



解釈のための因果推論

～因果媒介分析編～

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
2023年度タスクフォース3-1

Ver 1.0
2024年6月

目次

略語の定義	4
1. はじめに	5
1.1 背景と目的	5
1.2 本稿の構成	7
2. 回帰モデルを用いた古典的な因果媒介分析	8
3. 反事実フレームワークを用いた因果媒介分析	10
3.1 効果の定義と識別条件	10
3.2 推定方法	14
3.2.1 連続値アウトカム	14
3.2.2 2値アウトカム	16
3.2.2.1 リスク差からオッズ比への拡張	16
3.2.2.2 アウトカムの発生頻度及び媒介変数の種類別の導出	18
3.2.2.2.1 アウトカム (Y) がまれであり、媒介変数 (M) が連続量の場合	18
3.2.2.2.2 アウトカム (Y) がまれでなく、媒介変数 (M) が連続量の場合	19
3.2.2.2.3 アウトカム (Y) がまれであり、媒介変数 (M) が2値の場合	19
3.2.2.2.4 アウトカム (Y) がまれでなく、媒介変数 (M) が2値の場合	20
3.2.2.3 交互作用について	20
3.2.2.3.1 アウトカム (Y) がまれであり、媒介変数 (M) が連続量の場合	20
3.2.2.3.2 アウトカム (Y) がまれであり、媒介変数 (M) が2値の場合	21
3.2.2.4 因果媒介分析におけるロジスティック回帰モデルの特性	21
3.2.2.4.1 非線形性及び交互作用	21
3.2.2.4.2 Non-collapsibility	22
3.2.2.4.3 媒介変数と説明変数の影響	22
3.2.2.4.4 注意点と課題	22
3.2.3 Time-to-event アウトカム	23
3.2.3.1 生存時間アウトカムにおける効果指標	23
3.2.3.2 AFT モデルを用いた因果媒介分析	24
3.2.3.3 PH モデルを用いた因果媒介分析	25
3.3 感度分析	26
3.3.1 感度分析パラメータ γ	27
3.3.2 感度分析パラメータ λ	27
3.3.3 Bounding factor B	28
3.3.4 感度分析の報告	28
3.3.5 2値アウトカム以外の場合	29
3.4 媒介変数への介入効果指標	29
4. 解析コード	31

4.1	SAS	31
4.1.1	公開マクロ	31
4.1.2	プロシージャ	32
4.1.3	適用事例	34
4.1.3.1	シミュレーションデータでの媒介分析の適用	35
4.1.3.1.1	シミュレーションデータの生成	35
4.1.3.1.2	回帰モデルを用いた古典的な媒介分析の適用	38
4.1.3.1.3	公開マクロ%mediation の適用	41
4.1.3.1.4	PROC CAUSALMED の適用	45
4.1.3.2	SAS help 事例紹介	50
4.2	R	52
5.	付録	58
5.1	Cross-world conditional independence	58
5.2	2値アウトカムに対する推定方法の詳細	58
5.2.1	アウトカム (Y) がまれであり, 媒介変数 (M) が連続量の場合	58
5.2.2	アウトカム (Y) がまれでなく, 媒介変数 (M) が連続量の場合	61
5.2.3	アウトカム (Y) がまれであり, 媒介変数 (M) が2値の場合	61
5.2.4	アウトカム (Y) がまれな場合の Difference Method と Product Method の近似	63
5.3	AFTモデルにおけるNIEとNDE	64
5.4	PHモデルにおけるNIEとNDE	65
6.	参考文献	68
	執筆者・タスクフォースメンバー	71

図の目次

図 1-1	直接効果 (赤線) と間接効果 (青線) への分解を示した DAG	6
図 1-2	媒介変数 (M) とアウトカム (Y) 間の交絡変数 (L) が存在する場合の DAG	6
図 3-1	因果媒介分析における識別条件を検討するための DAG	11
図 3-2	媒介変数-アウトカム間の未測定交絡因子 U	27
図 4-1	シミュレーションデータで想定する DAG	35

表の目次

表 1-1	データサイエンス部会の主な因果推論に関連した報告書	5
表 4-1	SAS マクロの引数とその説明	31
表 4-2	Cognitive データセットにおける TE, DE, IE	52
表 4-3	子供の認知テストのスコアに対する回帰モデルの推定値	52
表 4-4	子供の学習意欲スコアに対する回帰モデルの推定値	52

