び医療機器の価格調整を目的に費用対効果評価制度が本格導入された。今後、本邦でも費用対効果評価を実践する機会が増加するにつれて、当該評価に非ランダム化比較試験データを利用する機会も増加することが予想される。一方で、非ランダム化比較試験では治療効果の推定に様々なバイアスを伴うため、研究デザインに応じてバイアスを最小化するための適切な解析手法を選択する必要がある。しかしながら、その際に重要となる因果推論に基づく解析手法に対する理解は、本邦において、これから実際に費用対効果評価を実践する担当者の中ではまだ十分に広まっていないと考えられる。そこで、本報告書では非ランダム化比較試験のデータ解析で一般に用いられる解析手法の解説を行う。なお、TSD 17では、NICEのアプレイザルにかかわる製薬企業の担当者やNICEのレビューアーを対象に、高度なプログラミング技術や生物統計学／計量経済学の

ⁱ TSD 17では観察研究を含めランダム化を伴わない試験を非ランダム化比較試験と呼んでいる。本報告書でも、観察研究を含めて非ランダム化比較試験と呼ぶことにする。

知識を必要とせずに、標準的な統計ソフトウェアを用いて適用可能な手法に焦点を当てている。本報告書は、本邦において、これから費用対効果評価を実践する担当者が因果推論に基づく解析手法に対する基本的な理解を得ることを目的としていることから、本報告書でもこれらの比較的实施が容易な手法を中心に解説する。

1. ランダム化比較試験と非ランダム化比較試験

ランダム化比較試験の場合、患者単位あるいはクラスター単位の無作為割付けにより、アウトカムに影響を及ぼす因子（未測定因子も含む）の分布は群間で平均的に同じになるため、例えば、単純なランダム化2群比較対照試験の場合、介入群ⁱⁱに割付けられた対象者集団のアウトカムと比較対照群に割付けられた対象者集団のアウトカムを比較することで、介入群の対照群に対する治療効果を推定することが可能となる。一方、非ランダム化比較試験の場合、各対象者は無作為割付け以外のメカニズムに基づいて治療を割付けられたり、治療を選択したりするため、単純にアウトカムを群間比較しただけでは、興味のある治療効果をバイアスなく推定することはできない。

非ランダム化試験を分類する方法は複数あるが、例えば、治療の割付けが研究実施者のコントロール下にあるか否か（実験研究か観察研究か）ⁱⁱⁱ、対照群を有するか否か、観察の時間軸の方向性（治療や曝露とアウトカムの時間的順序）が前向きか後ろ向きかなどによって分類することができる⁴。例えば、疑似ランダム化試験では、ランダム化のような洗練された方法には及ばないものの、研究実施者により患者は介入群か対照群の何れかの群に割付けられる（例：生年月日や医療記録番号による割付け）。疑似実験では、ランダム化は実施しないが、研究実施者が患者の割付けをコントロールする場合がある（例：研究実施者の選好による割付け）。一方、観察研究では、研究を意図した直接的な介入は行わず、例えば、日常診療などの経過観察を行う。次に、前後比較試験では、対照群を有さないため、治療による興味のあるアウトカムの変化を評価する。更に、この前後比較試験では、対照群を設定し、介入群と対照群間でアウトカムの変化を比較することができる。コホート研究には前向きの場合と後ろ向きの場合があり、同時的コホート研究では、同時期に介入を受けていない（または、異なる介入を受けている）対象者を経時的に追跡する。一方、歴史的コホート研究では、過去のある期間の介入を受けていなかったコホートと比較する。症例対照研究では、アウトカム（ケースとコントロール）を測定した下で後ろ向きに曝露の影響を評価する。また、縦断研究と異なり横断研究では、ある一時点におけるアウトカムと他の興味のある変数の関連を評価する。ケースシリーズ研究では、ある介入を受けた対象者集団のアウトカムを評価する。なお、これらの非ランダム化試験のデザインは医薬品開発では馴染みが薄いものの、疫学研究では一般的によく用いられているものであり、詳細については各書籍^{5,6}を参照いただきたい。また、リアルワールドデータの活用に関するガイダンスのAppendixには、実験研究と観察研究を分類するためのアルゴリズム^{iv}が紹介されており、そちらも合わせて参照いただきたい。

なお、本報告書では、対照群を有する非ランダム化比較試験に着目し、無作為割付け以外のメカニズム

ⁱⁱ 2群比較試験では、一般に両群とも何らかの介入を受けるが、本報告書では、興味のある治療を割付けられた群を介入群、もう一方の群を対照群と呼ぶことにする。また、観察研究の場合もこの呼び方を用いることとする。

ⁱⁱⁱ 非ランダム化試験と観察研究は一般に区別される場合が多いが、ここでの分類では、観察研究は非ランダム化試験に含まれる。

^{iv} 本項で述べた分類法と一部異なる。

に基づいて各対象者が治療を選択することにより生じる交絡バイアスを制御する方法を中心に説明する。

2. なぜ因果推論が必要か？

飲酒と心筋梗塞の関連を調べるため、前向きのコホート研究を実施した結果、Table 1 のデータ（仮想例）が得られたとする。飲酒する群、飲酒しない群で心筋梗塞を発症した対象者の割合はそれぞれ 6.7%、3.1%で、飲酒しない群の方が 3.6%低かった。

Table 1 心筋梗塞の発症状況の仮想例（佐藤ら⁷の書籍の表 1 を一部改変）

飲酒	心筋梗塞の発症		合計
	あり	なし	
する群	71 (6.7%)	988 (93.3%)	1059
しない群	29 (3.1%)	912 (96.9%)	941
合計	100	1900	2000

この結果を喫煙の有無で層別した結果、Table 2 のデータ（仮想例）が得られたとする。非喫煙者集団では、飲酒する群、飲酒しない群で心筋梗塞を発症した対象者の割合は何れも 2.6%、喫煙者集団では、飲酒する群、飲酒しない群で心筋梗塞を発症した対象者の割合は何れも 8.4%で、層ごとでは両群間に差は認められなかった。

Table 2 心筋梗塞の発症状況の仮想例（喫煙の有無別）（佐藤らの書籍⁷の表 2 を一部改変）

飲酒	非喫煙			喫煙		
	心筋梗塞の発症		合計	心筋梗塞の発症		合計
	あり	なし		あり	なし	
する群	8 (2.6%)	304 (97.4%)	312	63 (8.4%)	684 (91.6%)	747
しない群	22 (2.6%)	836 (97.4%)	858	7 (8.4%)	76 (91.6%)	83
合計	30	1140	1170	70	760	830

これらの結果から導かれる結論としては、飲酒をすると心筋梗塞の発症リスクが上昇する（因果関係がある）、それとも、飲酒をしてもしなくても心筋梗塞の発症リスクは変わらない（因果関係はない）のどちらが正しいのであろうか？

正しいのは後者である。つまり、喫煙は心筋梗塞の発症の原因であるため、Table 1 は、喫煙を介して飲酒と心筋梗塞の発症に疑似的な関連が生じている状況であり、実際には飲酒と心筋梗塞の発症に因果関

係はない (Table 2)。従って、Table 1 の結果から、飲酒をしたから心筋梗塞の発症リスクが上昇したと解釈するのは誤りである。この仮想例の場合は、考慮すべき因子が既知の因子一つだけ (喫煙のみ) であるので、比較的容易に結果を正しく解釈することができるが、実際の非ランダム化比較試験では未知の因子を含め考慮すべき因子が複数あり、それらが互いに影響し合う複雑な状況にある。そのような状況下で、ある治療とアウトカムの間に関連が認められた場合、それが疑似的な関連なのか (因果関係はない)、それとも、ある治療を受けたからアウトカムに変化が生じた (ある治療を受けていなかったとしたら、アウトカムに変化は生じていなかった) のかを明らかにするためには適切な解析手法を用いて検討する必要がある。その際に、因果推論に基づく考え方が重要となる。

3. 因果推論の考え方

本項では、因果推論の基礎となる考え方について概念を中心に解説する。なお、本項で述べる内容については、日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2016 年度タスクフォース 4 欠測のあるデータの解析チーム 因果推論サブチームによる成果物「臨床試験の estimand を理解するための因果推論」⁸や、日本製薬工業協会 医薬品評価委員会 データサイエンス部会シンポジウム⁹などで、数式を交えて解説がなされているので、そちらも合わせて参照いただきたい。また、本報告書で用いる用語は「疫学」の因果推論を基本とする。同じ概念であっても、分野によって用語が異なるため、注意いただきたい。

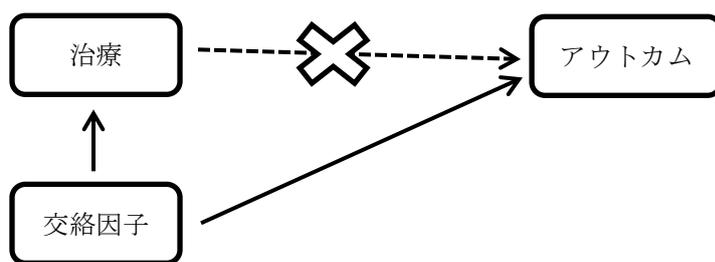
3.1 Rubin の因果モデルと潜在アウトカム

因果効果の推定に関する近代の方法論は、Rubin の因果モデルに基礎を置いている¹⁰。Rubin の因果モデルの基本的な考え方は、各個人に対して、治療を受けた場合と受けなかった場合の 2 つの潜在アウトカムを想定し、両方のアウトカムが観察できれば、その差を取ることで個人レベルでの治療効果を推定できるという点にある (ただし、幾つかの仮定は必要。詳細は 3.5 項参照)。しかし実際には、一方のアウトカムが観察されれば、もう一方のアウトカム (反事実アウトカム) を観察することはできないため (因果推論の根源問題と呼ばれる)、個人レベルでの治療効果を推定できない、という問題に直面する。その解決策の一つがランダム化である。すなわち、個人レベルでの治療効果の推定に着目するのではなく、集団レベルでの治療効果の推定に着目し、集団に対してランダム化を行い、群ごとにその個人レベルの治療効果の平均を取ることで、求めたい因果効果である、「集団のすべての対象者が治療を受けた場合」と「集団のすべての対象者が治療を受けなかった場合」の差を推定するというものである。しかし、適切なランダム化は必ずしも実施可能という訳ではなく、ランダム化を実施できない場合には、非ランダム化比較試験データを用いて集団レベルでの治療効果を推定する必要がある。

