



# ICH E9(R1)の理解に役立つ因果推論

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  
2020年度継続タスクフォース1  
因果推論サブチーム

Ver 1.0  
2022年7月

## 目次

1. はじめに .....	6
2. 因果推論の基本的事項と解析方法 .....	8
2.1 はじめに .....	8
2.2 因果推論の基本的事項 .....	10
2.2.1 潜在アウトカム及びそれを用いた治療効果及び交絡の定義 .....	10
2.2.2 治療効果の識別条件 .....	11
2.2.3 交絡因子と Directed Acyclic Graph (DAG) .....	11
2.2.4 交絡調整の基本的な方法の概説 .....	15
2.3 Inverse Probability Weighting (IPW) .....	17
2.3.1 ノンパラメトリック IPW .....	17
2.3.2 モデルへの拡張 (Marginal structural model (MSM) ) .....	20
2.3.3 事例紹介 .....	21
2.4 Standardization (標準化) .....	23
2.4.1 標準化による因果効果の推定 .....	23
2.4.2 SAS を用いた解析事例の紹介 .....	23
2.4.3 IPW と標準化の手法比較 .....	25
2.5 G-estimation .....	27
2.5.1 G-estimation を理解するための準備 .....	27
2.5.1.1 条件付き交換可能性 .....	27
2.5.1.2 構造ネストモデル .....	27
2.5.1.3 Rank preservation.....	28
2.5.2 推定方法と SAS 実装方法 (grid search) .....	29
2.5.2.1 1-パラメータの場合 .....	29
2.5.2.2 2-パラメータの場合 .....	30
2.5.3 推定方法と SAS 実装方法 (closed-form) .....	31
2.5.3.1 1-パラメータの場合 .....	31
2.5.3.2 2-パラメータの場合 .....	32
2.5.4 感度分析 .....	33
2.6 操作変数法 .....	34
2.6.1 操作変数とは .....	34
2.6.2 操作変数法による効果の推定 .....	36
2.6.3 IV estimand を解釈するための追加的条件 .....	37
2.6.3.1 効果の均一性 .....	38
2.6.3.2 単調性 .....	39
2.6.4 操作変数法によるバイアス .....	41
2.6.5 操作変数法とその他の手法 .....	43

2.6.6	付録 .....	44
2.6.6.1	操作変数のタイプ .....	44
2.6.6.2	線形回帰モデルからの IV estimand の導出 .....	45
2.6.6.3	IV estimand の変形 .....	45
2.6.6.4	二段階最小二乗推定量と Wald 推定量の関係 .....	46
2.6.6.5	Additive 構造平均モデル .....	47
2.6.6.6	単調性の条件と complier における効果 .....	48
3.	臨床試験データへの因果推論の適用 .....	50
3.1	系統的レビュー論文 .....	50
3.2	主要層別を適用した臨床試験事例 .....	54
3.2.1	主要層ストラテジーについて .....	54
3.2.2	主要層解析について .....	54
3.2.3	主要層解析に対するベイズ流アプローチ .....	57
3.2.3.1	主要層解析（ベイズ流の推測） .....	57
3.2.3.2	共変量と欠測の導入 .....	60
3.2.3.3	データ解析における事前分布の設定 .....	61
3.2.3.4	臨床試験事例の結果と議論 .....	62
4.	因果推論と estimand の関連 .....	65
4.1	臨床試験における因果推論と estimand.....	65
4.1.1	潜在アウトカム反応の定義 .....	65
4.1.2	中間事象 .....	66
4.1.3	治療変更の特定のための潜在アウトカムの利用 .....	68
4.1.4	アウトカムの潜在性と反事実的性質 .....	70
4.1.5	潜在アウトカムの枠組みにおける Estimand の定義 .....	71
4.1.5.1	序文 .....	71
4.1.5.2	治療方針ストラテジー.....	71
4.1.5.3	複合ストラテジー.....	71
4.1.5.3.1	2 値アウトカム Y.....	71
4.1.5.3.2	連続アウトカム Y.....	72
4.1.5.4	仮想ストラテジー.....	72
4.1.5.4.1	全患者遵守を想定 .....	72
4.1.5.4.2	中間事象後無治療を想定 .....	73
4.1.5.5	主要層ストラテジー.....	73
4.1.5.6	治療下ストラテジー.....	75
4.1.5.7	動的治療レジメンを伴うシナリオ .....	75
4.1.6	欠測データの取扱い .....	76
4.2	COVID-19 の estimand に対する影響 .....	77
4.2.1	導入 .....	77

4.2.2	データサイエンス部会声明 .....	77
4.2.3	COVID-19 パンデミック下での臨床試験実施に対する統計的課題と推奨 .....	78
4.2.3.1	中間事象の取扱い .....	79
4.2.3.2	その他（関心のある治療条件，対象集団，評価項目，治療効果の集団レベルでの要約）の取扱い .....	83
4.2.3.3	感度分析 .....	83
4.2.4	まとめ .....	84
5.	参考文献 .....	86
	執筆者・タスクフォースメンバー .....	89

## 図の目次

図 2.2-1	因果グラフの例 .....	12
図 2.2-2	裏口パスの例 .....	12
図 2.2-3	合流点の例 .....	13
図 2.2-4	他の因子が一つの場合の DAG.....	13
図 2.2-5	裏口テストの例 (1) .....	14
図 2.2-6	裏口テストの例 (2) .....	15
図 2.3-1	仮想観察研究における対象集団のデータの分布 .....	17
図 2.3-2	仮想観察研究における疑似集団のデータの分布 .....	18
図 2.3-3	対象集団の tree graph .....	18
図 2.3-4	疑似集団の tree graph .....	19
図 2.3-5	疑似集団の tree graph (stabilized weights を利用) .....	20
図 2.6-1	二重盲検下ランダム化試験における DAG .....	34
図 2.6-2	条件(ii)を満足しない例 (1) .....	42
図 2.6-3	条件(ii)を満足しない例 (2) .....	42
図 2.6-4	条件(iii)を満足しない例 .....	43
図 2.6-5	代替操作変数の例 (1) .....	45
図 2.6-6	代替操作変数の例 (2) .....	45
図 3.2-1	時点ごとの各層へ所属する事後確率 .....	63
図 3.2-2	Immune の層の事後リスク比 .....	64
図 4.1-1	因果ダイアグラム .....	68

## 表の目次

表 2.1-1	NHEFS データにおける背景因子の分布 .....	9
表 2.4-1	潜在変数を用いた処置変数のタイプ .....	39
表 3.2-1	中間事象の発生に関する層 .....	54
表 3.2-2	$(S, Z)=(s, z)$ の値と層の関係 .....	57
表 3.2-3	$(S, Z)$ ごとの disability model: $p(Y S, Z, \omega)$ .....	59
表 3.2-4	Disability model: $p_{Y S, Z, X, M} = 0, \omega$ .....	60
表 3.2-5	$t^*=12$ カ月の要約統計量 .....	63

## 1. はじめに

ICH E9(R1)<sup>1</sup>の出だしは次の通りである。“To properly inform decision making by pharmaceutical companies, regulators, patients, physicians and other stakeholders, clear descriptions of the benefits and risks of a treatment (medicine) for a given medical condition should be made available. Without such clarity, there is a concern that the reported “treatment effect” will be misunderstood.”である。この reported treatment effect を正確に理解するためには、本報告書で紹介する因果推論の考え方を理解しておくに役立つ。

特に中間事象 (intercurrent event) の理解は重要である。ICH E9(R1)において、中間事象は、治療の開始後に起こり、結果変数の観測を妨げる若しくはその解釈に影響を与えるイベントであると定義している。中間事象 (例えば、治療の中止やスイッチング) が起こった結果として生じる可能性がある欠測データや中間事象後に観測されたデータをどのように取り扱うかは、多くの議論や研究の対象となってきた<sup>2</sup>。臨床試験においては、ある介入の効果に焦点が当たることが多く、どのような介入の効果か推測の対象であるのか (すなわち、estimand) をまず明確に特定する必要があり、その estimand に沿って、適切な推定方法を選択する必要がある。また、必要に応じて estimand に則した欠測を取り扱う戦略を考える必要がある。観察研究のデータに対する因果推論の文献から、estimand に対応する上で有用な概念や解析方法を得ることができるが、多くの場合、そのような概念は、最近まで臨床試験の実践から切り離されたままであった。

因果推論の方法<sup>3,4,5,6</sup>は、当初は観察研究のデータに適用するために開発された。NRC によるレポート<sup>2</sup>が発出される前は、臨床試験の領域では、欠測データに対して、治療の中止から生じるもの、治療のスイッチングから生じるもの、又は救済治療の追加から生じるものなどを区別することはあまりなかった。その理由は、被験者が試験治療を中止したり救済治療を要したりする場合は試験を中止することが一般的であったためである。ランダム化後に起こる欠測の取扱いについて、曖昧なまま規制当局とスポンサーの間でコミュニケーションが行われてきてしまい、承認申請の際に論点となることもあった<sup>7,8</sup>。しかしながら、NRC によるレポート<sup>2</sup>において、最初に割り付けられた治療への遵守状況に関わらず、被験者は試験に参加し続けて評価し続けられることが促されている。よって、それに従うならば臨床試験では、治療の中止や変更があったときに、推定すべきものやそれをどのように推定するかを考えなければならない。一方で E9(R1)には次のような記述もある “It remains undisputed that randomisation is a cornerstone of controlled clinical trials and that analysis should aim at exploiting the advantages of randomisation to the greatest extent possible. However, the question remains whether estimating an effect in accordance with the ITT principle always represents the treatment effect of greatest relevance to regulatory and clinical decision making.” すなわち、ITT の原則が常に唯一無二のものであるかについては、考える必要があるとされている。

臨床試験の領域では、因果推論の概念を明示的に利用してこなかったものの、例えば、潜在アウトカムの枠組みは、ランダム化が行われない研究の解析では受け入れられており、estimand とは何かを理解するために役立つ可能性がある。潜在アウトカムが定義され、治療効果が被験者レベルで特定されれば、estimand は、標的集団におけるその期待値となる。詳細は、本報告書の

中で説明していく。

以上の背景から、臨床試験における estimand を設定する際に、因果推論の概念を理解しておくことが重要であるとの認識が広がってきていると考えられる。因果推論の枠組みを使用して estimand を議論することは、ICHE9(R1)に対する理解を深め、ICHE9(R1)を実装した臨床試験の計画と解析につながると考えられる。そこで、我々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2020年度継続タスクフォース1因果推論サブチームでは、因果推論の基本的事項と各解析方法を説明した上で、臨床試験データへの因果推論の適用事例を説明することとした。さらに、臨床試験における estimand と因果推論の関連にも触れ、最近の話題として COVID-19 の estimand に対する影響も含めて、説明を加えることとした。

本報告書の残りの部分の構成は以下の通りである。2章は、因果推論の基本的事項（2.2節）と代表的な解析方法（2.3節：Inverse Probability Weighting, 2.4節：Standardization, 2.5節：G-estimation, 2.6節：操作変数法）について説明し、各解析方法を理解することを目的とする。因果推論を初めて学ぶ統計家や非統計家の方に、2章をまず読んでいただきたい。3章は、因果推論のランダム化臨床試験の解析への適用事例を紹介し、因果推論の方法に対する理解を深めることを目的とする。4章は、臨床試験における estimand と因果推論の関連について説明し、estimand の構成要素の中でも特に盛んに議論されている中間事象の取扱いと因果推論との関連を理解することを目的とする。また、COVID-19が estimand に及ぼす影響について、因果推論の観点からの解釈を行う。

## 2. 因果推論の基本的事項と解析方法

### 2.1 はじめに

本報告書において、因果推論に関する複数の手法について紹介する。その際、理解を深めるためにいくつかの事例とともに説明する。本報告書で紹介する事例には、各手法において共通して用いるものがある。2.1章では、共通して用いた事例について、説明する。尚、2章では、下記を参照した。

- Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC. Chapter 12. <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/> (last access date 23-MAR-2021)

事例に用いたデータは、NHEFS (National Health and Nutrition Examination Survey Data I Epidemiologic Follow-up Study) データである。このデータは、米国において、臨床的、栄養的、行動的要因と、その後の罹患率、死亡率、病院利用率、危険因子の変化、機能制限、施設入所との関係を調査するために実施された観察研究によって得られた。NHEFS データは、下記に公開されている。

- <https://www.cdc.gov/nchs/nhanes/nhefs/>

本報告書では、この公開データの一部を事例紹介に用いた。NHEFS データに収められているデータは多岐にわたっており、事例では、次のデータを主に用いた。次節以降のプログラム例における変数名も合わせて記載する。

- 追跡期間中の喫煙／禁煙の状態にかかわらず、観察期間終了時に禁煙をしたかどうか (qsmk)  
[1 : Yes, 0 : No]
- 性別 (sex) [0 : Male, 1 : Female]
- 人種 (race) [0 : WHITE, 1 : BLACK OR OTHER]
- 教育レベル (education) [1 : 8TH GRADE OR LESS, 2: HS DROPOUT, 3: HS, 4: UNIVERSITY DROPOUT, 5: UNIVERSITY OR MORE]
- 1日あたりのタバコの本数 (smokeintensity)
- 喫煙年数 (smokeyrs)
- 体重 (wt71)
- 体重変化量 (wt82\_71)

上記データから、禁煙による体重増加への影響（因果効果）を推定することを目的とし、事例として用いている。

NHEFS データのうち、ベースライン時点と10年後（観察終了時）におけるデータが両方存在し、ベースライン時点の年齢が25歳から75歳の喫煙者である1566人を対象とする。主な被験者

背景を表 2.1-1 に示した。それぞれの個人における体重変化量  $Y$  は観察終了時の体重からベースライン時点の体重を引くことで得られる。ここで、観察終了時に喫煙していなかった人(禁煙した人)を  $A=1$ 、喫煙していた人(禁煙していない人)を  $A=0$  とする。  $A=1$  の集団の平均値は、 $E[Y|A=1] = 4.5\text{kg}$  であり、  $A=0$  の集団の平均値は、 $E[Y|A=0] = 2.0\text{kg}$  である。そして、その差は、 $E[Y|A=1] - E[Y|A=0] = 2.5\text{kg}$  と得られる。

表 2.1-1 NHEFS データにおける背景因子の分布

Mean baseline Characteristics	A	
	0	1
Age, years	46.2	42.8
Men, %	54.6	46.6
White, %	91.1	85.4
University, %	15.4	9.9
Weight, kg	72.4	70.3
Cigarettes, /day	18.6	21.2
Years smoking	26.0	24.1
Little exercise, %	40.7	37.9
Inactive life, %	11.2	8.9

[Hernán &amp; Robins (2020). Table 12.1]

ここで、もし、すべての参加者が観察終了時に喫煙していなかったとした場合の体重増加の平均値を  $E[Y^{a=1}]$  と定義し、すべての参加者が観察終了時に喫煙していたとした場合の体重増加の平均値を  $E[Y^{a=0}]$  と定義する。このとき、観察対象者における平均因果効果 (average causal effect) を  $E[Y^{a=1}] - E[Y^{a=0}]$  と定義する。本報告書では、この平均因果効果を求めることを目的として、因果推論のいくつかの手法を紹介する。

喫煙してなかった集団と喫煙していた集団で背景情報が異なる場合には、先に記載した  $E[Y|A=1] - E[Y|A=0]$  は、平均因果効果である  $E[Y^{a=1}] - E[Y^{a=0}]$  と一般に異なり、因果的な解釈ができない。

表 2.1-1 を見てみると、喫煙していなかったグループ ( $A=1$ ) は、喫煙していたグループ ( $A=0$ ) よりも年齢が約4歳低いことがわかる。もし、年齢が高くなると喫煙の有無にかかわらず体重が増加しにくいとすると、各集団の平均値の差である  $E[Y|A=1] - E[Y|A=0] = 2.5\text{kg}$  は真の平均因果効果を過小評価している可能性がある。このとき年齢を  $Y$  に対する  $A$  の効果の交絡因子と呼び、平均因果効果を求めるためにはその交絡の影響を調整する必要がある。改めて表2.1-1を見てみると、被験者背景の多くの因子が喫煙しなかった集団 ( $A=1$ ) と喫煙していた集団 ( $A=0$ ) の間で異なり、これらが交絡因子である場合には交絡の影響を調整する必要がある。

以降、本報告書では、平均因果効果や交絡の定義及びこれらの交絡の影響を調整した効果を推定するための手法について紹介する。

## 2.2 因果推論の基本的事項

### 2.2.1 潜在アウトカム及びそれを用いた治療効果及び交絡の定義

「禁煙」が「体重増加」に影響を与えるかどうかという臨床的な仮説を調べるために臨床研究を実施するという状況を考えよう。ここではランダム化を伴わない臨床研究を想定することとする。本臨床研究では、禁煙と体重増加以外にも年齢、性別、喫煙期間等のデータを収集しているとする。「禁煙」の「体重増加」に対する影響が興味のある因果効果であるが、ランダム化が伴っていないために年齢、性別、喫煙期間等も禁煙や体重増加と交絡している可能性がある。このような状況で、喫煙などといった処置に対する因果的な解釈が可能な効果を推定することが目的であり、その目的を達成するために因果推論が用いられている。

ここで、臨床研究において、禁煙した参加者と禁煙しなかった参加者の間で体重の増加量を比較することで、因果的な解釈が可能な効果を推定することができるかどうかの問題となる。残念ながら、これらの単純な比較では因果効果が正しく推定できず、禁煙した参加者が仮に禁煙しなかった状況を想定したときの体重の増加量と比較することができれば、因果的な解釈が可能な効果を推定することができる。しかし、現実には、禁煙した参加者と同じ参加者が仮に禁煙しなかった状況を観察することはできないので、因果的な解釈が可能な効果は推定することができなくなってしまった。そこで、現実の対応策として、禁煙した参加者では観察できなかったアウトカムが観察可能であることを期待して、禁煙した参加者と禁煙以外の背景因子ができるだけ似ている別の禁煙しなかった参加者に対して体重の増加量を観察し、それがもし禁煙した参加者と同じ参加者が仮に禁煙しなかった状況における体重の増加量と同じであれば、因果的な解釈が可能な効果を推定することができるようになる。逆に、現実には観察した、禁煙しなかった参加者の体重の増加量と、禁煙した参加者と同じ参加者が仮に禁煙しなかった状況における体重の増加量が異なっている状況を、交絡が生じているという。

正式に因果効果及び交絡の定義を導入するために、潜在アウトカムの考え方を説明する。Rubin<sup>1)</sup>が観察研究に応用したとされる反事実モデルでは、記法として、 $Y_i^{a=1}$ をもし被験者  $i$  が治療を受けた ( $a = 1$ ) なら観察されるであろう潜在アウトカム、 $Y_i^{a=0}$ をもし被験者  $i$  が治療を受けなかった ( $a = 0$ ) なら観察されるであろう潜在アウトカムとする。現実には被験者  $i$  が治療を受けていれば、 $Y_i^{a=1} = Y_i$ は観察されるが、 $Y_i^{a=0}$ は観察されないため反事実となる(詳細は後述するが、これも因果効果を識別するための仮定の一つとなっている)。以上のことから、個人における潜在アウトカムの差や比 ( $Y_i^{a=1} - Y_i^{a=0}$  や  $Y_i^{a=1}/Y_i^{a=0}$ ) などの個人レベルの因果効果は確認できないということがわかり、因果推論の根源問題といわれている。潜在アウトカムを用いて因果効果を定義すると、因果効果の指標は、対象母集団がもつ二変量潜在アウトカム ( $Y^{a=1}, Y^{a=0}$ ) のコントラストであり、例えば、0 (治療失敗など) または1 (治療成功など) の値をとる潜在アウトカムの場合、因果リスク差は、 $E[Y^{a=1}] - E[Y^{a=0}]$ として計算することができる。ここで、因果リスク差は、すべての対象者における期待値の差であることに注意していただきたい。一方で、例えば、 $E[Y|A = 1] - E[Y|A = 0]$ で計算することができる関連リスク差は関連の指標であり、得られている標本データによる単純比較から推定することができる。ここで、関連リスク差は、治療の水準内の対象者における期待値の差であることに注意していただ

きたい。このように因果と関連は、興味のある集団に対して、因果が集団全体を、関連が治療の水準ごとの一部の集団を対象にしているという点が異なっている。これらの二つの指標を用いると、交絡は、因果効果の指標と関連の指標が一致しないことと定義することができる。

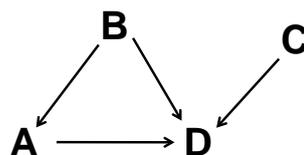
### 2.2.2 因果効果の識別条件

前節のように交絡を定義すると、禁煙しなかった参加者の体重の増加量の分布と、禁煙した参加者と同じ参加者が仮に禁煙しなかった状況における体重の増加量の分布が同じであれば交絡はなく、このとき両群は交換可能 (exchangeable) であるという。両群が交換可能であるとき、潜在アウトカム  $Y^a$  と治療  $A$  が確率的に独立であり ( $Y^a \perp\!\!\!\perp A$  for all  $a$ )、潜在アウトカムと治療に関係がない状況である。すなわち、 $E[Y^a|A=1] = E[Y^a|A=0] = E[Y^a]$  が成立している。両群の交換可能性は、治療の割付をランダム化した場合には平均的に保証されるので、ランダム化は因果効果の正しい推定に寄与している。また、交絡因子  $L$  で層別したとき、 $L$  で条件付ければ潜在アウトカム  $Y^a$  と治療  $A$  が確率的に独立である ( $Y^a \perp\!\!\!\perp A | L$  for all  $a$ ) とし、条件付き交換可能性が成立しているという。観察研究のデータから因果効果を識別するための条件 (交絡がない十分条件) には、上述した交換可能性以外に、正值性 (positivity) と治療の一致性 (consistency) がある。Positivity は、集団の中である治療しか受けないということはない (すべての治療について、 $\Pr(A=a|L=l) > 0$ ) という条件であり、例えば2.3で述べる IPW 推定量が well-defined であるためなどに必要となる。consistency は、潜在アウトカムのうちの一つは、現実に観察されるもの (事実) と一致する ( $Y = A \times Y^{a=1} + (1-A) \times Y^{a=0}$ ) という条件である。さらに因果効果の識別のために相互干渉がない (no interference) ことを仮定する。これは各被験者の潜在アウトカムは他の被験者の受ける治療に依存しないことを表す。これらに関しては4.1でも再び確認する。

### 2.2.3 交絡因子と Directed Acyclic Graph (DAG)

前節までで、因果効果を識別するための条件 (交絡がない十分条件) の一つに (条件付き) 交換可能性の成立があり、その成立に必要な交絡因子を特定する必要がある。交絡因子は、それぞれの疾患領域における医学的知識などに基づいて特定されるが、本報告書では、交絡因子を特定するための有益な方法の一つである DAG (非巡回有向グラフ) を紹介する。

まず、変数間の因果関係を表現したグラフを因果グラフといい、例えば、[図 2.2-1](#) のように変数  $A$  から変数  $D$  の間の因果関係を表現したものである。

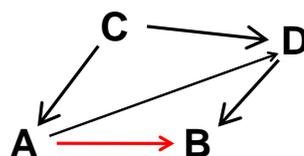


[Greenland et al. 1999]

図 2.2-1 因果グラフの例

続いて DAG を説明する。DAG は、原因から結果への影響を説明するために、変数間を矢線で結んだグラフである。DAG には二つのルールがあり、一つは矢線で結ばれた二変数の両方が原因かつ結果となることはないというもので、もう一つは原因の変数から始まり原因にかえるようなループはない（すなわち、矢線をたどって元の変数に戻るルートがない）というものである。DAG は、変数間の因果関係がはっきりしているときに有用なツールである。DAG を利用する目的は、変数間に関係が想定される場合にどの変数を調整すべきかを特定することである。研究計画時に DAG を作成するとき、治療と反応に影響を与えうるような変数をすべて DAG の中に含めることが重要であるが、その時点で個々の変数が交絡因子かどうかはわからない。DAG を作成することによって、交絡の調整が十分かどうかの判定方法を知り、調整が必要となる変数を選択することができる。また、視覚的な理解が可能なことも利点である。

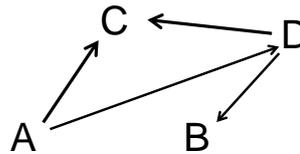
DAG において独特の用語である裏口パス (backdoor path) 及び合流点 (collider) を説明する。変数 A と変数 B の間の因果関係に興味があるとして、変数 A から変数 B への裏口パスの例を図 2.2-2 に示す。



[Greenland et al. 1999]

図 2.2-2 裏口パスの例

変数 A から変数 B への裏口パスは、変数 A から変数 B へのパスで、変数 A に矢線が入り、変数 B に矢線が入るようなパスで、変数 A と変数 B の関係に影響を与えうる因子が存在するパスを指す。図 2.2-2 では、A-C-D-B は変数 A から変数 B への裏口パスであるが、A-D-B は裏口パスではない。次に、合流点の例を図 2.2-3 に示す。



[Greenland et al. 1999]

図 2.2-3 合流点の例

図 2.2-3において、変数 A、変数 C 及び変数 D の関係に着目すると、合流点は、前の変数からも後ろの変数からも矢線が入る（すなわち、二つの変数から矢線が入っている）グラフに存在し、変数 C が合流点となっている。

もし、治療（変数 A）と結果（変数 Y）以外の因子が一つ（変数 Z）の場合、図 2.2-4 のような DAG を描くことができ、左の DAG では、変数 A から変数 Y への裏口パスが存在するので「交絡が生じている」、真ん中の DAG では、変数 A から変数 Y への裏口パスが存在しないので「交絡は生じていない」、右の DAG では、変数 A から変数 Y への裏口パスが存在しないので「交絡は生じていない」と判定することができる。この判定は、後述する裏口テストを実行することによって可能となる。

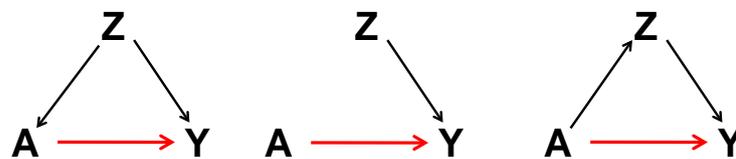


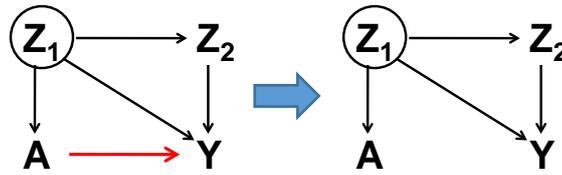
図 2.2-4 他の因子が一つの場合の DAG

最後に、交絡因子を特定するための裏口テスト（backdoor test）を説明する。裏口テストは、DAG で示した変数間の関係について、特定の変数を調整することにより交絡の調整が十分かどうかを判断するための手順であり、交絡を除去できているかどうかを調べることができる。裏口テストは、以下の三つの手順で実行する。

- 1) 治療の変数 A から出るすべての矢線を削除する。
- 2) 調整したい変数が他の二つの変数の共通の子孫（矢線の先の変数）であるならば、二つの変数を矢なし線で結ぶ。なお矢なし線は非因果なパスを表し、関連が生じることを意味する。
- 3) 1.及び 2.によって作成されたグラフにおいて、調整したい変数を含まずに、治療の変数 A から結果の変数 Y のパスで「途中に合流点がない裏口パス」があるかないかを調べる。

以上の手順より、「途中に合流点がない裏口パス」がない場合、交絡の調整はできており、「途中に合流点がない裏口パス」がある場合、交絡の調整は不十分（さらに交絡の調整が必要）と判

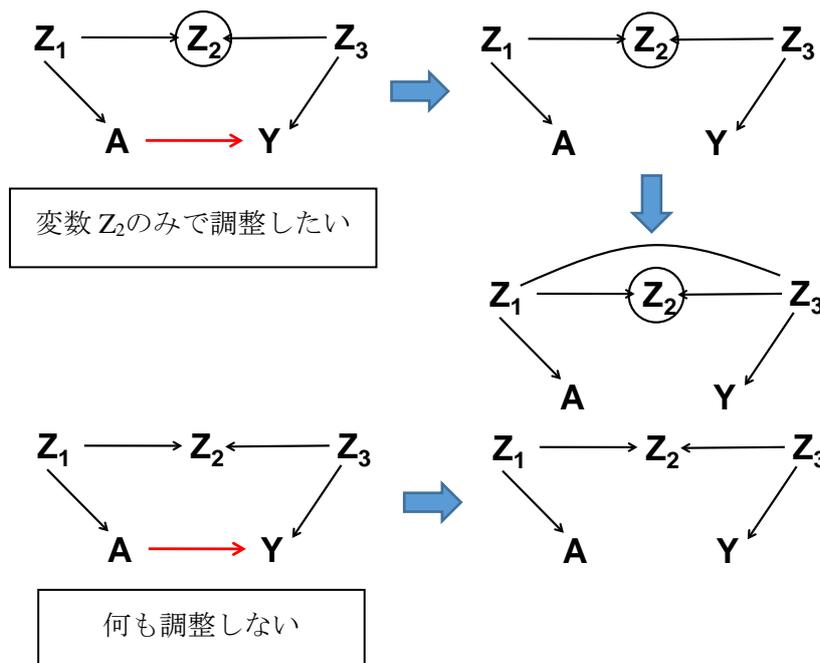
定する。以下の図 2.2-5に裏口テストの例 (1) を示す。



[Greenland et al. 1999]

図 2.2-5 裏口テストの例 (1)

変数  $Z_1$  のみで交絡を調整したいと考えたとする。まず、手順1)により、治療の変数  $A$  から出るすべての矢線を削除する。手順2)は、本 DAG では該当しない。手順3)により、治療の変数  $A$  から結果の変数  $Y$  のパスで「途中に合流点がない裏口パス」はない。よって、変数  $Z_1$  のみで交絡の調整はできており、変数  $Z_1$  で交絡を調整すれば十分であると判定できる。本 DAG では、変数  $Z_1$  で交絡を調整すれば十分であったが、変数  $Z_1$  及び  $Z_2$  で交絡を調整することもできる。一方で、変数  $Z_2$  のみで交絡を調整することはパス  $A-Z_1-Y$  が途中に合流点がない裏口パスとなっているため、不十分である。以上より、調整すべき変数の組み合わせは一通りではないということがわかる。例えば、変数  $Z_2$  の測定が高価なら測定しなくてもよいかもしれないし、変数  $Z_2$  に欠測が多い場合は変数  $Z_1$  のみで交絡を調整すればよい。このように DAG を用いた交絡因子の選択によって、既存の医学的知識を反映させて変数間の関係を示したグラフを構築でき、どの因子を調整すべきかを明確にすることができる。以下の図 2.2-6に別の裏口テストの例 (2) を示す。



[Greenland et al. 1999]

図 2.2-6 裏口テストの例 (2)

変数  $Z_2$  のみで交絡を調整したいと考えたとする。まず、手順1)により、治療の変数  $A$  から出るすべての矢線を削除する。手順2)は、本 DAG では該当し、共通の子孫をもつ変数  $Z_1$  と変数  $Z_3$  が矢なし線で結ばれ、変数  $Z_1$  と変数  $Z_3$  に関連が発生してしまう。手順3)により、治療の変数  $A$  から結果の変数  $Y$  のパスで「途中に合流点がない裏口パス」はある (パス:  $A-Z_1-Z_3-Y$ ) ので、変数  $Z_2$  のみで交絡を調整するのでは不十分であると判定される。一方で、同じ DAG を対象に何も調整しないと考えると、まず、手順1)により、治療の変数  $A$  から出るすべての矢線を削除する。手順2)は、本 DAG では該当しない。手順3)により、治療の変数  $A$  から結果の変数  $Y$  のパスで「途中に合流点がない裏口パス」はない。なお、変数  $Z_2$  は合流点になっている。合流点  $Z_2$  で調整しなければ変数  $Z_1$  と変数  $Z_3$  は周辺独立だが、合流点  $Z_2$  で調整すると変数  $Z_1$  と変数  $Z_3$  は従属になるため、バイアスの原因になってしまう。よって、本 DAG では交絡を調整する必要がないと判定できる。以上より、ある変数で交絡を調整することにより、逆にバイアスを導くこともあるので注意が必要である。