略語の定義

略語	定義	初出
DAG	Directed Acyclic Graph	1章
IE	Indirect Effect	1章
DE	Direct Effect	1章
TE	Total Effect	1章
NDE	Natural Direct Effect	3章
NIE	Natural Indirect Effect	3章
CDE	Controlled Direct Effect	3章
PDE	Pure Direct Effect	3章
PIE	Pure Indirect Effect	3章
TDE	Total Direct Effect	3章
TIE	Total Indirect Effect	3章
AFT	Accelerated Failure Time	3章
PH	Proportional Hazard	3章
PM	Proportion Mediated	3章
PE	Proportion Eliminated	3章
OR	Odds Ratio	3章
RR	Risk Ratio	3章
PI	Proportion due to Interaction	4章

1. はじめに

1.1 背景と目的

ICH E9(R1)により estimand の概念が広まったことに伴い、推定の対象を明確に定義したうえで効果を推定するために、Real World Data を利用した観察研究のみならず、介入を伴う臨床研究においても因果推論の枠組みを用いて estimand を議論する機会が増えている。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ではこれまで、estimand に対する理解が深まることを期待して作成された因果推論に関連する成果物として3つの報告書を公開している（表 1-1）。

表 1-1 データサイエンス部会の主な因果推論に関連した報告書

公開年月	2017年5月	2022年7月	2022年9月
タイトル	臨床試験の estimand を理解するための因果推論	ICH E9(R1)の理解に役立つ因果推論	ICH E9(R1)の理解に役立つ因果推論～時間依存性治療編～
概要	因果推論の枠組みを用いて estimand の議論を整理し、いくつかの状況における臨床試験の estimand について解説している。因果推論の解析方法の詳細については扱っていない。	Estimand を実装した臨床試験の計画と解析へのつながりを意識し、因果推論の基本的事項、解析方法及び臨床試験データへの因果推論の適用事例について説明している。	Estimand の枠組みで定義され得る時間依存性治療レジメンへの理解促進を期待し、時間依存性治療に対する因果推論の基本的概念理解、推定方法、適用事例、SAS コードについて紹介している。

介入を伴う臨床研究では治療（介入）とアウトカムの因果関係に焦点が当てられることが多く、上記の報告書は、関心のある治療とアウトカムとの因果関係の評価を目的とした因果推論の考え方を示している。しかし、これらの報告書では治療がアウトカムに対してどのようなメカニズムで影響を与えているかは評価の対象外であった。

臨床研究においては、治療とアウトカムの因果関係が主要な関心事である一方で、その因果関係のメカニズムや経路、治療とアウトカムの関係の間にある変数（媒介変数）の影響にも興味がある場合がある。薬剤開発の文脈では、薬力学マーカーや代替評価指標が媒介変数として考えられる。例えば、骨粗しょう症治療薬のランダム化比較試験では、2年後の骨折有無に対する治療効果だけでなく、治療によって変化した1年後の骨量の変化が2年後の骨折に与える影響についても関心事となる。このような場合、因果媒介分析が活用できる。因果媒介分析では、治療による骨折抑制効果を総合効果（Total Effect ; TE）と呼び、治療によって骨量が増加したことによる骨折抑制効果を意味する間接効果（Indirect Effect ; IE）と、骨量の変化による効果を除いた治療の骨折抑制効果を意味する直接効果（Direct Effect ; DE）に TE を分解することを考える。治療変数 A、媒介変数 M（上記の例における1年後の骨量に相当）、及びアウトカム変数 Y（上記の例における2年後の骨折有無に相当）の関係を図式的に示した [図 1-1](#) の Directed Acyclic Graph (DAG) では、青線が IE を、赤線が DE を指し、その和が TE に相当する。



[Hernán and Robins (2020) Figure 23.1を改変]

図 1-1 直接効果（赤線）と間接効果（青線）への分解を示した DAG

図 1-1を見て分かるように、因果媒介分析は時点の異なる二つの変数（治療 A と媒介変数 M）がアウトカム Y に与える影響を考えるため、2022年9月に公開された“ICH E9(R1)の理解に役立つ因果推論～時間依存性治療編～”（JPMA DS 部会 2022b）で取り扱った時間依存性治療に対する因果推論の特殊ケースとも考えられる。理想的なランダム化比較試験では治療 A とアウトカム Y 間に交絡は存在しないが、媒介変数 M への介入はできないため図 1-2の共変量 L のように媒介変数 M とアウトカム Y 間に交絡変数（ $M \leftarrow L \rightarrow Y$ ）が存在する可能性がある。このとき、媒介変数 M が合流点（ $A \rightarrow M \leftarrow L$ ）となっており、上記のような因果効果の分解を共変量 L の存在を考慮せずに媒介変数 M に対する単純な部分集団解析によって試みると選択バイアスを誘発するため、たとえランダム化比較試験であっても本稿で紹介するような因果推論手法が必要となる。また、これまでデータサイエンス部会で紹介してきた手法と同様に、各手法で推定可能な効果とそれぞれの手法に求められる条件や仮定を理解することが重要となる。