3.2 交絡因子

因果推論とは治療がアウトカムに影響を及ぼす (治療とアウトカムに因果関係がある) かどうかを推論

する方法論である。交絡因子とは、治療とアウトカムの両方に影響を及ぼす因子であり、治療効果の推定を妨げる因子となる。なお、交絡因子という用語は一般に疫学研究で用いられる。経済学の分野では一般に内生変数^vという用語が用いられ、それを正しく考慮しなければ欠落変数バイアス^{vi}を引き起こすことになる。交絡因子が存在するかどうかは、有向非巡回グラフ（Directed Acyclic Graph : DAG）によるグラフィカルなアルゴリズムを用いれば見通しが良くなる⁷。このアプローチは、手元にあるデータではなく、医学、疫学及び生物学的な背景情報から各変数間の直接・間接的な因果関係をグラフで表現するというもので、詳細は引用文献を参照いただきたいが、交絡は DAG における共通原因による裏口経路によるバイアスと定義される。Figure 1 は、治療からアウトカムへの直接の矢線がないため因果関係はないが、交絡因子があるため、交絡因子を介して見かけ上関連が生じている状況を表現した DAG である（2 項の仮想例の場合、飲酒が治療、心筋梗塞の発症がアウトカム、喫煙が交絡因子にあたる）。仮にすべての交絡因子が測定されていた場合、交絡因子により条件付けたり（DAG のルールに従えば、交絡因子の周りを四角で囲ったり）、交絡因子から伸びる矢線を削除することにより、交絡バイアスを制御することができる。前者の方法としては、回帰分析、層別解析及び対象者の限定など、後者の方法としては、標準化、Inverse Probability Weighting（IPW）法及びマッチング法などが挙げられる。一方、未測定の交絡因子がある場合、これらの手法では交絡バイアスを制御することができない。



注) 治療からアウトカムへの矢線がないため因果関係はないが、共通原因（交絡因子）の影響により治療とアウトカムに関連がある状況

Figure 1 DAG の例

3.3 治療の選択メカニズム

アウトカムモデルにおいて、治療変数とアウトカムモデルに含まれない変数間に相関がある場合、アウトカムに影響を及ぼす因子の分布が介入群と対照群間で異なり、バイアスが生じる。この相関が治療効果の推定に影響を及ぼすのは、治療とアウトカムの両方に影響を及ぼす観察可能な変数がモデルに含まれていないか、治療とアウトカムの両方に影響を及ぼす観察不可能な変数がある場合である。前者を観察可能な変数による選択（selection on observables）、後者を観察不可能な変数による選択（selection on

^v 交絡因子が存在する場合に原因となる変数で、誤差項との相関がある。

^{vi} アウトカムモデルに含めるべき変数が欠落していることにより生じるバイアス。

unobservables) と呼ぶ^{vii}。なお、観察不可能な変数とは、データとして収集されなかった変数と測定自体が不可能な変数の両方を含む。計量経済モデルでは、治療変数と観察不可能な変数間に相関がある場合、内生性があると言う。この場合、観察可能な変数の影響を調整しただけでは興味のある因果パラメータ^{viii}の一致推定量^{ix}を得ることができない。なお、観察可能な変数による選択とは、治療の選択が観察可能な変数のみに因る（観察不可能な変数に因らない）場合の他、治療の選択に影響を及ぼす観察不可能な変数はあるものの、それがアウトカムと相関しない場合も含む。また、観察不可能な変数による選択とは、治療の選択が観察不可能な変数のみに因る（観察可能な変数に因らない）場合の他、観察可能な変数にも観察不可能な変数にも因る場合も含む。

3.4 平均因果効果

治療効果の推定において興味があるのは、集団における個人レベルの治療効果の分布の特徴であり、その中でも一般に興味があるのは平均である。平均因果効果の種類としては、全集団の平均因果効果（Average Treatment Effect : ATE）、介入群の平均因果効果（Average Treatment Effect on the Treated : ATT）、局所平均因果効果（Local Average Treatment Effect : LATE）などが挙げられる。

ATE は、集団のすべての対象者に対する平均因果効果であり、ランダム化により ATE の一致推定量を得ることができる。この治療効果の推定では、他の治療効果の推定に比べて、より強い仮定を必要とするものの（4 項参照）、それらの仮定の多くは一般にデータから検証することはできない。ATT は、介入群の対象者に対する平均因果効果である。ATT は政策評価や経済学分野でよく用いられる。例えば、失業者に対する失業給付などの政策介入の効果に興味がある場合、介入群の対象者に対して政策介入を実施した場合と実施しなかった場合の効果を推定する方が、非介入群の対象者も含めた全集団に対する ATE を推定するよりも意味があると考えられるからである。ATT は差の差分分析により推定可能である（4.3.1 項参照）。LATE は、類似した特性を有する対象者から成る特定の部分集団の平均因果効果である。LATE は操作変数法や回帰不連続デザインにより推定可能である（4.2.1 項及び 4.3.2 項参照）。以上、3 つの平均因果効果について述べたが、治療効果を推定する上で前提となる仮定はそれぞれ異なるため、まずは目的とする平均因果効果を決める必要がある。ただし、NICE のアプレイザルでは一般に、集団全体に対して介入を行ったときの治療効果、すなわち、周辺因果効果（Marginal Treatment Effect : MTE）である ATE の推定に興味がある。

治療効果には均一な（homogeneous）治療効果と異質な（heterogeneous）治療効果がある。対象者特性（観察されなかった対象者特性を含む）の違いに関わらず、すべての対象者に対する治療効果が同じである場合、その治療効果は均一であると言い、ATE、ATT 及び LATE は同じになる。一方、対象者間で治療効果が異なる場合、その治療効果は異質であると言い、異質性は対象者特性の違いにより生じる。治療の選択はこの異質性の決定因子によりなされるため、ATE、ATT 及び LATE に通常違いが生じる。例えば、ATE と ATT の推定を行い、両者に相違が生じた場合、治療効果の異質性が疑われるということにな

^{vii} TSD 17 では、治療の選択メカニズムを観察不可能な変数の有無により区分しており、本報告書でもこの用語を用いることとする。ほぼ同義であるが、疫学研究では一般に未測定の変数の有無による区分が用いられる。

^{viii} 同じ集団に対する異なる介入により推定されるパラメータ。

^{ix} サンプルサイズが大きくなるにつれて、真値に確率的に収束する推定量。

る。

3.5 一般的事項

様々な解析手法の中から適切な解析手法を選択するためには、各対象者に対する治療の選択メカニズムを十分に理解する必要がある。成立する仮定が異なれば、適用可能な解析手法も異なる。また、治療効果の推定値は、前提となる仮定や構築したモデルに大きく依存するため（詳細は4項参照）、これらの仮定やモデルの適切性については、十分に吟味する必要がある。また、前提となる仮定やモデルが正しいかどうかを検証することは一般に困難な場合が多いため、医学研究で一般によく用いられる感度分析¹¹を行い、異なる仮定やモデルに対して結果がどの程度頑健であるかを定量的に確認することが重要となる。このように、因果効果を推定する場合には、1) 目的とする因果効果を決める（推定対象の他、因果リスク差や因果リスク比などの指標の決定も含む）、2) 前提となる仮定を置く、3) 因果効果を推定するためのモデルを構築する、4) データから因果効果を推定する、5) 感度分析を行い、結果の頑健性を確認する、という一連のステップに従う必要がある。また、当てはめたモデルの頑健性を確認するため、**placebo test^x**（**anti test**とも呼ばれる）が有用な場合がある^{12,13}。その他にも、同じデータを用いた他の研究結果と比較することが有用な場合がある。

アウトカムモデルにパラメトリックモデルを用いる場合には、適切にモデルを特定する必要がある。モデルを用いる場合、例えば、アウトカムと説明変数の間に線形性や加法性などを含め何らかの関係性を仮定することになるため、モデルを誤特定した場合、たとえその他の仮定が成立していたとしても、因果効果を正しく推定することはできない。Kearnsら¹⁴は、過去のアプレイザルで回帰モデルを利用する際に適切な報告を行っていたかどうかを確認するとともに、回帰モデルを利用する際の留意点をチェックリストに纏めているので、必要に応じて参照いただきたい。

因果効果の推定では、反事実モデルに対して仮定を置く必要がある。ある対象者*i*の潜在アウトカムは他の対象者*j*が受けた治療には依存しないこと、並びにどの治療においても、受ける治療の詳細（用量や種類など）は対象者によって異なることを仮定する **Stable Unit Treatment Value Assumption (SUTVA)**、各対象者は2つの潜在アウトカムのうちの一方のみが実際に観察されるという **consistency assumption** などが挙げられる。4.1項で述べる観察可能な変数による治療の選択では、これらの仮定に加え、条件付き交換可能性及び正值性の仮定 (**positivity assumption**) を置く必要がある。なお、これらの反事実モデルに対する仮定の詳細は、各書籍^{15,16}で数式を交えて説明がなされているので、必要に応じて参照いただきたい。

4. 因果効果の推定

4.1 観察可能な変数による治療の選択

^x あるアウトカムに対する介入の効果を評価するために当てはめたアウトカムモデルの頑健性を確認するため、その介入と関連がないアウトカムに対して同一のモデルを当てはめ、介入効果の有無を確認する手法。

観察可能な変数による治療の選択では、条件付き交換可能性及び正值性の仮定が必要となる。条件付き交換可能性の仮定は、no unmeasured confounder（未測定の変数がない）の仮定とか治療の無視できる割り当ての仮定とも呼ばれ、「適切に選択した共変量で条件付けた場合、潜在アウトカムは割付け治療とは独立になる」ことを仮定する。平均因果効果を推定するためには、平均での交換可能性（弱く無視できる割り当てとも呼ばれる）の仮定が成立していればよい。数式で表すと以下のとおりである。

$$E[Y_a | A = 1, C] = E[Y_a | A = 0, C] \quad \text{for } a = 0, 1$$

ただし、 Y_a は潜在アウトカム（ Y_1 ：治療を受けた場合の潜在アウトカム、 Y_0 ：無治療の場合の潜在アウトカム）、 A は割付け治療（1：治療あり、0：治療なし）、 C は共変量を表す。治療の選択に影響を及ぼすと考えられる共変量が十分に測定されており、それらが調整可能である場合、潜在アウトカムの平均は治療と独立とみなすことは適切であると考えることができ、その仮定の下で、因果効果の推定が可能となる。

ただし、共変量としては治療の割付け前に測定された変数を含める必要があり、治療の影響を受ける変数を含めると交換可能性の仮定が成立しなくなる。また、治療の選択に影響を及ぼす変数であったとしても、アウトカムに影響を及ぼす未測定の変数とは独立した変数（操作変数と呼ばれる。4.2.1項参照）を用いてはならない。この操作変数を含めると、それが外生変数^{xi}である場合を除き、治療効果の推定にバイアスを伴う。また、外生変数である場合でも、治療効果のばらつきが大きくなり、効率が悪くなる¹⁷。この交換可能性の仮定の成立は、基本的にデータからは検証することができないことから、文献や専門家の意見を参考に仮定の成否を慎重に吟味する必要がある。また、すべての交絡因子を測定することは現実的には不可能であるため、たとえ未測定の変数が存在したとしても、推定した治療効果は大きく異なることを確認することが重要となる。