## 2.2.4 交絡調整の基本的な方法の概説

本節では交絡調整の基本的な方法を概説するが、必要最小限の記載に留めるため、網羅的な解説については教科書や総説論文<sup>12, 13, 14</sup>等を参照していただきたい。

興味のある因果関係に対して、結果 (outcome) に対する交絡因子の影響を調整することを目的に回帰モデルが使用される。回帰モデルは、一般的に以下の式で表現される。

$$E[Y|A, L] = \alpha_0 + \alpha_A A + \alpha_1 L_1 + \dots + \alpha_k L_k$$

ここで、 $L_k$  は交絡因子、 $\alpha_k$  はパラメータである。もちろん回帰モデルに交互作用項を入れることも可能である。結果がどのような変数かによって、さまざまな回帰モデルが使用される。結果が連続値であれば重回帰モデル、結果が二値であればロジスティック回帰モデル、結果が生存時間であれば Cox 回帰モデルが代表的なモデルである。回帰モデルによる交絡調整の前提条件は、測定されていない交絡因子がないこととモデルが正しく特定されていることである。

興味のある因果関係に対して、交絡因子の治療に対する影響を調整することを目的に傾向スコアが使用される。傾向スコアは、一般的には治療を受ける確率として知られており、治療を受けたかどうかを結果変数として交絡因子を用いて推定される。このことにより、各被験者の治療を受ける確率そのものを推定するというよりは、交絡因子の各層における治療を受ける確率を推定することができる。推定の方法はさまざまであるが、例えば、ロジスティック回帰モデルが用いられる。最近では、機械学習を利用することも議論されている<sup>15</sup>。注意点として、交絡を調整するために、説明変数に十分な量の交絡因子を含める必要があることが挙げられる。

前述の前提条件の下、傾向スコアが同じ複数の対象者は、治療を受ける確率が同じであり、治療を受けるかどうかをランダムに割り付けた場合と同じと考えることができ、傾向スコアで条件付けたときに条件付き交換可能性を達成することができる。また、複数の交絡因子を一つのスコアにまとめて次元を落とすことによって、複数の交絡因子の組み合わせによる層別に比べ、

傾向スコアによる層別は被験者数が少なくなる可能性を低くすることができることも傾向スコアの利点として挙げられる。

傾向スコアを用いた解析の基本的な方法として、傾向スコアでの層別、傾向スコアでのマッチング、傾向スコアを説明変数とした回帰モデルが挙げられるが、傾向スコアが連続値であることにより、厳格な交絡調整が難しく交絡が残ることやマッチングでは、マッチをとれない被験者がいることにより対象集団が変わることや被験者数が減ることが解析方法の限界として指摘されている。2章の以降の節及び5章では、それらの限界に対応しうる発展的な解析方法を紹介していく。

## 2.3 Inverse Probability Weighting (IPW)

本節では、Inverse Probability Weighting (IPW) について説明する。観察研究では交絡が起こっているのが一般的であり、興味のある治療効果に対し交絡の調整をせずに推定するとバイアスを含む推定値が得られることが知られている。そこで何らかの方法で交絡の調整をする必要がある。本節では、代表的な交絡の調整方法の一つである IPW について紹介する。

まずは、2.3.1にてモデルを用いないノンパラメトリック IPW を説明する。次に2.3.2にてモデルへの拡張をした周辺構造モデル (Marginal structural model (MSM)) を紹介し、2.3.3にて事例を用いて SAS 実装方法を紹介する。

### 2.3.1 ノンパラメトリック IPW

想定する状況として、ある観察研究における治療A (1:治療あり, 0:治療なし) による効果を知りたいとする。20人の患者に対し、死亡の有無Y (1:死亡, 0:生存) が結果変数であり、交絡因子L (1:重体, 0:非重体) が考えられるとする。このとき図 2.3-1のようなデータが得られたとする。

#### 対象集団

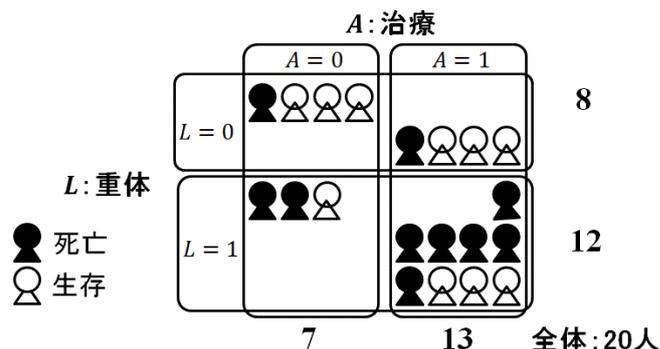


図 2.3-1 仮想観察研究における対象集団のデータの分布

重体患者の方が非重体患者よりも死亡割合が高く、重体患者の割合が治療ありで多かったことから、交絡因子である重体Lを考慮せずに死亡割合を求めた場合、治療ありで死亡割合は7/13、治療なしで死亡割合は3/7となり、治療ありの方が治療なしより死亡割合が高いという結果が得られる。しかし、重体L別に見た場合、非重体L = 0での死亡割合は治療の有無に限らずいずれも3/4となり、重体L = 1でも同様に死亡割合は治療の有無によらず同じとなる。すなわち、重体Lで条件付けた交換可能性 ( $Y^a \perp\!\!\!\perp A | L$  for all  $a$ ) が成り立つ。ここでは、重体Lが治療Aに影響を与えない交換可能性 ( $Y^a \perp\!\!\!\perp A$  for all  $a$ ) が成り立つような治療Aと結果 $Y^a$ が確率的に独立となる疑似集団を作ることを考える。例えば、元々の対象集団である非重体L = 0で治療なし群A = 0の4人を2倍の8人にする。他の3つの集団も同様に“膨らませる”事により重体Lが治療Aに影響を与えない図 2.3-2のような疑似集団を作成することができる。

疑似集団

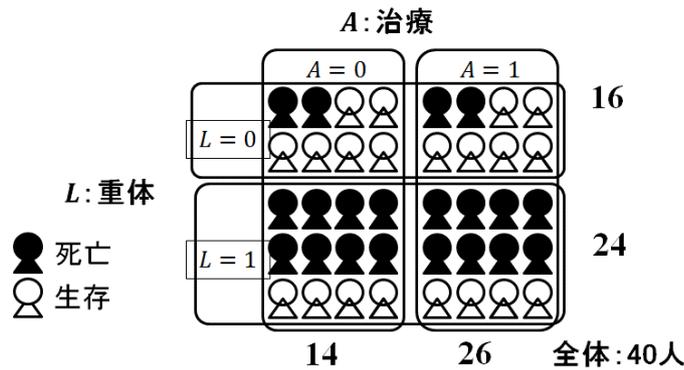
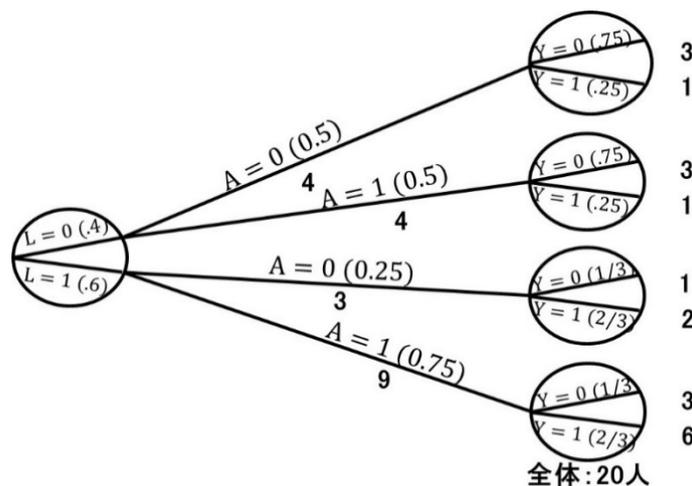


図 2.3-2 仮想観察研究における疑似集団のデータの分布

この疑似集団では関連の指標 (例えば, 関連リスク比) が因果の指標 (例えば, 因果リスク比) に一致する.

次に, 上記イメージ図をもう少し数学的に tree graph (fully randomized causally interpreted structured tree graph or FRCISTG<sup>6, 16</sup>) を用いて疑似集団の作成の仕方について説明する. 対象集団 (図 2.3-1) を tree で表記したものが以下の図2.3-3になる. ここで, 左の縁から出ている線 (branch)の下, 及び一番右の列に示されている数値はそれぞれ属している患者数を表す. また, branch 上の数値は治療の有無の割合を示す. 例えば, 非重体 $L = 0$ の患者である8人は全体の40%を占め, このうち50%の4人が治療なしとなり, うち生存者は3人 (75%), 死亡者は1人 (25%) である.

対象集団

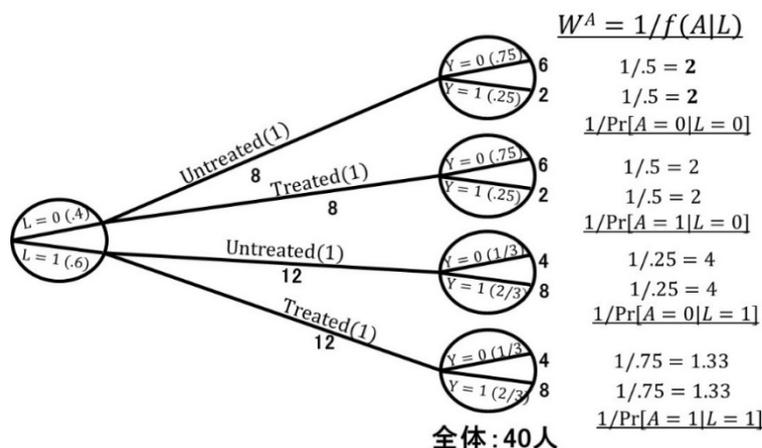


[Hernán & Robins (2020). Chapter 2, Figure 2.1改編]

図 2.3-3 対象集団の tree graph

次に疑似集団 (図 2.3-2) を tree で表したものが以下の図 2.3-4になる. ここで,  $W^A = 1/f(A|L)$  ( $f(\cdot|L)$ は $L$ で条件付けた確率密度関数 (確率関数) を表す) は重みを表し, 交絡因子が与えられた下での観察された治療確率の逆数である. 疑似集団は対象集団に対し, それぞれの層を重みで膨らませた集団である. これにより,  $L$ によらず割付確率が一定となるような, 交絡のない交換可能性が成立する状況を作り出すことができる. なお, 治療なしの重みは,  $1/\Pr[A = 0|L = 0] = 1/(1 - \Pr[A = 1|L = 0])$  であることに注意されたい.

疑似集団



[Hernán & Robins (2020). Chapter 2, Figure 2.2改編]

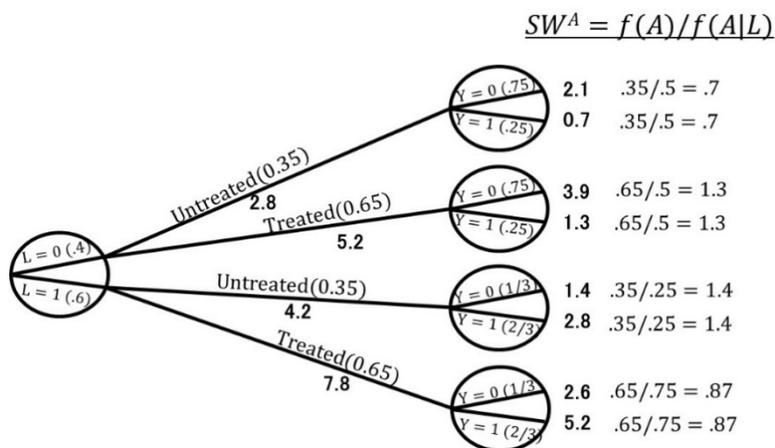
図 2.3-4 疑似集団の tree graph

ここで注目したい点は, 疑似集団の治療あり及び治療なしの人数である. それぞれ20人ずつ存在する. 対象集団が全体で20人であったことを考えると, これは仮に全員が治療していたら観察されたであろう死亡割合 $E[Y^{a=1}]$ , 及び仮に全員が治療していなかったら観察されたであろう死亡割合 $E[Y^{a=0}]$ をそれぞれ求めており, 疑似集団における関連の指標の推定は因果効果 $E[Y^{a=1}] - E[Y^{a=0}]$ を推定していることがわかる. また, ここでの重みの期待値は2となるが, これは2群あり, 患者数全体が20人から40人と2倍になったことと一致している. なお, この場合の治療 $A = a$ の効果の平均は,  $E[I(A = a)Y/f(A|L)]$  (ここで $I(\cdot)$ は指示関数を表す) で求まる.

これまで重みに各個人が治療を受ける確率の逆数を用いてきたが, 重みは交絡を調整した疑似集団を作成できればよく, これだけに限らない. ここでは拡張した重み stabilized weights :  $SW^A = f(A)/f(A|L)$ について紹介する. これは先ほどの重みに対し, 分子に何らかの重みを掛けており,  $L$ によらない $A$ の関数であれば何でも良い. 例えば, 対象集団における各群の治療確率を掛け合わせた場合の tree graph を以下の図 2.3-5に示す. この時, 重みの期待値は1となり, 疑似集団全体の患者数は20人となる. Stabilized weights の特徴としては, 分子に重みを掛けることによって, 重みの幅が狭まり, 効果の推定値は $W^A$  (non-stabilized weights) を用いるよりも安定する. ただし, 平均因果効果の点推定値はいずれの重みでも同じである. 変わるのは95%信頼区間であるが, 時間依存性や治療の有無が連続量の場合等, 飽和モデルでない時は信頼区間が狭くなる. しかし, ここではモデルを用いないノンパラメトリックを想定しているため, 信頼区

間の幅は変わらない.

疑似集団 (stabilized weights を利用)



[Hernán & Robins (2020). Chapter 12, Figure 12.1]

図 2.3-5 疑似集団の tree graph (stabilized weights を利用)

2.3.2 モデルへの拡張 (周辺構造モデル)

前節では、治療が二つ、交絡因子が一つのみで2水準であるような比較的シンプルな状況を想定していたため、ノンパラメトリックで対応可能であった。しかし、現実にはもう少し複雑な状況が多い。例えば、治療では多群や連続量となったり、複数の交絡因子を考えたりすることも多く、このような場合ノンパラメトリックでは対応できない。そこで、本節では前節の考え方を拡張し、モデルで対応する方法周辺構造モデルを紹介する。

周辺構造モデル (Marginal structural model) とは、 $E[Y^a] = \beta_0 + \beta_1 a$  のような形で記載されるモデルである。左辺は潜在アウトカムに対する期待値で表現し、観測されない値のため推定できない ("structural" を指す)。右辺は、交絡因子を含む必要がなく、全体集団に対する推定を想定しているような marginal model で表す。仮に全員が  $a = 0$  時のモデルは  $E[Y^a] = \beta_0$ 、仮に全員が  $a = 1$  時のモデルは  $E[Y^a] = \beta_0 + \beta_1$  となり、平均因果効果は  $E[Y^{a=1}] - E[Y^{a=0}] = \beta_1$  で求めることができる。

次に推定方法について紹介する。推定には、IPW を用いて疑似集団を作成するが、実際にここでは、 $E[Y|A] = \theta_0 + \theta_1 A$  に対し IPW の重み付き最小二乗法を用いる。重み付き最小二乗法で求められた  $\theta_1$  の推定値は、周辺構造モデルのパラメータ  $\beta_1$  に等しく一致推定量となる。なお、以降は治療  $a$  が2値の場合を想定して説明する。治療が連続量の場合も算出可能であるものの、重みに確率密度関数を仮定した  $f(A|L)$  を推定する必要がある。この場合、正確な治療確率を求めるのは難しく、誤特定はパラメータ推定に影響を与えることが知られているので、用いる際には注意が必要である。

周辺構造モデルでは興味のある効果が平均因果効果である場合、周辺モデルを想定し共変量をモデルに含めないが、治療効果が共変量の水準によって異なることが疑われ、むしろ各水準での効果に興味がある場合は、効果修飾因子  $V$  (例えば、性別) をモデルに考慮することができ

る。この場合のモデルは以下ようになる。

$$E[Y^a|V] = \beta_0 + \beta_1 a + \beta_2 Va + \beta_3 V$$

この時 $\beta_2 \neq 0$ となれば、性別による効果修飾があると解釈できる。推定には、重み $SW^A(V) = f(A|V)/f(A|L)$ を用いる。この際、交絡因子 $L$ に効果修飾因子 $V$ が含まれていることに注意する。これにより推定値の信頼区間が狭くなるが、この理由は、直観的には、分母及び分子に効果修飾因子 $V$ を含めることによりばらつきが吸収され、重み $SW^A$ の幅が狭くなるためと解釈できる。また、偏りのない平均因果効果の推定のために効果修飾因子 $V$ の水準ごとに交換可能性が保たれる必要があることも注意されたい。

### 2.3.3 事例紹介

本節では、禁煙効果の事例を通じて SAS の実装方法について紹介する。ここで、治療 $A$ として禁煙の有無 (qsmk)、効果修飾因子 $V$ として性別 (sex) とし、交絡因子は $L$ で表す。性別による効果修飾が疑われるため、重みは stabilized weights :  $SW^A(V) = f(A|V)/f(A|L)$ を用いる。

- 1) 重み ( $f(A|L)$ : 分母) の推定 : ロジスティック回帰より推定

```
proc logistic data=データセット名 descending;
ods exclude ClassLevelInfo ModelAnova Association FitStatistics GlobalTests;
class exercise active education;
model qsmk = sex race age age*age education smokeintensity smokeintensity*smokeintensity
smokeyrs smokeyrs*smokeyrs exercise active wt71 wt71*wt71;
output out=データセット名 p=変数名 (禁煙確率の推定値) ;
run;
```

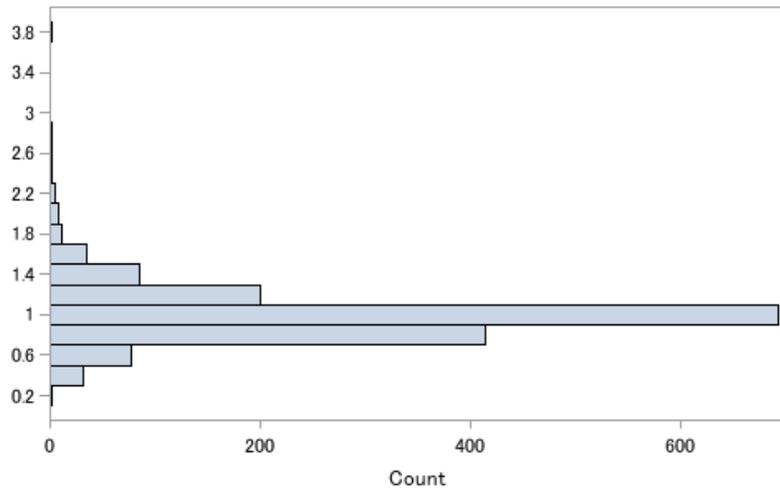
- 2) 重み ( $f(A|V)$ : 分子) の推定 : ロジスティック回帰より推定

```
proc logistic data=データセット名 descending;
ods exclude ClassLevelInfo ModelAnova Association FitStatistics GlobalTests;
model qsmk = sex;
output out=データセット名 p=変数名 (禁煙確率の推定値) ;
run;
```

- 3) 重み ( $SW^A(V) = f(A|V)/f(A|L)$ ) を算出

```
data データセット名;
merge データセット名 (重み : 分母) データセット名 (重み : 分子) ;
by seqn;
if qsmk=1 then 重み=(重み : 分子)/(重み : 分母);
else if qsmk=0 then 重み=(1-重み : 分子)/(1-重み : 分母);
run;
```

## 重みの分布 (平均 : 0.999 [0.29, 3.80])



重みの平均値が約1であることが確認できる。

## 4) 重み付き最小二乗法で治療効果を推定

```

proc genmod data=データセット名;
  class seqn sex;
  weight 重み;
  model model wt82_71= qsmk sex qsmk*sex;
  repeated subject=seqn / type=ind;
run;

```

## 求められたパラメータ推定値

パラメータ		推定値	標準誤差	95% 信頼限界		Z	Pr >  Z
<b>Intercept</b>		1.7757	0.3247	1.1393	2.4121	5.47	<.0001
<b>qsmk</b>		3.3625	0.8140	1.7672	4.9578	4.13	<.0001
<b>sex</b>	0	0.0087	0.4488	-0.8709	0.8884	0.02	0.9845
<b>sex</b>	1	0.0000	0.0000	0.0000	0.0000	.	.
<b>qsmk*sex</b>	0	0.1595	1.0461	-1.8908	2.2098	0.15	0.8788

禁煙の効果は約3.4kgの体重増加と推定された。なお、信頼区間の算出には保守的な推定量としてロバスト分散を用いた。禁煙の効果に対する性差はわずか約0.16kgであり、また得られたp値も大きいことから、効果の修飾はほぼないと解釈できる。

## 2.4 Standardization (標準化)

本節では、因果効果を推定したい場合に交絡バイアスを調整する手法の一つである標準化について説明する。

標準化では、まず交絡因子の影響を除去するために交絡因子のレベルで層別を行う。そして、各層内において因果効果を推定したのち、各層の重み付き和を求めることで全体の因果効果の推定を行う。

本節では、まず推定方法についてさらに詳細を補足する。その後、標準化を用いた解析事例を紹介する。最後に2.3で述べられたIPWとの比較を行う。なお、標準化という名称は、適切な標準集団を定めて、処置群と対照群の交絡因子の分布が標準集団の交絡因子の分布と同じだと仮定した場合に期待されるアウトカムの比較をすることに由来する<sup>47</sup>。標準集団として全体集団を考える（なお、以降ではそのもとで記載している）ことで、“全体集団に対する治療による因果効果、すなわち $E[Y^{a=1}] - E[Y^{a=0}]$ を推定したい”という目的に対応するが、標準集団として処置群など別の集団を考えることで、別の集団に対する因果効果を推定することもできる。標準集団としてどの集団を定めるかは、我々が今何を推定したいかに依存する。

### 2.4.1 標準化による因果効果の推定

標準化を用いた因果効果の推定量は以下のように定義される：

$$\begin{aligned}\widehat{E}[Y^{a=1}] - \widehat{E}[Y^{a=0}] &= \sum_l (\widehat{E}[Y|A=1, L=l] - \widehat{E}[Y|A=0, L=l]) \times \Pr[L=l] \\ &= 1/n \sum_{i=1}^n (\widehat{E}[Y|A=1, L_i] - \widehat{E}[Y|A=0, L_i])\end{aligned}$$

右辺を見ると、各層 $l$ に対して $E[Y|A=a, L=l]$ を推定し、 $\Pr[L=l]$ で重み付け和をとることで因果効果を推定している。

重み付け和になっていることから明らかであるが、多くのサンプルが含まれる層の因果効果は全体の因果効果への寄与が大きい。また、上の定義では交絡因子 $L$ を離散変数として表現しているが、交絡因子 $L$ に連続変数が含まれる場合は分布を確率密度関数 $f_L(l)$ に、和を積分に置き換えることで同様に定義できる。2.4.3で示すが、標準化による因果効果の推定量は交換可能性とconsistencyの仮定のもとで不偏推定量となる。

$\sum$ の中の $E[Y|A=a, L_i]$ の推定の方法に着目する。交絡因子の個数やそのカテゴリ数が多い状況では、各層 $l$ を実際に構成してノンパラメトリックに推定することは、層数が膨大になることから難しい。そのためモデルを当てはめてパラメトリックに推定することが多い。

### 2.4.2 SASを用いた解析事例の紹介

SASを用いた禁煙効果の解析事例を紹介する。治療 $A$ は禁煙の有無(qsmk)でありアウトカムは体重である。禁煙による体重増加量が興味のある因果効果である。

標準化を用いた推定を実装する場合、データセットの拡張、アウトカムのモデリング、予測、平均化による標準化の4段階から成る。

(1<sup>st</sup> step) データセットを三つに拡張する。つまり今あるデータセット (1症例1レコード) について、同じデータセットをもう二つ用意する。一つはそのまま何も編集せず置いておくが、拡張した二つには編集を行う。一方には治療の値をすべて  $A = 0$  とし、もう一方はすべて  $A = 1$  に修正する。また、これら二つのデータセットに対してアウトカム  $Y$  のデータは削除 (つまりすべて欠測の状態に) する。この修正したデータセットは各群の標準化した平均を推定するために活用される。

(2<sup>nd</sup> step)  $E[Y|A = a, L_i]$  に線形モデルを当てはめ、パラメータを推定する (実装例は以下)。

```
proc genmod data=1st step でそのまま何も編集していないデータセット名;
  class exercise active education;
  model wt82_71 = qsmk sex race age age*age education
    smokeintensity smokeintensity*smokeintensity smokeyrs smokeyrs*smokeyrs
    exercise active wt71 wt71*wt71 qsmk*smokeintensity;
  output out = predicted_mean p = meanY;
run;
quit;
```

出力の一部を以下に示す：

最大尤度パラメータ推定値の分析

パラメータ	自由度	推定値	標準誤差	Wald 95% 信頼限界	Wald カイ 2 乗	Pr > ChiSq
Intercept	1	-1.6906	4.3594	-10.2348 6.8536	0.15	0.6982
qsmk	1	2.5596	0.8037	0.9844 4.1348	10.14	0.0014
sex	1	-1.4303	0.4658	-2.3432 -0.5173	9.43	0.0021
race	1	0.5601	0.5780	-0.5727 1.6929	0.94	0.3325
age	1	0.3596	0.1622	0.0417 0.6776	4.91	0.0266
age*age	1	-0.0061	0.0017	-0.0095 -0.0027	12.66	0.0004
education	1	0.1950	0.7364	-1.2483 1.6383	0.07	0.7912
education	2	0.9854	0.6965	-0.3797 2.3505	2.00	0.1571
education	3	0.7513	0.6297	-0.4828 1.9854	1.42	0.2328
education	4	1.6865	0.8658	-0.0104 3.3835	3.79	0.0514
education	5	0.0000	0.0000	0.0000 0.0000	.	.
smokeintensity	1	0.0491	0.0514	-0.0516 0.1498	0.91	0.3389
smokeinte*smokeinten	1	-0.0010	0.0009	-0.0028 0.0008	1.13	0.2877

(3<sup>rd</sup> step) モデルのパラメータが推定できたので、得られた回帰モデルを用いて 1st step でアウトカム  $Y$  を欠測にした二つ目、三つ目のデータセットのアウトカムを予測、補完する。

```
proc means data = predicted_mean mean noprint;
  class interv ;
  var meanY ;
  types interv ;
  output out = results (keep = interv mean ) mean = mean ;
run;
```

(4<sup>th</sup> step) 最後に上記プログラムのように二つ目、三つ目のデータセットにおける補完したアウトカムの平均の差を考慮することで因果効果の推定値が導かれる。なお“interv”とは1st stepで作成した三つのデータセットのカテゴリ変数（interv=-1は何も編集していないオリジナルデータセットに対応）であり、プログラム実行により以下の出力が得られる。

Intervention	Predicted Value
-1	2.56319
0	1.66027
1	5.17884

実際の解析では interv=1と interv=0の平均値の差が因果効果の点推定値であり、3.5kg（5.18-1.66=3.52）となる。95%信頼区間をブートストラップ法により算出（Hernan and Robins<sup>42)</sup>のPROGRAM13.4を参照）した結果、禁煙により3.5kg（95%CI: 2.6~4.5kg）の体重増加と推定される。

### 2.4.3 IPW と標準化の手法比較

因果効果の推定手法である IPW と標準化について、解析事例の紹介ではいずれの手法も推定値がほとんど似たような値を示した。実際、交換可能性と一致性が成立し、かつ仮定したモデルが正しい場合には両手法は以下のように同等である：

$$\begin{aligned} E[I(A = a)Y/f(A|L)] &= \sum_l 1/f(A|L) \{E[Y|A = a, L = l]f(a|l) \Pr(L = l)\} \\ &= \sum_l \{E[Y|A = a, L = l] \Pr(L = l)\} \end{aligned}$$

また、標準化の不偏性は以下のように示せる：

$$\begin{aligned} E[Y^a] &= \sum_l \{E[Y^a|L = l] \Pr(L = l)\} \\ &= \sum_l \{E[Y^a|A = a, L = l] \Pr(L = l)\} \text{ (交換可能性)} \\ &= \sum_l \{E[Y|A = a, L = l] \Pr(L = l)\} \text{ (一致性)} \end{aligned}$$

同等であるならばどのように使い分ければよいのか疑問が残る。両手法を見返すと IPW は  $f(A|L)$ （治療モデルと呼ぶこととする）にモデルを置く一方で、標準化は  $E[Y|A = a, L]$ （アウトカムモデルと呼ぶこととする）にモデルを置いている。つまり、IPW と標準化はそれぞれ異なる部分にモデルを仮定している。どちらかを選択する際にはモデルの妥当性に関して信頼が高い方を用いればよいのであろうが、ここでは両方とも用いることで因果効果の推定において重要な情報が得られる可能性を紹介したい。

どのようなモデルについてもある程度の誤特定、それに伴うバイアスの発生は現実的に避けることのできないものである。しかし、治療モデルの誤特定とアウトカムモデルの誤特定によ

るバイアスは一般的には同じ大きさ、同じ方向性のバイアスにはならない。つまり IPW と標準化のモデルの誤特定による各点推定値への影響は一般的に異なる<sup>46</sup>。そのため、IPW と標準化の推定値が大きく異なる場合、少なくともいずれかが深刻なモデルの誤特定をしている可能性があるという知見を得ることができ、モデリングを見直すチャンスを獲得できるわけである。

以上の観点から、使用可能であればモデルの妥当性の確認のために両手法とも推定に用いるのがよいといえる。

また、他の推定手法として **doubly robust** 推定量を考えることもできる。これは因果効果の推定に治療モデルとアウトカムモデルの両方を用いるものであり、いずれかのモデルが正しく特定できていれば、因果効果に対して一致性をもち（この性質が“**doubly robust**”と呼ばれる理由である）、どちらのモデルとも正しく特定できていれば、**doubly robust** 推定量の漸近分散は IPW よりも小さくなり、さらにセミパラメトリック漸近有効な推定量である。またアウトカムモデルが正しく特定できているとき、標準化による推定量と比べてさほど効率損失は生じない<sup>48</sup>。ただし、どちらのモデルとも誤特定した場合に一致性は失われ、かつ推定量の分散が大きくなってしまふ特徴があるので注意が必要である<sup>49</sup>。詳細は上記した論文を参照していただきたい。

## 2.5 G-estimation

本節では、G-estimation について説明する。G-estimation は、前節までに見てきた IPW (MSM) や標準化 (g-formula) と同じく、「一般化 (generalized)」した方法で、(経時変化する処置を含む) 処置の因果効果を推定する g-methods の一つである。G-estimation の特徴は、構造ネストモデルを用いた交絡調整による推定方法にある。ただし、経時変化する処置に対する具体的な解析は本報告書の対象外とする。興味のある場合は、Hernan and Robins<sup>46</sup>の Part 3を参照されたい。

まずは、2.5.1にて利用するモデル・必要な仮定を説明する。次に、2.5.2, 2.5.3にて grid search や closed-form を用いた場合の因果効果の推定方法と SAS 実装方法を紹介する。また、2.5.4にて感度分析の考察を行う。

### 2.5.1 G-estimation を理解するための準備

#### 2.5.1.1 条件付き交換可能性

Consistency (一貫性) の下で  $Y = Y^a$  として、潜在アウトカムを観察できたとし、A に対して以下のロジスティック回帰モデルを仮定する。

$$\text{logit Pr}[A = 1|Y^a, L] = \alpha_0 + \alpha_1 Y^a + \alpha_2 L$$

このとき、条件付き交換可能性が成立すれば、L の条件付きの下で  $Y^a$  と A が独立なので、次のようになる。

$$\Pr(A = 1|Y^a, L) = \Pr(A = 1|L)$$

従って、 $\alpha_1$  の推定値は0である。G-estimation では、この性質を因果効果の推定に利用する。なお、モデルの誤特定はパラメータ推定に影響する。

#### 2.5.1.2 構造ネストモデル

A が Y に与える因果効果に興味があるとき、以下の structural model (SM) で表現する (ただし、打切りが無いことを仮定する)。

$$E[Y^{a=1}|L] - E[Y^{a=0}|L] = E[Y^{a=1} - Y^{a=0}|L]$$

仮に L の水準間で因果効果に違いが無ければ、各水準で一定の平均因果効果と考え以下のように表現できる ( $\beta_1$  は L で条件づけない集団全体の平均因果効果でもある)。

$$E[Y^{a=1} - Y^{a=0}|L] = \beta_1$$

つまり、条件付き因果効果の構造モデルは、以下のようにも表現できる。

$$E[Y^a - Y^{a=0}|L] = \beta_1 a$$

ただし、L の水準間で因果効果に違いがあるならば、A と L の交互作用項を追加しないと、L の水準内での正しい因果効果の推定値が得られないため、以下のように表現する。

$$E[Y^a - Y^{a=0}|L] = \beta_1 a + \beta_2 aL$$

また、条件付き交換可能性が成立すれば、L の条件付きの下で  $Y^a$  と A が独立なので、以下のように表現できる。

$$E[Y^a - Y^{a=0}|L] = E[Y^a - Y^{a=0}|A = a, L] = \beta_1 a + \beta_2 aL$$

このモデル式から  $\beta_1$  と  $\beta_2$  を G-estimation (2.5.2, 2.5.3で後述) により推定する。なお、経時

変化する処置の場合にこのモデルは「ネスト」するため<sup>50</sup>、構造ネスト平均モデル (structural nested mean models, SNMM) と呼ばれている。

なお、打切りがある場合、興味があるのは  $E[Y^{a=1} - Y^{a=0}|A, L]$  ではなく、全員が打ち切られていない場合の平均因果効果  $E[Y^{a=1, c=0} - Y^{a=0, c=0}|A, L]$  である。censor と uncensor に条件付き交換可能性を仮定すれば、以下のように SNMM で表現できる。

$$E[Y^{a, c=0} - Y^{a=0, c=0}|A = a, L] = \beta_1 a + \beta_2 aL$$

実際の解析では、G-estimation は交絡因子のみ調整可能であるため、以下の手順で行う。

- (i) IPW を用いて疑似集団 (pseudo population) を作成して打切りによる選択バイアスを調整する。
- (ii) 疑似集団内で G-estimation を実施する。

なお、周辺構造モデル (MSM) は以下のように表現された。

$$E[Y^a|V] = \beta_0 + \beta_1 a + \beta_2 aV + \beta_3 V$$

処置との交互作用にならない無処置下の  $E[Y^{a=0}|V = v] = \beta_0 + \beta_3 v$  の推定には興味がなく、以下のように SNMM の形式で表現できる。

$$E[Y^a - Y^{a=0}|A = a, V] = E[Y^a|V] - E[Y^{a=0}|V] = \beta_1 a + \beta_2 aV$$

MSM に比べてパラメータが少なく、モデルの誤特定に対してよりロバストである可能性がある。

### 2.5.1.3 Rank preservation

本節では rank preservation の仮定について説明する。各被験者において、処置ありのときの潜在アウトカム、処置無しの潜在アウトカムの両方を考える。それぞれの潜在アウトカム内で各被験者に降順または昇順で順位をつける。この順位を以降 rank と呼ぶ。各被験者の処置ありの潜在アウトカムの rank と処置無しの潜在アウトカムの rank が全ての被験者で同じとき、rank preservation が成立しているといい、以下のように表現できる。

$$Y_i^a - Y_i^{a=0} = \psi_1 a \text{ for all subjects } i$$

このとき一貫性より、以下のように表現できる。つまり、すべての被験者の潜在アウトカム  $Y^{a=0}$  を明示可能である。

$$Y^{a=0} = Y - \psi_1 A$$

なお、(L が同じ値を持つ) すべての被験者で処置効果 ( $Y^{a=1} - Y^{a=0}$ ) が同じ値であるとき、additive (conditional) rank preservation が成立しているといい、以下のように表現できる。また同様に、すべての被験者の潜在アウトカムを明示可能である。

$$Y_i^a - Y_i^{a=0} = \psi_1 a + \psi_2 aL_i \text{ for all subjects } i$$

ちなみに、rank preservation の成立を仮定することはもっともらしくない<sup>46</sup>。本仮定は G-estimation の理解を促進するため紹介しており、後で述べるが、G-estimation の手順は、rank preservation の仮定の成立によらない。

## 2.5.2 推定方法と SAS 実装方法 (grid search)

grid search と呼ばれるアルゴリズムを用いた推定方法を説明する。

- 1) 潜在アウトカム $Y^{a=0}$ を以下のようにモデル化すると、 $\psi^+$ の中から正しい $\psi_1$ を選ぶことと、正しい潜在アウトカム $H(\psi_1)$ を選ぶことは等価となる。

$$H(\psi^+) = Y - \psi^+ A$$

- 2) 2.5.1.1 で紹介したとおり、以下のロジスティックモデルも正しければ、条件付き交換可能性より、 $\psi^+ = \psi_1$ を代入すると $\alpha_1 = 0$ となる。つまり、 $\psi^+$ を少しずつ変えてロジスティック回帰を繰り返し (grid search)、 $\alpha_1 = 0$ となった時の $\psi^+$ が、求めたい g-estimate  $\widehat{\psi}_1$ である。

$$\text{logit Pr}[A = 1|H(\psi^+), L] = \alpha_0 + \alpha_1 H(\psi^+) + \alpha_2 L$$

- 3)  $\psi_1$ の推定値の 95%信頼区間は、 $\alpha_1 = 0$ の検定で p 値が 0.05 を上回った $\psi^+$ のセットである。なお、分散の推定には、打ち切りによる選択バイアスの調整のために IPW を用いることを考慮する必要がある。SAS 実装方法の紹介ではロバスト分散を利用するが、大標本では保守的な信頼区間を構成する。Bootstrapping を利用する場合は、ロバスト分散を用いる場合よりも信頼区間が狭くなるが、計算量が増える。

$\psi_1$ に対する上記 G-estimation アルゴリズムは、 $\text{logit Pr}[A = 1|H(\psi^+), L]$ と $H(\psi^+)$ が正しく特定されていれば、SNMM の  $\beta_1$  の一致推定値を与える。2.5.1.3で述べたが、これは rank preservation の仮定の成否によらない。 $H(\beta_1)$ と $Y^{a=0}$ が同一の条件付き平均値を持つことのみを必要としており、すべての被験者で $H(\beta_1) = Y^{a=0}$ である必要はない。

### 2.5.2.1 1-パラメータの場合

禁煙したことによる、体重への影響を検討したいとする。

- (i) まず、打ち切りによる選択バイアス調整用重み  $w_c$  を算出する。

```
proc logistic data=解析データセット名;
  class 変数名;
  model [Censor 変数] = qsmk sex race age age*age education smokeintensity
    smokeintensity*smokeintensity smokeyrs smokeyrs*smokeyrs exercise active wt71 wt71*wt71;
  output out=結果データセット名 p=[Censor 確率の推定値変数];
run;
data 解析データセット名 A;
  merge 解析データセット名 結果データセット名;
  by [被験者番号];
  if [Censor 変数] = <打ち切り無の水準値>; /*打ち切り例は重み w_c の算出のみに寄与*/
  w_c = 1 / [Censor 確率の推定値変数];
run;
```

- (ii) G-estimation by grid search

```
%macro gest_gsearch;
%do loop=200 %to 500 %by 10; /*grid search : Ψ を2.0~5.0まで0.1刻みで変化*/
%let psi=%eval(&loop)/100;
data 解析データセット名 B;
set 解析データセット名 A;
Hpsi = wt82_71-&psi*qsmk; /*H(ψ+) = Y - ψ+A*/
run;
proc genmod data 解析データセット名 A;
class [被験者番号] exercise active education;
weight w_c; /*重み w_c の利用*/
model qsmk = sex race age age*age educationsmokeintensity
smokeintensity*smokeintensity smokeyrs smokeyrs*smokeyrs exercise active wt71 wt71*wt71
Hpsi / dist = bin link = logit;
repeated subject = [被験者番号] / type = ind; /*ロバスト分散の利用*/
run;quit;
%end;
%mend;
```

(iii) grid search の結果と解釈

ψ+が3.4と3.5の時、H(ψ)に対する係数α1が最も0に近いことから、禁煙した場合、体重が3.4~3.5kg 増加する。

Testing ψ+	H(ψ) 推定値α1	ロバスト分散を利用			ナイーブ分散を利用		
		標準誤差	95%信頼区間		標準誤差	95%信頼区間	
2.0	0.0267	0.0085	0.0100	0.0435	0.0080	0.0111	0.0423
...							
3.3	0.0027	0.0088	-0.0145	0.0199	0.0080	-0.0130	0.0184
3.4	0.0009	0.0088	-0.0164	0.0181	0.0080	-0.0149	0.0166
3.5	-0.0010	0.0089	-0.0184	0.0164	0.0080	-0.0168	0.0147
...							

2.5.2.2 2-パラメータの場合

V の水準によって平均因果効果が異なる (交互作用がある) 場合, SNMM は以下のように表現できる。

$$E[Y^a - Y^{a=0} | A = a, L] = \beta_1 a + \beta_2 aV$$

対応する rank preservation model は以下の通りである。

$$Y_i^a - Y_i^{a=0} = \psi_1 a + \psi_2 aV \text{ for all subjects } i$$

H(ψ+)を以下のように定義し,

$$H(\psi^+) = Y - \psi_1^+ A - \psi_2^+ AV$$

以下のロジスティックモデルを仮定して, α1 = 0かつα2 = 0となるψ+を見つける。

$$\text{logit Pr}[A = 1|H(\psi^+), L] = \alpha_0 + \alpha_1 H(\psi^+) + \alpha_2 H(\psi^+)V + \alpha_3 L$$

ただし、パラメータ $(\alpha_1, \alpha_2)$ の数が1つ増えると grid search で探索するのは大変である。

### 2.5.3 推定方法と SAS 実装方法 (closed-form)

推定方程式が線形モデルであれば、解析的に計算可能 (closed form) である。以下に、Hernan and Robins<sup>46</sup>の program 14.3を参考に、SAS での実装方法を紹介する。

ただし、一般に推定方程式は、closed form で表せず、ソフトウェアで実装可能な近似的方法で探索する。SAS/IML では「Nelder-Mead simplex method」(NLPNMS Call) が利用でき、推定方程式を微分できない場合が扱える。Hernan and Robins<sup>46</sup>の17章 (program 17.5) を参照されたい。

#### 2.5.3.1 1-パラメータの場合

例えば、1-パラメータの SNMM を考える。

$$\text{logit Pr}[A = 1|H(\psi^+), L] = \alpha_0 + \alpha_1 H(\psi^+) + \alpha_2 L$$

スコア検定を考えると、下記の推定方程式を得ることができ、 $I[C_i = 0]W_i^C$ はデータから推定可能である。

$$\sum_{i=1}^N I[C_i = 0]W_i^C H_i(\psi^+) (A_i - E[A|L_i]) = 0$$

特に、 $H_i(\psi^+) = Y_i - \psi_1^+ A_i$ とした場合、残りは線形モデル (closed form) であり解析的に計算可能である。

$$\widehat{\psi}_1 = \frac{\sum_{i=1}^N I[C_i = 0]W_i^C Y_i (A_i - E[A|L_i])}{\sum_{i=1}^N I[C_i = 0]W_i^C A_i (A_i - E[A|L_i])}$$

例えば、禁煙したことによる、体重への影響を検討したいとする。

- (i) データから $E[A|L_i]$ を算出する。なお、 $w\_c$ は2.5.2.1の(i)で算出した値である。

```
proc logistic data=解析データセット名 descending;
  class 変数名;
  weight w_c;
  model qsmk = sex race age age*age educationsmokeintensity
  smokeintensity*smokeintensity smokeyrs smokeyrs*smokeyrsexercise active wt71 wt71*wt71;
  output out=結果データセット名 p=[禁煙確率の推定値]; /*p が算出したE[A|L_i]である*/
run;
```

- (ii) 1-parameter closed form を計算する。

なお、grid search 同様に打ち切り例は寄与しない。入力データを打ち切り例 $I(C=0)$ に限定する。

```
proc iml;
  use 入力データ;
  read all var {qsmk [禁煙確率の推定値]wt82_71 w_c};
  close [禁煙確率の推定値];
```

```

[psi] = sum(w_c#wt82_71#(qsmk- [禁煙確率の推定値]))
        / sum(w_c #qsmk#(qsmk- [禁煙確率の推定値]));
create 結果データセット名 A from [psi];
append from [psi];
close 結果データセット名 A;
print [psi];
quit;

```

(iii) 1-parameter closed form の結果と解釈

禁煙した場合、体重が3.4kg 増える。

Psi
3.4458871

### 2.5.3.2 2-パラメータの場合

V の水準（1日あたりのタバコの喫煙本数）によって因果効果が異なる場合を検討したいとする。

(i) 2-parameter closed form を計算する。

各パラメータに対するスコア検定の推定方程式を連立方程式とし、配列 lhs（2行2列）と配列 rhs（2行1列）で指定し、solve 関数で計算した結果を t 関数で転置して表示する。

```

proc iml;
  use 結果データセット名;
  read all var { qsmk [禁煙確率の推定値] wt82_71 w_c wt82_71 };
  /*score test の推定方程式を連立方程式として lhs と rhs に設定*/
  lhs = J(2,2,0);
  lhs[1,1] = sum(qsmk#w_c#(qsmk-[禁煙確率の推定値]));
  lhs[1,2] = sum(qsmk#smokeintensity# w_c#(qsmk-[禁煙確率の推定値]));
  lhs[2,1] = sum(qsmk#smokeintensity# w_c#(qsmk-[禁煙確率の推定値]));
  lhs[2,2] = sum(qsmk#smokeintensity#smokeintensity# w_c#(qsmk-[禁煙確率の推定値]));
  rhs = J(2,1,0);
  rhs[1] = sum(wt82_71#w_c#(qsmk - [禁煙確率の推定値]));
  rhs[2] = sum(wt82_71#smokeintensity#w_c#(qsmk - [禁煙確率の推定値]));
  [psi] = t (solve (lhs, rhs));
  create 結果データセット名 A from [psi];
  append from [psi];
  close 結果データセット名 A;
  print [psi];
quit;

```

## (ii) 2-parameter closed form の結果と解釈

元々のタバコが0本で禁煙すると、体重が2.86kg 増える。また、タバコが1本増えると、体重が0.03kg 増える。

Psi	
2.8594631	0.0300411

## 2.5.4 感度分析

条件付き交換可能性の仮定についての感度分析を考える。つまり、未測定交絡に対する感度分析の検討である。

G-estimation は $\alpha_1 = 0$ となる性質を用いたが、 $\alpha_1 = 0$ 以外についても計算アルゴリズムを適用可能であることを利用する。具体的には、ある範囲の複数の $\alpha_1$ について G-estimation を繰り返し、効果の推定値をプロットすることにより、未測定交絡の向きや大きさに対してどの程度安定するか検討可能である。ただし、未測定交絡の大きさを $\alpha_1$ のスケールで解釈するのは難しい。

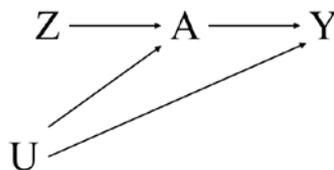
## 2.6 操作変数法

本報告書の他節にて紹介している因果推論の方法は、「交絡及び選択バイアスを調整するために必要な変数がすべて特定され、正しく測定されている」という、重要でありながら検証できない仮定を置いている。この仮定がもっともらしくない状況は一般に起こり得ることであり、仮定からの逸脱がある状況下での推定値には偏りが残る（残差バイアス：residual bias）。

本節では、調整に必要なすべての変数を測定する必要がなく、別の仮定の下で因果効果の妥当な推定値を与える方法として、操作変数法について説明する。

### 2.6.1 操作変数とは

二重盲検下ランダム化試験において割付治療不遵守を起こした被験者がいる状況を図 2.6-1 の DAG の下で考える。ここで、 $Z$  はランダム割付（1：被験治療群，0：プラセボ治療群）， $A$  は実際に受けた処置（1：被験治療，0：プラセボ治療）， $Y$  はアウトカム， $U$  はアウトカム  $Y$  と処置  $A$  の共通原因（未測定因子も含む）である。



[Hernán & Robins (2020). Chapter 16, Figure 16.1]

図 2.6-1 二重盲検下ランダム化試験における DAG

$Y$  に対する  $A$  の平均因果効果について、一致推定したいとする。このとき、他節にて紹介したような因果推論のための手法（IPW，標準化，g-estimation）はいずれも、裏口パス（ $A \leftarrow U \rightarrow Y$ ）をブロックする変数を正しく測定し、調整する必要がある。しかし、調整に必要ないくつかの変数が未測定もしくは不完全な測定であったり、モデル内で調整すべき変数を誤特定していたりすると、いずれの方法でも偏った推定値を与える。

操作変数（Instrument Variable：IV）法では、他節で紹介した手法とは全く異なる仮定の下で、 $U$  による交絡を調整するために通常必要となる変数が未測定であっても  $Y$  に対する  $A$  の平均因果効果を推定することができる（推定方法の詳細は2.6.2を参照のこと）。そのためには、操作変数と呼ばれる  $Z$  が必要となる。変数  $Z$  が次の3つの条件（以降、操作変数3条件）を満たすとき、変数  $Z$  を操作変数と呼ぶ。

- (i)  $Z$  は  $A$  と関連する
- (ii)  $Z$  は  $A$  を通してのみ  $Y$  に影響を与える
- (iii)  $Z$  と  $Y$  の共通原因はない

図 2.6-1の二重盲検下ランダム化試験においては、ランダム割付  $Z$  が操作変数となる。ほとんどの被験者は割り付けられた治療を受けるため、条件(i)を満足すると考えられる。条件(ii)は、二重盲検下であるため満足すると期待できる。条件(iii)は、 $Z$ がランダムに決まるため満足する。

次に観察研究の例として、体重変化に対する禁煙の効果を観察した例を考える。ここでも、変数  $A$  を処置 (1: 禁煙, 0: 喫煙継続), 変数  $Y$  をアウトカム (体重変化) とする。このとき、操作変数の候補として、各対象者の居住地におけるタバコ1箱の値段が150円以上であれば「1」を取り、150円未満の場合は「0」を取る変数  $Z$  を考える。操作変数3条件のうち、条件(i)は得られているデータから確認できる。実際に条件(i)を満足しているかを示すためには、変数  $Z$  と処置  $A$  が関連していることとして  $\Pr[A = 1|Z = 1] - \Pr[A = 1|Z = 0] > 0$  となることを示す必要がある。この観察研究の事例では、左辺第1項の確率が25.8%、左辺第2項の確率が19.5%で、その差は6.3%であった。変数  $Z$  と処置  $A$  の関連が弱い  $Z$  を“弱い操作変数”と呼び、一般に  $A$  を  $Z$  で回帰した際の  $F$  値が10以下の場合に弱い操作変数と言われている。この例では、 $F$  値がわずかに0.8であった。弱い操作変数における問題については2.6.4を参照のこと。

一方、条件(ii)及び条件(iii)を満足することはデータからは確認できない。条件(ii)を満足していることを示すならば、変数  $Z$  が処置  $A$  を介してのみアウトカム  $Y$  の原因となることを示す必要がある。しかし、これをデータから示すことはできない。例えば、処置  $A$  が与えられた下でのアウトカム  $Y$  に対する変数  $Z$  の効果を確認することで、条件(ii)を満足するかを検証しようとしたとしよう。しかし、処置  $A$  はパス  $Z \rightarrow A \leftarrow U \rightarrow Y$  上の合流点であるため、たとえ条件(ii)を満足していたとしても処置  $A$  が与えられた下では  $Z$  と  $Y$  の間に関連が生じる (選択バイアス) ため、この方法では条件(ii)を満足することを検証することはできないことがわかる。また、条件(iii)を満足することもデータから示すことはできない。つまり、条件(ii)及び条件(iii)は満足することを仮定するしかなく、本報告書で紹介した他節の手法と同様に、操作変数法も検証できない仮定に基づいていることがわかる。

上記の例のように、一般に観察研究においては、操作変数として利用する変数  $Z$  が真に操作変数3条件を満足しているかは証明することができない。そのため、変数  $Z$  が条件(ii)及び条件(iii)を満足するという仮定を置き、その仮定のもっともらしさを周辺の知識に基づいて説明するしかない。これは、本報告書で紹介してきた他節の手法における仮定を正当化しようとする場合と同様である。

ランダム化比較試験か観察研究かによらず、変数  $Z$  が真に操作変数であると仮定してみる。このとき、交絡因子を特定することなく、そして測定することなく、アウトカム  $Y$  に対する処置  $A$  の平均因果効果を一致推定できるかという、実は追加的条件なくしては平均因果効果の一致推定はできず、平均因果効果が存在する (取り得ない値を除いた) 範囲しか特定できない。詳細は、2.6.6.2, 3.2.2を参照のこと。さらに、その範囲は非常に広いことが多い。追加的条件を置かず操作変数3条件のみで特定可能な範囲の詳細については、Swanson et al.<sup>17</sup>を参照のこと。体重変化に対する禁煙の効果を観察した例において、追加的条件を置かず操作変数3条件のみで特定可能な範囲は、体重増加、体重減少、体重変化なしのいずれの結果も取り得る範囲となり、まったく意味をなさないものであった。残念ながら、検証不可能な追加的条件が成立するという更なる仮定を置かない限り、操作変数法で解釈できる結果はここまでとなる。

操作変数法における推定方法を2.6.2で紹介し、2.6.3では操作変数法により推定された結果を平均因果効果と解釈可能にする追加的条件について紹介する。

## 2.6.2 操作変数法による効果の推定

2値変数  $Z$  が操作変数3条件を満たし、さらに2.6.3で提示する追加的条件(iv)または(v)も満たす状況を考える。どちらの追加的条件の成立を仮定するかによって推定の対象 (ICH E9(R1)<sup>1</sup>の文脈での estimand) が異なることに注意が必要である。このとき、アウトカム  $Y$  に対する処置  $A$  のリスク差における平均因果効果は、

$$\frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[A|Z = 1] - E[A|Z = 0]}$$

で求めることができ、これが2値の操作変数に対する操作変数法における推定量（以降、IV estimand ; ICH E9(R1)<sup>1</sup>の文脈での estimand とは意味が異なる）となる。本式の期待値を標本平均で置き換えたものを Wald 推定量と呼ぶ。処置  $A$  が2値変数であるため、 $E[A|Z = 1] = \text{Pr}[A = 1|Z = 1]$  と書ける。また、操作変数  $Z$  が連続値の場合の IV estimand は、

$$\frac{\text{Cov}(Y, Z)}{\text{Cov}(A, Z)}$$

となる。ここで、 $\text{Cov}$  は共分散を意味する。本式の IV estimand の導出及び詳細は、2.6.6.2及び2.6.6.3を参照のこと。

この IV estimand を直感的に理解するために、再度二重盲検下ランダム化試験を考える。IV estimand の分子は、 $Y$  に対する  $Z$  の平均因果効果であり ITT 効果（処置方針によって得られる効果）となっている。一方で、分母は、 $A$  に対する  $Z$  の平均因果効果であり割り付けられた処置の遵守程度を評価していることになる。そのため、完全に割り付けられた処置が遵守されれば、IV estimand の分母は1となり、 $Y$  に対する  $A$  の効果と  $Y$  に対する  $Z$  の効果は等しくなる。しかし、遵守状況が悪くなるにつれて分母は0に近づいていくため、 $Y$  に対する  $A$  の効果は  $Y$  に対する  $Z$  の効果よりも大きくなる。つまり、遵守率が低ければ低いほど、 $Y$  に対する  $A$  の効果（IV estimand）と  $Y$  に対する  $Z$  の効果の差は大きくなることを意味する。

IV estimand は、分子の ITT 効果を分母で増幅させることで交絡因子の調整を不要としている。処置遵守率が低下することは  $A$  に対する  $Z$  の効果（リスク差）が0に近づくことを意味し、IV estimand の分子がより増幅されることになる。これは、遵守率が低くなると ITT 効果は処置の平均因果効果よりも小さくなることを考えると、IV estimand の構造を理解しやすくなるかもしれない。

体重変化に対する禁煙の効果を観察した例において IV estimand を推定するために、期待値を標本平均で置き換えた Wald 推定量を算出してみると、 $\hat{E}[Y|Z = 1] - \hat{E}[Y|Z = 0] = 2.686 - 2.536 = 0.1503$  及び  $\hat{E}[A|Z = 1] - \hat{E}[A|Z = 0] = 0.2578 - 0.1951 = 0.0627$  より、 $0.1503 / 0.0627 = 2.4\text{kg}$  となる。この推定値は、操作変数  $Z$  が操作変数3条件を満たし、さらに2.6.3で提示する追加的条件(iv)または条件(v)を満たすとき、体重変化に対する禁煙の平均因果効果と解釈できる。ただし、本節冒頭でも述べたように追加的条件によって推定の対象が異なることに注

意が必要である。詳細は2.6.3を参照のこと。

その他の推定方法として、二段階推定法が挙げられる。二段階推定法には、主に2-stage least squares (TSLS), 2-stage prediction substitution (TSPS), 2-stage residual inclusion (TSRI)がある。これらの二段階推定法は、操作変数  $Z$  又は処置変数  $A$  が2値以外の場合や興味のある処置効果の指標が「アウトカムの平均の差」以外の場合、測定した共変量を用いた場合等で選択される<sup>18</sup>。ここでは TSLS の詳細を紹介する。今、 $Y$  に対する  $A$  の効果を調べたいため、線形回帰モデル  $Y = \beta_0 + \beta_1 A + \varepsilon$  を考え、 $Y$  に対する  $A$  の効果（リスク差）として  $\beta_1$  を推定することを考える。しかし、 $Y$  と  $A$  の間に交絡がある状況では  $A$  と  $\varepsilon$  の間に相関があり、線形回帰モデルに対する最小二乗推定量は一致推定量とはならない。そこで、パラメータ  $\beta_1$  に対する変数が  $\varepsilon$  と無相関となるように、線形回帰モデル  $A = \alpha_0 + \alpha_1 Z + u$  による予測値を用いることを考える。これが、第一段階 (first-stage) である。ここでは、モデルによる予測値に興味があるためパラメータの推定量の性質を無視して最小二乗推定量を用いてよい。予測値は  $\hat{A} = \hat{\alpha}_0 + \hat{\alpha}_1 Z$  から得られる。ここで、 $\hat{\alpha}_1 = \sum(Z - \bar{Z})(A - \bar{A}) / \sum(Z - \bar{Z})^2$  である。実際、条件(ii)及び条件(iii)より  $Cov(\varepsilon, Z) = 0$  であるため、 $Cov(\varepsilon, \hat{A}) = 0$  となることがわかる。また、条件(i)の関連の強さがこの予測精度に直結することに注意する。次に、第二段階 (second-stage) として、予測値  $\hat{A}$  に置き換えた線形回帰モデル  $Y = \beta_0 + \beta_1 \hat{A} + \varepsilon$  を考える。追加的条件(iv)より  $\beta_1$  を一つのパラメータとみなすことができ、 $\hat{A}$  と  $\varepsilon$  は無相関であることから最小二乗推定量  $\hat{\beta}_1$  は一致推定量となる。この推定量  $\hat{\beta}_1$  を操作変数法における二段階最小二乗推定量と呼ぶ。なお、この二段階最小二乗推定量は、操作変数法が一つであり、かつ第二段階におけるモデルに共変量を含めなければ、数値的には Wald 推定量と一致する。詳細は2.6.6.4を参照のこと。実際、体重変化に対する禁煙の効果を観察した例では上述の二つのモデルによる二段階最小二乗推定量も2.4kgとなる。しかし、体重変化に対する禁煙の効果を観察した例においては、2.4kg という点推定値の95%信頼区間が (-36.5, 41.3) であり、非常に広がっている。これは、 $Z$  と  $A$  の関連が弱く、第一段階における  $A$  の予測が不十分であったことが一因と考えられる。

二段階最小二乗推定量及びその分散には、強いパラメトリックな仮定を置いている。一方、構造平均モデルを用いる場合は、それらのいくつかの仮定を置かなくても良い。詳細は、2.6.6.5を参照のこと。構造平均モデルのパラメータは g-estimation を用いて推定できる。二段階最小二乗推定法もしくは構造平均モデルの g-estimation のどちらを選択するかはトレードオフの関係は、モデルに対する仮定の強さという点で、アウトカム回帰モデルもしくは構造ネストモデルのどちらを選択するかはトレードオフと類似していると言える。詳細は2.5を参照のこと。

繰り返しにはなるが、ここまで紹介してきたすべての操作変数法における推定量は、追加的条件のいずれかを満足する場合のみ平均因果効果として解釈可能となる。詳細について次節で紹介する。

### 2.6.3 IV estimand を解釈するための追加的条件

IV estimand を  $Y$  に対する  $A$  の平均因果効果と解釈するためには、操作変数3条件だけでは不十分であり、四つ目の追加的条件が必要となる。その一つが「効果の均一性」であり、この条件にはいくつか考え方があがるが、その詳細を2.6.3.1に示す。また、「効果の均一性」という条件を

仮定することがもっともらしくない状況も多くあると考えられ、そのときは代替条件として「単調性」という条件を仮定する。この条件の詳細は2.6.3.2に示す。

### 2.6.3.1 効果の均一性

四つ目の追加的条件として以下を考え、いくつかの考え方を示す。

#### (iv) 効果の均一性

(iv-1) 一つ目の考え方として、条件(iv)を最も極端に考えると、どの個人でもアウトカム Y に対する処置 A の効果が一定、ということになる。体重変化に対する禁煙の効果を観察した例に当てはめると、禁煙することで観察研究の対象となったどの個人も正確に体重が2.4kg 増加することが条件(iv)になる。しかし、実際には禁煙後に体重が増加した者もいれば、減少した者もいるであろう。体重がまったく変わらない者もいれば、禁煙せずとも体重が変動する者もいるはずである。この例からも、どの個人でも効果が一定であるとしてこの条件(iv)を仮定することはもっともらしいとは言えない。

(iv-2) 二つ目の考え方として、一つ目の考え方 (iv-1) よりも極端な程度をやや下げてみる。Z と A は2値変数とし、A=1の集団及び A=0の集団それぞれで、Y に対する A の平均因果効果が Z 間で等しい、つまり加法的 (additive) に考えると  $E[Y^{a=1} - Y^{a=0} | Z = 1, A = a] = E[Y^{a=1} - Y^{a=0} | Z = 0, A = a]$  for  $a = 0, 1$  という条件(iv)であり、乗法的 (multiplicative) に考えると、 $E[Y^{a=1} | Z = 1, A = a] / E[Y^{a=0} | Z = 1, A = a] = E[Y^{a=1} | Z = 0, A = a] / E[Y^{a=0} | Z = 0, A = a]$  for  $a = 0, 1$  という条件(iv)になる。

(iv-3) 三つ目の考え方は、交絡因子 U が効果修飾因子ではない、つまり Y に対する A の平均因果効果が交絡因子 U のどのレベルでも同じである  $E[Y^{a=1} | U] - E[Y^{a=0} | U] = E[Y^{a=1}] - E[Y^{a=0}]$  ということである。二つ目の考え方 (iv-2) 及び三つ目の考え方 (iv-3) はどの交絡因子 U に対しても潜在アウトカムを用いた上記の条件が成立する必要があるため、この条件を仮定することももっともらしいとは言えないことがわかる。

(iv-4) 最後に四つ目の考え方は、交絡因子 U 間において Z と A の相関が一定である、つまり  $E[A | Z = 1, U] - E[A | Z = 0, U] = E[A | Z = 1] - E[A | Z = 0]$  という条件である。これまでの三つの考え方 (iv-1, iv-2, iv-3) とは異なり、一部検証可能な条件である。例えば、いくつかの測定されている交絡因子については、そのすべてのレベルで Z と A の相関が一定でなければ、四つ目の考え方 (iv-4) では条件(iv)を満足しないことになる。つまり、条件(iv)が成立することを完全に保証することはすべての交絡因子 U で確認するため難しいが、条件(iv)が成立しないことを検証することは可能である。