ATEの推定では、条件付き交換可能性の仮定とともに正值性の仮定が必要となる。正值性の仮定はオーバーラップの仮定とかマッチングの仮定とも呼ばれ、共変量のどの組み合わせにおいても、介入群と対照群の両群に必ず対象者が存在する可能性があることを仮定する。この仮定は、ある特性を有する対象者が一方の群のみに存在し、もう一方の群には存在しない可能性を排除するために置かれ、この可能性を排除できない場合、ATEを推定することはできない。その場合、3.5項で述べた一連のステップの1)に立ち返り、推定対象を含め、目的とする因果効果を再検討したり、この仮定が成立する集団の抽出を検討したりする必要がある。

Rosenbaumら¹⁸は、交換可能性の仮定とオーバーラップの仮定を合わせて、無視できる割り当ての仮定と呼んでおり、この仮定はATEの推定の基礎となる。一方、ATTの推定では、ATEの推定よりも弱い仮定が成立していればよく、無治療の潜在アウトカムに対する平均での無視できる割り当ての仮定の成立と、介入群の対象者が対照群にも存在する可能性があることが条件となる。

因果効果の推定において最も重要となるのが交換可能性の仮定であるため、まずはこの仮定の適切性について検討する必要がある。次に重要となるのがオーバーラップの仮定である。仮に共変量の分布が

^{xi} アウトカムモデルにおいて誤差項との相関のない変数。

オーバーラップする部分（コモンサポート^{xii}と呼ばれる）が十分でない場合、トリミングやマッチング（4.1.5 項参照）を行い、オーバーラップの仮定が成立する集団を抽出する方法の適用が考えられる。

以下では、観察可能な変数による治療の選択を仮定した手法として、TSD 17 で紹介されている 6 つの手法について解説する。このうち 5 つが傾向スコアを用いた手法である。その他の比較的よく用いられる傾向スコアを用いた手法としては、傾向スコアによる層別解析が挙げられる。なお、傾向スコアの Kategorisierung が不適切だと残差交絡によるバイアスが生じる可能性がある。Rosenbaum¹⁹は、傾向スコアの五分位点を用いた層別解析により、測定された交絡因子により生じるバイアスの約 90% の補正が可能であることを示した。

4.1.1 回帰モデルに基づく方法

回帰モデルに基づく方法^{20,21}では、アウトカムに対して回帰モデルを当てはめ（1 段階目）、モデルに含めた共変量から各対象者が治療を受けた場合と受けなかった場合の潜在アウトカムの予測値を推定し、その予測値の差に基づき平均因果効果を推定する（2 段階目）。その際、治療の無視できる割り当ての仮定が成立するように、適切な共変量を用いて予測モデルを構築することが重要である。平均因果効果の推定には、通常のパラメトリック回帰モデルに加えて、カーネル推定量やシリーズ推定量の利用も可能である^{xiii}。なお、この 2 段階法に基づいて平均因果効果を推定する場合には、潜在アウトカムの予測値の推定に伴うばらつきを考慮して、平均因果効果の標準誤差を適切に調整する必要がある。

回帰モデルに基づく方法では、オーバーラップの仮定の成立が前提となるもの、回帰モデルに含める共変量の分布のコモンサポートの程度に関わらず適用自体は可能である。ノンパラメトリック回帰モデルを用いる場合には、モデルのパラメータを推定する過程で、モデルに含める共変量の分布のコモンサポートの程度を必然的に確認することになるが、パラメトリック回帰モデルを用いる場合には、コモンサポートの程度が十分であるかどうかの確認をしばしば怠りがちになるので、注意する必要がある。

通常回帰分析は、回帰モデルに基づく方法の単純な場合に相当する。通常回帰分析では、均一な治療効果を仮定するケースが多いが、治療効果の異質性を考慮するため、回帰モデルに治療と共変量の交互作用を含めることは可能である。ただし、交互作用がある場合、結果の要約が困難になる可能性がある。また、治療効果の異質性は観察された共変量のみ依存し、観察されていない異質性はないことを前提とする。なお、本手法により推定されるのは一般に因果パラメータではなく関連パラメータ^{xiv}となる。

4.1.2 Inverse Probability Weighting 法

バランススコア（balancing score）とは、それを条件付けることにより、共変量と治療の割り当てが独立となるような共変量の関数のことである。共変量そのものもバランススコアの一つであるが、傾向スコ

^{xii} 一般に介入群と対照群間で傾向スコアの分布が重なる部分を指すことが多い。

^{xiii} カーネル推定量及びシリーズ推定量は何れもノンパラメトリック回帰モデルに基づく推定量。前者はカーネル関数を用いた推定量、後者はスプライン関数や多項式回帰などの柔軟なモデルを用いた推定量。

^{xiv} 違う集団に対する異なる介入により推定されるパラメータ。一般に因果パラメータとは異なる。

アとは、その中で最も粗い次元のバランススコアであり、共変量で条件付けた下での治療を受ける条件付き確率のことである。

IPW 法^{22, 23}に基づく ATE は、測定された共変量から傾向スコアを推定し、その逆数を利用して群ごとに当該治療を受ける確率の逆数で重み付けした重み付き平均の差として表される。IPW 推定量は欠測データの影響を調整した治療効果として解釈可能である。すなわち、測定された共変量により治療の選択がなされる場合、介入群では観察されやすいが、対照群では観察されにくい（対照群では欠測となりやすい）特性を有する対象者や、逆に、介入群では観察されにくい（介入群では欠測となりやすい）が、対照群では観察されやすい特性を有する対象者が存在する。IPW 法では、それぞれの群において、このような当該群での観察確率が低い少数の対象者に大きな重みを与えることで治療効果の補正を行う。例えば、治療を受ける確率が 0.25 である場合、背景に同じような特性を持つ対象者が 3 名いるので、治療を受ける確率の逆数で重み付けることにより、そのような対象者も解析で考慮するというものである。IPW 法は回帰モデルに基づく標準化とも捉えることができる。

IPW 法を用いる場合、傾向スコアに対してモデル化を行い、傾向スコアを適切に予測する必要がある。また、オーバーラップの仮定が成立するように、傾向スコアに対しては十分に柔軟なモデルの当てはめの検討が重要となる。例えば、非線形な関連を考慮した柔軟なモデルを当てはめることで、可能な限り傾向スコアが 0 や 1 になることを防ぐことができる。傾向スコアに対してパラメトリックモデルを当てはめる場合も、共変量の 2 乗項や多項式あるいは交互作用項などを含めることで、モデルの柔軟性を考慮することが可能となる。傾向スコアの推定方法としては、従来のロジスティック回帰モデルを用いた方法から、近年は機械学習アルゴリズムを用いた方法^{24, 25}が提案されている。

傾向スコアの分布のコモンサポートの程度が十分でない場合、一部の対象者の重みが極端に大きくなり（傾向スコアの予測値が 0 に近くなり）、推定が不安定になるという問題が生じる可能性がある。この問題に対処するため、安定化した重みが提案されている^{23, 26}。また、2 値データに対してパラメトリックモデル（プロビットモデル及びロジットモデル）を当てはめる場合、0 や 1 に近い傾向スコアについては、当てはめたモデルに応じてその予測値は異なる傾向にあるため、この点も考慮してモデルを選択する必要がある。

4.1.3 2 重ロバスト法

回帰モデルに基づく方法では、アウトカムに対するモデル化（outcome modeling）は行うが、治療の選択に対するモデル化（exposure modeling）は行わない。一方、IPW 法では、治療を受ける確率に対するモデル化は行うが、アウトカムに対するモデル化は行わない。2 重ロバスト法²⁷とは回帰モデルに基づく方法と IPW 法を組み合わせた方法であり、その利点は 2 つのモデルのうち少なくともどちらか一方を正しく特定できてさえいれば、治療効果の一致推定量を得ることができる点にある。また、両者を正しく特定できていれば、通常の IPW 法よりも漸近分散が小さくなり効率が良くなる。しかしながら、両方のモデルが誤特定されている場合、治療効果の推定にバイアスを伴う。

2 重ロバスト法には IPW 法の拡張（Augmented IPW : A-IPW）法²⁶と回帰モデルに基づく方法の拡張（Regression Adjustment IPW : RA-IPW）法²⁸の 2 つの推定法がある。A-IPW 法は、治療の選択に対する

モデルの誤特定に対して、拡張項 (augmentation term) を用いて補正を行う。IPW 法と同様、傾向スコアの予測値が 0 や 1 に近い場合、推定は不安定になる。一方、RA-IPW 法は、回帰モデルに基づく方法に基づくが、アウトカムに対するモデルの誤特定に対しては、治療を受ける確率の逆数による重みを利用して補正を行う。

4.1.4 傾向スコアを説明変数に含めた回帰分析

傾向スコアを説明変数に含めた回帰分析では、治療と共変量との相関は傾向スコアにより十分に説明可能であると仮定する。最も単純な方法は、アウトカムに対するパラメトリック回帰モデルの当てはめであるが、カーネル推定量やシリース推定量なども利用可能である。傾向スコアに対するモデルを正しく特定できていれば、通常の最小 2 乗法により一致推定量を得ることができる。しかしながら、傾向スコアを説明変数に含めた回帰分析は、一般に効率が悪いことが知られている²⁸。

4.1.5 マッチング法

マッチング法とは、介入群の対象者と 1 つ以上の観察された対象者特性が類似した対照群の対象者を同定することで、ランダム化の状況を模倣する方法である。治療効果に対してパラメトリックな仮定を置くことなく、観察された対象者特性は類似し、割付け群のみ異なる対象集団間で比較を行い、治療効果を推定することが可能となる。マッチング法が適しているのは、対象者数が多く、(傾向スコアを用いる場合には) 傾向スコアの推定に利用可能な共変量が十分にあり、また、介入群の対象者と対照群の対象者が同一の集団に属する (例: 同一の試験で両群共通の様々な共変量を有する) 場合である。ただし、介入群の対象者と対照群の対象者が属する集団が異なる場合でも、仮定を満たすことができれば、マッチングの実施は可能である。

マッチング法には正確法と不正確法がある。正確法が適しているのは、対象者数が多く、少数の離散型の共変量のみを考慮すれば十分な場合である。例えば、性別、人種及び医療機関の 3 つの共変量のみが完全に一致する対象者を対照群から選択してペアを構成する場合である。しかしながら、考慮する共変量が少ない場合、一般にマッチングした対象者は十分に類似していない可能性が高く、無視できる割り当ての仮定の成立が疑わしくなる。一方、不正確法では一般に、考慮する共変量の関数に基づくスカラー量を用いてマッチングを行う^{xv}。最も有名な不正確法の一つが最近傍法である。最近傍法では一般に、治療の選択に影響を及ぼす複数の共変量の距離を利用して、可能な限り類似した対象者を同定する。考慮する共変量が多い場合、次元の問題に対処するため、傾向スコアを利用することも可能である。傾向スコアを利用した最近傍法では、マッチングする際にどの程度の傾向スコアの差までを許容するかを設定することができ (キャリパーマッチングと呼ばれる)、その基準には、傾向スコアあるいは傾向スコアのロジット変換値の標準偏差を用いることが多い。Rosenbaum²⁹は傾向スコアのロジット変換値の標準偏差に 0.25 を掛けた値を用いることを推奨し、Austin³⁰は 0.20 を掛けた値を用いることを推奨している。なお、マッチングを実施した後の共変量の分布が十分に類似していなかった場合、キャリパーの基

^{xv} 複数の共変量の一つの指標に纏めて対象者特性の類似性を検討する方法。例えば、複数の共変量の全体的な距離に近い対象者ほど類似性があるとして、最も距離に近い対象者を対照群から選択してペアを構成する。

準値を再考する必要がある。また、傾向スコアに基づくマッチングを行う場合、治療効果の推定では傾向スコアの推定に伴う誤差を考慮する必要がある³¹。

マッチングの抽出方法には非復元法と復元法がある。非復元法では、同じ対象者は最大でも1度しか抽出されないのに対し、復元法では、同じ対象者が複数回抽出される可能性がある。ただし、非復元法を用いてマッチングを行う場合、対照群の対象者数が少ないと十分なマッチングを実施できない可能性がある。また、1対1のマッチングのみではなく、対照群からk名を選択するk対1のマッチングを実施することも可能である。

マッチング法を用いる場合にも無視できる割り当てとオーバーラップの仮定が必要となる。ATEの推定では、全集団に対する反事実アウトカムを予測する必要があるため、どちらの群の対象者が脱落しても対象集団が変化し、推定した治療効果の解釈が困難になる可能性がある。一方、ATTの推定では、介入群に対する反事実アウトカムのみを予測すればよいので、介入群の対象者が脱落した場合のみ、推定した治療効果の解釈が困難になる可能性がある。この対象者の脱落による影響は、復元法を用いることによりある程度緩和できるが、対照群の対象者数が少ない場合には、対照群の同一対象者が何度も抽出される可能性がある。また、最近傍の対象者のみを許容してマッチングを行えばバイアスを減らすことができるのに対して、その条件を緩めて対照群から抽出する対象者数を増やせば分散を小さくすることができる。このように、バイアスと分散はトレードオフの関係にあり、この点も考慮してマッチングの抽出回数を決定する必要がある。また、追加で抽出した対象者が最初に抽出した対象者ほど十分には類似していなかった場合はバイアスが大きくなるため、キャリパーマッチングがよく用いられる。カーネルを利用した重みで対照群のすべての測定値を重み付けした値を用いてマッチングを行う場合もある（カーネルマッチングと呼ばれる）。

無視できる割り当ての仮定について確認した後は、マッチングを実施した後の共変量の群間でのコモンサポートの程度が十分であるかどうかを確認する必要がある。マッチングを実施した後でもコモンサポートの程度が十分でない場合には、キャリパーの基準値を変更することの他、トリミングを行い、対象集団を再定義する方法も考えられる。

このように、マッチング法には幾つかのオプションがあるため、状況に応じて適切なオプションを決める必要がある。例えば、傾向スコアを利用した最近傍法の枠組みでキャリパーマッチングの実施を計画する場合、傾向スコアに基づきマッチングを実施した後、共変量の群間でのバランスを確認し、バランスが取れていない場合にはキャリパーの基準値の変更などを実施した上で再度マッチングを実施し、共変量の群間でのバランスの取れた2群を構築するという一連の手順を繰り返すことになる。その際、結果に基づく恣意的な傾向スコアモデルの選択を防ぐため、この一連の手順はアウトカムの情報を用いることなく実施すべきである。つまり、マッチングを実施し、解析に用いる対象者を同定してから、群間でのアウトカムの比較を一度だけ実施すべきである。

4.1.6 マッチングした対象集団に対する回帰分析

マッチング法は、回帰モデルに基づく調整解析と組み合わせることができる。傾向スコアを用いたマッチングにより新たに構成した対象集団に対する回帰分析により、マッチングで考慮しきれなかったアン

バランスを調整することができる。Kreif ら³²は、状況によっては、マッチングした対象集団に対する回帰分析の方が2重ロバスト法よりも治療効果の推定に伴うバイアスが小さくなることを示した。

4.1.7 実装

4.1.1 項～4.1.6 項で解説した手法は既存のソフトウェア（SAS や R など）を用いて比較的容易に実装可能である。特に SAS/STAT 14.2 (SAS 9.4)³³からは、CAUSALTRT プロシジャ及び PSMATCH プロシジャが新たに追加され、前者により平均因果効果の推定が、後者によりマッチングの実施が可能となった³⁴。³⁵。医薬品の開発において各規制当局への提出を目的とした解析は、通常 SAS を用いて実施するため、実施可能な手法の制限はあるものの、SAS による実装が可能となった点は非常に大きいといえる。本項でも SAS による実装を中心に解説する。

CAUSALTRT プロシジャ

CAUSALTRT プロシジャ (SAS/STAT 14.2) では、アウトカムが連続型あるいは離散型で、治療変数が2値である場合の平均因果効果の推定が可能である。PROC CAUSALTRT ステートメントにおいて、METHOD=により、回帰モデルに基づく方法 (4.1.1 項)、IPW 法 (4.1.2 項) 及び2重ロバスト法 (4.1.3 項) の実施が可能である。また、ATT を指定することで ATT の推定が可能である (LATE の指定は不可)。

OUTPUT ステートメントにおいて、KEYWORD=により、推定した傾向スコア、逆確率による重み及び各群での潜在アウトカムの予測値などを SAS データセットに出力することができるので、そのデータセットを用いて、例えば、傾向スコアを説明変数に含めた回帰分析 (4.1.4 項) や傾向スコアによる層別解析 (4.1 項) を容易に実施することができる。

更に、CAUSALTRT プロシジャでは、傾向スコアモデル (SAS で実施できるのは、ロジスティック回帰モデル) の診断をグラフィカルに実施できる。PROC CAUSALTRT ステートメントにおいて、COVDIFFFPS を指定することで、傾向スコアモデルの各共変量に対して重み付けする前と後の標準化平均値の (介入群と対照群の) 差と (介入群の対照群に対する) 分散比の算出が可能である。また、PLOTS=で、PSCOVDEN や PSDIST を指定することで、重み付けする前と後の連続変数の各共変量の密度関数や傾向スコアの箱ひげ図を容易に作成可能である。以下には、SAS/STAT 14.2 user's guide, the causaltrt procedure の example 33.1 (SAS/STAT 15.1 user's guide, the causaltrt procedure の example 36.1 と同じ) の example データに対して通常の IPW 法 [なお、SAS のデフォルトは IPW 法の亜型である IPWR (IPW with ratio adjustment) 法である] を適用し ATE を推定する過程で、これらの診断を実施した結果の一部を示す。

SAS Code:

`ods graphics on;`

```
proc causaltrt data=smokingweight covdiffps method=IPW;
```

```
  class Sex Race Education Exercise Activity Quit /desc;
```

```
  psmodel Quit = Sex Age Education Exercise Activity YearsSmoke PerDay
```

```
    /plots=(PSDist pscovden(effects(Age YearsSmoke)));
```

```
  model Change;
```

run;

データの詳細は user's guide を参照いただきたいが、Table 3 より、重み付けを行った結果、考慮したすべての共変量の標準化平均値の差の絶対値は小さくなり、1 つの共変量を除くすべての共変量の分散比も 1 に近づいたことを確認できる。

Table 3 重み付けする前と後の標準化平均値の差と分散比 (SAS の出力結果)

Covariate Differences for Propensity Score Model					
Parameter	Standardized Difference		Variance Ratio		
	Unweighted	Weighted	Unweighted	Weighted	
Sex	1	-0.1603	-0.0200	0.9962	1.0006
Sex	0				
Age		0.2820	0.0318	1.0731	0.9847
Education	5	0.1660	0.0111	1.4610	1.0268
Education	4	-0.0270	0.0196	0.9167	1.0624
Education	3	-0.0472	-0.0015	0.9811	0.9994
Education	2	-0.1116	-0.0034	0.8498	0.9953
Education	1				
Exercise	2	0.0568	-0.0029	1.0252	0.9986
Exercise	1	0.0398	0.0166	1.0119	1.0049
Exercise	0				
Activity	2	0.0740	-0.0074	1.2182	0.9796
Activity	1	0.0268	0.0196	1.0043	1.0029
Activity	0				
YearsSmoke		0.1589	0.0253	1.1846	1.0894
PerDay		-0.2167	0.0027	1.1679	1.3323

Figure 2 より、重み付けを行った結果、連続変数の共変量である年齢 (Age) 及び喫煙年数 (YearsSmoke) の密度関数の群間の重なりが大きくなったことを確認できる。

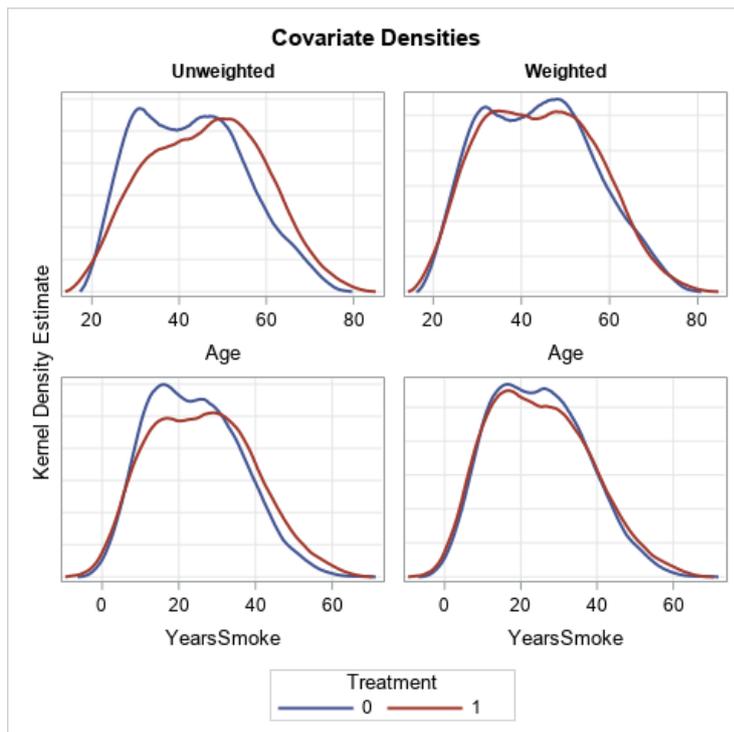


Figure 2 重み付けする前と後の連続変数の共変量の密度関数 (SAS の出力結果)