ここまで、条件(iv)に対して四つの考え方 (iv-1, iv-2, iv-3, iv-4) を見てきたが、いずれの考え方においてもその条件を仮定することは非現実的でもっともらしいとはいえないものであり、処置の平均因果効果として解釈可能な推定値を操作変数法によって導くことができるかには疑問が残るかもしれない。特に、二段階最小二乗推定量が一致推定量となるには、一つ目の考え方の条件(iv)が必要となる。そのため、二段階最小二乗推定法よりもパラメトリックな仮定が少

ない構造平均モデルを利用することは、選択肢の一つとなるだろう。操作変数法における構造平均モデルについては、2.6.6.5を参照のこと。

### 2.6.3.2 単調性

「効果の均一性」の代替条件として、次の条件を仮定する場合がある。

#### (v) 単調性

この条件を整理するために、もう一度二重盲検下ランダム化試験を考え、 $Z$  をランダム割付の指示変数 ( $Z=1$ が被験治療群,  $Z=0$ がプラセボ治療群),  $A$  を処置変数 ( $A=1$ が被験治療を受ける,  $A=0$ が被験治療を受けない),  $Y$  をアウトカム変数とする。  $A^{z=1}$  及び  $A^{z=0}$  をそれぞれ、ある個人が被験治療群 ( $z=1$ ) 及びプラセボ治療群 ( $z=0$ ) に割り付けられたときに取る潜在処置変数 (1か0を取る2値変数) とする。

仮に各個人の2つの潜在処置変数  $A^{z=1}$  及び  $A^{z=0}$  がわかるとすると、この試験集団におけるすべての個人を互いに排反な次の4つの部分集団を構成できる (表 2.6-1)。

#### 1) Always-taker :

どの治療群に割り付けられたかによらず、常に被験治療を受ける集団。つまり、 $A^{z=1} = 1$  かつ  $A^{z=0} = 1$  となる集団。

#### 2) Never-taker :

どの治療群に割り付けられたかによらず、決して被験治療を受けない集団。つまり、 $A^{z=1} = 0$  かつ  $A^{z=0} = 0$  となる集団。

#### 3) Complier :

割付に従って、被験治療群に割り付けられたときは被験治療を受け、プラセボ群に割り付けられたときは被験治療を受けない集団。つまり、 $A^{z=1} = 1$  かつ  $A^{z=0} = 0$  となる集団。

#### 4) Defier :

割付に従わず、被験治療群に割り付けられたときは被験治療を受けず、プラセボ群に割り付けられたときは被験治療を受ける集団。つまり、 $A^{z=1} = 0$  かつ  $A^{z=0} = 1$  となる集団。

表 2.6-1 潜在変数を用いた処置変数のタイプ

	$A^{z=1}$	$A^{z=0}$
Always-taker	1	1
Never-taker	0	0
Complier	1	0
Defier	0	1

表 2.6-1の部分集団の区分は、compliance type や principal strata (主要層) 等と呼ばれ、実際には各個人がどの部分集団に属するかを特定することはできない。例えば、仮に、被験治療群 ( $Z=1$ ) に割り付けられ実際に被験治療を受けた ( $A=1$ ) 個人がいても、その個人が complier なのか

always-taker なのかわからない。

表 2.6-1 の部分集団のうち, defier が存在しないとき, すべての個人に対して  $A^{z=1} \geq A^{z=0}$  となることから, 条件(v)「単調性」が成立していると言う。

条件(iv)を条件(v)で代替していると考えると, IV estimand は全体集団における平均因果効果  $E[Y^{a=1}] - E[Y^{a=0}]$  ではなく, complier (割付に従う集団) における平均因果効果  $E[Y^{a=1} - Y^{a=0} | A^{z=1} = 1, A^{z=0} = 0]$  (Compliers Average Causal Effect; CACE) となる。CACE は局所平均因果効果 (Local Average Causal Effect; LACE) の一つである。ここでは, 単調性の仮定の下で, IV estimand が CACE となることの証明の概略を示すことにする。IV estimand の分子は, Y に対する Z の効果であり, 4つの各主要層における Z の効果の重み付き平均となる。しかし, always-taker 及び never-taker においては, A の値は各部分集団内では一定となり, 条件(ii)より Z の効果は完全に A を介するため, Z の効果は0となる。そして, 条件(v)の下では defier が存在しないので, IV estimand における分子は complier における Y に対する Z の効果であり, これは complier における Y に対する A の効果と同値である。また, IV estimand における分母は, 全体集団における complier の割合となる。以上から, 条件(v)の下では, IV estimand は complier における処置の平均因果効果となる。詳細は2.6.6.6を参照のこと。

観察研究では, defier が存在しない状況での complier における効果を推定するために IV estimand が用いられることがある。体重変化に対する禁煙の効果を観察した例では, タバコの値段が高ければ (Z=1) 禁煙し (A=1), タバコの値段が安ければ (Z=0) 喫煙を継続する (A=0) 集団を complier とする。逆を言えば, タバコの値段が高ければ (Z=1) 喫煙を継続し (A=0), タバコの値段が安ければ (Z=0) 禁煙する (A=1) 集団が defier となる。もし defier が存在せず, 操作変数 Z が因果的操作変数 (定義は2.6.6.1を参照) であれば, IV estimand の Wald 推定量であった2.4kg は complier における平均因果効果の一致推定量となる。

条件(iv)を条件(v)で代替することで, complier という限られた集団における効果しか推定できないが, 条件(iv)を仮定するよりも条件(v)を仮定した方がもっともらしい場合もあるため, そのような状況下では条件(v)を仮定した下で操作変数法を用いて効果を推定することは有用である。しかし, 条件(v)を仮定することに対する限界もいくつかある。

一つ目は, 観察データからは complier という集団が特定できないという点である。仮に, 全体集団における complier の割合 (IV estimand における分母) を求めることができて, 操作変数として選択した変数次第でその割合は変動し, complier という集団がどういう集団か一般化することができない。つまり, complier における効果として解釈できる推定値が得られたとしても, 実際にはどういった集団に対する効果なのかを特定できないことになり, 意思決定に用いることが難しいことがわかる。

二つ目は, 特に観察研究において単調性の仮定はほとんどの場合で妥当でないという点である。ランダム化試験であれば, プラセボ治療群 (Z=0) に割り付けられながら被験治療を受けること (A=1) はほとんどないと考えられるため, defier の不存在という仮定は受け入れられやすい。同様に always-taker も存在しないことになるが, complier か never-taker かの区別はできない。一方, 観察研究においては, 単調性の仮定を正当化するのは難しい。例えば, 医師の好みを反映する変数 Z を考える。ある医師は被験治療薬を処方する割合が高い (Z=1) が, 糖尿病の患者に

は処方避ける傾向にあり、また別の医師は被験治療薬を処方する割合が低い ( $Z=0$ ) が、疾患活動が高い場合は処方をする傾向にある、という2人の医者がいたとする。このとき、糖尿病を合併し疾患活動が高い患者は  $Z=1$  の医師の診療を受けると被験治療を受けず ( $A=0$ )、 $Z=0$  の医師の診療を受けると被験治療を受ける ( $A=1$ ) ことになり、**defier** ( $A^{Z=1} = 0$  かつ  $A^{Z=0} = 1$ ) であることがわかる。つまり、リスクとベネフィットの両面で複数の判断基準によって処置が選択される状況下では **defier** の不existenceは考えにくい。

2.6.6.1で示した代替操作変数  $Z$  の場合は、さらに複雑になる。因果的操作変数  $U_z$  が連続量、 $Z$  が2値であり、この  $Z$  を用いて操作変数法により推定された推定値は **complier** における効果とはならない。むしろ、この推定値は単に全集団における重み付き平均となっており、解釈が難しい。これは、変数  $Z$  が因果的操作変数でない場合は、 $U_z$  で単調性が成立していても、 $Z$  に対して単調性が成立することを保証できないためである。なお、 $U_z$  が2値の場合は、 $U_z$  が与えられた下で  $Z$  が  $A$  及び  $Y$  と独立のとき、操作変数法による推定量を **complier** における平均因果効果と解釈可能になる<sup>19</sup>。

最後に、特に重要なことは、母集団を四つの部分集団または主要層に分割することが正当化されない場合があることである。多くの現実的な状況では、**complier** の部分集団は母集団に対して **ill-defined** な集団である。例えば、複数の医師がいる状況で「医師の好み」を反映する変数  $Z$  を用いると、ある患者を診察した可能性のある同じ嗜好レベルのすべての医師が、全く同じ処置を選択しなければならないことになる。これは非現実的な仮定であるだけでなく、どの医師が患者を診ることができたかが不明な多くの観察研究では、定義することが本質的に不可能である。

## 2.6.4 操作変数法によるバイアス

前節までは四つ目の追加的条件として条件(iv)と条件(v)のそれぞれの長所短所について述べてきた。その中では、変数  $Z$  が真に操作変数であることは暗に仮定していた。しかし、観察研究においては、変数  $Z$  が操作変数3条件を満足しておらず、有用な操作変数ではないかもしれない。また、条件(i)だけが満足していたとしても、それは弱い操作変数かもしれない。つまり、前節で述べた四つ目の追加的条件が成立していたとしても、操作変数法による推定値は大きなバイアスを持っているかもしれない。本節では、操作変数3条件それぞれの仮定が結果に与える影響について考える。

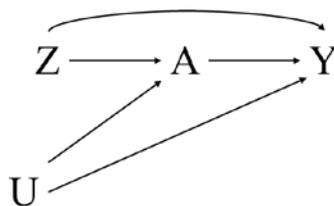
条件(i)はデータから確認できるため、 $Z$  が操作変数になり得るかを確認する際には、まずは  $Z$  と処置  $A$  の関連を調べる。しかし、体重変化に対する禁煙の効果を観察した例のようにその関連が弱いとき、 $Z$  は弱い操作変数となり、これには3つの問題がある。

一つ目は、体重変化に対する禁煙の効果を観察した例のように、信頼区間が広がることである。二つ目は、条件(ii)及び条件(iii)を満足しない場合に起こるバイアスを増幅させることである。これは、弱い操作変数の場合、IV estimandの分母を小さくすることになるため、IV estimandの分子に影響する条件(ii)及び条件(iii)の不満足によるバイアスをさらに増幅させることになるためである。三つ目は、たとえ大標本の場合でも弱い操作変数はバイアスをもたらすため、大標本のために狭い信頼区間を持つ点推定値は誤った箇所であり、分散が過小評価されていること

になる。

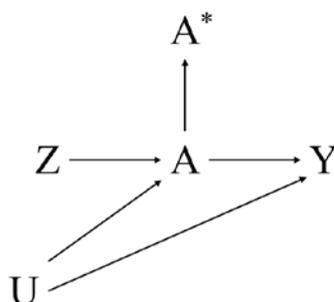
この三つ目の問題をさらに理解するために、ランダムに生成される2値変数  $Z$  を考える。無限標本の下では、 $Z$  がランダムに生成されるため  $A$  との関連は見られず、IV estimand の分母はゼロになるはずである。しかし、有限標本の下では、偶然にも弱い相関という形で、 $Z$  と  $A$  の間に関連が見られるかもしれない。もしこのランダムに生成される変数  $Z$  を操作変数としたとき、IV estimand の分母は0にはならないがとても小さい値になる。結果として、IV estimand の分子を不正確に増幅させることになり、潜在的に大きなバイアスを持つことになる。このように弱い操作変数をもたらすバイアスや不確かさを考慮すると、条件(ii)及び条件(iii)を満足する程度が少々低くても、関連の強い操作変数を用いた方が望ましい場合があるかもしれない。

条件(ii)は、アウトカムに対して操作変数の直接効果はなく、 $A$  を介してのみ操作変数の効果があるというもので、これはデータから確認することができない条件であった。図 2.6-2 のような場合は、条件(ii)は満足しないことになる。処置  $A$  を介することのない  $Y$  に対する操作変数の効果は IV estimand の分子に影響を及ぼし、 $Y$  に対する  $A$  の効果であるかのように分母によって不正確に増幅されることになる。 $A$  が連続量や多変量であった場合に、図 2.6-3 のようにそれを2値変数  $A^*$  で置き換えることがあるかもしれないが、このとき条件(ii)は満足しない。なぜなら、 $A$  に対しては条件(ii)に満足していたとしても、 $Z$  は  $A^*$  を介することのない  $Y$  に対する効果 ( $Z \rightarrow A \rightarrow Y$ ) を持つことになるためである。



[Hernán & Robins (2020). Chapter 16, Figure 16.8]

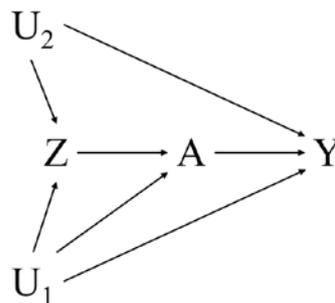
図 2.6-2 条件(ii)を満足しない例 (1)



[Hernán & Robins (2020). Chapter 16, Figure 16.9]

図 2.6-3 条件(ii)を満足しない例 (2)

条件(iii)はアウトカムに対する操作変数の効果を見る場合において交絡がないというもので、これもデータから確認することができない条件であった。図 2.6-4は操作変数  $Z$  とアウトカム  $Y$  の共通原因によって交絡が起きている状況を示している。観察研究においては、一般にこのような  $Z$  と  $Y$  の共通原因が存在する。このような共通原因が存在する状況では、IV estimand の分子が  $Y$  に対する  $Z$  の効果とならず、分母によって不正確に分子が増幅されることになる。



[Hernán & Robins (2020). Chapter 16, Figure 16.10]

図 2.6-4 条件(iii)を満足しない例

$Y$  に対する  $Z$  の効果に対して完全に交絡がないことを仮定するのではなく、共変量  $V$  の各レベルでは  $Y$  に対する  $Z$  の効果に対して交絡がないと仮定する方が自然な場合もあるだろう。このとき、共変量  $V$  の各レベルに対して操作変数法を用いて推定することができ、二段階最小二乗推定法では共変量として変数  $V$  を含めることで推定できる。

$Y$  に対する  $Z$  の効果に対して交絡がないことを確認するために、測定されている共変量に関してバランスが取れているかを確認することがあるかもしれない。測定されている共変量でバランスが取れていれば、未測定の共変量に対しても同じように考えようとするだろう。しかし、必ずしも未測定の共変量までバランスが取れているとは言えず、仮にわずかなバランスの崩れでも、ここまで述べてきたように大きなバイアスをもたらす可能性もある。

## 2.6.5 操作変数法とその他の手法

操作変数法は、他の節で説明されている因果推論の方法 (IPW, 標準化, g-estimation) のいずれとも少なくとも次の三つの側面で異なるものと言える。

一つ目は、例え無限のデータが利用可能であっても、操作変数法はモデルの仮定が必要ということである。IPW や標準化では、処置, アウトカム, 交絡因子に関するデータが母集団のすべての個人に対して得られていれば、ノンパラメトリックに平均因果効果を推定することができた。一方で、操作変数, 処置, アウトカムに関するデータが母集団のすべての個人に対して得られたとしても、操作変数法を用いて平均因果効果を推定するためにはパラメトリックな仮定が必要となる。詳細は、2.6.6.2を参照のこと。

二つ目は、操作変数法で要求される条件(i)~(iii), 及び(iv)または(v)の4つの条件において比較

的わずかなずれがあっても、予測できない、そして方向もわからない大きなバイアスをもたらす可能性があることである。IV estimand は、分子をコンプライアンスの程度である分母によって増幅させている。そのため、操作変数の条件が完全に満足していなかったり、弱い操作変数であったりすると、いずれの方向にもバイアスをもたらすことになる。結果として、調整しない解析結果よりも操作変数法による推定法の方がバイアスを持つことになるかもしれない。対照的に、他の節で説明されているその他の因果推論の方法は、要求される条件からわずかに外れた場合には、それに応じたバイアスをもたらす傾向にあり、大きなバイアスを導くことは少ない。つまり、操作変数法は要求される条件を満足しないことに対して、非常に弱いことになり、他の手法と比較して注意が必要になる。

三つ目は、操作変数法を利用できるような理想的状況が限られていることである。操作変数法は、未測定交絡因子が多く、処置 A が2値であり、強い操作変数 Z が存在し、「効果の均一性」または「単調性」の仮定が合理的である状況に限られるためである。このことから、操作変数法を時間依存的に処置が変化する状況や複雑な処置方針に対して用いることは望ましくないことがわかるだろう。

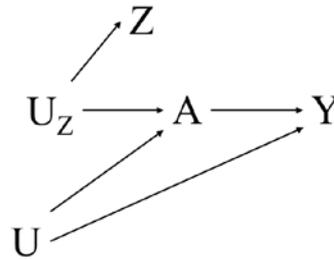
因果推論では、仮定のもっともらしさから結果を解釈することになるため、異なる仮定を用いて得られる結果が複数あることは望ましいことであり、その観点では操作変数法は魅力的な手法の一つと言える。因果推論におけるいずれの手法においても、それぞれの解析方法の限界を認識することが重要であり、特に操作変数法においては予測できない方向、大きさのバイアスを持つ可能性があるため、特に注意が必要である。

## 2.6.6 付録

2.6.1節から2.6.5節の内容に関する補足説明を本節にまとめた。読者の興味や必要性に応じて参照のこと。各付録を引用している箇所は2.6.1節から2.6.5節の本文から確認されたい。

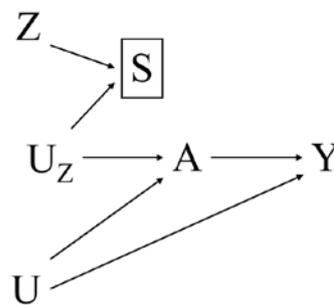
### 2.6.6.1 操作変数のタイプ

図 2.6-1は操作変数 Z が処置 A に対して直接的な因果関係をもつ状況を表現していた。このとき、Z を因果的操作変数 (causal instrument) と呼ぶ。因果的操作変数以外の操作変数は、処置 A に対して直接的な因果関係を持たない。図 2.6-5における Z は、Z と A の共通原因  $U_z$  による関連から操作変数の条件(i)を満足する。このとき、 $U_z$  を未測定の因果的操作変数と呼び、Z を測定された代替 (surrogate または proxy) 操作変数と呼ぶ。ここで、 $U_z$  が Z と Y の共通原因になっており、条件(iii)を満足しないように見える。実は、条件(iii)の本質は、潜在アウトカム  $Y^{a,z}$  と操作変数 Z が独立であり、Z に関して交換可能性が成立することである。図 2.6-5に対応する Single World Intervention Graph (SWIG) を用いると、その条件が成立することが分かり、操作変数 Z が条件(iii)の逸脱とはならないことがわかる。SWIG の詳細は Richardson and Robins<sup>20</sup>を参照のこと。図 2.6-6は、Z がある部分集団において代替操作変数となる例である。代替操作変数 Z と因果的操作変数  $U_z$  の共通の子孫である S で条件付けることによって、Z と A の間に関連が生まれる。



[Hernán &amp; Robins (2020). Chapter 16, Figure 16.2]

図 2.6-5 代替操作変数の例 (1)



[Hernán &amp; Robins (2020). Chapter 16, Figure 16.3]

図 2.6-6 代替操作変数の例 (2)

### 2.6.6.2 線形回帰モデルからの IV estimand の導出

変数  $Z$  を操作変数3条件及び追加的条件(iv)を満足する操作変数とする。今、 $Y$  に対する  $A$  の効果を調べたいため、線形回帰モデル  $Y = \beta_0 + \beta_1 A + \varepsilon$  を考え、 $Y$  に対する  $A$  の効果（リスク差）として  $\beta_1$  を推定したい。しかし、 $Y$  と  $A$  の間に交絡がある状況では  $A$  と  $\varepsilon$  の間に相関があり、線形回帰モデルに対する最小二乗推定量は一致推定量とはならない。そこで、操作変数  $Z$  を用いる。条件(i)より  $Cov(A, Z) \neq 0$ 、条件(ii)及び条件(iii)より  $Cov(\varepsilon, Z) = 0$  となり、追加的条件(iv)より  $\beta_1$  を一つのパラメータと見なすことができるため、

$$\begin{aligned} Cov(Y, Z) &= Cov(\beta_0 + \beta_1 A + \varepsilon, Z) \\ &= 0 + \beta_1 Cov(A, Z) + 0 \\ \beta_1 &= \frac{Cov(Y, Z)}{Cov(A, Z)} \end{aligned}$$

となり、上式の IV estimand が導出される。

### 2.6.6.3 IV estimand の変形

操作変数  $Z$  及び処置  $A$  は2値とする。このとき、 $E[YZ]$  は

$$\begin{aligned}
E[YZ] &= \sum_y \sum_z yz \cdot \Pr[Y = y, Z = z] \\
&= \sum_y \{y \cdot 1 \cdot \Pr[Y = y, Z = 1] + y \cdot 0 \cdot \Pr[Y = y, Z = 0]\} \\
&= \Pr[Z = 1] \sum_y y \cdot \frac{\Pr[Y = y, Z = 1]}{\Pr[Z = 1]} \\
&= \Pr[Z = 1] \cdot E[Y|Z = 1]
\end{aligned}$$

と書ける。また、 $E[Y]E[Z]$  は

$$\begin{aligned}
E[Y]E[Z] &= \left( \sum_y y \cdot \Pr[Y = y] \right) \left( \sum_z z \cdot \Pr[Z = z] \right) \\
&= \left\{ \sum_y (y \cdot \Pr[Y = y, Z = 1] + y \cdot \Pr[Y = y, Z = 0]) \right\} \times (1 \cdot \Pr[Z = 1] + 0 \cdot \Pr[Z = 0]) \\
&= \left( \Pr[Z = 1] \sum_y y \cdot \frac{\Pr[Y = y, Z = 1]}{\Pr[Z = 1]} + \Pr[Z = 0] \sum_y y \cdot \frac{\Pr[Y = y, Z = 0]}{\Pr[Z = 0]} \right) \times \Pr[Z = 1] \\
&= (\Pr[Z = 1] \cdot E[Y|Z = 1] + \Pr[Z = 0] \cdot E[Y|Z = 0]) \Pr[Z = 1]
\end{aligned}$$

と書ける。これより、 $E[YZ] - E[Y]E[Z]$  は

$$\begin{aligned}
E[YZ] - E[Y]E[Z] &= \Pr[Z = 1] \{(1 - \Pr[Z = 1]) \cdot E[Y|Z = 1] - \Pr[Z = 0] \cdot E[Y|Z = 0]\} \\
&= \Pr[Z = 1] \Pr[Z = 0] (E[Y|Z = 1] - E[Y|Z = 0])
\end{aligned}$$

となる。同様にして、 $E[AZ] - E[A]E[Z]$  は

$$E[AZ] - E[A]E[Z] = \Pr[Z = 1] \Pr[Z = 0] (E[Y|Z = 1] - E[Y|Z = 0])$$

となる。よって、

$$\begin{aligned}
\frac{Cov(Y, Z)}{Cov(A, Z)} &= \frac{E[YZ] - E[Y]E[Z]}{E[AZ] - E[A]E[Z]} \\
&= \frac{\Pr[Z = 1] \Pr[Z = 0] (E[Y|Z = 1] - E[Y|Z = 0])}{\Pr[Z = 1] \Pr[Z = 0] (E[Y|Z = 1] - E[Y|Z = 0])} \\
&= \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[A|Z = 1] - E[A|Z = 0]}
\end{aligned}$$

のように表現できる。

#### 2.6.6.4 二段階最小二乗推定量と Wald 推定量の関係

二段階最小二乗推定量は、線形回帰モデル  $Y = \beta_0 + \beta_1 \hat{A} + \varepsilon$  における最小二乗推定量であった。ここで、 $\hat{A} = \hat{\alpha}_0 + \hat{\alpha}_1 Z$ 、 $\hat{\alpha}_1 = \Sigma(Z - \bar{Z})(A - \bar{A}) / \Sigma(Z - \bar{Z})^2$  である。よって、

$$\hat{\beta}_1 = \frac{\Sigma(\hat{A} - \bar{\hat{A}})(Y - \bar{Y})}{\Sigma(\hat{A} - \bar{\hat{A}})^2}$$

$$\begin{aligned}
&= \frac{\sum(\hat{\alpha}_0 + \hat{\alpha}_1 Z - (\hat{\alpha}_0 + \hat{\alpha}_1 \bar{Z}))(Y - \bar{Y})}{\sum(\hat{\alpha}_0 + \hat{\alpha}_1 Z - (\hat{\alpha}_0 + \hat{\alpha}_1 \bar{Z}))^2} \\
&= \frac{\sum(\hat{\alpha}_1(Z - \bar{Z}))(Y - \bar{Y})}{\sum(\hat{\alpha}_1(Z - \bar{Z}))^2} \\
&= \frac{\sum(Z - \bar{Z})(Y - \bar{Y})}{\hat{\alpha}_1 \sum(Z - \bar{Z})^2} \\
&= \frac{\sum(Z - \bar{Z})(Y - \bar{Y})/n}{\sum(Z - \bar{Z})(A - \bar{A})/n}
\end{aligned}$$

となり、 $\hat{\beta}_1$  が IV estimand を標本共分散で置き換えた Wald 推定量と一致することがわかる。また、標本共分散は母共分散に確率収束することから、

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum(Z - \bar{Z})(Y - \bar{Y})/n}{\sum(Z - \bar{Z})(A - \bar{A})/n} \\
&= \frac{\sum(Z - \bar{Z})\{(\beta_0 + \beta_1 \hat{A} + \varepsilon) - (\beta_0 + \beta_1 \bar{A} + \bar{\varepsilon})\}/n}{\sum(Z - \bar{Z})(A - \bar{A})/n} \\
&= \frac{\sum(Z - \bar{Z})\{\beta_1(\hat{A} - \bar{A}) + (\varepsilon - \bar{\varepsilon})\}/n}{\sum(Z - \bar{Z})(A - \bar{A})/n} \\
&= \beta_1 + \frac{\sum(Z - \bar{Z})(\varepsilon - \bar{\varepsilon})/n}{\sum(Z - \bar{Z})(A - \bar{A})/n} \\
&\rightarrow \beta_1 + \frac{\text{Cov}(\varepsilon, Z)}{\text{Cov}(A, Z)} = \beta_1 + \frac{0}{\text{Cov}(A, Z)} = \beta_1 \quad (n \rightarrow \infty)
\end{aligned}$$

となり、Wald 推定量は一致推定量であることがわかる。なお、 $\text{Cov}(\varepsilon, A) \neq 0$  であるため、不偏推定量とはならない。

### 2.6.6.5 Additive 構造平均モデル

処置 A 及び操作変数 Z は2値とし、加法的な構造平均モデル（飽和モデル）として、 $E[Y^{a=1} - Y^{a=0} | A = 1, Z] = \beta_0 + \beta_1 Z$  を考える。このモデルは、 $E[Y - Y^{a=0} | A, Z] = A(\beta_0 + \beta_1 Z)$  とも表現できる。このとき、A=1の集団において、パラメータ  $\beta_0$  は Z=0の個人における平均因果効果、パラメータ  $\beta_0 + \beta_1$  は Z=1の個人における平均因果効果となる。つまり、パラメータ  $\beta_1$  は Z による効果修飾の大きさとなる。

このモデルにおいて、Z による効果修飾はないと仮定すると、 $\beta_1 = 0$  となり、 $E[Y - Y^{a=0} | A, Z] = A(\beta_0 + \beta_1 Z) = A\beta_0$  の  $\beta_0$  は IV estimand に一致する。つまり、IV estimand は、Z による効果修飾がない A=1の集団に対する処置効果を表現した加法的な構造平均モデルのパラメータであると言える。

証明は単純である。Z は操作変数であるため条件(ii)が成立し、 $E[Y^{a=0} | Z = 1] = E[Y^{a=0} | Z = 0]$  となる。また、構造平均モデル  $E[Y - Y^{a=0} | A, Z] = A\beta_0$  に対して、Z が与えられた下での条件付き期待値を考える。このとき、

$$E[E[Y - Y^{a=0} | A, Z] | Z] = \sum_A E[Y - Y^{a=0} | A, Z] \cdot f_{A|Z}$$

$$\begin{aligned}
&= \sum_A \left[ \sum_Y (Y - Y^{a=0}) \cdot f_{Y|A,Z} \right] \cdot f_{A|Z} \\
&= \sum_Y (Y - Y^{a=0}) \sum_A f_{Y|A,Z} \cdot f_{A|Z} \\
&= \sum_Y (Y - Y^{a=0}) \sum_A f_{(Y,A)|Z} \\
&= \sum_Y (Y - Y^{a=0}) \cdot f_{Y|Z} \\
&= E[Y - Y^{a=0}|Z] \\
&= E[Y|Z] - E[Y^{a=0}|Z],
\end{aligned}$$

$$E[A\beta_0|Z] = \beta_0 \cdot E[A|Z]$$

となるため、 $E[Y^{a=0}|Z] = E[Y|Z] - \beta_0 \cdot E[A|Z] = E[Y - A\beta_0|Z]$  と書ける。よって、

$$E[Y^{a=0}|Z = 1] = E[Y^{a=0}|Z = 0]$$

$$E[Y - A\beta_0|Z = 1] = E[Y - A\beta_0|Z = 0]$$

$$E[Y|Z = 1] - \beta_0 \cdot E[A|Z = 1] = E[Y|Z = 0] - \beta_0 \cdot E[A|Z = 0]$$

$$\beta_0 \cdot (E[A|Z = 1] - E[A|Z = 0]) = E[Y|Z = 1] - E[Y|Z = 0]$$

$$\beta_0 = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[A|Z = 1] - E[A|Z = 0]}$$

となり、 $\beta_0$  は IV estimand に一致することがわかる。

Zによる効果修飾はないと仮定しない、つまり $\beta_1 \neq 0$ の構造平均モデルは、操作変数3条件の制約を課しているものであるが、なお未知のパラメータが2つあり、操作変数3条件のみでは平均因果効果を特定することが難しいことを示唆している。

また、上記で $\beta_0$ はIV estimandに一致することを示したが、 $\beta_0$ はA=1の集団に対する平均因果効果であって、全体集団に対する平均因果効果 $E[Y^{a=1} - Y^{a=0}]$ ではない。 $\beta_0$ を全体集団に対する平均因果効果と解釈するには、A=1の集団に対する平均因果効果とA=0の集団に対する平均因果効果が同一であるという仮定、つまりA=0の集団においても $\beta_1 \neq 0$ の構造平均モデルを仮定する必要がある、これもまた検証不可能な仮定である。

### 2.6.6.6 単調性の条件と complier における効果

処置 A 及び操作変数 Z は2値とする。Z がランダム割付であるとき、IV estimand が単調性の仮定の下で complier における平均因果効果となることを示す。

まず、ITT 効果を、各主要層の ITT 効果の重み付き平均と考えて次のように展開する。

$$\begin{aligned}
E[Y^{z=1} - Y^{z=0}] &= E[Y^{z=1} - Y^{z=0}|A^{z=1} = 1, A^{z=0} = 1] \cdot \Pr[A^{z=1} = 1, A^{z=0} = 1] \\
&\quad + E[Y^{z=1} - Y^{z=0}|A^{z=1} = 0, A^{z=0} = 0] \cdot \Pr[A^{z=1} = 0, A^{z=0} = 0] \\
&\quad + E[Y^{z=1} - Y^{z=0}|A^{z=1} = 1, A^{z=0} = 0] \cdot \Pr[A^{z=1} = 1, A^{z=0} = 0] \\
&\quad + E[Y^{z=1} - Y^{z=0}|A^{z=1} = 0, A^{z=0} = 1] \cdot \Pr[A^{z=1} = 0, A^{z=0} = 1].
\end{aligned}$$