Figure 3 より、群間での傾向スコアの分布の重なりを確認できる。

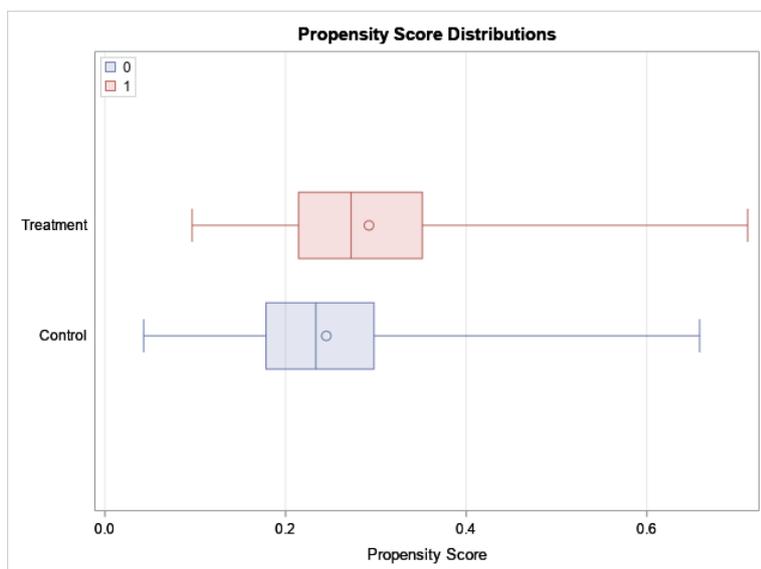


Figure 3 各群の傾向スコアの分布 (SAS の出力結果)

今後、アウトカムが生存期間の場合や2値以外の治療の場合への拡張が望まれる。また、機械学習アル

ゴリズムを用いた方法は R にて実行可能である。

PSMATCH プロシジャ

PSMATCH プロシジャにより、傾向スコアを用いたマッチング (4.1.5 項) が容易に実装可能である。PSMATCH プロシジャでは、傾向スコアモデルに含める共変量を指定し、傾向スコアを推定し、マッチングを実施する。また、マッチングを実施した後、CAUSALTRT プロシジャと同様、共変量の群間でのバランスをグラフィカルに確認することができる。

PSMATCH プロシジャでは、マッチングの条件や方法を設定することができる。PROC PSMATCH ステートメントにおいて、REGION=により、マッチングを行う際のサポート範囲を指定できる。MATCH ステートメントにおいて、METHOD=により最近傍法や復元抽出法、EXACT=により正確法の指定が可能であり、オプションで k=を指定することで、k:1 マッチングの実施が可能である。また、CALIPER=でキャリパー範囲の指定が可能である。

OUTPUT ステートメントにおいて、CAUSALTRT プロシジャと同様、推定した傾向スコアや逆確率による重みなどのデータセットへの出力が可能である。また、マッチング後の対象データを出力することもできるので、例えば、マッチングした対象集団に対する回帰分析 (4.1.6 項) を容易に実施できる。ASSESS ステートメントにおいて、WEIGHT=により、考慮した共変量に対して重み付けした標準化群間差を算出することが可能であり、また、CAUSALTRT プロシジャと同様、PLOTS=により、マッチング前後における各共変量の分布の群間での重なりをグラフィカルに確認することが可能である。また、これらは R のパッケージ「Match」を用いても実施可能である。

4.2 観察不可能な変数による治療の選択

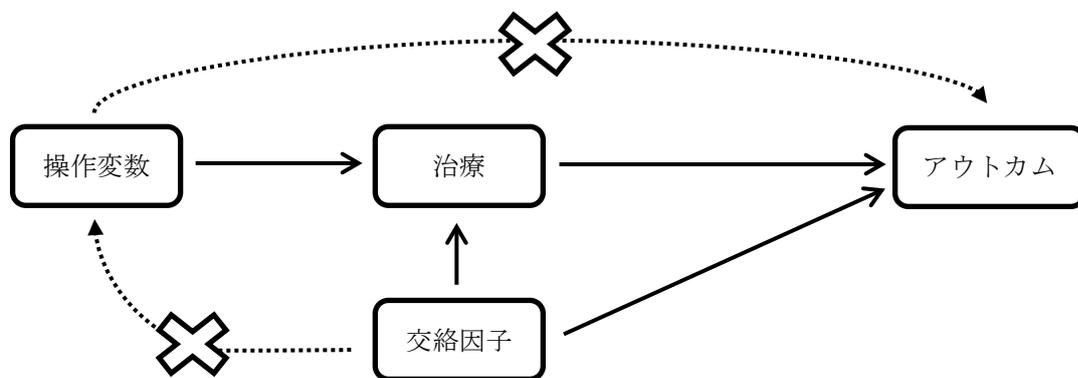
観察不可能な変数とは、上述のとおり、データとして収集されなかった変数や測定自体が不可能な変数を含む。観察不可能な変数が治療の選択に影響を及ぼす（つまり、未測定の交絡因子がある）場合、4.1 項で述べた手法では因果効果を推定することはできない。このように観察不可能な変数が治療の選択に影響を及ぼす場合でも因果効果を推定するための一般的な手法として、ランダム化以外では、操作変数法と回帰不連続デザインが挙げられる。ただし、これらの手法を用いて因果効果を推定するためには、4.1 項とは異なる仮定が必要となる（詳細は 4.2.1 項及び 4.3.2 項参照）。本項では操作変数法について解説し、回帰不連続デザインは自然実験の項で解説する。

4.2.1 操作変数法

治療と関連する変数で、かつ治療を通してのみアウトカムと関連する変数は操作変数と呼ばれる。操作変数法とは、操作変数を利用して、治療のアウトカムに対する因果効果を推定する方法である。操作変数法では適切な操作変数を選択することが非常に重要となるが、今に至るまで、様々な操作変数を利用した事例が報告されている³⁶。例えば、McClellan ら³⁷は、急性心筋梗塞に対する集中治療が高齢者の死

亡率に及ぼす影響を調べるために、自宅と集中治療の実施できる病院との距離を操作変数として用いた。自宅と病院との距離が近いほど、集中治療を受ける確率が高くなるため、自宅と病院との距離は、集中治療を受ける確率に影響を及ぼすが、死亡率には直接影響を及ぼさないと考えられる。この仮定が成立していれば、アウトカムとは関連しない因子（自宅と病院との距離）の影響により、治療の受けやすさ（集中治療を受ける確率）のみ異なる対象者間で比較を行うことが可能である。従って、交絡因子のデータが全くないような状況下でも、操作変数法を適用することで、未測定の交絡因子の影響も調整した因果効果を推定することが可能となる。

操作変数は幾つかの条件を満たす必要がある（Figure 4）。まずは上述したとおり、操作変数は、治療と関連する変数で、かつ治療を通してのみアウトカムと関連する変数である必要があり、この条件は *exclusion restriction* と呼ばれる。次に、操作変数とアウトカムの両方に影響を及ぼす共通の因子（交絡因子）は存在しないことを前提とする。次は操作変数の関連性であり、操作変数は治療と強く関連している必要がある。直感的には、操作変数は治療と関連するばらつきの原因となる変数であり、操作変数が外生的なばらつきを与えることで、ランダム化に近い状況を作り出すことになる。従って、操作変数自体が十分なばらつきを有さないか、あるいは、治療に十分なばらつきをもたらさない場合（弱い操作変数と呼ばれる）、モデルは弱く同定されているため、治療効果の推定に伴うバイアスが増幅することが知られている。このような弱い操作変数に対しては外生性の条件を満たすことが重要となり、上述した集中治療と死亡率の研究では、アウトカムモデル（死亡率に対するモデル）において、自宅と病院との距離が外生変数となる。これらの3条件のうち、操作変数の関連性については比較的容易にデータを用いて評価することができるが、*exclusion restriction* や「共通の因子はない」の仮定の成立はデータから検証することができない。



注) 操作変数からアウトカムへの直接の矢線はない。操作変数とアウトカムの両方に影響を及ぼす共通の因子はない。

Figure 4 操作変数が満たすべき条件

これらの3条件に加え、因果効果を推定するためには更なる仮定が必要となる。治療効果が均一である（集団のすべての対象者に対する治療効果の大きさが同じである）ことを仮定した場合や、治療効果は不均一であるものの、観察されない対象者固有の効果が治療の選択に影響を及ぼさないと仮定した場合、標準的な操作変数法により ATE の推定が可能となる。

一方, *monotonicity* の条件が満たされれば, 観察されない対象者固有の効果が治療の選択に影響を及ぼす場合でも, 操作変数法により *LATE* の推定が可能となる。*Monotonicity* の仮定では, 治療の割り当てと操作変数の単調性が成立しており, 操作変数が治療効果に対して逆効果にならないことを仮定する。例えば, 操作変数を 0, 1 の 2 値変数とした場合, 各対象者を 4 つの集団, すなわち, 1) 操作変数が 1 ならば治療を受け, 操作変数が 0 ならば治療を受けない従順な対象者集団 (*complier*), 2) 操作変数が 0 ならば治療を受け, 操作変数が 1 ならば治療を受けない天邪鬼な対象者集団 (*defier*), 3) 操作変数が 0 でも 1 でも常に治療を受ける対象者集団 (*always-taker*), 4) 操作変数が 0 でも 1 でも常に治療を受けない対象者集団 (*never-taker*) の何れかに分類することができる。*Monotonicity* の仮定では, このうち *defier* がいないことを仮定する。

この仮定の下で, 操作変数法により推定することができる *LATE* とは, 集団全体の平均因果効果ではなく, 操作変数が変化すれば, その操作変数の変化に従って治療を受けるかどうかを決定する従順な対象者集団における治療効果であり, *Complier Average Treatment Effect (CATE)* とも呼ばれる。従って, 操作変数が異なれば, 従順な対象者集団も変化するため, *LATE* も異なる。また, ランダム化比較試験などで得られる *ATE* とは定義が異なるため, それらの結果と単純に直接比較できない点には注意が必要である。

操作変数法による因果効果の推定では, 一般に 2 段階推定法や, それを更に一般化した一般化モーメント法^{xvi}が用いられる。2 段階推定法では 2 つのモデルが必要となり, 1 つ目のモデルでは, 操作変数と測定された交絡因子が治療に与える影響を評価し, 2 つ目のモデルでは, 測定された交絡因子と治療がアウトカムに与える影響を評価する。

操作変数法と傾向スコアに基づく方法 (4.1 項参照) とでは前提となる条件が互いに異なるため, 仮に両方の方法を適用して結果が異なった場合, それらの前提となる条件の違いを確認することが重要となる。上述のとおり, 前提となる仮定の多くはデータからは検証することができないため, 事前知識や外部情報に基づいて個々の前提条件が正しいかどうかを検討し, どちらの手法がより適切であるかを考察する必要がある。