第1項から第4項まで、それぞれ always-taker, never-taker, complier, defier の ITT 効果を示している。

ここで、条件(ii)より  $E[Y^{z=1}|A^{z=1} = A^{z=0}] = E[Y^{z=0}|A^{z=1} = A^{z=0}]$  が成立するため、always-taker 及び never-taker の ITT 効果は0になる。また、単調性の仮定から defier は存在しないため、defier の ITT 効果の項も0になる。よって、全体の ITT 効果は、単に

$$E[Y^{z=1} - Y^{z=0}] = E[Y^{z=1} - Y^{z=0}|A^{z=1} = 1, A^{z=0} = 0] \cdot \Pr[A^{z=1} = 1, A^{z=0} = 0]$$

と書ける。

Complier においては、Y に対する Z の効果は Y に対する A の効果と等しいため、

$$E[Y^{z=1} - Y^{z=0}|A^{z=1} = 1, A^{z=0} = 0] = E[Y^{a=1} - Y^{a=0}|A^{z=1} = 1, A^{z=0} = 0]$$

となるため、

$$E[Y^{a=1} - Y^{a=0}|A^{z=1} = 1, A^{z=0} = 0] = \frac{E[Y^{z=1} - Y^{z=0}]}{\Pr[A^{z=1} = 1, A^{z=0} = 0]}$$

と書ける。

ここで、Z はランダム割付であるため、任意の z 及び a に対して  $Y^{a,z}$  と Z は独立であると言える。また、任意の z に対して  $Y^z = Y$  (一致性) が成立するとすると、

$$\begin{aligned} E[Y^{z=1} - Y^{z=0}] &= E[Y^{z=1}] - E[Y^{z=0}] \\ &= E[Y^{z=1}|Z = 1] - E[Y^{z=0}|Z = 0] \\ &= E[Y|Z = 1] - E[Y|Z = 0] \end{aligned}$$

となる。さらに、任意の z に対して  $A^z = A$  (一致性) が成立するとし、表 2.6-1、単調性の仮定、処置 A が2値であること、任意の z に対して  $A^z$  と Z は独立であることから、

$$\begin{aligned} \Pr[A^{z=1} = 1, A^{z=0} = 0] &= 1 - \Pr[A^{z=1} - A^{z=0} = 0] \quad (\text{単調性より}) \\ &= 1 - \Pr[A^{z=0} = 1] - \Pr[A^{z=1} = 0] \\ &= 1 - E[A^{z=0}] - \{1 - E[A^{z=1}]\} \\ &= E[A^{z=1}|Z = 1] - E[A^{z=0}|Z = 0] \\ &= E[A|Z = 1] - E[A|Z = 0] \quad (\text{一致性より}) \end{aligned}$$

となる。以上のことから、

$$E[Y^{a=1} - Y^{a=0}|A^{z=1} = 1, A^{z=0} = 0] = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[A|Z = 1] - E[A|Z = 0]}$$

となり、IV estimand が complier における平均因果効果となることがわかる。

### 3. 臨床試験データへの因果推論の適用

#### 3.1 系統的レビュー論文

Farmer ら<sup>21</sup>は、ランダム化比較試験の解析への因果推論の適用について系統的レビューを実施している。ただし、本レビュー論文では、ランダム化された治療のノンコンプライアンスの調整は、レビューの対象外となっている。

ランダム化された治療群間比較以外の比較に関心があることがあり、その際の解析では、ランダム化比較試験であっても観察研究と同様に交絡等の問題が生じうる。例えば、HIV 感染者における生存に対する抗レトロウイルス療法の効果を調べたいと考えたとき、CD4陽性リンパ球数は、生存に対する時間依存性のリスク因子であり、治療の開始とも関連しているため、時間依存性の交絡因子となる。しかし、抗レトロウイルス療法は、治療後に CD4陽性リンパ球数も改善させる。時間依存性の交絡因子が治療の影響を受けるとき、標準的な回帰モデル等ではこのような交絡を適切に調整できない。そこで、因果推論の適用を考えることとなる。

本レビュー論文の目的は、ランダム化比較試験の解析の中で因果推論がどれくらい広く適用されているかを把握し、それらの方法の使用によって得られる価値を理解することである。ランダム化された治療群間比較以外の比較として、本レビュー論文では、併用治療、治療のスイッチング、動的な治療戦略等の有用性の評価に焦点を当てている。

臨床試験の検索方法として、まずキーワード検索を実施している。期間は1986年から2014年9月（英語論文のみ）とし、データベースは、Medline, Premedline, Embase, Cochrane Library 及び Web of Science を用いている。具体的なキーワードは、((time-varying confounding OR causal effect or parameter OR causal inference) AND (marginal structural models OR inverse probability weighting OR g-estimation OR g-formula OR structural nested models))としている。次に5個の影響力の大きい論文<sup>6,22,23,24,25</sup>の被引用による検索を実施している。臨床試験のスクリーニングとして、まず要約（アブストラクト）での1次スクリーニングを実施し、検索で定義された因果推論の方法を使用しかつ臨床試験データを使用している論文を選択している。次に、論文全文での2次スクリーニングを実施し、因果推論の方法を使用していない論文、理論のみの論文、方法の詳細が不明な学会の要約、チュートリアル論文または観察研究の論文、ベースライン共変量等の不均衡の調整に IPW のみを適用した論文、ランダム化された治療のノンコンプライアンスの調整のみに因果推論の方法を適用した論文、曝露が割付できないものを対象とした論文、シミュレーションのみの理論的な論文、適用が観察研究のデータのみの理論的な論文または Sequential multiple assignment randomised trial (SMART) デザインを解析した論文を除外している。

因果的な臨床疑問について、本レビュー論文では、併用治療、一連の治療、治療のタイミング、治療の用量や期間の4つのタイプに分類している。併用治療とは、ランダム化された治療以外で使用された治療の効果に関心がある場合に該当する。一連の治療とは、ランダム化後の治療レジメンの効果や1次治療への反応に依存する2次治療の比較もしくは調整に関心がある場合に該当する。治療のタイミングとは、2次もしくはランダム化後の治療のタイミングに関心がある場合に該当する。治療の用量や期間とは、ランダム化されない用量の戦略や治療期間の効果に関心がある場合に該当する。

臨床試験の検索の結果、始めに2773個の論文が同定され、1次スクリーニングにより1032個の論文が、2次スクリーニングにより、26個の論文が残されている。さらに、2報の重複を削除し、1報の新規論文を追加することによって、合計25個の論文が特定されている。因果的な臨床疑問についての内訳は、併用治療が11報、一連の治療が5報、治療のタイミングが2報、治療の用量や期間が7報であり、併用治療に関する論文がもっとも多かった。疾患領域については6つの領域で使用され、その内訳は、Cardiovascular disease が5報（その内、diabetes が3報）、HIV が7報、がん領域が6報、mental disorders が4報、paediatrics が2報、transfusion medicine が1報であった。

併用治療に関する論文11報のうち、Cardiovascular disease で5報、HIV で5報、mental disorders で1報が該当している。ある HIV を対象とする論文では、殺菌剤を開始した患者におけるホルモン避妊薬の併用の HIV 感染予防効果（こちらが本論文における因果的な臨床疑問の対象）の評価のために、因果推論の方法として、IPW を用いた周辺構造モデルが用いられ、ホルモン避妊薬に予防効果はなかったという結果が示されていた。別の HIV を対象とする論文では、コンドームの使用という中間変数を調整した殺菌剤の効果とコンドーム使用の効果を IPW にて評価し、また別の論文では、抗レトロウイルス療法を開始した患者における抗生剤の併用の死亡等への効果を IPW にて評価していた。このように、併用治療の因果効果は、ランダム化された治療以外にさまざまな治療が実施される疾患領域で想定される。

また、ある Cardiovascular disease を対象とする論文は、心房細動患者に対するアピキサバンとワルファリンの比較試験（ARISTOTLE 試験）のデータを解析した論文であり、本論文における因果的な臨床疑問は、アスピリンの併用効果を評価することとされていた。本試験では、アスピリンは医師の裁量でランダム化された治療に上乘せすることとされていた。因果推論の方法として、IPW を用いた周辺構造モデルが用いられたが、標準的な解析では、脳卒中や大出血に対するアスピリンのリスクを過小評価するような結果となっていた。別の Cardiovascular disease を対象とする論文では、さまざまな心血管系のアウトカムに対するアトルバスタチンの効果（こちらが本論文における因果的な臨床疑問の対象）を評価しており、本試験でも、アトルバスタチンは医師の裁量で使用することとされていた。因果推論の方法として、IPW を用いた周辺構造モデルや G 推定が用いられていた。本 Cardiovascular disease を対象とする論文では、複数の方法で類似した結果が得られたとのことで、標準的な解析や複数の因果推論の方法を用いて結果を比較しているような因果推論の適用論文も見受けられる。

一連の治療に関する論文として、がん領域で5報が該当している。がん領域の臨床試験ではしばしば、ランダム化された1次治療への反応に依存して、治療のスイッチングや2次治療の開始が生じる。一般的には、Grade の高い有害事象や疾患の増悪が起こった時点でランダム化されていない治療を開始することとなり、その時点での患者の状態を考慮して解析する必要がある。5報中2報は最適な治療の組み合わせを評価することを目的とし、周辺構造モデルを応用した方法や G formula が用いられていた。5報中3報は後治療を調整してランダム化した治療を評価することを目的とし、G 推定や周辺構造モデルが用いられていた。本適用例では例えば、すべての被験者が同じ後治療を受けたらという仮定を置いて解析を実施していた。

治療のタイミングに関する論文2報のうち、HIV で1報、mental disorders で1報が該当している。HIV を対象とする論文では、HIV 感染者を対象に、初回のウイルス学的失敗後の治療のスイッ

チのタイミングを評価していた。因果推論の方法として、動的な治療戦略を比較するためのIPW推定が用いられていた。解析の結果、早期にスイッチする方がよいとの結果が得られたが、ウイルス学的失敗を経験した被験者に限定した（調整しない）解析では、治療戦略間に差はなかったとのことで、標準的な未調整の解析と因果推論の方法で結果が異なっている事例である。また、mental disordersを対象とする論文では、統合失調症患者を対象に、標準的な抗精神病薬と新規の非定型抗精神病薬を比較した試験に適用しており、本試験では治療のスイッチングを許容していた。そこで、12カ月後の症状を最小化するためのスイッチングの閾値を評価することを目的として、周辺構造モデルを応用した方法が用いられていた。

治療の用量や期間に関する論文7報のうち、がん領域で1報、mental disordersで2報、transfusion medicineで1報、HIVで1報、paediatricsで2報が該当している。がん領域を対象とする論文では、長期の維持療法における、血球数に応じて用量を選択する戦略が検討されており、因果推論の方法として、周辺構造モデルを応用した方法が用いられていた。臨床試験の実施計画書で規定されていた用量選択の戦略に比して、因果推論の方法を用いた場合、より高い用量を最適の選択とする傾向にあった。mental disordersを対象とする論文では、医師の裁量による用量調節が許容されており、因果推論の方法として、IPWを用いた周辺構造モデルが用いられていた。transfusion medicineを対象とする論文では、輸血の回数について、IPWにて評価されていた。HIVを対象とする論文では、殺菌ジェルの累積投与量について評価するために因果推論の方法が用いられており、本論文の著者らは、殺菌ジェルの病変発生に対する因果効果を調べるために、性交回数を調整した構造加速失敗時間モデルを応用した方法を用いていた。paediatricsを対象とする論文では、12カ月後の乳児の成長に対する授乳の期間が評価されており、因果推論の方法として、周辺構造モデルやG推定が用いられていた。本論文では両方法間で解析結果が異なったとのことで、欠測データの取扱いの違いも原因の一つではないかと本paediatricsを対象とする論文で考察されていた。

以上のように、多くの疾患領域のランダム化臨床試験の解析に因果推論が適用されている。また、対象とする臨床疑問もさまざまである。しかし、適用事例の数自体は少ないと考えられる。併用治療の効果についての臨床疑問や周辺構造モデルの適用が多いが、理由として、特に慢性疾患では治療薬の数も多くまた割り付けられた治療のみで治療することはあまりなく合併症の治療を行うこともあること、取り組みやすい臨床疑問であること、周辺構造モデルは比較的直観的で実行しやすいことが挙げられる。周辺構造モデルを適用する際には、処方の際のばらつきが十分あれば、その臨床疑問について適用可能性が高いことに留意する。一連の治療や治療の用量や期間については、臨床試験の実施計画書の規定によりばらつきが小さく、また、実施計画書からの逸脱として多くのデータが取得されないため、適用事例は少なかった。治療のタイミングについても、実施計画書でランダム化治療の失敗の閾値が明確に定義されている場合には、結果として観察されるタイミングのばらつきが小さい。

25報の論文中、12報は医学雑誌に掲載され、そのうち10報は、周辺構造モデルを用いた併用治療の評価のために因果推論の方法が用いられていた。逆にそれら以外は統計的な方法論に関心があると受け取られる。適用事例の数自体が少ないのは、方法論が認知されていないか、臨床的な結論を導くには限界があると捉えられているか等の理由が考えられる。また、HIVやがん

領域では因果推論の方法の必要性が認識されているが、それ以外の領域では啓発が必要と考えられる。結論として、因果推論の方法を広めるためにより広範な啓発が必要で、ランダム化比較試験の解析に付加価値を与えるためにどのように適用するかを考えなければならないと結ばれている。

## 3.2 主要層別を適用した臨床試験事例

### 3.2.1 主要層ストラテジーについて

ICH E9(R1)<sup>1</sup>において、主要層ストラテジーについて触れられている。すべての治療において、中間事象が発現する可能性に応じて被験者集団を区分する。治療群が二つの場合には、ある中間事象の発現有無に対して、いずれの治療群でも中間事象が発現しない被験者集団、治療 A では発現するが治療 B では発現しない被験者集団、治療 B では発現するが治療 A では発現しない被験者集団、及びいずれの治療でも発現する被験者集団の4つの主要層が考えられる。

ここで、割付について、本節では A=1 を実薬群、A=0 をプラセボ群とする。投与群 A のときにアウトカム（例えば、有無の2値を想定）が発生するかしないかの潜在変数を  $Y^a$  とすると、 $Y^a = 1$  をアウトカム発生あり、 $Y^a = 0$  をアウトカム発生なしとする。さらに、投与群 A のときに中間事象が発生するかしないかの潜在変数を  $S^a$  とすると、 $S^a = 1$  を中間事象発生あり、 $S^a = 0$  を中間事象発生なしとする。観測される測定値については、一致性の仮定より、 $S = A \times S^{a=1} + (1 - A) \times S^{a=0}$ 、 $Y = A \times Y^{a=1} + (1 - A) \times Y^{a=0}$  となる。整理すると、以下の表 3.2-1 のようになる。ここでは、それぞれの層を Immune(I), Doomed(D), Benefiter(B), Harmed(H) とよぶ。

表 3.2-1 中間事象の発生に関する層

	$S^{a=1}$	$S^{a=0}$	備考
Immune(I)	0	0	いずれの投与群でも中間事象なし
Doomed(D)	1	1	いずれの投与群でも中間事象あり
Benefiter(B)	0	1	プラセボ群のときのみ中間事象あり
Harmed(H)	1	0	実薬群のときのみ中間事象あり

### 3.2.2 主要層解析について

主要層に関する事例として、本報告書では、ある臨床試験に関する論文<sup>26</sup>に記載されている内容を紹介する。本試験は、二次性進行型多発性硬化症に対する障害進行抑制効果の評価を目的とした、無作為化、二重盲検、プラセボ対照第 III 相臨床試験であり、実薬とプラセボは2:1に割付けられ、主要層を適用した評価項目は進行障害の発現の有無/発現割合である。ただし、試験の途中で（多発性硬化症の）症状の再発を起こす患者もいる状況である。ここで、再発なし集団における進行障害が認められるまでの期間に対しての薬効を評価したい。すなわち、再発が中間事象と考えられる。

ここでの推定の目的は、いずれの投与群でも再発が起こらない集団（Immune）における因果リスク比 ( $P[Y^{a=1} = 1|I]/P[Y^{a=0} = 1|I]$ ) である。まず、次の仮定をおく。

- 交換可能性  
 $A | S^{a=1}, S^{a=0}, Y^{a=1}, Y^{a=0}$
- 単調性  
 $S^{a=0} = 0 \Rightarrow S^{a=1} = 0$

交換可能性は、ランダム化しているため、自然な仮定と考えられる。単調性は、プラセボ群で再発しないのであれば、実薬群でも再発しないことを表しており、 $S^{a=1} = 1 \Rightarrow S^{a=0} = 1$ 、すなわち、実薬群で再発するのであれば、プラセボ群でも再発することと同値である。これにより、 $\text{Harmed}(S^{a=0} = 0, S^{a=1} = 1)$ 、すなわち、実薬群のときのみ再発するということはない、という仮定となる。この仮定はデータから成り立っていることを確認することはできないが、成立していないことが確認できる場合がある。

単調性から、 $P(S^{a=0} = 0) \leq P(S^{a=1} = 0)$ が成り立つ。また、交換可能性、一致性から $P(S^{a=0} = 0) \leq P(S^{a=1} = 0) \Leftrightarrow P(S^{a=0} = 0|A = 0) \leq P(S^{a=1} = 0|A = 1) \Leftrightarrow P(S = 0|A = 0) \leq P(S = 0|A = 1)$ を導くことができる。すなわち、観測データから単調性を検討することができ、 $P(S = 0|A = 0) \leq P(S = 0|A = 1)$ が成り立っていないければ、単調性は不成立であると判定できる

さて、ここからは上記二つの仮定の下で、 $P[Y^{a=1} = 1|I]/P[Y^{a=0} = 1|I]$ を求めていく。分母、分子それぞれについて考えていくが、まずは、分母から考える。

$$\begin{aligned} P[Y^{a=0} = 1|I] &= P[Y^{a=0} = 1|S^{a=0} = 0, S^{a=1} = 0] \\ &= P[Y^{a=0} = 1|S^{a=0} = 0] \quad \because \text{単調性} \\ &= P[Y^{a=0} = 1, S^{a=0} = 0]/P[S^{a=0} = 0] \\ &= P[Y^{a=0} = 1, S^{a=0} = 0|A = 0]/P[S^{a=0} = 0|A = 0] \quad \because \text{交換可能性} \\ &= P[Y = 1, S = 0|A = 0]/P[S = 0|A = 0] \quad \because \text{一致性} \\ &= P[Y = 1|S = 0, A = 0] \end{aligned}$$

以上より、分母はデータから推測可能であることがわかる。次に、分子について考える。

$$\begin{aligned} P[Y^{a=1} = 1|I] &= P[Y^{a=1} = 1|S^{a=0} = 0, S^{a=1} = 0] \\ &= P[Y^{a=1} = 1|S^{a=0} = 0] \quad \because \text{単調性} \\ &= P[Y^{a=1} = 1, S^{a=0} = 0]/P[S^{a=0} = 0] \\ &= P[Y^{a=1} = 1, S^{a=0} = 0|A = a]/P[S^{a=0} = 0|A = 0] \end{aligned}$$

ここから先が分母のように式変形を進めることができないので、分子について点推定が困難である。これは、前述の仮定の下では、プラセボ群での再発なしの人は **Immune** と特定できるが、実薬群の再発なしの人は、**Immune** と **Benefier** が混在しているため特定できないことを意味している。そこで、以降では、データから推定可能な $P[Y^{a=1} = 1|I \text{ or } B]$ の式を変形し、 $P[Y^{a=1} = 1|B]$ を感度パラメータのように利用して、 $P[Y^{a=1} = 1|I]$ について点推定ではなく範囲を推定することを考える。なお、ここで言う範囲とは信頼区間を推定することではないことに注意が必要である。

ここで、

$$\begin{aligned} P(D) &= P[S^{a=1} = 1, S^{a=0} = 1] \\ &= P[S^{a=1} = 1] \quad \because \text{単調性} \\ &= P[S^{a=1} = 1|A = 1] \quad \because \text{交換可能性} \\ &= P[S = 1|A = 1] \quad \because \text{一致性} \end{aligned}$$

すなわち、 $P(D)$ は、実薬群で再発した人の割合で推定できる。

$$\begin{aligned} P(I) &= P[S^{a=1} = 0, S^{a=0} = 1] \\ &= P[S^{a=0} = 0] \quad \because \text{単調性} \end{aligned}$$

$$= P[S^{a=0} = 0|A = 0] \because \text{交換可能性}$$

$$= P[S = 0|A = 0] \because \text{一致性}$$

すなわち、 $P(I)$ は、プラセボ群で再発していない人の割合で推定できる。

$$P(H) = P[S^{a=0} = 0, S^{a=1} = 1] = 0 \because \text{単調性}$$

$$P(B) = 1 - P(D) - P(I)$$

$$P[Y^{a=1} = 1|D] = P[Y^{a=1} = 1|S^{a=0} = 1, S^{a=1} = 1]$$

$$= P[Y^{a=1} = 1|S^{a=1} = 1] \because \text{単調性}$$

$$= P[Y^{a=1} = 1, S^{a=1} = 1]/P[S^{a=1} = 1]$$

$$= P[Y^{a=1} = 1, S^{a=1} = 1|A = 1]/P[S^{a=1} = 1|A = 1] \because \text{交換可能性}$$

$$= P[Y = 1, S = 1|A = 1]/P[S = 1|A = 1] \because \text{一致性}$$

$$= P[Y = 1|S = 1, A = 1]$$

すなわち、 $P[Y^{a=1} = 1|D]$ は、推定できる。

$$P[Y^{a=1} = 1|I \text{ or } B] = P[Y^{a=1} = 1|S^{a=1} = 0]$$

$$= P[Y^{a=1} = 1|S^{a=1} = 0, S^{a=0} = 0] \times P[S^{a=0} = 0|S^{a=1} = 0]$$

$$+ P[Y^{a=1} = 1|S^{a=1} = 0, S^{a=0} = 1] \times P[S^{a=0} = 1|S^{a=1} = 0]$$

$$= P[Y^{a=1} = 1|I] \times P[I|I \text{ or } B] + P[Y^{a=1} = 1|B] \times P[B|I \text{ or } B]$$

この式を整理すると、

$$P[Y^{a=1} = 1|I] = P[Y^{a=1} = 1|I \text{ or } B]/P[I|I \text{ or } B] - (P[B|I \text{ or } B]/P[I|I \text{ or } B]) \times P[Y^{a=1} = 1|B]$$

ここで、 $P(I), P(B), P(D), P[I|I \text{ or } B] = P(I)/\{P(I) + P(B)\}$ ,  $P[B|I \text{ or } B] = P(B)/\{P(I) + P(B)\}$ は推定可能であり、また、

$$P[Y^{a=1} = 1|I \text{ or } B] = P[Y^{a=1} = 1|S^{a=1} = 0]$$

$$= P[Y^{a=1} = 1, S^{a=1} = 0]/P[S^{a=1} = 0]$$

$$= P[Y^{a=1} = 1, S^{a=1} = 0|A = 1]/P[S^{a=1} = 0|A = 1] \because \text{交換可能性}$$

$$= P[Y = 1, S = 0|A = 1]/P[S = 0|A = 1] \because \text{一致性}$$

$$= P[Y = 1|S = 0, A = 1]$$

であり、 $P[Y^{a=1} = 1|I \text{ or } B]$ も推定可能である。

改めて、分子についてまとめると、

$$P[Y^{a=1} = 1|I] = P[Y^{a=1} = 1|I \text{ or } B]/P[I|I \text{ or } B]$$

$$- (P[B|I \text{ or } B]/P[I|I \text{ or } B]) \times P[Y^{a=1} = 1|B]$$

であるが、第2項の $P[Y^{a=1} = 1|B]$ のみ推定することができない。しかし、確率であるため0から1の値であるので、 $P[Y^{a=1} = 1|B]$ について0または1を代入することにより、範囲としては推定することができる。

$$\frac{P[Y^{a=1} = 1|I \text{ or } B]}{P[I|I \text{ or } B]} - \frac{P[B|I \text{ or } B]}{P[I|I \text{ or } B]} \leq P[Y^{a=1} = 1|I]$$

$$\leq P[Y^{a=1} = 1|I \text{ or } B]/P[I|I \text{ or } B]$$

範囲で推定する場合、その範囲は狭い方が望ましい。追加の仮定をおくことにより、上記の範囲よりも狭いものが構成できる可能性はあるが、本報告書では、その方法の一つとして、新たなアプローチを以降の節で紹介する。

### 3.2.3 主要層解析に対するベイズ流アプローチ

#### 3.2.3.1 主要層解析（ベイズ流の推測）

$St=\{I, D, B, H\}$ を層全体の集合,  $G$ を $St$ の値を取る確率変数とし, 対象集団の被験者が層 $g$ に属する確率 $\pi_g$ を,

$$\pi_g = P(G = g) \quad (g \in \{St\})$$

とする. ここで, 前節の仮定とは異なり $g$ にはH (Harmed) も含め, 「強く情報を持つ事前分布」を用いることで単調性の仮定を実現させることを考える. 層 $g$ の集団が, 処置 $a$ を受けたときの, 主要評価項目 (主要なイベント) の発現確率 $\theta_g^a$  (ロジットスケール) を,

$$\theta_g^a = \text{logit}\{P[Y^a = 1] | G = g\} \quad (g \in \{I, D, B, H\})$$

とする. 興味ある estimand ( $P[Y^{a=1} = 1 | I] / P[Y^{a=0} = 1 | I]$ ) は, expit 関数<sup>1</sup>を用いて,

$$\frac{P[Y^{a=1} = 1 | I]}{P[Y^{a=0} = 1 | I]} = \frac{\text{expit}(\theta_I^{a=1})}{\text{expit}(\theta_I^{a=0})}$$

と示される.

$\pi_g, \theta_g^a$ のパラメータ全体のモデルパラメータベクトルを $\omega$ とする. ベイズ流の推測にあたって,  $p(\omega)$ を $\omega$ の事前分布,  $p(Y, S | A, \omega)$ を投与群 $A$ におけるアウトカムと中間事象 ( $Y, S$ )の同時確率関数とする. ベイズの定理より,

$$\begin{aligned} p(\omega | Y, S, A) &\propto p(Y, S | A, \omega) \cdot p(\omega) \\ &= p(Y | S, A, \omega) \cdot p(S | A, \omega) \cdot p(\omega) \end{aligned}$$

となる.  $p(Y | S, A, \omega)$ は,  $S$  (再発) で条件づけたアウトカムモデル (以下, disability model),  $p(S | A, \omega)$ は, (周辺) 再発モデル (以下, relapse model) である.

まず, relapse model について説明する. 観測される $(S, A) = (s, a)$ の値と, 層の関係を表 3.2-2 に示す.  $(S, A)$  のそれぞれについて, 観測データから, 属する層を2つに絞ることができる<sup>2</sup>.

表 3.2-2  $(S, A) = (s, a)$ の値と層の関係

$(S, A)$	情報	層
(0,0)	$S^{a=0} = 0$	$I$ or $H$
(0,1)	$S^{a=1} = 0$	$I$ or $B$
(1,0)	$S^{a=0} = 1$	$D$ or $B$
(1,1)	$S^{a=1} = 1$	$D$ or $H$

relapse model について,

$$p(S | A, \omega) = (1 - A) \cdot \text{Bernoulli}(S | \pi_D + \pi_B) + A \cdot \text{Bernoulli}(S | \pi_D + \pi_H)$$

と示すことができる.  $\text{Bernoulli}(S | \pi_D + \pi_B)$ は,  $A = 0$ の時の再発確率 ( $S^{a=0} = 1$ , すなわち  $D$

<sup>1</sup> expit 関数は, logit の逆関数, すなわち $y = \text{logit}(x) \Leftrightarrow \text{expit}(y) = x$ .  $\text{expit}(x) = \exp(x) / (1 + \exp(x))$ .

<sup>2</sup> 層 H を除外していないことに注意

or B), Bernoulli( $S|\pi_D + \pi_H$ )は,  $A = 1$ の時の再発確率 ( $S^{a=1} = 1$ , すなわち D or H) に対応する.

事前分布に関して, 各層に所属する確率をカテゴリカル分布 ( $n=1$ の多項分布) とし, 事前分布を設定するパラメータを $\alpha = (\alpha_I, \alpha_D, \alpha_B, \alpha_H)$ とする. 各層に属する確率を,

$$\pi_g = P(G = g) = \sigma_g(\alpha) \quad g \in \{I, D, B, H\}$$

とする. ここで,  $\sigma_g(\alpha)$ に softmax 関数を用いると,

$$\sigma_I(\alpha) = \frac{\exp(\alpha_I)}{\exp(\alpha_I) + \exp(\alpha_D) + \exp(\alpha_B) + \exp(\alpha_H)}$$

$$\sigma_D(\alpha) = \frac{\exp(\alpha_D)}{\exp(\alpha_I) + \exp(\alpha_D) + \exp(\alpha_B) + \exp(\alpha_H)}$$

$$\sigma_B(\alpha) = \frac{\exp(\alpha_B)}{\exp(\alpha_I) + \exp(\alpha_D) + \exp(\alpha_B) + \exp(\alpha_H)}$$

$$\sigma_H(\alpha) = \frac{\exp(\alpha_H)}{\exp(\alpha_I) + \exp(\alpha_D) + \exp(\alpha_B) + \exp(\alpha_H)}$$

となる. すべての右辺の分子分母を,  $\exp(\alpha_B)$ で割ると,

$$\sigma_I(\alpha) = \frac{\exp(\alpha_I - \alpha_B)}{\exp(\alpha_I - \alpha_B) + \exp(\alpha_D - \alpha_B) + 1 + \exp(\alpha_H - \alpha_B)}$$

$$\sigma_D(\alpha) = \frac{\exp(\alpha_D - \alpha_B)}{\exp(\alpha_I - \alpha_B) + \exp(\alpha_D - \alpha_B) + 1 + \exp(\alpha_H - \alpha_B)}$$

$$\sigma_B(\alpha) = \frac{1}{\exp(\alpha_I - \alpha_B) + \exp(\alpha_D - \alpha_B) + 1 + \exp(\alpha_H - \alpha_B)}$$

$$\sigma_H(\alpha) = \frac{\exp(\alpha_H - \alpha_B)}{\exp(\alpha_I - \alpha_B) + \exp(\alpha_D - \alpha_B) + 1 + \exp(\alpha_H - \alpha_B)}$$

となり,  $\alpha'_I := \alpha_I - \alpha_B, \alpha'_D := \alpha_D - \alpha_B, \alpha'_H := \alpha_H - \alpha_B$ と取り直すと,  $\alpha' := (\alpha'_I, \alpha'_D, 0, \alpha'_H)$ とできる<sup>3</sup>. 本論文では,  $\alpha'$ を $\alpha$ と記載していることから, 本報告書でもこれ以降同様に,  $\alpha'_I, \alpha'_D, \alpha'_H$ も $\alpha_I, \alpha_D, \alpha_H$ と記載する. すなわち,  $\alpha = (\alpha_I, \alpha_D, 0, \alpha_H)$ とすると,

$$\sigma_I(\alpha) = \frac{\exp(\alpha_I)}{\exp(\alpha_I) + \exp(\alpha_D) + 1 + \exp(\alpha_H)}$$

$$\sigma_D(\alpha) = \frac{\exp(\alpha_D)}{\exp(\alpha_I) + \exp(\alpha_D) + 1 + \exp(\alpha_H)}$$

$$\sigma_B(\alpha) = \frac{1}{\exp(\alpha_I) + \exp(\alpha_D) + 1 + \exp(\alpha_H)}$$

$$\sigma_H(\alpha) = \frac{\exp(\alpha_H)}{\exp(\alpha_I) + \exp(\alpha_D) + 1 + \exp(\alpha_H)}$$

となり,  $\alpha$ に事前分布を設定する. 本変換により, 共変量を組み込むことが容易になり, ま

<sup>3</sup> Softmax 関数の次元の制約から, 0は消せない.