4.2.2 実装

SAS/STAT 14.2 (SAS 9.4) において, 操作変数法による解析を実施するためのプロシジャはない。しかし, 上述のとおり, 操作変数推定量は 2 段階最小 2 乗法などの 2 段階推定法を用いて推定することができる。2 段階最小 2 乗法は *SAS/ETS* に属する *SYSLIN* プロシジャで実施可能である³⁸。また, *SYSLIN* プロシジャを使用できない環境下でも, 通常の *REG* プロシジャとデータステップを組み合わせることで比較的容易に実施可能である。その際, 2 段階目のモデルを当てはめる際に, モデルステートメントで *hcc* を指定することで, 1 段階目のモデルに基づくパラメータ推定に伴う誤差を考慮することができる。また, *R* のパッケージ「*Sem*」の *tsls()*関数や「*Aer*」の *ivreg()*関数にて 2 段階最小 2 乗法を用いることができる。また, *Gatsonis* らの書籍³⁶では「*ivpack*」を利用した *R* コードが紹介されている。

^{xvi} モーメント法とは, 母集団のモーメントをパラメータを利用して表現し, これと標本モーメントが一致するように推定する方法。このモーメント法を一般化したのが一般化モーメント法である。

4.3 自然実験

自然実験では外生性の事象（経済学では外生的なショックと呼ばれる）を利用する。外生性の事象とは、法律や制度の変更、自然災害など、対象者の無作為割付けを引き起こす事象であり、各対象者は偶然性により介入群あるいは対照群に割付けられることになる。自然実験は一般に公衆衛生における介入の評価に用いられており、**Medical Research Council**により、自然実験の利用に関する指針が公表されている³⁹。本項では差の差分析と回帰不連続デザインについて解説する。なお、操作変数法を含め、これらの手法は疑似実験の研究デザインであるため、自然実験が当てはまらない状況でも利用可能である。また、自然実験や疑似実験のエビデンスレベルは、複数のランダム化比較試験を統合したメタアナリシス、ランダム化比較試験に続く階層に位置し、通常回帰分析などよりもエビデンスレベルは高い。一般にエビデンスレベルが低いと、因果関係と相関関係を誤認してしまう可能性が高くなる。また、本項で解説する自然実験では、通常解析手法を用いて容易に治療効果を推定できるため、実装の解説は省略する。

4.3.1 差の差分析

差の差分析は、操作変数法や回帰不連続デザインに比べて弱い（因果効果を適切に推定するのが困難な）研究デザインではあるものの、医療経済学の分野では幅広く利用されている。自然実験デザインにおける差の差分析では、外生性の事象の前後で介入群と対照群の差を比較する。その際、同一対象者の介入前後の経時データを用いる場合と、同一対象集団から介入前後で抽出した調査対象の横断データを用いる場合がある。例えば、**Bellou**ら⁴⁰は、差の差分析を用いて、身分証明書のデザイン変更がアルコール摂取やタバコ使用に及ぼす影響を評価した。米国では1994年から2009年にかけて43の州で、未成年者のアルコール／タバコの消費量を減らすことを目的に身分証明書のデザイン変更が行われた。**Bellou**らは、州ごとに異なる時点で身分証明書のデザイン変更が行われたことを利用して、介入群と対照群を設定し、差の差分析を実施した。その結果、デザイン変更を行ってから最初の数年間は、アルコール／タバコの消費量の大幅な減少が認められることを示した。

差の差分析で前提となる重要な仮定は2つある。第1に、介入群と対照群で共通のトレンド（時間とともに起こる自然な変化）を仮定する。この仮定では、外生性の事象がなかった場合、介入群と対照群のアウトカムは並行したトレンドを示すことを仮定する。外生性の事象が起こる前に、仮にアウトカムが複数時点で測定されており、それらのデータから介入群と対照群のアウトカムのトレンドが並行であることを示すことができれば、この共通のトレンドの仮定は妥当と考えられる。一方、外生性の事象が起こる前のトレンドが両群間で異なる場合や、介入群と対照群の対象者がそれぞれ異なる外生性の事象の影響を受けやすい集団に属する場合、この共通のトレンドの仮定は成立しない可能性が高い。この仮定が成立しない場合、介入前の類似したトレンドを有するヒストリカルデータの利用が有用な場合がある。第2は共通のショックの仮定であり、外生性の事象の前後で、アウトカムに影響を及ぼすような別の事象が起きていない、もしくは起きているとしても両群に対して同じように影響を及ぼすことを仮定する。これらの2つの仮定に加えて、同一対象者を繰り返し調査する経時データではなく、異なる対象者を複数時点で調査する横断データを用いる場合、集団の属性に系統的な変化が生じないことも前提条件とし

て必要である。このように、差の差分析では、条件付き交換可能性などよりも緩い仮定の下で因果効果を推定することが可能である。なお、一般に差の差分析で推定する治療効果は ATT である。

4.3.2 回帰不連続デザイン

回帰不連続デザインは、治療の割付けを規定する連続変数（強制変数とも呼ばれる）による不連続性を利用して因果効果を推定する方法である。詳細は以下で述べるが、適切な連続変数を選択することができれば、連続変数に対して恣意的に決定された閾値により分断された両側の対象者集団は、アウトカムに影響を及ぼす未測定因子の分布は同じだが、事前規定した規則に基づく割付け治療のみ異なる集団とみなすことができるため、疑似的なランダム化の下で因果効果を推定することができる。回帰不連続デザインは、操作変数法と同様、「未測定因子の交絡がない」の仮定が成立しない場合にも適用可能な手法である。また、回帰不連続デザインによって得られる治療効果は LATE である。

治療の割付けが連続変数の決定関数で表される場合はシャープな不連続性と呼ばれる。例えば、治療の割付けを規定する連続変数が年齢であり、ある年齢の閾値を上回る対象者はすべて治療を受け、閾値を下回る対象者はすべて治療を受けない場合である。一方、連続変数による治療の割付けが不連続であるものの、治療の割付けを完全には規定しない場合はファジーな不連続性と呼ばれる。上述の例で言えば、年齢以外の変数が治療の割付けの決定に影響を及ぼす場合で、介入を受けた対象者と介入を受けなかった対象者がそれぞれ閾値の片側のみではなく両側に混在することになる。

シャープな不連続性の場合、割付け治療は強制変数により外生的に決定されるため、少なくとも閾値の近傍では、アウトカムに影響を及ぼす未測定因子による治療の選択は生じないと考えられる。なお、この状況下では、不連続性を決定する重要な連続変数は介入群と対照群の間でオーバーラップしないため、マッチング法は利用できないし、回帰分析を用いると、勝手に外挿を実施して治療効果を推定してしまうことになる。一方、ファジーな不連続性の場合、閾値の近傍では、アウトカムに影響を及ぼす未測定因子による治療の選択は生じておらず、閾値を上回る対象者集団と閾値を下回る対象者集団間で未測定因子の分布は類似しているという仮定が必要である。

以上より、回帰不連続デザインを用いる前提条件としては、治療の割付けの決定規則と連続変数の閾値が既知であること、連続変数は閾値の近傍で連続であること、連続変数の閾値の近傍では潜在アウトカムは連続であり、アウトカムに影響を及ぼす因子の分布は類似していることが必要となる。

Figure 5 にシャープな不連続性及びファジーな不連続性の場合の治療を受ける確率の分布の例を示す。詳細は Moscoe らの論文⁴¹を参照いただきたい。シャープな不連続性の場合、治療の割付けを規定する連続変数 Z が 40 以上の対象者は全員治療を受け、40 未満の対象者は全員治療を受けない。一方、ファジーな不連続性の場合、連続変数 Z が 40 以上の対象者は治療を受ける確率が高いものの、連続変数 Z のみで治療を受けるかどうかが決まる訳ではないため、治療を受けない対象者も存在する。同様に、連続変数 Z が 40 未満であっても治療を受ける対象者が存在する。

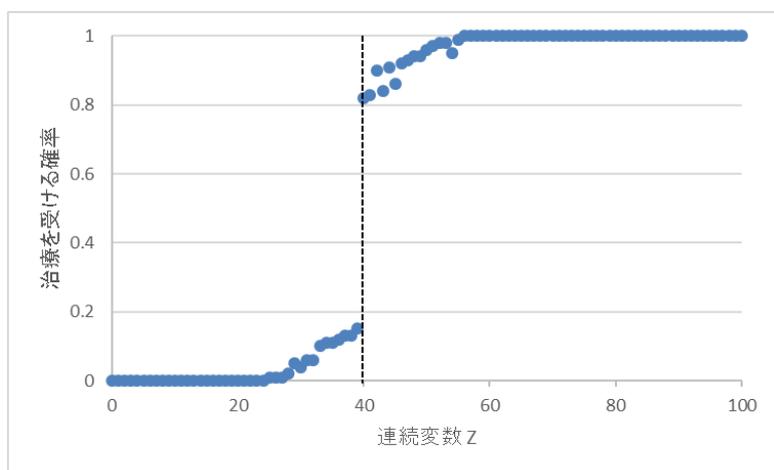
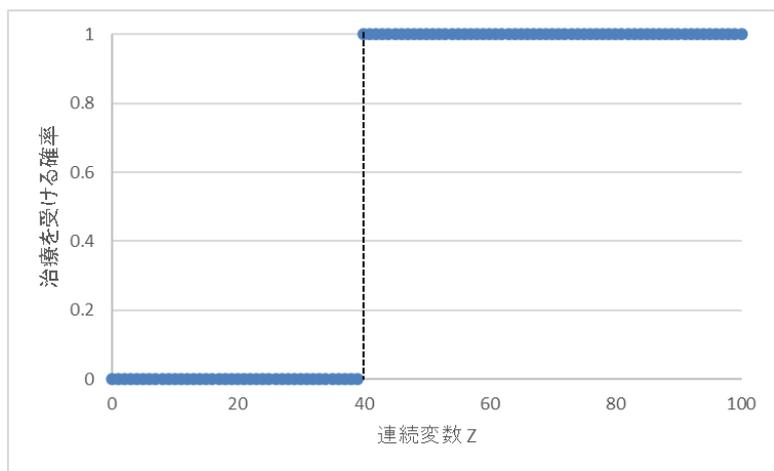


Figure 5 連続変数 Z に対する治療を受ける確率の分布 (上図: シャープな不連続性の場合, 下図: ファジーな不連続性の場合) (Moscoe らの論文⁴¹の Figure 1 を一部改変)

Figure 6 にシャープな不連続性及びファジーな不連続性の場合のアウトカムの分布の例を示す。シャープな不連続性の場合、 $Z=40$ の近傍の対象者に限れば、 $Z \geq 40$ の対象者集団と $Z < 40$ の対象者集団間で未測定の変数を含めた対象者特性は類似しており、唯一の違いは治療を受けたかどうかのみである（疑似的なランダム化が行われている）とみなすことができる。その場合、 $Z=40$ で生じるアウトカムの不連続が LATE となる。実際のデータ解析では、使用するデータのバンド幅を決定し^{xvii}、例えば、閾値の近傍の対象者集団に対して局所線形回帰モデル（ノンパラメトリック回帰モデルを用いる場合）を当てはめ LATE を推定する。また、ファジーな不連続性の場合、閾値の反対側の治療を受ける確率も考慮して LATE を推定する必要がある。