た、効率がよいとされている<sup>4</sup>。

次に、disability model ( $S$  (再発) で条件づけたアウトカムモデル) について説明する。

例えば、 $(S, A) = (1, 1)$  の場合、disability model:  $p(Y | S, A, \omega)$  は、

$$\begin{aligned} p(Y = 1 | S = 1, A = 1, \omega) &= p(Y = 1 | S = 1, A = 1, G = D, \omega) \times P(G = D | S = 1, A = 1, \omega) \\ &\quad + p(Y = 1 | S = 1, A = 1, G = H, \omega) \times P(G = H | S = 1, A = 1, \omega) \\ &= p(Y = 1 | A = 1, G = D, \omega) \times P(G = D | G \in \{D, H\}, \omega) \\ &\quad + p(Y = 1 | A = 1, G = H, \omega) \times P(G = H | G \in \{D, H\}, \omega) \end{aligned}$$

一貫性の仮定より、

$$\begin{aligned} &= p(Y^{a=1} = 1 | A = 1, G = D, \omega) \cdot \frac{\pi_D}{\pi_D + \pi_H} + p(Y^{a=1} = 1 | A = 1, G = H, \omega) \cdot \frac{\pi_H}{\pi_D + \pi_H} \\ &= p(Y^{a=1} = 1 | G = D, \omega) \cdot \frac{\pi_D}{\pi_D + \pi_H} + p(Y^{a=1} = 1 | G = H, \omega) \cdot \frac{\pi_H}{\pi_D + \pi_H} \quad (*^{56}) \end{aligned}$$

これより、以下が成り立つ。

$$p(Y = 1 | S = 1, A = 1, \omega) = \text{expit}(\theta_D^{a=1}) \cdot \frac{\pi_D}{\pi_D + \pi_H} + \text{expit}(\theta_H^{a=1}) \cdot \frac{\pi_H}{\pi_D + \pi_H}$$

$(S, A)$  ごとの disability model:  $p(Y | S, A, \omega)$  については表 3.2-3 のように示される。

表 3.2-3  $(S, A)$  ごとの disability model:  $p(Y | S, A, \omega)$

$(S, A)$	層	Disability model
$(0, 0)$	$I$ or $H$	$\frac{\pi_I}{\pi_I + \pi_H} \text{Bernoulli}(\text{expit}(\theta_I^{a=0})) + \frac{\pi_H}{\pi_I + \pi_H} \text{Bernoulli}(\text{expit}(\theta_H^{a=0}))$
$(0, 1)$	$I$ or $B$	$\frac{\pi_I}{\pi_I + \pi_B} \text{Bernoulli}(\text{expit}(\theta_I^{a=1})) + \frac{\pi_B}{\pi_I + \pi_B} \text{Bernoulli}(\text{expit}(\theta_B^{a=1}))$
$(1, 0)$	$D$ or $B$	$\frac{\pi_D}{\pi_D + \pi_B} \text{Bernoulli}(\text{expit}(\theta_D^{a=0})) + \frac{\pi_B}{\pi_D + \pi_B} \text{Bernoulli}(\text{expit}(\theta_B^{a=0}))$

<sup>4</sup> Carpenter et al.<sup>23</sup>

<sup>5</sup> (\*) の式変形の補足。Joint exchangeability,  $A \perp\!\!\!\perp \{S^{a=1}, S^{a=0}, Y^{a=1}, Y^{a=0}\}$  より  $A \perp\!\!\!\perp Y^{a=1} | S^{a=0}, S^{a=1}$ .  $G = D: S^{a=0} = S^{a=1} = 1$  などから、 $G$  は  $S^{a=0}, S^{a=1}$  で決まるから、 $A \perp\!\!\!\perp Y^{a=1} | G$ . よって、 $p(Y^{a=1} = 1 | A = 1, G = D, \omega) = p(Y^{a=1} = 1 | G = D, \omega)$

<sup>6</sup>  $(X, Y) \perp\!\!\!\perp A$  が成り立つとき、 $f(X, Y, A) = f(X, Y) \cdot g(A)$  が成り立つ。このとき、 $f(X, A | Y) = \frac{f(X, Y, A)}{f(Y)} = \frac{f(X, Y) \cdot g(A)}{f(Y)} = f(X | Y) \cdot g(A) = f(X | Y) \cdot g(A | Y)$  ( $\because Y \perp\!\!\!\perp A$ ) より、 $X \perp\!\!\!\perp A | Y$  が成り立つ。すなわち、 $(X, Y) \perp\!\!\!\perp A \Rightarrow X \perp\!\!\!\perp A | Y$ .

(1,1)	<i>D or H</i>	$\frac{\pi_D}{\pi_D + \pi_H} \text{Bernoulli}(\text{expit}(\theta_D^{a=1})) + \frac{\pi_H}{\pi_D + \pi_H} \text{Bernoulli}(\text{expit}(\theta_H^{a=1}))$
-------	---------------	---

次に、治療効果と estimand の関係について整理する。  $\Delta_g$  を治療効果とすると、

$$\theta_g^{a=1} = \theta_g^{a=0} + \Delta_g \quad (g \in \{I, D, B, H\})$$

となる。  $\theta_g^a = \text{logit}\{P[Y^a = 1] | G = g\}$  より、

$$\text{logit}\{P[Y^{a=1} = 1 | G = g]\} = \text{logit}\{P[Y^{a=0} = 1 | G = g]\} + \Delta_g$$

$$P[Y^{a=0} = 1 | G = g] = \frac{\exp(\theta_g^{a=0})}{1 + \exp(\theta_g^{a=0})}$$

$$P[Y^{a=1} = 1 | G = g] = \frac{\exp(\theta_g^{a=1})}{1 + \exp(\theta_g^{a=1})} = \frac{\exp(\theta_g^{a=0} + \Delta_g)}{1 + \exp(\theta_g^{a=0} + \Delta_g)}$$

Estimand について、治療効果を用いて示すと、

$$\begin{aligned} \frac{P[Y^{a=1} = 1 | I]}{P[Y^{a=0} = 1 | I]} &= \frac{\exp(\theta_I^{a=0} + \Delta_I)}{1 + \exp(\theta_I^{a=0} + \Delta_I)} \cdot \frac{1 + \exp(\theta_I^{a=0})}{\exp(\theta_I^{a=0})} \\ &= \frac{\exp(\Delta_I) \cdot (1 + \exp(\theta_I^{a=0}))}{1 + \exp(\theta_I^{a=0} + \Delta_I)} \end{aligned}$$

となる。

ここまで出てきたパラメータを整理すると、relapse model のパラメータ  $\omega_{rel}$  は、  $\omega_{rel} = (\alpha_I, \alpha_D, \alpha_H)'$  , disability model は、  $\omega_{dis} = (\theta_I^{a=0}, \Delta_I, \theta_I^{a=0}, \Delta_I, \theta_D^{a=0}, \Delta_D, \theta_B^{a=0}, \Delta_B, \theta_H^{a=0}, \Delta_H)'$  , 全体のパラメータは、  $\omega = (\omega'_{rel}, \omega'_{dis})'$  となる。  $\omega_{dis}$  について、  $\theta_I^{a=1}, \theta_D^{a=1}, \theta_B^{a=1}, \theta_H^{a=1}$  は他の2つから決まるため省略することができる。

### 3.2.3.2 共変量と欠測の導入

$X$  をベースライン共変量ベクトル、  $M$  を  $(Y, S)$  の欠測指示変数 ( $M = 1$  : 欠測,  $M = 0$  : 観測) とし、  $M \perp\!\!\!\perp (Y, G) | X, A$  と仮定する。ベースライン共変量  $X$  をモデルに組み込むと、

$$\begin{aligned} \pi_{g,x} &= P[G = g | X = x] \\ \text{expit}(\theta_{g,x}^a) &= P[Y^a = 1 | G = g, X = x] \end{aligned}$$

となり、relapse model は、

$$p(S | A, X, M = 0, \omega) = (1 - A) \cdot \text{Bernoulli}(S | \pi_{D,X} + \pi_{B,X}) + A \cdot \text{Bernoulli}(S | \pi_{D,X} + \pi_{H,X})$$

となる。disability model:  $p(Y | S, A, X, M = 0, \omega)$  は、表3.2-4のように示される。

表 3.2-4 Disability model:  $p(Y | S, A, X, M = 0, \omega)$

$(S, Z)$	層	Disability model
(0,0)	<i>I or H</i>	$\frac{\pi_{I,X}}{\pi_{I,X} + \pi_{H,X}} \text{Bernoulli}(\text{expit}(\theta_{I,X}^{a=0})) + \frac{\pi_{H,X}}{\pi_{I,X} + \pi_{H,X}} \text{Bernoulli}(\text{expit}(\theta_{H,X}^{a=0}))$

(0,1)	$I$ or $B$	$\frac{\pi_{I,X}}{\pi_{I,X} + \pi_{B,X}} \text{Bernoulli}(\text{expit}(\theta_{I,X}^{a=1})) + \frac{\pi_{B,X}}{\pi_{I,X} + \pi_{B,X}} \text{Bernoulli}(\text{expit}(\theta_{B,X}^{a=1}))$
(1,0)	$D$ or $B$	$\frac{\pi_{D,X}}{\pi_{D,X} + \pi_{B,X}} \text{Bernoulli}(\text{expit}(\theta_{D,X}^{a=0})) + \frac{\pi_{B,X}}{\pi_{D,X} + \pi_{B,X}} \text{Bernoulli}(\text{expit}(\theta_{B,X}^{a=0}))$
(1,1)	$D$ or $H$	$\frac{\pi_{D,X}}{\pi_{D,X} + \pi_{H,X}} \text{Bernoulli}(\text{expit}(\theta_{D,X}^{a=1})) + \frac{\pi_{H,X}}{\pi_{D,X} + \pi_{H,X}} \text{Bernoulli}(\text{expit}(\theta_{H,X}^{a=1}))$

例えば、 $S = 1, A = 1$  の場合は、

$$p(Y = 1 | S = 1, A = 1, X, M = 0, \omega) = \text{expit}(\theta_{D,X}^{a=1}) \cdot \frac{\pi_{D,X}}{\pi_{D,X} + \pi_{H,X}} + \text{expit}(\theta_{H,X}^{a=1}) \cdot \frac{\pi_{H,X}}{\pi_{D,X} + \pi_{H,X}}$$

となる。

治療効果については、 $\theta_{g,x}^{a=1} = \theta_{g,x}^{a=0} + \Delta_{g,x}$  として、 $x$ ごとに異なる治療効果を考える。集団の群間差は、「 $x$ ごとの群間差」を、「 $x$ の分布」で重み付けて平均したものと考えると、 $F(x)$ :  $X$  の累積分布関数とし、 $\pi_g = \int \pi_{g,x} dF(x)$  より、

$$\text{expit}[\theta_g^a] = \frac{1}{\pi_g} \int \text{expit}[\theta_{g,x}^a] \pi_{g,x} dF(x)$$

となる。本論文の事例では、共変量は2種類の2カテゴリ、すなわち、 $2 \times 2$ の4つの部分集団ごとに別々のパラメータを与える。一般に、連続量やカテゴリ数が多くなった場合などでは、回帰モデル等<sup>7</sup>が必要となる。

事例の臨床試験の解析では、 $t^* = 12, 18, 24$ カ月の3時点の解析を実施している。使用した共変量は、ベースラインの EDSS スコア（「6以上」と「6未満」で2値化）、試験開始前2年以内の再発（有無の2値）であり、共変量のカテゴリの組み合わせで、計4つの部分集団ができる。

### 3.2.3.3 データ解析における事前分布の設定

$\alpha_{g,x} \sim N(0, 2^2)$  for  $g \in \{I, D\}$  については、以下を設定根拠とする。まず、共変量がないとき、 $\alpha_I, \alpha_D$  の事前分布に  $N(0, 1)$  を仮定した場合、 $\pi_I, \pi_D$  の事前分布の中央値が約0.31で95%信頼区間が(0.04, 0.80)となる。特に極端な値もとらなければ、ある層に極端に偏ったりもしていない。次に、共変量があるとき、共変量のグループに関して大体一様分布すると仮定すると、“effective variance”が、共変量の組み合わせの数に比例して減少する。これを考慮して、弱い情報をもつ正規事前分布として、 $\alpha_{g,x} \sim N(0, 2^2)$  としている。

$\alpha_{H,x} \sim N(-50, 0.1^2)$  は、強い情報をもつ事前分布（他の $\alpha_g$ とくらべて、極端な位置にあり、SDが小さい）を考えることで、単調性の仮定に比較的近い状況に持ち込むことができる。すなわち、ほぼ $P(\pi_H = 0) = 1$ となり、事後確率もほぼ1となる。事前分布の情報量を弱める（位置をずらす、SDを増やす等）ことで、単調性の仮定に対する感度分析ができる。

$\theta_{g,x}^{a=0} \sim N(\text{logit}(0.3), 2^2)$   $g \in \{I, B, D, H\}$  については、事例の臨床試験（Expand trial）のプロトコル中の untreated patients の2年間の disability rate を考慮して、平均を選択している。分散につ

<sup>7</sup> 例)  $\alpha_{g,x} = x' \beta_g$  ( $g \in \{I, D, B, H\}$ )

いては $\alpha_{g,x}$ と同様の議論で、共変量なしのときの disability rate の中央値である0.3で95%信頼区間が(0.06, 0.75)より適切であり、共変量からできる部分集団が4つなので分散を4倍にしている。

$\Delta_{g,x} \sim N(0, 2^2)$ については、事前分布として平均的には効果なしとしている。分散は $\alpha_{g,x}$ と同様の議論に基づき設定している。

$\pi_I, \pi_D$ の分布が極端に偏るため、 $\alpha_I, \alpha_D \sim N(0, 10^6)$ のような、一様な事前分布は用いなかった。同様に、 $\theta_{g,x}^{a=0}$ に関しても、 $N(0, 50^2)$ のような moderately diffuse prior でも、(0.038, 0.75)に入る事前確率は0.03である。

感度分析として、 $\alpha_{H,x}$ の分布を変更することで、単調性からの乖離を検討している。弱い単調性があるシナリオとして、 $\alpha_{H,x} \sim N(-2, 0.5^2)$ としている。共変量がないとき、 $\pi_H$ の中央値が0.04で95%信頼区間は(0.01, 0.13)となる。Harmedに属することも、低い確率でありうるという設定となる。単調性がないシナリオとして、 $\alpha_{I,x}, \alpha_{D,x}$ と同じ $\alpha_{H,x} \sim N(0, 2^2)$ としている。他の層と同様、Harmedに属することがありうるという設定となる。

Benefiterについては、 $\alpha_B = 0$ としたため、 $\pi_B$ の事前分布は $\pi_I, \pi_D$ とは異なることに注意が必要である。事前分布の中央値が0.29で、95%信頼区間が狭くなっている。 $(\alpha_I, \alpha_D)$ に別の分布を考えることも可能で、例えば相関係数0.5の2変量正規分布などは、ほぼ同じ事前分布となる。また、 $(\pi_I, \pi_D, \pi_B) \sim \text{Dirichlet}(1, 1, 1)$ といった設定変更も可能で、両方を試した結果、3.2.3.4の結果と大きく変わらなかった。

解析には解析ソフトの Stan を使用している。No-U-Turn sampler (NUTS) を用いた MCMC (Carpenter et al.<sup>27</sup>参照) により事後分布を推定した。

### 3.2.3.4 臨床試験事例の結果と議論

表 3.2-5に、 $t^*=12$ カ月の要約統計量を、共変量のグループ×投与群ごとに示す。

<sup>8</sup> 論文中に、(0.03, 0.75)が「上で示した95%信頼区間」とあるが、上で示した95%信頼区間は(0.06, 0.75)。確認はとれていないものの、0.03が0.06の誤植の可能性あり。

表 3.2-5  $t^*=12$ カ月の要約統計量

Month	Covariate pattern	Treatment	Relapse	# patients	# CDPs
12	1	Placebo	No	68	15
12	1	Placebo	Yes	13	7
12	1	Siponimod	No	145	25
12	1	Siponimod	Yes	22	5
12	2	Placebo	No	109	26
12	2	Placebo	Yes	17	9
12	2	Siponimod	No	221	44
12	2	Siponimod	Yes	15	7
12	3	Placebo	No	63	14
12	3	Placebo	Yes	11	4
12	3	Siponimod	No	125	13
12	3	Siponimod	Yes	20	7
12	4	Placebo	No	130	19
12	4	Placebo	Yes	7	1
12	4	Siponimod	No	304	56
12	4	Siponimod	Yes	13	5

(Magnusson, et al. Statist Med. 2019;38:4761-4771, SIM\_8333-Supp-0001-supplemental tables.pdf, Table 2)

図 3.2-1に、時点ごとの各層へ所属する事後確率を示す。Immune に属する確率は、他と比べてかなり大きく、単調性の仮定するとき、中央値は0.8以上である。単調性の仮定でなくても、Immune に属する確率が最も高く、単調性の仮定でないときは、Harmed と Doomed が同じ程度となっている。

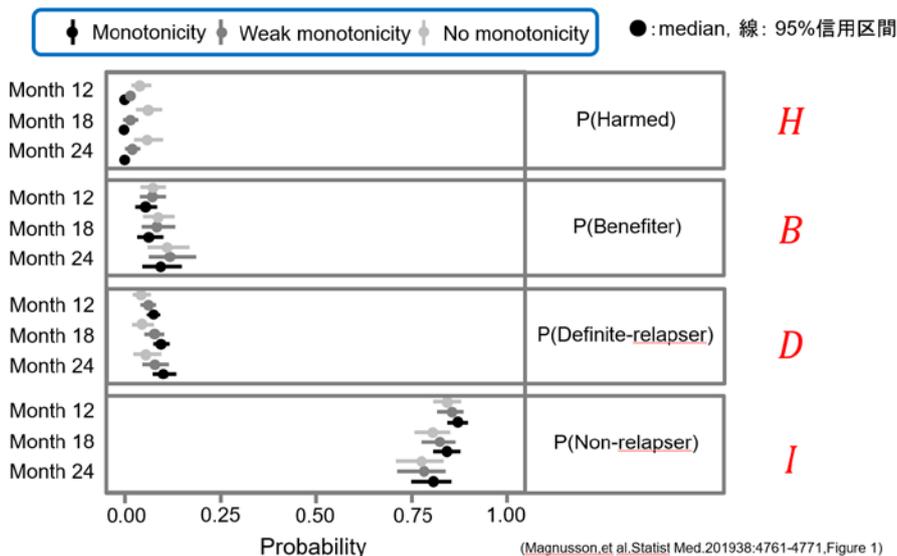


図 3.2-1 時点ごとの各層へ所属する事後確率

図 3.2-2に Immune の層の事後リスク比を示す。

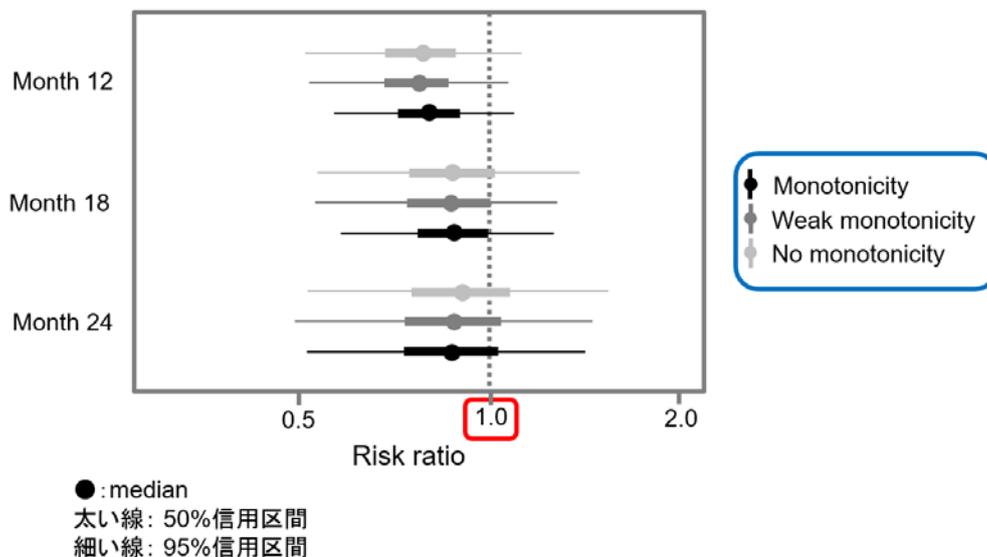


図 3.2-2 Immune の層の事後リスク比

因果リスク比  $< 1$  は治療効果ありと解釈される。単調性を仮定すると、因果リスク比の事後分布の中央値は  $0.80 \sim 0.85$  であり、また、単調性の仮定、時点によらず、因果リスク比  $< 1$  となる確率が少なくとも約  $70\%$  以上である。95%信用区間は、 $t^*$ が増えるほど広がっており、被験者の減少の影響と解釈できる。なお、追跡期間の中央値は、約18カ月である。

単調性の仮定を弱めても、結果はあまり変わらなかった。点推定値は少し動いただけで、信用区間は、ほとんど重なっている。単調性の仮定を弱めても、信用区間が広がるが、仮定は、結論に対して質的に強い影響を与えることはなかった。

頻度論では、例えば3.2.2節の事例で述べたように、実薬群における中間事象が発生しなかった患者が、Immune層かBenefiter層に属するか識別不可能であった。ベイズ流の解析では、情報のある事前分布を用いて単調性の仮定を表現することで、合理的な事前分布を用いる限り事後分布を導出できる。本論文で提案されている解析手法の利点としては、各層に属する確率を表すパラメータが取り得る、尤もらしい範囲を事前分布で表現できることが挙げられる。また、事前分布の設定によって結果がどの程度異なるかが重要であるが、事前分布のパラメータの設定を変えることで、感度分析を実施できるということも利点であると考えられる。

本論文では欠測にMARを仮定しており、Cree et al.<sup>28</sup>はMCARを仮定している。ここで述べているestimandは、ICH E9(R1)<sup>1</sup>における“主要層ストラテジー”に関連する方法であるが、事例の臨床試験(Expand trial)の主要estimandは、治療方針ストラテジーに基づいている。主要層は、Permutt<sup>29</sup>では、「推定しない場合、主要層での効果を正確に定義するのは容易」、Akacha et al.<sup>30</sup>では、「臨床統計家のより正式なトレーニングと実践的な経験が必要」とされている。本論文は、主要層ストラテジーを用いた1つの事例であり、本報告書では、因果推論を用いた事例として紹介した。

## 4. 因果推論と estimand の関連

本章は、臨床試験における estimand や COVID-19 の estimand に対する影響と因果推論の関連について説明し、estimand の構成要素の中でも特に盛んに議論されている中間事象に対する取扱いと因果推論との関連を理解することを目的とする。4.1 では、Lipkovich らの論文<sup>31</sup>を参照し、潜在アウトカムを用いた枠組みにおける estimand について説明する。4.2 では、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からの声明<sup>32</sup>や COVID-19 パンデミック下での臨床試験実施に対する統計的課題と推奨をまとめた論文<sup>33</sup>を参照し、COVID-19 の estimand (特に中間事象) に対する影響を説明した上で、それに対して因果推論の枠組みを用いて説明を加える。尚、本章は主に論文等の紹介であるため、ここまでの章と一部重複している説明がある点、記号が異なる点に注意していただきたい。

### 4.1 臨床試験における因果推論と estimand

#### 4.1.1 潜在アウトカム反応の定義

estimand の定義に対して因果効果の定義を導入することを考える (因果 estimand)。標準的な統計モデルでは、因果 estimand は、他の変数 (例えば、潜在的な交絡変数) で条件付けた構造パラメータ (例えば、治療の変数に関する係数) として理解されている。しかし、観察研究においては、治療の指示変数 ( $Z = 0, 1$ ) 及び共変量  $X$  で条件付けて連続量の結果変数をモデル化する (すなわち、 $E(Y|Z, X)$ ) という単純な状況でも、すべての交絡変数が測定されモデルで正しく特定されるという検証できない条件下でのみ治療の変数に関する係数は因果的に解釈することができる。

Lipkovich らの論文<sup>31</sup>では、因果 estimand を定義するために、潜在結果変数を用いている。潜在結果変数の枠組みは、実験的及び観察的な状況設定の両方を記述するための統一的な言語を提供してくれるので実用的である。重要なことは、潜在結果変数により、治療の割付メカニズムとは独立に、また特定のモデルの仮定に依存することなく、因果効果を定義することができることである。ここで、 $Y_i^z$  を被験者  $i$  がもし治療  $Z = z$  を受けた場合に観察されるであろう潜在アウトカムと表記すると、例えば、 $Y_i^z$  の差 ( $Y_i^z - Y_i^{z'}$ ) は、変数  $Z$  に関して、単純な観察値ではなく、その値を介入や操作によって “set” したときのそれぞれの実現値の差と考えることができる。このとき、潜在アウトカムは、実際にどの治療に割り付けられたかや実際にどの治療を受けたかに関わらず、すべての治療に対して定義することができる。例えば、並行群間試験において、ある被験者が治療  $Z = 0$  にランダム化されたとき、潜在アウトカム  $Y_i^0$  は観察可能である。一方、潜在アウトカム  $Y_i^1$  は観察不可能だが、原理的には定義することができる。因果推論の文献では、そのような観察不可能な結果は、しばしば反事実と呼ばれる。試験結果の解釈のことを考えると、潜在アウトカムは、試験や予期される中間事象を踏まえて、臨床的な観点から明確に定義される必要がある。

潜在アウトカムは、観測された結果  $Y_i$  と以下の式のように関連付けられる。

$$Y_i = Y_i^1 \cdot Z_i + Y_i^0 \cdot (1 - Z_i)$$

この式は、ある治療  $Z$  の下での潜在アウトカムは、当該治療を受けた被験者で観察される結果

と一致することを意味する。重要なことは、この関係はいつでも正しいものではなく、「一致性の仮定」と呼ばれる仮定であるということである。より一般的には、“stable unit treatment value assumption (SUTVA)”という、治療群が受ける治療は、非治療群に影響を与えないという仮定として知られている。相互干渉がない (no interference) という仮定としても知られている。

因果 estimand は、潜在アウトカムの期待値で定義することが可能である。平均因果効果 (Average Treatment Effect ; ATE) は、潜在アウトカムを用いて以下の式のように定義することができる。

$$\delta = E(Y^1 - Y^0)$$

ここで、治療の割付がランダムかどうかに関わらず、 $\delta$ は因果的な解釈をもっていることに注意する。因果 estimand  $\delta$ は、ランダム化試験において以下のように識別することができる。

$$\begin{aligned}\delta &= E(Y^1 - Y^0) = E(Y^1) - E(Y^0) \\ &= E(Y^1|Z = 1) - E(Y^0|Z = 0) \\ &= E(Y|Z = 1) - E(Y|Z = 0)\end{aligned}$$

2行目は、潜在アウトカムとランダム化される治療の割付は独立 ( $Z \perp \{Y^0, Y^1\}$ ) という仮定 (本報告書で説明した交換可能性の仮定と同義) から導かれ、3行目は、一致性の仮定から導かれる。このようにランダム化された治療群間で被験者は交換可能とみなすことができる。

ランダム化がされていなくても、未測定の変数がないと仮定できれば、 $\delta$ を推定できる可能性がある。すなわち、測定された変数  $X$  で条件付ければ、潜在アウトカムと治療の割付  $Z$  は独立 ( $Z \perp \{Y^0, Y^1\}|X$ ) と仮定することができるので、平均因果効果を推定することが可能となる。

$$\begin{aligned}E(Y^1) &= E_x(E(Y^1|X)) \\ &= E_x(E(Y^1|X, Z = 1)) \\ &= E_x(E(Y|X, Z = 1))\end{aligned}$$

ここで、外側の期待値  $E_x$  は、全体集団の  $X$  の分布についてのものである。2行目は、条件付き独立の仮定から導かれる。 $E(Y^0)$  についても同様に導出できる。

以上から考察すると、未測定の変数がないという仮定と欠測データに対する missing at random (MAR) の仮定は類似していると考えられる。欠測データが中間事象による交絡に対する取扱いに伴って生じることを考えると、MAR の下で、欠測データによる交絡は、交絡イベントの発生を完全に説明するすべての観測された因子で条件付けることにより除去することが可能となる。欠測が未測定の結果に依存しないということは、中間事象後の未来の潜在アウトカムが中間事象とは独立ということと同じであり、ランダム化していない研究で治療の割付に関する交絡を調整することと MAR の下で欠測に関する交絡を調整することは類似していると考えられる。

解析によって交絡を調整するアプローチは多く存在し、回帰分析、層別、マッチング、逆確率重み付け (2.2.4及び2.3参照) やこれらの組み合わせなどが考えられる。

#### 4.1.2 中間事象

中間事象は Ratitch et al. (2020) に従って二つに分類される。1) アウトカムの観測を阻害する、または estimand にとってアウトカムの意味を歪めるような事象及び2) 治療レジメンからの逸脱

である。以下のリストは中間事象の分類1または分類2を含む。

- 死亡
- 割り付けられた治療への遵守の欠如
- 様々な理由による治療の中止（試験を中止しない場合、または中止をとまなう場合）
- 他の治療への切り替え（例：レスキュー治療、または最初の治療が失敗したことにより開始されるセカンドライン治療）
- 軽度な副作用を緩和するための併用薬の使用
- 割り付けられた治療の用量の変更

ランダム化後の事象は、興味のある **estimand** によっては中間事象となる場合がある。例えば、「死亡」は、痛みの症状緩和のための治療の試験においては分類1の中間事象になるが、抗がん剤の臨床試験においては主要評価項目となる。

因果推論の枠組みにおいては、介入とアウトカムは区別されている。介入 (**R** で表される) は、たとえ医学的に必要なものであっても（必要不可欠なもので、行わざるを得ないものであっても）、決定を行うことである。アウトカム (**S** で表される) は、事象（起きたこと）である。例えば、患者が重度の有害事象を経験し治療を中止した場合、有害事象はアウトカムであり、治療の中止は介入となる。以下、「介入」と「治療」は交換可能な用語として使用する。また、「治療」は、複数の治療を含む「治療レジメン」を示す用語として用いられることもある。

ランダム化後の、アウトカムと介入の区別は、Robins らによって開発された、時間依存性の交絡因子が存在する下での、時間依存治療の下での経時的データに対する **estimand** を評価するための因果推論の枠組みにおいては中心的なものである<sup>34</sup>。一般的に、初期（あるいはランダム化された）治療 **Z**（本節では、治療を「**Z**」と表記している）がアウトカム **S** に影響し、それが後続する治療の変更 **R** に影響し、さらに最終的なアウトカム **Y** に影響する。このアウトカム **S** のように、過去の治療の影響を受け、後続する治療に影響し、最終的なアウトカムに影響する中間の結果は、時間依存性交絡因子とよばれる。

より正式に述べると、試験治療の集合 **Z** の要素の一つを選んで、最初に割り付けられた治療 **z** とする。興味のあるアウトカム **Y** を評価する前に、被験者は他の治療  $r \in \mathbf{R}$  を受ける可能性がある。ここで集合 **R** は、割付後に行われる可能性のある、あらゆる治療あるいは割り付けられた治療の変更が含まれる。例えば、集合 **R** には、**Z** と同じ治療（すなわち割り付けられた治療）の継続、または新しい治療が含まれるかもしれない。新しい治療は状況に応じて、レスキュー治療、割り付けられた治療への上乗せ治療としての併用薬、他の治療への切り替え、あるいは無治療などがあるだろう。治療の変更は、事前に規定されていることもある（治療方針やレジメンの一部として）かもしれないし、より自然発生的なもの、例えば、治療を中止したいという被験者の決定、あるいはプロトコルで定められた治療に対する不遵守のこともあるだろう。

治療 **R** は、図 4.1-1 に示される、時間依存性の交絡因子 **S** に影響されるかもしれない。中間アウトカムによる時間依存性交絡の概念は重要である。なぜなら、**S** は交絡因子（観察研究においてベースラインの共変量が初期の治療に影響を与える要因と似たもの）であり、かつ初期の治

療 Z からアウトカム Y への媒介変数でもあるからである。わかりやすくいうと、Z の Y に対する総合効果の一部が、「間接的に」M を通して現れるとき、変数 M は「媒介」とみることができる。図 4.1-1 においては、S および R が「媒介」変数となる。このアイデアは、「直接効果」というものを想定している。直接効果とは、Z から Y に向かう直接のリンク以外の、すべての Z から Y に向かうパスの変数を固定したときに得られる Z の Y に対する効果である。