^{xvii} 閾値の近傍のデータのみを用いた解析の他、利用可能なすべてのデータを用いた解析の実施も可能である。

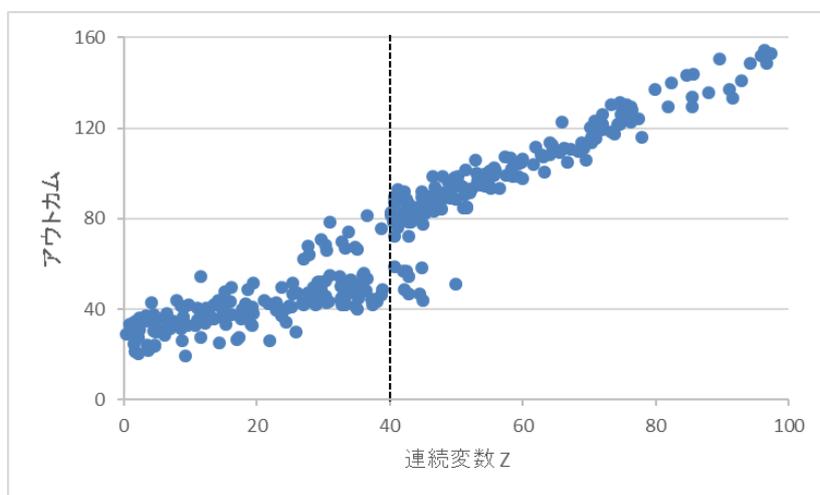
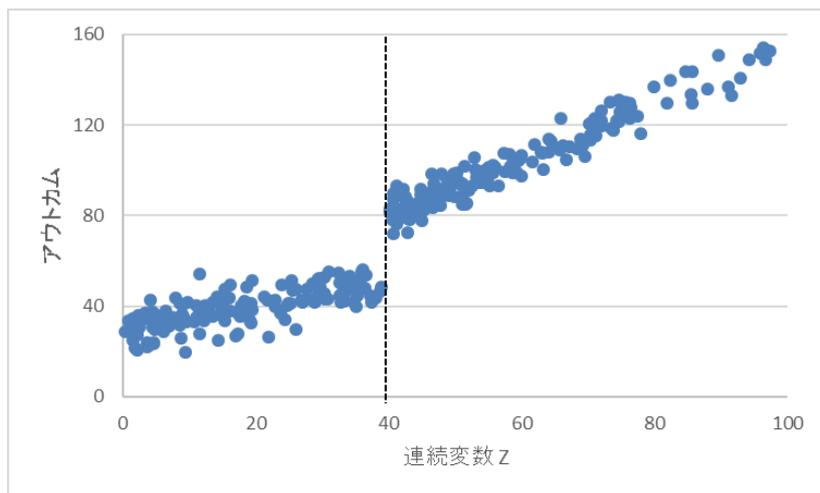


Figure 6 連続変数 Z に対するアウトカムの分布（上図：シャープな不連続性の場合，下図：ファジーな不連続性の場合）（Moscoe らの論文⁴¹の Figure 3 を一部改変）

回帰不連続デザインでは適切な連続変数を選択することが重要となる⁴²。例えば，Shigeoka⁴³は，本邦において 70 歳を超えると医療費の自己負担額が下がることを利用して，70 歳の前後で医療サービスの利用頻度や健康状態が変わるかどうかの研究を行った。また，Almond ら⁴⁴は，低出生体重児の診断基準が 1500 g であり，それを下回ると多くの医療サービスが投入されることを利用して，医療費が健康アウトカムに及ぼす影響について検討した。

5. 適切な解析手法の選択

4.1 項では，観察可能な変数による治療の選択を仮定した手法として，回帰モデルに基づく方法，IPW 法，2 重ロバスト法，傾向スコアを説明変数に含めた回帰分析，マッチング法及びマッチングした対象集団に対する回帰分析，4.2 項では，観察不可能な変数による治療の選択を考慮した手法として，操作変数法，4.3 項では，自然実験のデザインとして，差の差分析及び回帰不連続デザインについて解説した。こ

れらはすべて TSD 17 で紹介されている手法である。TSD 17 ではその他にパネルデータ分析について紹介されている。パネルデータとは、通常複数時点で同一対象を繰り返し測定したときに得られるデータであり、解析手法としては一般に各対象者固有の効果を考慮した一階差分モデルや固定効果モデル、あるいはランダム効果モデルなどが用いられる。パネル調査やパネルデータ分析は経済学や政策評価などの領域で盛んに行われている。

4 項で解説した Rubin の因果モデルに基づく手法を用いる場合、SUTVA や consistency assumption などの本手法共通の仮定が必要となるのに加えて、用いる手法に応じた仮定が更に必要となる。例えば、観察可能な変数による治療の選択を仮定した手法では、条件付き交換可能性とオーバーラップの仮定が必要となり、操作変数法では、これらの仮定は必要ないものの、その代わりに exclusion restriction などの仮定が必要となる。何れも強い仮定ではあるものの、その多くはデータから検証することが困難である。因果効果の推定では、治療の選択過程を十分に理解し、どの仮定を置くことが妥当であるかをよく吟味した上で、それに応じた適切な解析手法を選択することが重要となる。TSD 17 では、前提となる仮定の違いに着目しながら、適切な解析手法を選択するためのアルゴリズムが提案されている。一部改変して、本報告書の Appendix に記載したので、参照いただきたい。また、本報告書では 4 項でその内容を網羅できているため割愛するが、TSD 17 では、各解析手法を用いる際に考慮すべき項目を記載したチェックリストを開発しているので、必要に応じてそちらも合わせて参照いただきたい。

おわりに

本報告書では、因果推論に基づく解析手法について解説した。NICE appraisal committee の chair らが final appraisal document を用いて実施したプラグマティックレビューによれば、治療効果を推定するために非ランダム化比較試験データを用いたアプレイザルでは、その殆どが公開されている試験の集計データをそのまま用いており、潜在的な交絡因子の調整は行っていなかった。また、比較可能な IPD を用いたアプレイザルでは、観察された共変量の分布の群間での相違を調整するため、多変量回帰あるいは傾向スコアによるマッチング法を用いた事例があったのみで、過去のアプレイザルを確認した限りは、本報告書で解説した手法の適用事例はまだ殆どないようである²。しかしながら、今後、医療経済評価における医療情報データベースやリアルワールドデータなどの活用に伴い、これらの手法を適用する機会は増えてくることが予想されたため、今後に備えて、今の段階でこれらの手法を理解しておくことは重要であると思われる。

本報告書では、TSD 17 で紹介されていた、比較的容易に実行可能な手法の解説に主眼を置いた。しかし、例えば、1 回限りの治療効果ではなく、治療が複数回繰り返し行われる場合の治療効果を推定したい場合には、周辺構造モデルや構造ネストモデルなどのより高度な手法が必要となる⁴⁵。これらの手法については、日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2017 年度タスクフォース 5 による成果物「Pragmatic Trials のススメ」にも触れられているので、参照いただきたい。

昨今領域を問わず医療情報データベースやリアルワールドデータの活用の議論が盛んに行われている。しかしながら、ランダム化比較試験ではなく非ランダム化比較試験のデータを用いた治療効果の推定では様々なバイアスを伴うため、適切な手法を用いてそれらの影響を最小化する必要がある。その際、非

ランダム化比較試験のデータから、仮にアウトカムや傾向スコアに対して当てはまりの良いモデルを構築できたとしても、臨床的に重要な交絡因子をモデルに含めていなければ、誤った結論を導く可能性が高い。我々が知りたいのは相関関係ではなく因果関係であり、因果関係を推測するには、手元にあるデータの他、最新の知見も駆使して、適切な仮定を置いた下で適切な解析手法を選択する必要がある。そのためには、本報告書で解説したとおり、各手法の特徴や前提となる仮定を十分に理解しておくことが重要である。