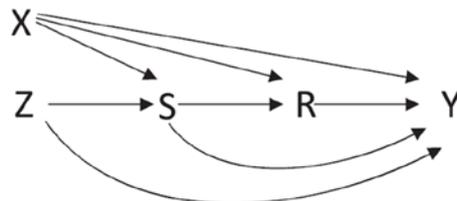


図 4.1-1 因果ダイアグラム

図 4.1-1 に示したダイアグラムには、ランダム化後に起きた中間アウトカム S と治療 R が示されているが、これらは複数の時点における、複数のアウトカム及び治療の変更に一般化することができる。これらの複数の変更はランダム化比較試験よりも観察研究においてより一般的だろう。あるいは、ランダム化後の治療の変更 R が起きないけれども中間のアウトカムが交絡因子となるシナリオが考えられる。例えば、禁煙の研究において、治療の変更がなかった場合であっても、タバコの中断（中間アウトカム S）が、ニコチンの離脱症状の重症度（アウトカム Y）への交絡因子となる。

因果推論の枠組みを用いて、以下のように、さまざまな因果 estimand を記述することができる。なお、4 章ではこれ以降、潜在結果変数について、表記の都合から、 $Y(z)$  のように括弧を用いた表記を採用する。

- 初期の治療 Z の効果（総合効果及び、それを S または R によって直接効果と間接効果に分解したもの）。後述する治療方針ストラテジーに対応する。
- Z と R（S による交絡の影響を除いた）を合わせた同時効果
- 中間アウトカムの潜在アウトカム  $\{S(0), S(1)\}$ （または潜在治療  $\{R(0), R(1)\}$ ）で条件付けられた Z の効果。後述する主要層ストラテジーに対応する。
- Z と R（S の関数として）によって記述される動的治療レジメンの効果

#### 4.1.3 治療変更の特定のための潜在アウトカムの利用

中間アウトカムや治療変更への対処方法は、今我々が何に興味があるのかによって変わりうる。例えば、検証試験では、ランダム化後の治療変更の多くは中間事象として何かしらの対処が必要となる。なぜなら検証試験の主目的は、特定の介入に対する因果効果を適切に評価することだからである。また興味の対象として、中間アウトカムに基づく主要層に対する因果効果に焦点が当てられることもある。これは例えば、治療方法は問わず生存している者に対する QOL への影響に興味がある場合である。この例では中間アウトカムが“生存か否か”となるわけで

ある。

中間アウトカムや治療変更への対処方法がその時々で変わるということは、潜在アウトカムの定義の拡張が必要である。具体的にはこれまでの潜在アウトカムは、最初の治療 $Z = z$ に対する最終時点の潜在アウトカム $Y(z)$ としてしか表現してこなかったが、中間アウトカムや治療変更も考慮した潜在アウトカム $Y(z, s, r)$ として表現する必要が生じることとなる。試験治療の投与量などの変更をプロトコルで許容している試験でも、固定したレジメンの評価に興味がある試験でも、治療変更（新規治療の追加やスイッチング、中止なども含む）は現実的に避けられないものである。治療変更とはそれまでの治療に対する反応を受けてなされるため、割り付けられた最初の治療はその後の反応に応じて変わりうることを考慮して *estimand* や潜在アウトカムを定めておくことは重要であることから、潜在アウトカムの記載表現の拡張は自然である。

以上より、本節以降では、潜在アウトカム $Y(z, s, r)$ は、最初に治療 $Z = z$ を受けた患者がその後中間事象 $S = s$ を発生し、それに伴う介入 $R = r$ を受けたときに観測されたであろうアウトカムとする。なお、 $S$ や $R$ 自体が潜在的な変数であってもよい。また、観測データと潜在データを関連付けるために *Composition assumption* と呼ばれる仮定を置く。仮定の紹介のため記号をシンプルに、最初の治療 $Z$ 、潜在的な中間事象 $S$ 、潜在的な介入 $R$ のみに絞ると、*Composition assumption* とは $R(z) = R(z, S(z))$ を意味する。ここで、 $R(z, S(z))$ とは最初に治療 $Z = z$ を受け、その後中間事象 $S(z)$ が生じた場合に得られたであろう介入である。*Composition assumption* は構造的因果モデルのデータ生成過程から導かれる。詳細については VanderVeele and Vansteelandt<sup>35</sup>を参照されたい。

この仮定より潜在アウトカムを、

$$Y(z, S(z), R(z, S(z))) \equiv Y(z, R(z, S(z))) = Y(z, R(z)) = Y(z)$$

のように略記することがある。

以上の記載拡張により、前節にて紹介があった $Z$ の $Y$ への総合効果や同時効果についても明確に定義できるようになる。ここで、総合効果とは最初の治療 $Z$ の効果、つまり $Z$ や $Z$ 由来の中間アウトカムの影響を受けている $R$ の効果も含めた効果である。つまり総合効果とは“ $Z$ の直接効果” + “ $Z$ の $R$ を媒介する間接効果”であり、 $S$ や $R$ に特別な対処をしないことから、いわゆる *Pure ITT Estimand*（治療方針ストラテジーの概念）に合致する。一方で、同時効果とは $Z$ と $R$ の直接効果のみの集合であり、固定した治療レジメンによる因果効果に興味がある場合に有用である。

また、同時効果に対し固定された治療レジメンではなく、中間アウトカムの値に基づいてその後続治療をどのようにするか定める動的レジメン（例：レスキュー薬の投与や治療用量の増加を必要とする“レスキュー条件”を満たすまでの $t$ 日間治療を実施）の因果効果に対しても、上記の潜在アウトカムの記載拡張により同様に表すことができる。つまり、動的レジメン $r = g(z, S(z))$ による潜在アウトカムは $Y(z, g(z, S(z)))$ で書ける。なお $r$ とは、 $Z = z$ を受け、その後中間アウトカム $S(z)$ が生じた患者に、それに基づいた後続治療を返す関数（レジメンルール）である。いくつかの *estimand* は動的レジメンの考え方をういた方が解釈しやすく、詳細は後節を参照されたい。

本節の最後として簡潔であるが、最適なレジメンの推定が主目的である場合の話を紹介する。

このとき動的レジメンの効果の推定、最適なレジメンの同定のために **Sequential Multiple Assignment Randomized Trials (SMART)** と呼ばれる試験が実施されることがある。なお、通常の検証的なランダム化試験は最適なレジメンの同定には不向きである。なぜならランダム化試験ではプロトコルで細かく事前に規定する事項が多く制約的で、医師などの意思決定の自由度が低いためである。例えば、レスキュー条件を厳格に事前に定める試験においては条件に合致しない症例がレスキュー治療を受ける確率は0であるはずである。これは **IPW** 推定量を用いるときに必要な **positivity** の仮定が成立しないことに繋がりうる。SMART についての詳細は **Chakraborty and Moodie<sup>36</sup>**や **Tsiatis et al.<sup>37</sup>**を参照されたいが、ここでは実際に導出された最適なレジメンの具体例として **Clinical Antipsychotic Trials of Intervention and Effectiveness (CATIE)** 試験 (**Lieberman et al.<sup>38</sup>**) を紹介する。この試験は初回のランダム化された治療を受けた後の第2ステージにおいてさまざまな治療オプションへの再ランダム化がなされた試験である。この試験データの再解析により、閾値パラメータ $\theta$ を用いた以下のような最適なレジメンが推定された (**Shortreed and Moodie<sup>39</sup>**)。

“ベースライン時点で **PANSS** スコアが $\theta$ 以上であれば、**perphenazine** を投与する。その後 **PANSS** スコアが $\theta$ を下回った場合には **atypical antipsychotic** へとスイッチする。ベースライン時点で **PANSS** スコアが $\theta$ を下回っている場合は全12か月間、**atypical antipsychotic** を投与する。”

#### 4.1.4 アウトカムの潜在性と反事実的性質

因果 estimand を定義する際にそれぞれの臨床的な文脈における意義と解釈可能性を検討することが重要であるのと同じく、潜在アウトカムを定義する際にも同様の検討が重要である。定義が可能であるということがすなわち、それが適切であることを意味するわけではない。

反事実の程度の尤もらしさについて、複数のシナリオを想定し、並べて比較検討することが有用である。以下、アドヒアランス (割り付けられた治療の遵守) を中間事象とする状況を用いて例示する。実際の治療  $R \in \{0, 1\}$  と同じ水準を用いて、実薬治療及びプラセボへの遵守を表すこととする。例えば、実薬治療にランダム化され、その後不遵守となった患者  $i$  の潜在アウトカムは  $Y_i(1, 0)$  と表現される。第1の引数は割り付けられた治療を示し、第2の引数は実際の治療を示す。

上記の状況について、以下3つのシナリオを想定し、各シナリオにて  $Y_i(1, 0)$  が観察された場合に潜在アウトカム  $Y_i(1, 1)$  に意義があるかどうかを考察する。

- 1) 治療とは無関係な理由による不遵守。例えば、治験への参加の継続が不可能となるような転居等。
- 2) 治験薬の投与中止に至るような有害事象 (重症度が類似している当該事象を発現した、すべての患者ではないが一部の患者で投与中止しているような状況)。
- 3) 試験実施計画書により投与を中止することが義務付けられた患者。例えば、肝酵素が一定の閾値を超えた場合等。

シナリオ1は、これらの患者が遵守していたであろうシナリオを想定することが容易であり、

通常は適切であるため、最も反事実的でない。シナリオ3は、これらのアウトカムを持つすべての患者が継続する確率が0であるため、最も反事実的である。このような状況で割り付けられた治療を遵守した場合の潜在アウトカムは、おそらく臨床的に適切ではない。シナリオ2は、1と3の中間の状況である。不遵守が確実ではない状況では、遵守下での潜在的な結果を考慮することが適切である。なお、例えば、実際の臨床現場でのアドヒアランスが、患者が薬剤に対して盲検化されプラセボを投与されている可能性のある臨床試験よりも優れている場合には、これらの（遵守下の）反事実アウトカムがより適切である可能性がある。

#### 4.1.5 潜在アウトカムの枠組みにおける Estimand の定義

##### 4.1.5.1 序文

本節では、一般的なシナリオに関して、潜在アウトカムの枠組みで因果 estimand を定義できるかを示す。ICH E9(R1)<sup>1</sup>に概説されている中間事象を扱うための五つのストラテジーとともに示す。

前節の、実薬治療とプラセボの2群によるランダム化試験でのアドヒアランスの中間事象を例示に用いる。(Z,R)を、最初にランダム化された治療 Z と、中間事象により Z と潜在的に異なる可能性のある「最終」治療 R からなる、研究中の治療レジメンとする。遵守患者については、 $R(0)=0$ ,  $R(1)=1$ と表現できる。簡単のため、連続/2値結果変数 Y のみを考える。

##### 4.1.5.2 治療方針ストラテジー

Estimand は、初期治療に従ったかどうかにかかわらずランダム化群間のアウトカムを比較することであり、平均因果効果として、潜在アウトカムの枠組みにおいて以下のように定義することができる。

$$\begin{aligned} \text{ATE} &= E[Y(1, S(1), R(1, S(1))) - Y(0, S(0), R(0, S(0)))] \\ &\equiv E[Y(1) - Y(0)] \end{aligned}$$

4.1.3で考察したように、ここで、S 及び (S に対する介入) R は非侵襲的であり、(最初にランダム化された治療) Z の潜在的なアウトカムでもある。したがって、治療方針ストラテジーは、規定された興味のある治療レジメンの一部であるいくつかの治療の効果の影響を評価している。

前述のように、治療方針ストラテジーは、任意の S または R に関連する直接的及び間接的効果に分割することのない、Z の総合効果の推定に対応している。

##### 4.1.5.3 複合ストラテジー

中間事象が関心のある結果の一部となるように、さまざまな方法で実装できる。

##### 4.1.5.3.1 2値アウトカム Y

まずは、2値アウトカム Y を想定する。Y=1は再発またはノンレスポンスなどの不良なアウトカムを示す。有害事象による治験薬投与中止などの、中間アウトカム S=1（中間事象）の患者は、他の臨床アウトカムの状態にかかわらず、不良アウトカム Y=1が割り当てられる。潜在アウトカムは、

$$\tilde{Y}(z) = I(Y(z) \vee S(z)), \quad z = 0, 1$$

として再定義することができる。ここで  $I(\cdot)$  は指示関数であり、条件式が真のとき（つまり、 $Y(z) = 1$  または  $S(z) = 1$ ）に1、偽のとき（つまり、 $Y(z) = 0$  かつ  $S(z) = 0$ ）に0となる。

Estimand は、

$$E[\tilde{Y}(1) - \tilde{Y}(0)]$$

である。中間アウトカム  $S$  を制御することにより  $Z$  が  $Y$  に及ぼす直接的な影響を「引き出す」ことを試みているわけではない。また、この場合、中間事象に続く新たな介入  $R$  は考慮していない。

#### 4.1.5.3.2 連続アウトカム $Y$

次に、連続アウトカム  $Y$  を想定する。単一の estimand 内で連続的な  $Y$  および2値の  $S$  を組み合わせる必要があるため、状況はより複雑になる。

2値アウトカムの場合と同様に、 $Y$  の連続尺度を粗くすることによって中間事象をアウトカムに組み込む方法がある。たとえば、 $Y$  をランクまたはさらに粗いカテゴリに変換し、（望ましくない）イベント  $S$  を最も低いランク（カテゴリ）に割り当てることができる。

$S$  についてのより詳細な情報は、イベント  $S$  の発生のタイミングまたは重症度に基づいて  $S$  を有する患者をランク付けすることによって組み込むことは可能である。

別の戦略として、イベント  $S$  によって切り捨てられた  $Y$  の曲線下面積を使用することがあり、これは、イベントが起こるまでの時間とその時間中のアウトカム  $Y$  との組み合わせを反映する（例えば、質調整生存時間）。

さらに別の戦略として、 $S$  の発生を条件とした  $Y$  の分布についての（検証不可能である）仮定を立てることにより、 $Y$  の母集団パラメータについての仮説の中に  $S$  を統合することができる。例えば、estimand は、 $S$  を経験した患者の潜在アウトカム  $Y$  が分位点またはトリミング閾値よりも下（悪い位置）にランク付けされるという仮定を用いて、トリミングされた平均値または分位点における治療の差であってもよい。この estimand は、部分集団を「 $S(z)$  の関数として正確に定義する」ことなく適用される。

#### 4.1.5.4 仮想戦略

##### 4.1.5.4.1 全患者の治療遵守を想定

興味のある estimand の一つは、たとえ一部の患者がノンアドヒアラントでも、すべての患者が治療を遵守したと仮定した場合の（反事実の状況の）、ランダム化されたすべての患者における平均因果効果である。この場合、平均因果効果は、いわゆる初期治療に対する「制御された直接効果」を標的とする。

$$ATE = E[Y(1, 1) - Y(0, 0)]$$

中間治療 ( $R$ ) は、中間結果  $S(z)$  にかかわらず、あるレベルで、ここでは初期治療と同じになるように制御または固定されるという意味で「制御」される。ここでは、略記  $Y(z, S(z), r) = Y(z, r)$  を利用していることに留意されたい。本略記については4.1.3を参照されたい。

#### 4.1.5.4.2 中間事象後無治療を想定

もう一つの仮想ストラテジーについて検討する。ここでは、時点  $t$  まで割り付けられた薬剤により治療を行うが、割り付けられた薬剤による治療が不可能な中間事象を有する患者については、治療中止をもたらす中間事象まで治療し、中間事象の発生から時点  $t$  までは治療を行わない状況での治療効果を考える。たとえ倫理的配慮により無治療が許されない場合でも、最初の治療による時点  $t$  での効果を評価することが適切な場合があるかもしれない。

中間事象後介入は、 $R \in \{0, 1, 2\}$  を想定する。ここで、 $0$  と  $1$  は  $Z \in \{0, 1\}$  と同じ治療水準を意味し、 $R=2$  は「無治療」を意味する。そして、ランダムに選択された患者  $i$  は、 $Z=1$  にランダム化され中間事象の非存在下 ( $S_i(1)=0$ ) でランダム化治療を続けた場合には、潜在アウトカム  $Y_i(1, 1)$  を持ち、もしくは中間事象の発生後 ( $S_i(1)=1$ ) に治療へのアクセスを拒否された場合には、 $Y_i(1, 2)$  となる。潜在アウトカムは、両者の混合として、

$$Y_i(1, 1)(1 - S_i(1)) + Y_i(1, 2)S_i(1)$$

と書くことができる。

仮想ストラテジーにおける平均因果効果 estimand は、(中間事象  $S$  の条件付けによって) 以下のように表すことができる。

$$\begin{aligned} \text{ATE} = & E[Y(1,1)|S(1) = 0] \Pr(S(1) = 0) \\ & + E[Y(1,2)|S(1) = 1] \Pr(S(1) = 1) \\ & - \{E[Y(0,0)|S(0) = 0] \Pr(S(0) = 0) \\ & - E[Y(0,2)|S(0) = 1] \Pr(S(0) = 1)\}. \end{aligned}$$

この仮想ストラテジーは、中間事象が治療  $R=2$  への切り替えのトリガーとなる動的治療レジメンの一例とみなすことができることに留意されたい。

これらの複雑な状況では、estimand を言葉で表現することは煩わしいことがある。潜在アウトカムの表記を用いることで、多少複雑ではあるが、あいまいさを排した表現が可能となる。このように、仮想ストラテジーを使用する場合には、関心のあるシナリオを正確に定義することが重要であり、これは典型的には、「何が起こらないか」及び「特定の条件が満たされた場合に何が起こるか」の両方を述べることを意味する。前の例では、レスキュー治療の条件を満たした患者が以前のランダム化治療を継続するか無治療で継続するかを明記する必要がある。

#### 4.1.5.5 主要層ストラテジー

ここまでは、すべての被験者についての因果効果について考えてきたが、本節では、部分集団における因果効果を考える。この効果を local average treatment effect (LATE) とよぶ。この estimand は潜在変数  $Y(z, S(z), R(z, S(z))) \equiv Y(z)$  を用いて、 $\text{LATE} = E[Y(1) - Y(0) | R(1) - R(0) = 1]$  と表現する。ここで、 $R(1) - R(0) = 1$  は、 $R(1) = 1, R(0) = 0$  となる状況においてのみ成り立つ。LATE estimand は Angrist, Imbens, and Rubin<sup>40</sup> によって紹介され、次の仮定の下、操作変数型の推定量である。なお、操作変数法に関しては2.6を参照していただきたい。

- 仮定 1: 逆の治療を受ける “defiers” がいない。すなわち “monotonicity” が成立している。

- 仮定 2：“exclusion restrictions”として，“never-takers” ( $R(1)=R(0)=0$ ) と “always-takers” ( $R(1)=R(0)=1$ ) において同一の潜在アウトカムを持つ ( $Y(0)=Y(1)$ )。

Imbens and Rubin<sup>41</sup>は、この estimand を “complier average causal effect” (CACE) とよんでいる。

LATE の他のバリエーションとして、対照となる治療の潜在的な遵守に関わらず、実治療は遵守する患者における治療効果に対する estimand がある。これは、 $E[Y(1) - Y(0) | R(1) = 1]$ として表現できる。

ここまでは、ランダム化後の治療によって主要層を構成することを考えてきたが、ランダム化後の中間アウトカムによって主要層を構成することも可能である。例えば、どちらの治療を割り付けても生存する患者の層における治療効果を評価することが考えられる (survivor average causal effect (SACE) : Zhang and Rubin 2003<sup>42</sup>; Hayden, Pauler, and Schoenfeld<sup>43</sup>; Imai<sup>44</sup>)。他に、連続値のアウトカムの閾値によって規定される主要層が考えられる。例えば、プラセボ（心理療法のみ）では反応しない大うつ病性障害の患者の層を考えてみる。興味のあるイベント（反応）を  $S=1$ を表すと、estimand は  $E[Y(1) - Y(0) | S(0)=0]$ である。この主要層 estimand は割付の直接的な効果となるが、より限定された集団（プラセボでは改善しないであろう患者）での効果である。

より広い視点から見ると、主要層は、臨床的に意味のある疑問に対応するために、多くの基準に基づいて行うことができる。例えば、Mehrotra, Li, and Gilbert<sup>45</sup>は、ワクチンとプラセボのどちらにランダム化されたかに関わらず、（ランダム化後に）感染するであろう患者によって定義された主要層のウイルス負荷に対するワクチン接種の効果を評価している。このアウトカムは感染した患者のみで測定されることになる。

先に述べた CACE は2つの仮定をおいていた（仮定1及び仮定2）。より一般的な設定では、これらの仮定は問題があるかもしれない。例えば、仮定2の “exclusion restrictions” について、ランダム化後の介入 ( $R$ ) ではなく、ランダム化後のアウトカム ( $S$ ) におきかえた場合、妥当ではないかもしれない。また、仮定1の “defiers” がいないという仮定は、実薬とプラセボを比較する場合には妥当であるが（例えば、プラセボでは再発しない患者は実薬では再発しない；実薬では AE を発症しない患者はプラセボでは AE を発症しない、など）、実薬同士を比較する場合には妥当性が低いかもしれない。

一般的に主要層 estimand は識別性が問題となり、強い仮定を置かなければ観察データから推定することはできない。本報告書では詳細は述べないが、主要層 estimand の識別性の問題については、様々なアプローチが提案されている。一つのアプローチは、識別性についての感度分析を行うことである。また、別のアプローチとして、Magnusson et al.<sup>26</sup>は、ベイズ流の枠組みを開発した (3.2参照)。

主要層 estimand の識別可能性に対応するためのもう一つの選択肢は、ベースライン共変量 ( $X$ ) に含まれる情報を用いることである。これには強い仮定が必要であり、特に、 $X$  が与えられた場合、層  $S(0)$ は、潜在的な転帰  $Y(1)$ を予測するための追加情報を提供しないという仮定が必要である。すなわち、 $S(0) \perp Y(1) | X$  である。この仮定は “principal ignorability” として知られている。

#### 4.1.5.6 治療下ストラテジー

治療下ストラテジーにおける estimand は、治療期間を考慮することなく、治療へのアドヒアランスの終了を示す時刻  $t$  及びそれ以前での治療効果を評価する。

治療下ストラテジーを説明する例として、生存している間の症状の重症度に対する緩和治療の効果について、死亡を中間事象とした場合である。対照群と実薬群に無作為に割り付けられた患者の潜在的な治療期間を、試験が無限に継続すると仮定して、それぞれ  $t_{dr}(0)$ ,  $t_{dr}(1)$  と定義する。ここで、 $\tilde{t}(z) = \min(t, t_{dr}(z))$ ,  $z=0,1$  とし、 $Y_u(z)$ ,  $z=0,1$  は時点  $u$  における潜在アウトカムであるので、治療下ストラテジーにおける estimand は、 $E[Y_{\tilde{t}(1)}(1) - Y_{\tilde{t}(0)}(0)]$  と表現できる。

#### 4.1.5.7 動的治療レジメンを伴うシナリオ

ICH E9(R1)<sup>1</sup>では動的治療レジメンの効果はカバーしていないが、ここでは動的治療レジメンについて検討する。

例として、 $S(z)$ を「レスキュー薬の使用 ( $R=1$ ) が必要なアウトカムか否か」とし、その判定条件は  $L(z)$ が閾値  $\theta$  を超えているかどうかとする。このとき、 $S(z)=g_\theta(z, L(z))=I(L(z)>\theta)$  と考えられる。 $I(\cdot)$ は指示関数であり、 $\theta$ は臨床的に関連する症状の重症度のカットオフであり、レスキューの必要性を定義し、 $L$ の値が低いほど、より良い転帰を表す。治療レジメンを制御するパラメータ  $\theta$  は、あらかじめ決められたものであり、データから決定したものではない。

動的治療レジメンの治療効果は、中間連続アウトカム  $L$  の関数である  $E[Y(1, g_\theta(L)) - Y(0, g_\theta(L))]$  である。

他の例として、患者が最初に低用量群 ( $Z=0$ ) または高用量群 ( $Z=1$ ) に無作為に割り付けられ、低用量群の患者は、特定の条件 (有効性が不十分) が満たされた場合に (一種のレスキューとして) 高用量群に切り替えることができる動的な治療レジメンを考える。“rescue” (dose escalation) の必要性を  $I(L>\theta)$  と定義し、治療レジメン  $g_\theta(z, L(z))$  とする。 $g_\theta(z, L(z))=I(L(z)>\theta)(1-z)+z$  で表すことができる。 $Y(z, r)$ を投与量  $z$  で開始し、後の投与量  $r$  が規則  $r=g_\theta(z, L(z))$  に基づいて割り当てられた患者の潜在アウトカムとすると動的レジメンの効果は  $E[Y(1, g_\theta(1, L)) - Y(0, g_\theta(0, L))]=E[Y(1, 1) - Y(0, g_\theta(0, L))]$  として表すことができる。治療期間全体を通して高用量を投与する治療レジメンと、用量漸増を行う治療レジメン、すなわち、低用量から開始し、有効性が不十分であるという事前に定められた基準を満たした場合には高用量にする治療レジメンの2つの治療レジメンを比較していることになる。すべての患者が正確にレジメン  $g_\theta(z, L(z))$  に関連するルールに従っていれば、この estimand は、4.1.5.2で説明した治療方針ストラテジー  $E[Y(1, R(1)) - Y(0, R(0))]$  によって与えられたものと同じになる。治療変更のルール (用量調整、レスキュー薬の追加または切り替え、背景療法への変更など) が決定論的に規定されていないか、正確に守られていない場合、治療方針ストラテジーは、治療レジメンの加重平均である。これに対し、動的治療レジメンは、ベースライン後の各状態を特定の治療介入にマッピングするため、治療方針ストラテジーよりも具体的である。動的治療レジメンの課題は、例えば、国ごとの臨床実践がまったく異なる可能性のある異なる地域では、一般化できないということかもしれない。

#### 4.1.6 欠測データの取扱い

潜在アウトカムの枠組みにおける estimand は、各被験者に対して  $Y(1)$  や  $Y(0)$  等の複数の潜在アウトカムが存在し、それらを観察できるものと仮定して、その期待値で定義される。そのため、estimand を定義する段階では、欠測の問題を気にする必要はなく、欠測に対する対処は推定の段階で検討することになる。

一致性の仮定が成り立っていたとしても、一人の被験者に対して観察可能な潜在アウトカムは一つのみである。そのため、仮にすべての被験者に対して試験終了まで観察できたとしても、残りの潜在アウトカムは欠測していると言える。ここでの潜在アウトカムの欠測は、試験デザイン上観察可能なアウトカムが欠測している訳ではない。そこで、潜在アウトカムの期待値を、estimand に対応する形で観察可能なアウトカムの期待値に置き換えて解析を行う必要がある。このことから、デザイン上観察可能なアウトカムに欠測がある場合は、推定の段階でその欠測に対して考慮する必要があることがわかる。つまり、潜在アウトカムの枠組みを用いた場合でも、ICH E9(R1)<sup>1</sup>と整合する手順で解析手法を検討することが可能になる。すなわち、科学的な疑問が試験の目的と estimand を定義し、その後 estimand と一致した方法で解析手法を選択することになる。

欠測の問題に対処するにあたり、治療中止や試験中止は、アウトカムの欠測とは対処方法の考え方が異なることを認識することが重要になる。治療又は試験の中止は、中間事象（試験治療以外の治療への切り替え、又は治療なしへの切り替え）と見なすことができ、ICH E9(R1)<sup>1</sup>に記載されているいずれかのストラテジーを用いて estimand の枠組みで対処することが可能となる。例えば、治療方針ストラテジーでは治療の切り替え等の中間事象が起きたとしてもそれ以降に観察されたアウトカムが欠測とはならず、中間事象発現以降のデータが収集されていない場合にのみ欠測となる。選択したストラテジーによっては、中間事象発現以降のアウトカムが観察されていたとしても欠測として扱われる場合もある。当該中間事象が発現した被験者内で、欠測有無（中間事象発現以降のデータ収集有無）のグループ間で収集されている時点までのアウトカムプロファイル及びベースライン共変量が同じ場合は、MAR の仮定の下で議論が可能となる。割付時の治療を継続した場合の効果に焦点を当てた仮想ストラテジーでは、治療切り替え後の観察されたアウトカムは関心のあるパラメータの推定には関係がないため、治療の切り替え以降のデータは欠測とみなされる。このとき、MAR の仮定の下で欠測データを処理する場合、当該中間事象発現以降の仮想的アウトカムは、収集されている時点までのアウトカムプロファイルプロファイル及びベースライン共変量が同じ当該中間事象未発現の被験者と同様であったと仮定することになる。ただし、MAR の仮定が満たされているかは観察されたデータからは検証できないため、感度分析により MAR の仮定の下での推定結果に対する頑健性を評価することが重要である。

一般には、複合ストラテジー、治療下ストラテジー、及び治療方針ストラテジーによって中間事象を取り扱う場合は、中間事象によるデータの欠測はない（中間事象によらない欠測は起こり得る）。一方で、主要層ストラテジーや仮想ストラテジーによって中間事象を取り扱う場合は、中間事象によるデータの欠測が起こることを想定する必要がある。

## 4.2 COVID-19の estimand に対する影響

### 4.2.1 導入

2019年12月に中国で発生した COVID-19は、2020年には世界的なパンデミックとなった。この疾患と公衆衛生上の対策によって、世界的に臨床試験に非常に大きな影響を与えており、治験の実施医療機関の制限などから、治験実施計画書の遵守の困難さを含む運営上の問題が発生している。進行中の試験について、治験依頼者は、被験者の安全を確保し、運営上の問題に対処するために、さまざまな治験実施計画書の緩和策を実施してきている。

COVID-19の影響による治験実施計画書からの逸脱と試験実施手順の変更は、臨床試験のデータへさまざまな影響をもたらし、重要な統計的問題も発生している。極端な場合には、試験の完全性 (integrity)、解釈可能性、又は試験の目的の達成を脅かす可能性がある。試験結果の綿密かつ適切な解釈を提供するために、計画された統計解析を修正したり、補足したりする必要があるかもしれない。議論されている問題の多くは、パンデミック前に開始された、COVID-19に関連しない治療に対する進行中の臨床試験に関連するものであるが、問題点の多くは、パンデミック中又は後に開始された、新しい試験にも適用できる可能性がある。因果推論との関連という意味では、議論されている統計的問題の一つとして、estimand を構成するときに重要な考慮事項となる中間事象として、COVID-19の影響にどのように対応するかということ挙げることができる。本節では、4.2.2で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からの声明「COVID-19が治験に及ぼす影響について—統計解析・データマネジメントの視点—」<sup>32</sup>及び4.2.3で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KT-1がMeyer et al.<sup>33</sup>を翻訳した成果物を紹介した上で、4.1で説明した因果推論と臨床試験の estimand とも関連付けて、estimand に対する COVID-19の影響について、特に中間事象の取扱いに焦点を当てて説明する。

### 4.2.2 データサイエンス部会声明

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は、COVID-19が治験に及ぼす影響について、統計解析やデータマネジメントの視点から、今後起こりうる問題やこれらへの対処についてまとめ、広く共有することを目的として、2020年5月11日付で声明を公表した。

声明では、COVID-19の感染拡大により、統計解析の内容を再検討する際、試験目的に沿った当該試験で知りたい治療効果は何かという estimand に関係する点に着目することが重要であるとしている。また、中間事象の取扱いを含む estimand に関係する統計解析に関連する問題として、以下の内容が記載されている。

【以下、声明から抜粋】

- 「その試験で知りたい治療効果」を第一に考慮します。一般に、COVID-19の感染拡大があっても、試験計画時に決めた「その試験で知りたい治療効果」そのものは変わらないと考えます。
- 「その試験で知りたい治療効果」を明確にするためには中間事象の対応法を規定する必要があります。COVID-19感染拡大により、試験開始時には想定していなかった中間事象が生じる

可能性があるため、試験中であっても当該中間事象の対応法を考え、必要に応じて解析方法を変更します。中間事象の対応を検討するにあたり、疾患領域や、薬剤の特性、試験デザイン、試験実施地域などを考慮することが重要になります。