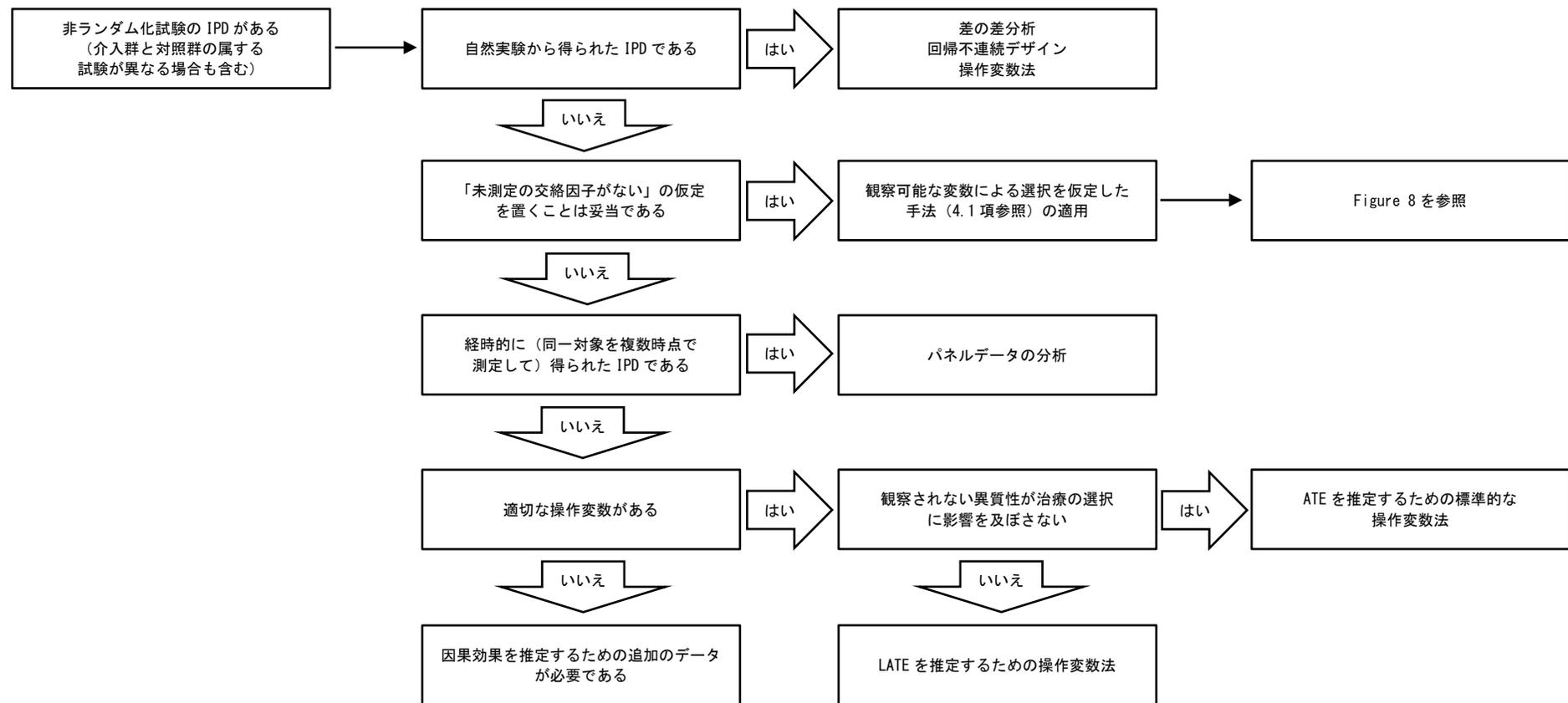
引用文献

- ¹ NICE. Guide to the methods of technology appraisal. 2013.
Available from <https://www.nice.org.uk/process/pmg9/chapter/foreword>
- ² Faria R, Hernandez Alava M, Manca A, Wailoo AJ. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for technology appraisal: Methods for comparative individual patient data. 2015.
Available from <http://www.nicesdu.org.uk>
- ³ Bell H, Wailoo AJ, Hernandez M, Grieve R, Faria R, Gibson L, Grimm S. The use of real world data for the estimation of treatment effects in NICE decision making. 2016.
Available from <http://nicesdu.org.uk/methods-development/real-world-data/>
- ⁴ Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, Petticrew M, Altman DG; International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technology Assessment* 2003; 7(27): 1-173.
- ⁵ Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology Third Edition*. 2008.
- ⁶ Rosenbaum PR. *Observational Studies Second Edition*. 2002.
- ⁷ 佐藤俊哉・松山裕. 「疫学・臨床研究における因果推論」. *統計科学のフロンティア* 5. 2002; 131-175.
- ⁸ 日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2016年度タスクフォース4 欠測のあるデータの解析チーム 因果推論サブチーム. 臨床試験の estimand を理解するための因果推論. 2016.
Available from <http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/estimand.html>
- ⁹ 臨床試験の estimand を理解するための因果推論－因果推論の基礎と causal estimand－シンポジウム. 2018.
Available from <http://www.jpma.or.jp/medicine/shinyaku/tiken/symposium/>
- ¹⁰ Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 1974; 66(5): 688-701.
- ¹¹ Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, Thabane M, Giangregorio L, Dennis B, Kosa D, Borg Debono V, Dillenburg R, Fruci V, Bawor M, Lee J, Wells G, Goldsmith CH. A tutorial on sensitivity analyses in clinical trials: The what, why, when and how. *BMC Medical Research Methodology* 2013; 16: 13-92.
- ¹² Jones AM. Identification of treatment effects in health economics. *Health Economics* 2007; 16(11): 1127-1131.
- ¹³ Galiani S, Gertler P, Schargrodsky E. Water for life: The impact of the privatization of water services on child morality. *Journal of Political Economy* 2005; 113: 83-120.

-
- ¹⁴ Kearns B, Ara R, Wailoo A, Manca A, Alava MH, Abrams K, Campbell M. Good practice guidelines for the use of statistical regression models in economic evaluations. *Pharmacoeconomics* 2013; 31(8): 643-652.
- ¹⁵ Imbens GW, Rubin DB. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. 2015.
- ¹⁶ 星野崇宏. 『調査観察データの統計科学：因果推論・選択バイアス・データ融合』. 岩波書店. 2009.
- ¹⁷ Wooldridge J. Should instrumental variables be used as matching variables. *Research in Economics* 2016; 70(2): 232-237.
- ¹⁸ Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70(1): 41-55.
- ¹⁹ Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 1984; 79(387): 516-524.
- ²⁰ Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period: Application to control of the healthy worker survivor effect. *Mathematical Modelling* 1986; 7: 1393-1512.
- ²¹ Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: Demonstration of a causal inference technique. *American Journal of Epidemiology* 2011; 173(7): 731-738.
- ²² Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 1995; 90: 106-121.
- ²³ Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11(5): 550-560.
- ²⁴ Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. *Statistics in Medicine* 2010; 29(3): 337-346.
- ²⁵ van der Laan MJ, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. 2011.
- ²⁶ Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine* 2004; 23: 2937-2960.
- ²⁷ Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics* 2005; 61(4): 962-973.
- ²⁸ Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. 2010.
- ²⁹ Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 1985; 39(1), 33-38.
- ³⁰ Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics* 2011; 10(2): 150-161.
- ³¹ Imbens GW, Abadie A. Matching on the estimated propensity score. *Econometrica* 2016; 84(2): 781-807.
- ³² Kreif N, Grieve R, Radice R, Sekhon JS. Regression-adjusted matching and double-robust methods for estimating average treatment effects in health economic evaluation. *Health Services and Outcomes Research Methodology* 2013; 13(2-4): 174-202.
- ³³ SAS Institute Inc. *SAS/STAT(R) 14.2 User's Guide*. SAS Institute Inc., Cary NC, 2016.

-
- ³⁴ 田栗正隆. SAS による因果推論: CAUSALTRT プロシジャの紹介. SAS ユーザー総会. 2017.
- ³⁵ 魚住龍史・矢田真城・山本倫生・川口淳. SAS による傾向スコアマッチング. SAS ユーザー総会. 2017.
- ³⁶ Gatsonis C, Morton SC. *Methods in Comparative Effectiveness Research*. 2017.
- ³⁷ McClellan M, McNeill BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *The Journal of the American Medical Association* 1994; 272(11): 859-866.
- ³⁸ Faries DE, Leon AC, Haro JM, Obenchain RL. *Analysis of Observational Health Care Data Using SAS*. 2010.
- ³⁹ Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, Ogilvie D, Petticrew M, Reeves B, Sutton M, Thompson S. Using natural experiments to evaluate population health interventions: New medical research council guidance. *Journal of Epidemiology and Community Health* 2012; 66(12): 1182-1186.
- ⁴⁰ Bellou A, Bhatt R. Reducing underage alcohol and tobacco use: Evidence from the introduction of vertical identification cards. *Journal of Health Economics* 2013; 32(2): 353-366.
- ⁴¹ Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: A review of current and best practice. *Journal of Clinical Epidemiology* 2015; 68(2): 122-133.
- ⁴² 中室牧子・津田友介. 「原因と結果」の経済学. ダイヤモンド社. 2007.
- ⁴³ Shigeoka H. The effect of patient cost sharing on utilization, health, and risk protection. *American Economic Review* 2014; 104(7): 2152-2184.
- ⁴⁴ Almond D, Doyle JJ, Kowalski AE, Williams H. Estimating marginal returns to medical care: Evidence from at-risk newborns. *The Quarterly Journal of Economics* 2010; 125(2): 591-634.
- ⁴⁵ Robins JM. Marginal structural models versus structural nested models as tools for causal inference. *Statistical Models in Epidemiology, the Environment, and Clinical Trial* 1999; 116: 95-134.

Appendix



IPD : 個別患者データ, ATE : 全集団の平均因果効果, LATE : 局所平均因果効果

Figure 7 適切な手法を選択するためのアルゴリズム (TSD 17² の Figure 2 を一部改変)

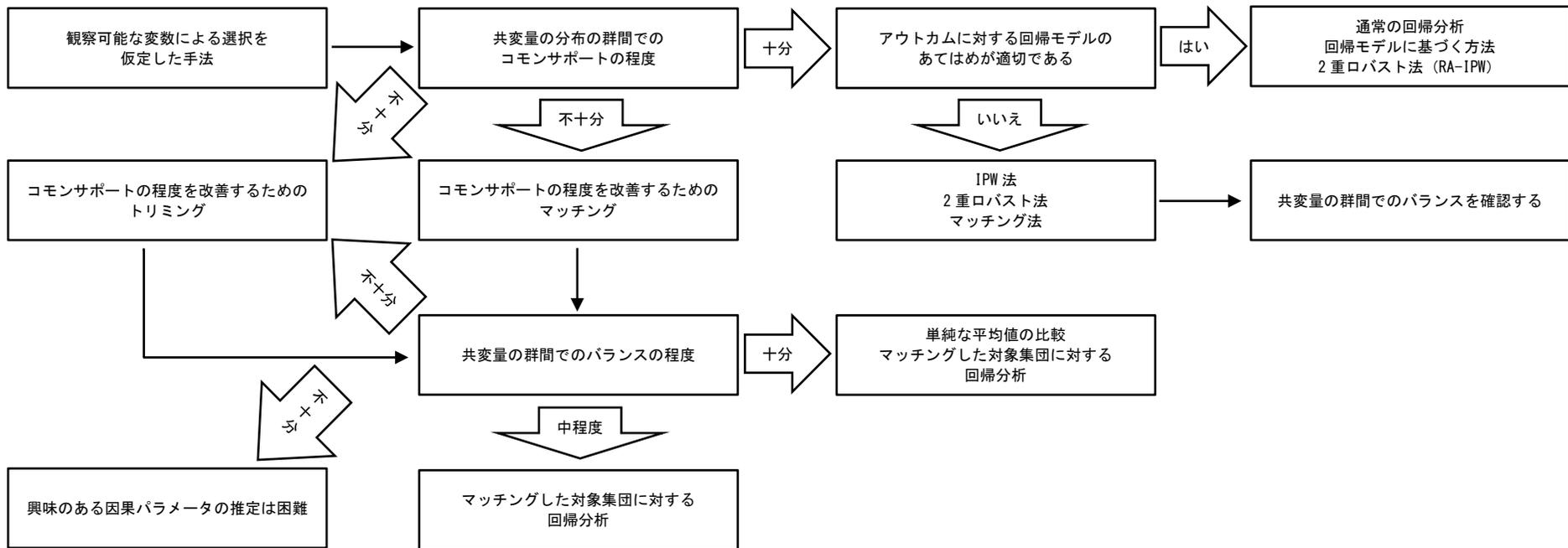


Figure 8 観察可能な変数による選択を仮定した手法の中から適切な手法を選択するためのアルゴリズム (TSD 17²の Figure 3 を一部改変)

2019年度データサイエンス部会 KT5

作成担当者

吉田 瑞樹 ファイザーR&D 合同会社

レビュー担当者

東 美恵 エーザイ株式会社
渥美 淳 東レ株式会社
荒西 利彦 日本イーライリリー株式会社
石渡 量太 サノフィ株式会社
奥山 ことば MSD 株式会社*
河田 祐一 中外製薬株式会社*
直井 一郎 大日本住友製薬株式会社
中島 章博 帝人ファーマ株式会社
野島 俊秋 興和株式会社
町田 光陽 塩野義製薬株式会社
松下 泰之 第一三共株式会社*

*TF リーダー

担当副部長

酒井 弘憲 エーザイ株式会社