- 中間事象は、COVID-19に被験者自身が感染して起こる中間事象と、感染リスクを避けるために行われる緊急措置や行為に起因する中間事象に大別できます。多くの場合（特に後者の場合には）、COVID-19による中間事象の影響が、「その試験で知りたい治療効果」に含まれないと考えられます。この時には仮想ストラテジー（「中間事象が発生しなかった」状況を考えます）を適応し、当該中間事象が発生しなかった治療効果を推測することが考えられます。一方で中間事象が「その試験で知りたい治療効果」に影響を与えると想定される場合は、当該中間事象の対応方法を特に慎重に検討する必要があります。

【抜粋終わり】

上記の通り、データサイエンス部会の声明では、多くの場合にCOVID-19による中間事象に対して仮想ストラテジーを適用することが記載されており、この点について、後の節で因果推論との関連を含めて説明を加える。

#### 4.2.3 COVID-19パンデミック下での臨床試験実施に対する統計的課題と推奨

2020年7月9日に、アメリカ統計学会の機関誌の一つである *Statistics in Biopharmaceutical Research* 誌にCOVID-19パンデミック下での臨床試験に対する統計学的考察の論文<sup>33</sup>が公表された。また、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KT-1は、「翻訳：COVID-19パンデミック下での臨床試験実施に対する統計的課題と推奨」を2020年7月22日に公開した。その中で特に *estimand* の取扱いに関して、以下の内容が記載されている。

COVID-19パンデミックの影響を受けた試験では、最初の主要な論点は、主要な目的、すなわち、主要な *estimand* がCOVID-19の潜在的な交絡の影響を除いた治療効果を対象とすべきかどうかとなる。パンデミック前に開始されたほとんどの試験については、当初の主要な目的を計画通りに維持すること、すなわちパンデミック関連の混乱によって交絡されない治療効果を対象とすることを（論文の著者らは）推奨する。パンデミック後の新しい試験では、この治療効果の定義も妥当であろうが、その妥当性は、試験計画の多くの側面に依存する。これは自動的に、すべての可能性のあるパンデミックと関連する中間事象について同じ仮想シナリオを持つ広範な「*hypothetical estimand*」を意味するものではない。試験の状況に応じ、異なるタイプの中間事象に対して異なる方法で対処する必要があるかもしれない。また、パンデミックと関連する混乱を考慮するために、当初の *estimand* の定義に若干の変更が必要となるかもしれない。COVID-19パンデミック中に出現した特定の状態又は部分集団における治療効果を特徴づけるために、探索的な目的で補足的 *estimand* を検討してもよい。試験実施計画書で *estimand* が明示的に定義されている試験では、試験依頼者はパンデミックと関連する混乱が *estimand* の要素に及ぼす影響を検討し、必要に応じて修正すべきである。Estimand が明示的に定義されていない試験の場合、パンデミックと関連する影響が試験目的、実施、推測及び統計解析に与える影響について検討する必要がある。

以上から、4.2.2で説明した部会声明と同様に、本論文では、COVID-19による中間事象に対して仮想ストラテジーを適応する可能性について記載されている。以降の節で、estimand の構成要素ごとに、COVID-19が与える影響を説明する。

#### 4.2.3.1 中間事象の取扱い

パンデミックと関連しない中間事象に対するストラテジーは変わらないはずである。本論文では、パンデミックと関連する中間事象の取扱いについてのみ議論されている。

中間事象のうち、パンデミックと関連する要因の結果として発生し、他のパンデミックと関連しない理由（例えば、有効性の欠如や毒性による治療中止など）に起因しない場合には、パンデミックと関連する中間事象とみなすべきである。パンデミックと関連する重要な中間事象は、まず、試験治療のアドヒアランス（試験治療の中止など）や最終的なアウトカム（死亡など）の確認に影響するかどうかの観点から分類すべきである。次に、これらの中間事象を取り扱うための適切なストラテジーを選択するために、中間事象をパンデミックと関連する要因に応じてさらに分類する必要がある。

関連する要因には、試験治療の利用可能性（試験治療へのアドヒアランスやアクセス）、被験者の COVID-19の感染状態、COVID-19に感染した場合の併用治療がある。要因が組み合わされることによって生じる中間事象の種類は潜在的に多いが、すべて又は大部分の中間事象を同様の方法で扱うことができるであろう。ある種の試験治療のノンアドヒアランスは、通常は中間事象として考慮されないかもしれないが、パンデミックの状況下では中間事象として考慮する必要があるかもしれない。例えば、著しい遵守状況の低下や一時的な治療中断は、試験計画時には想定されていなかったかもしれないが、これらがパンデミックと関連する混乱に起因する可能性が高いと考えられる場合には、中間事象となりうる。また、死亡の発生が当初予想されていなかった試験では、COVID-19による死亡は潜在的な中間事象として考慮されるべきである。

パンデミックと関連しない理由に起因する中間事象（例：試験治療の中止）の種類が多くは、パンデミック発生以前の主要な estimand において対処されているであろう。パンデミックが原因でこれらの中間事象が発生した場合においても、当初のストラテジーが正当であるかどうかを検討することから始めることを（論文の著者らは）推奨する。当初のストラテジーが正当化できない場合は、別のストラテジーを選択すべきである。

治療方針ストラテジーは、得られた結論をパンデミックが存在しない場合の結論として一般化できないため、ほとんどのパンデミックと関連する中間事象について、治療方針ストラテジーに対しては一般的に科学的な関心がないであろう。例えば、治療方針ストラテジーの下で推定される治療効果は、パンデミックと関連する混乱に起因する早期の治療中止に対して、パンデミックと関連する要因（試験薬供給の中断など）に起因する治療の中止や変更が生じるレジメンの影響を反映しており、主要な試験目的に沿わないと考えられる。治療方針ストラテジーの下では、推定される治療効果は感染症及びそれに対する治療効果を反映している可能性があり、これは主要な目的における関心事ではないであろう。パンデミックと関連する中間事象に対して治療方針ストラテジーを適用することは、そのような事象を発現した被験者の割合が低く、パンデミックと関連しない同様の中間事象（例えば、同じ治療の変更という中間事象）に対

してこのストラテジーが計画されていた場合には、正当化される可能性がある。比較的短期間の治療中断に対応する中間事象の取扱いにおいても検討されるであろう。また、非劣性試験及び同等性試験における、パンデミックと関連する早期の試験治療中止の中間事象については、そのような事象数の増加に応じて治療群間の類似性が高まる可能性があるため、治療方針ストラテジーの使用は回避されるべきである。

#### 【4.1との関連：治療方針ストラテジーについての考察】

治療方針ストラテジーに関して、4.1において、平均因果効果は以下の式で計算されると説明した。

$$ATE = E \left[ Y \left( 1, S(1), R(1, S(1)) \right) - Y \left( 0, S(0), R(0, S(0)) \right) \right] \equiv E[Y(1) - Y(0)]$$

本式によると、パンデミックと関連する要因（式中の  $S$  に相当）及び早期の治療中止や変更（式中の  $R$  に相当）のような中間事象が最終的な平均因果効果の式（最右辺の式）に含まれていないため、パンデミックと関連する要因を明示的に考慮することなく、それらを反映したものとして平均因果効果が推定されることがわかる。因果推論の観点から考えても、「治療方針ストラテジーは、得られた結論をパンデミックが存在しない場合の結論として一般化できないため、ほとんどのパンデミックと関連する中間事象について、治療方針ストラテジーに対しては一般的に科学的な関心がないであろう」という本論文の主張を理解することができる。

#### 【治療方針ストラテジーについての考察終わり】

複合ストラテジーについても同様の考えが適用される。複合ストラテジーでは、結果変数の定義に中間事象が組み込まれているが、パンデミックと関連する多くの中間事象には適切ではないと考えられる。例えば、パンデミックと関連する混乱による試験治療の中止は、有効性の欠如や副作用による中止と同様に治療の失敗としてカウントされるべきではない。なお、呼吸器疾患を対象とした試験では、COVID-19の合併症が好ましくない転帰の一形態として複合ストラテジーで考慮される可能性があるため、より詳細な検討が必要となるかもしれない。

#### 【4.1との関連：複合ストラテジーについての考察】

複合ストラテジーに関して、4.1において、特に結果変数が2値の場合、平均因果効果は以下のように定義されると説明した。結果変数は、疾患の再発又はノンレスポンスなどの不良なアウトカムを示すと想定する。このとき、治験薬投与中止の原因となるような副作用の発現などの、中間事象あり ( $S=1$ ) の被験者は、他の臨床アウトカムの状態に関わらず、不良なアウトカムを割り当てる。このとき潜在結果変数は、

$$\tilde{Y}(0) = I(Y(0) \vee S(0)), \tilde{Y}(1) = I(Y(1) \vee S(1))$$

で再定義され、平均因果効果は以下の式で計算される。

$$ATE = E[\tilde{Y}(1) - \tilde{Y}(0)]$$

本式によると、パンデミックと関連する要因（式中の  $S$  に相当）が直接的に潜在結果変数に組み込まれることにより平均因果効果が推定されることがわかる。中間事象がパンデミックと関

連する混乱に起因する場合、それを潜在結果変数として取り扱うことの妥当性をよく議論しなければならないと考えられる。

【複合ストラテジーについての考察終わり】

COVID-19と関連する事象（例えば、COVID-19による重篤な合併症又は死亡）で層別化する主要層ストラテジーは、将来の実臨床での状況と関連しない要因に基づいて定義された被験者の部分集団に結論を限定することになるため、主要な estimand として関心を持たれる可能性は低いと考えられる。

【4.1との関連：主要層ストラテジーについての考察】

主要層ストラテジーに関して、4.1において、部分集団における平均因果効果に興味があるときに適用されるストラテジーとして説明した。また、4.1では、そのような平均因果効果として local average treatment effect (LATE) が以下の式で計算されると説明した。

$$\text{LATE} = E[Y(1) - Y(0) | R(1) - R(0) = 1]$$

ただし、部分集団の規定（上の数式の「 $R(1) - R(0) = 1$ 」の部分であり、 $R$ で規定される場合も  $S$ で規定される場合も考えられる。）はどの部分集団に興味があるかによって変わるため、注意していただきたい。例えば、COVID-19による重篤な合併症を中間事象として考えるなら、どちらの治療を受けても COVID-19による重篤な合併症を発現しない部分集団を対象とすることが一例として考えられる。

本式によると、LATE は中間事象で条件付けた平均因果効果が推定されることから、COVID-19と関連する事象で構成する部分集団に対して興味があるかどうか、主要層ストラテジーを適用するかどうかを検討する際に重要であることがわかる。

【主要層ストラテジーについての考察終わり】

パンデミックと関連しない中間事象に対して、当初から治療下ストラテジーの適用が計画されている場合、特定の条件下では、治療下ストラテジーが引き続き適切な場合がある。治療効果の確立に治療期間が関係ない場合（平均因果効果に対して治療下にある時間を考慮する必要がない場合。例えば、緩和ケアにおける疼痛の治療。）には一般的に正当化できる。一方で、試験治療の結果を確実に測定するために必要な最小の治療期間など、特定の条件を考慮する必要があるかもしれない。

【4.1との関連：治療下ストラテジーについての考察】

治療下ストラテジーに関して、4.1において、平均因果効果は以下の式で計算されると説明した。

$$\text{ATE} = E[Y_{t(1)}(1) - Y_{t(0)}(0)]$$

本式によると、治療下にある潜在的な時間によって平均因果効果が推定されることから、試験治療の結果を確実に測定するために必要な最小の治療期間などがパンデミックと関連する要因によって影響されるかどうかを検討する必要があることがわかる。

## 【治療下ストラテジーについての考察終わり】

中間事象が発生しなかった場合の治療効果に関心を持つ仮想ストラテジーは、ほとんどのパンデミックと関連する中間事象に対して自然な選択である。特にパンデミックを理由として試験治療が中断された場合の中間事象に適用される。このような被験者については、治療へのアクセスが中断されていない類似の被験者と同様の方法で試験を継続したという仮想的な状況を考えることが合理的である。仮想的な状況としては、試験治療を遵守する被験者と、パンデミックと関連しない理由で試験治療を遵守しない被験者が混在している状況を含むであろう。推定方法は estimand の検討には含まれないが、利用可能なデータに基づいて、仮想ストラテジーの下で安定した方法で治療効果を推定できることは当然と考えるべきではなく、estimand の定義を最終化する際に推定手法の性能を評価すべきである。

## 【4.1との関連：仮想ストラテジーについての考察】

仮想ストラテジーに関して、4.1において、いくつかの仮想の例を説明した。中間事象が発生しなかった場合の治療効果を考えるとき、仮想ストラテジーのうちの一つは、中間事象が評価を妨げなかった場合に、どのような結果が得られたであろうかを推定することであると考えられる。具体的には、たとえ一部の被験者においてパンデミックを理由として試験治療が中断されたとしても、すべての被験者が治療を遵守したと仮定した場合の、ランダム化されたすべての被験者における平均因果効果が考えられる。この場合、平均因果効果は、ランダム化された治療に対して制御された直接効果と考えることができ、4.1において、以下の式で計算されると説明した。

$$ATE = E[Y(1, 1) - Y(0, 0)]$$

中間事象後の治療は、中間事象に関わらずある水準で、ここではランダム化された治療と同じになるように制御又は固定されるという意味で「制御」されると表現している。

## 【仮想ストラテジーについての考察終わり】

COVID-19による死亡の中間事象については慎重に検討する必要がある。適切なストラテジーは試験の対象疾患や臨床評価項目によって異なる。死亡が評価項目の構成要素とならない死亡率の低い疾患領域では、COVID-19の感染に関連する死亡に対する仮想ストラテジーが推奨される可能性がある。死亡が評価項目の一部である、より重篤な疾患を対象とした試験では、規制の観点から、治療の有益性を評価する際に2つ以上の estimand に対して関心を持つことは避けられない。COVID-19と関連する死亡数が少ない場合や、パンデミックの影響を治療効果の推定値に反映させたい場合には、結果に COVID-19と関連する死亡を含める実用的な方法、すなわち複合ストラテジーを用いる方法が適している。COVID-19の感染に関連する死亡に対して仮想ストラテジーを利用することは、COVID-19が存在しない状況（例えば、当該疾患（COVID-19を含む、治験実施中にパンデミックを起こすような感染症）が根絶された場合や有効な治療法が出現した場合など）での治療の有益性を評価する上で重要である。このような試験には高齢、虚弱、又は免疫不全の被験者がしばしば含まれており、死亡が COVID-19によって引き起こされた

ものであるか、又は被験者が COVID-19に感染した状態で死亡したかを判断することが困難となる可能性があることが認識されている。

治療下ストラテジー、複合ストラテジー、及び主要層ストラテジーは主要な estimand としては関心がないかもしれないが、COVID-19の感染と転帰に基づいて層別化された被験者の部分集団、又は COVID-19に対して投与される併用治療について検討する科学的根拠がある場合には、補足的な estimand として関心を持たれるかもしれない。

#### 4.2.3.2 その他（関心のある治療条件、対象集団、評価項目、治療効果の集団レベルでの要約）の取扱い

一般的に、関心のある治療条件は当初の意図したものと同じであるべきである。しかしながら、運用上、パンデミックと関連する理由により、治療の実施方法を変更する必要があるかもしれない。試験治療のアドヒアランスや併用薬に伴うパンデミックと関連する合併症は、中間事象として考慮し、適切なストラテジーで対処すべきである。

主要な試験目的に沿うために、対象集団は当初の計画通りに維持すべきであり、パンデミックのためだけに変更すべきではない。

評価項目は一般的に当初の計画通りに維持されるべきである。COVID-19による死亡などのパンデミックと関連する中間事象を複合ストラテジーで扱う場合は、評価項目の定義を調整する必要があるかもしれない。ランダム化から治療開始までの間に多数の遅延がある場合、評価項目がランダム化日から相対的に定義されている（例えば、イベントまでの時間が評価項目）場合には、治療開始日を評価開始日として再定義することを検討してもよい。しかし、非盲検試験の状況では、これは好ましくないかもしれない。

各治療法の結果及び治療法間の比較を記述した集団レベルの要約は、一般的には変更されるべきでない。まれな状況では、例えば、当初計画されていた評価項目が数値であり、COVID-19の死亡に対して生存した被験者の結果よりも悪いランク付けをする複合ストラテジーが使用される場合などは、要約指標を変更する必要があるかもしれない。この場合、要約指標を平均値から中央値に変更するかもしれない。ログランク検定の妥当性は、生存確率が試験の初期と後期に組み入れられた被験者で同じであるという仮定、及びイベントが記録された時間に起こったという仮定に依存している。このような仮定は評価する必要があるかもしれない。評価した結果によっては、代替の要約指標を用いた感度分析的な estimand が検討される。

#### 4.2.3.3 感度分析

パンデミックが試験に与える影響を評価するために、追加の感度分析が必要となる場合がある。パンデミックに関連しない中間事象や欠測データについては、事前に計画していた感度分析を実施すべきである。パンデミックに関連するものを含むすべての中間事象及び欠測データに対して、事前に計画していた感度分析を単純に適用することは問題があるかもしれない。

COVID-19によるパンデミックが発生していなかった場合における治療効果の推定には、パンデミックとの関連性により異なるストラテジーが必要となる可能性がある。COVID-19に関連した中間事象では MCAR 又は MAR の仮定が妥当であることが多いが、この仮定は他の欠測デー

タにはあてはまらない可能性がある。パンデミックと関連する欠測データが多い場合、欠測データに対する感度分析が過度に保守的になる可能性がある。欠測データが比較的多いということは、通常であれば、試験のデザインや実施方法に問題があるという指標になり得て、試験結果の不確実性を高めることにつながってしまうが、パンデミックに起因する欠測データが多い場合は、この前提は弱くなる。

パンデミックが試験結果に与える影響を理解することが重要になる。そのための方法として以下の三つが考えられる。

方法1：パンデミックと関連した中間事象の要因を説明する情報の使用

収集されたデータは、パンデミックと関連した中間事象と欠測データの定義、これらを取り扱うストラテジーで考慮される。パンデミックの影響が適切に調整されるため、多くの点で理想的なアプローチであり、臨床試験で避けられない変動を統計解析で調整するための標準的なアプローチである。しかし、詳細なデータを収集する必要がある、状況によっては実施可能ではない場合がある。

方法2：パンデミック期間（例：パンデミック前／期間中／パンデミック後）の使用

パンデミック期間を正確かつ精度高く定義する必要があり、パンデミック期間を定義する根拠を文書化しなければならない。単一国における試験であれば、パンデミックの影響や対策が被験者間で比較的同質であり適用するのが簡単である。しかし、パンデミックは緩やかに進展する可能性があり、パンデミック期間の定義が複雑になり、被験者に与える影響は一様ではない可能性が高い。さらに、感染症のアウトブレイクにはいくつかの波があるかもしれない。国際共同試験となると、パンデミック期間の定義及び被験者に与える影響は地域により異なる可能性があるため、困難になる可能性がある。地域ごとにCOVID-19の状況（被験者数、死亡者数）又は地域ごとの対策時期に基づく標準化された実用的な定義が必要となる。パンデミック期間が定義されれば、異なるパンデミック期間における来院に対する時間依存性の変数をモデルや中間事象の定義に適切に組み込むことができる。状況は急速に変化しており、現時点では、パンデミック期間の定義、実施、解釈に関する決定的な推奨事項を提供することはできない。

方法3：各被験者を治療や収集された評価に対する影響の程度に応じて分類（治験実施計画書からの逸脱の詳細、中間事象、評価の欠測、パンデミックと関連した中止の理由など）

追跡期間が固定され、追跡不能が少ない試験の場合、例えば部分集団を定義した標準的な部分集団解析ができる。この問題に対する単一のアプローチを支持するには十分なエビデンスがない。最初の二つのアプローチは、規制上のガイダンスの対象となっている。パンデミックがどのように発展するかなどを見るまでは、パンデミックの影響を複数の方法で検討するのが賢明である。

#### 4.2.4 まとめ

COVID-19の影響で治験実施計画書からの逸脱や試験実施手順の変更が生じたことによって、臨床試験のデータへさまざまな影響をもたらし、重要な統計的問題も発生している。しかし、特にパンデミック前に開始された、COVID-19に関連しない治療に対する進行中の臨床試験については、当初の主要な目的を計画通りに維持すること、すなわちパンデミック関連の混乱によ

って交絡されない治療効果を estimand とすることを、上記したデータサイエンス部会声明及び *Statistics in Biopharmaceutical Research* 誌に公表された論文ともに推奨している。本論点に関して、現時点では COVID-19の影響によって当初の計画通りに試験を実施・運営することは困難であったとしても、試験の主要な目的としては当初計画したランダム化比較試験の結果から導かれるものを維持するという点で、観察研究データに基づく因果推論における、仮想的なランダム化比較試験 (the target trial<sup>46</sup>) を模倣する試みと重ね合わせて考えることができると考えられる。観察研究データの統計解析において、どのような target trial を模倣しようとしているのかは重要な問題であり、その target trial における estimand にアプローチするために因果推論が用いられると解釈することもできる。COVID-19の影響を受けた臨床試験では、治療条件が関心のあつたものとは異なってしまったり、追跡の方法が変わってしまったり、パンデミック関連の混乱によって想定外の間接事象が起こったりするかもしれない。そのような場合においても、当初計画したランダム化比較試験が target trial と考えられることが多く、当初の計画通りの estimand を維持すること、つまりパンデミックと関連しない間接事象に対する戦略は変更せず、多くのパンデミックと関連する間接事象に対しては当該間接事象が発生しなかった場合の治療効果に関心を持つことの妥当性は、target trial の考え方からも考察することが可能である。

また、transportability の考え方から考察することも可能である。Transportability とは、Hernán and Robins<sup>46</sup>によると、一つの集団 (例えば、COVID-19の影響を受けた集団) で推定した平均因果効果を、他の集団 (例えば、当初の計画通りの集団) での意思決定に用いることを意図したときに用いる概念である。一つの集団で関心のある治療条件の平均因果効果を正しく推定できた場合、その効果が他の集団でも同じかどうかは両方の集団の特徴に依存するとされており、効果修飾因子の分布、集団内における個人間の干渉の程度 (個体間の独立性)、及び関心のある治療条件が well-defined かどうか (例えば、ある検査がある閾値以下のときレスキュー治療を許容する治療条件に関心がある状況で、当初の計画通りの集団では閾値を下回ってすぐにレスキュー治療を受けた一方、COVID-19の影響を受けた集団では閾値を下回った後一定時間経過後にレスキュー治療を受けていた場合には、transportability が低くなるは低いかもしれない) の三つに依存するとされている。これらの三つの要因が COVID-19の影響を受けうるのかどうかを考察することによって、transportability を検討することが可能かもしれない。

以上より、COVID-19の影響を受けた臨床試験において間接事象などの取扱いを検討する際に、本4.2で紹介した因果推論の考え方や概念が、本来推定すべき estimand を特定するために有用である可能性があると考えられる。

## 5. 参考文献

- <sup>1</sup> International Conference on Harmonisation Expert Working Group. *The common technical document for the registration of pharmaceuticals for human use: efficacy E9(R1)*, [https://database.ich.org/sites/default/files/E9-R1\\_Step4\\_Guideline\\_2019\\_1203.pdf](https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf) (accessed 23 July 2021).
- <sup>2</sup> National Research Council (NRC). *The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials*, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press, 2010.
- <sup>3</sup> Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiological research. *Epidemiology* 1999; 10: 37-48.
- <sup>4</sup> Pearl J. Causal inference in the health sciences: A conceptual introduction. *Health Serv Outcomes Res Methodol* 2001; 2: 189-220.
- <sup>5</sup> Rubin DB. Bayesian inference for causal effects: The role of randomization. *Ann Stat* 1978; 6: 34-58.
- <sup>6</sup> 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. *Math Model* 1986; 7: 1393-1512 (errata in *Comput Math Appl* 1987; 14: 917-921).
- <sup>7</sup> 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. 審査資料で欠測データが議論された事例, [https://www.accessdata.fda.gov/drugsatfda\\_docs/nda/2014/202293Orig1s000StatR.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/202293Orig1s000StatR.pdf) (accessed 9 March 2022).
- <sup>8</sup> U.S. Department of Health and Human Services Food and Drug Administration. STATISTICAL REVIEW(S), [https://www.accessdata.fda.gov/drugsatfda\\_docs/nda/2014/202293Orig1s000StatR.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/202293Orig1s000StatR.pdf) (accessed 9 March 2022).
- <sup>9</sup> Robins JM. Correction for non-compliance in equivalence trials. *Stat Med* 1998; 17: 269-302.
- <sup>10</sup> Rubin DB. More powerful randomization-based P-values in double-blind trials with noncompliance (With Discussion). *Stat Med* 1998; 17: 371-389.
- <sup>11</sup> Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974; 66: 688-701.
- <sup>12</sup> 岩崎学. 統計的因果推論. 朝倉書店, 2015.
- <sup>13</sup> 宮川雅巳. 統計的因果推論—回帰分析の新しい枠組み—. 朝倉書店, 2004.
- <sup>14</sup> 田中司朗. 医学のための因果推論の基礎概念. 計量生物学 2019; 40: 35-62.
- <sup>15</sup> Van der Lann MJ, Rose S. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer Series in Statistics. New York, NY: Springer-Verlag; 2011
- <sup>16</sup> Robins JM. Addendum to “A new approach to causal inference in mortality studies with sustained exposure periods -application to control of the healthy worker survivor effect”. *Comput Math Appl* 1987; 14: 923-945 (errata in *Comput Math Appl* 1987; 18: 477).
- <sup>17</sup> Swanson SA, Hernán MA, Miller M, et al. Partial identification of the average treatment effect using instrumental variables: review of methods for binary instruments, treatments, and outcomes. *J Am Stat*

- Assoc* 2018; 113: 933-947.
- <sup>18</sup> 篠崎智大. 東京理科大学 因果推論講義資料 第8回操作変数法. 2020.
- <sup>19</sup> Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; 17: 360-372.
- <sup>20</sup> Richardson TS, Robins JM. Single world intervention graphs (SWIGs): A unification of counterfactual and graphical approaches to causality. Working Paper Number 128, 2013 Available at <http://www.csss.washington.edu/Papers/>.
- <sup>21</sup> Farmer RE, Kounali D, Walker AS et al. Application of causal inference methods in the analyses of randomised controlled trials: a systematic review. *Trials* 2018; 19: 23.
- <sup>22</sup> Robins JM, Hernán M, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11: 550-560.
- <sup>23</sup> Hernán M, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; 11: 561-570.
- <sup>24</sup> Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008; 168 :656-664.
- <sup>25</sup> Robins JM. Association, causation, and marginal structural models. *Synthese* 1999; 121: 151-179.
- <sup>26</sup> Magnusson BP, Schmidli H, Rouyrre N et al. Bayesian inference for a principal stratum estimand to assess the treatment effect in a subgroup characterized by postrandomization event occurrence. *Stat Med* 2019; 38: 4761-4771.
- <sup>27</sup> Carpenter B, Gelman A, Hoffman MD et al. Stan: a probabilistic programming language. *J Stat Softw* 2017; 76: 1-32.
- <sup>28</sup> Cree B, Magnusson B, Rouyrre N et al. Disentangling treatment effects on disability and relapses: analysis of siponimod in secondary progressive multiple sclerosis. 2019. In preparation.
- <sup>29</sup> Permutt T. A taxonomy of estimands for regulatory clinical trials with discontinuations. *Stat Med* 2016; 35: 2865-2875.
- <sup>30</sup> Akacha M, Bretz F, Ruberg S. Estimands in clinical trials-broadening the perspective. *Stat Med* 2017; 36: 5-19.
- <sup>31</sup> Lipkovich I, Ratitch B, Mallinckrodt CH. Causal inference and estimands in clinical trials. *Stat Biopharm Res* 2020; 12, 54-67.
- <sup>32</sup> 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. COVID-19が治験に及ぼす影響についてー統計解析・データマネジメントの視点ー, [http://www.jpma.or.jp/medicine/shinyaku/tiken/message/pdf/covid-19\\_statement.pdf](http://www.jpma.or.jp/medicine/shinyaku/tiken/message/pdf/covid-19_statement.pdf) (accessed 23 July 2021).
- <sup>33</sup> Meyer RD, Ratitch B, Wolbers M et al. Statistical issues and recommendations for clinical trials conducted during the COVID-19 pandemic. *Stat Biopharm Res* 2020; 12, 399-411.
- <sup>34</sup> Daniel RM, Cousens SN, De Stavola BL et al. Methods for dealing with time-dependent confounding. *Stat Med* 2013; 32, 1584-1618.
- <sup>35</sup> VanderWeele TJ, Vansteelandt, S. Conceptual issues concerning mediation, interventions and composition.

- Stat Interface* 2009; 2: 457-468.
- <sup>36</sup> Chakraborty B, Moodie EEM. *Statistical methods for dynamic treatment regimes* New York: Springer-Verlag, 2013.
- <sup>37</sup> Tsiatis AA, Davidian M, Holloway ST et al. *Dynamic treatment regimes statistical methods for precision medicine* Boca Raton: Chapman & Hall/CRC Press, 2019.
- <sup>38</sup> Lieberman JA, Stroup TS, McEvoy JP et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New Eng J Med* 2005; 353: 1209-1223.
- <sup>39</sup> Shortreed SM, Moodie EEM. Estimating the optimal dynamic antipsychotic treatment regime: Evidence from the sequential multiple assignment randomized CATIE schizophrenia study. *J R Stat Soc Ser C Appl Stat* 2012; 61: 577-599.
- <sup>40</sup> Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996; 91: 444-455.
- <sup>41</sup> Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann Stat* 1997; 25, 305-327.
- <sup>42</sup> Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by ‘death’. *J Educ Behav Stat*, 2003; 28: 353-368.
- <sup>43</sup> Hayden D, Pauler DK, Schoenfeld D. An estimator for treatment comparisons among survivors in randomized trials. *Biometrics* 2005; 61: 305-310.
- <sup>44</sup> Imai K. Sharp bounds on causal effects in randomized experiments with ‘truncation-by-death’. *Stat Probab Lett* 2008; 78: 144-149.
- <sup>45</sup> Mehrotra D, Li X, Gilbert P. A comparison of eight methods for the dual-endpoint evaluation of efficacy in a proof-of-concept HIV vaccine trial. *Biometrics* 2006; 62: 893-900.
- <sup>46</sup> Hernán MA, Robins JM. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.
- <sup>47</sup> 佐藤俊哉. 疫学研究における交絡と効果の修飾. *統計数理* 1994; 42: 83-101.
- <sup>48</sup> Bang H and Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics* 2005; 61: 962-973.
- <sup>49</sup> Kang JDY and Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with discussion). *Stat Sci* 2007; 22: 523-580.
- <sup>50</sup> Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Commun Stat Theory Methods* 1991; 20: 2609-2631.

**執筆者・タスクフォースメンバー**

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  
2020年度継続タスクフォース1 因果推論サブチーム

**【サブチームメンバー（本稿執筆者，50音順）】**

大浦 智紀	日本イーライリリー株式会社	担当：4.1
小谷 基 (~2021/9/16)	小野薬品工業株式会社	担当：2.2, 3.1, 4.1, 4.2
尾崎 凌斗	中外製薬株式会社	担当：2.6, 4.1
黒田 晋吾	武田薬品工業株式会社	担当：4.2
竹ノ内 一雅	日本イーライリリー株式会社	担当：3.2
土川 克	ゼリア新薬工業株式会社	担当：4.1, 2.5
手島 梨恵	ノバルティスファーマ株式会社	担当：2.3, 4.2
寺田 亘	アステラス製薬株式会社	担当：4.1, 2.4
中西 豊支	興和株式会社	担当：2.1, 3.2, 4.1
山崎 広徳	大正製薬株式会社	担当：4.2, 2.5
渡邊 大丞	サノフィ株式会社	担当：4.2

**【サブチームリーダー】**

小谷 基 (~2021/9/16)	小野薬品工業株式会社
中西 豊支 (2021/9/17~)	興和株式会社

**【タスクフォースリーダー，推進委員】**

竹ノ内 一雅	日本イーライリリー株式会社
--------	---------------

**【担当副部長】**

菅波 秀規	興和株式会社
土屋 悟	住友ファーマ株式会社