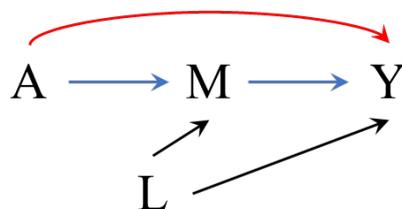


図 1-2 媒介変数（M）とアウトカム（Y）間の交絡変数（L）が存在する場合の DAG

因果媒介分析は統計的観点から見ても比較的高度な内容となるが、その適用にあたっては DAG の特定（特にランダム化比較試験では図 1-2の L）が推定の肝となり、臨床担当者には各変数間の関連について理解していることが求められる。そして統計担当者には、因果媒介分析に必要な仮定とその下での解釈性を十分に把握し、臨床担当者との協議を通じて各適用事例における解釈を適切に提示できることが求められる。そこで本稿では、統計担当者を対象に、因果媒介分析における基本的事項を整理し、アウトカムの種類に応じた推定方法を紹介する。加えて、因果媒介分析ではより強い仮定を伴うことから、その仮定に対する感度分析手法も紹介する。更に、理論の理解を助けるために、数値例と解析コードを提示し、読者が手元で実行し理解を深めることを期待する。

なお、本稿は表 1-1に示した公開済みの成果物の内容を前提とし、因果媒介分析の基本的事項の紹介に焦点を当て、媒介変数が1つの状況に限定している。複数の媒介変数が存在する状

況での因果効果の分解に関心がある場合は、[VanderWeele and Vansteelandt \(2014\)](#)を参照されたい。本稿で紹介する識別条件や仮定と異なる状況下で推定する Path-specific effect ([Avin et al. 2005](#)) や Interventional effect ([VanderWeele and Daniel 2017](#)) については、より複雑な内容となるため各自文献を参照されたい。

1.2 本稿の構成

本稿の構成は以下の通りである。2章では、古典的な媒介分析手法として回帰モデルを用いた手法を、その限界とともに紹介する。3章では、2章で紹介した手法の限界に対処するために提案された、反事実フレームワークを用いた因果媒介分析について、効果の定義や識別条件とともに基本的な推定方法を提示する。3章は、主に [VanderWeele \(2016\)](#), [Valeri and VanderWeel \(2013\)](#), 及び [Hernán and Robins \(2020\)](#)を参照している。4章では、解析時に参考になるであろう SAS コードと R コードを紹介し、3章で紹介した推定方法を適用するための SAS コードを利用した事例も紹介する。5章には、3章で紹介した推定方法の数理的詳細を提示する。

2. 回帰モデルを用いた古典的な因果媒介分析

因果媒介分析には2つの古典的なアプローチが存在している。一つが **Difference Method**、もう一つが **Product Method** と呼ばれており、本節ではこれらの方法を順に解説していく ([VanderWeele 2016](#))。

Difference Method は疫学及び生物医科学の分野でより頻繁に使用されてきたが、2つの線形回帰モデルを適用することにより構成される。ここで、関心のある治療を A 、媒介変数を M 、アウトカムを Y 、ベースライン共変量を L と表記する。**Difference Method** における1つ目の回帰モデルは、以下のように単純にアウトカム Y に対して治療 A と共変量 L を説明変数とした回帰モデルを当てはめる。

$$E[Y|a, l] = \phi_0 + \phi_1 a + \phi_4' l$$

係数 ϕ_1 はアウトカム Y に対する治療 A の **TE** として解釈される。2つ目の回帰モデルは、説明変数として媒介変数 M も含めること以外は同様である。

$$E[Y|a, m, l] = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' l$$

2つ目の回帰モデルの治療の係数 θ_1 (媒介変数が追加されたモデルの係数) が1つ目の回帰モデルの治療の係数 ϕ_1 (媒介変数がないモデルの係数) と比較したときに変化していれば、媒介変数の影響が示唆されると考えることができる。理由は、媒介変数が治療のアウトカムに対する効果の一部を説明していると考えられるためである。この2つの係数の差が媒介効果もしくは **IE** として解釈されることから、**Difference Method** と呼ばれている。

$$IE = \phi_1 - \theta_1$$

媒介変数を含めたモデルにおける治療の係数 θ_1 は、媒介変数を制御した後のアウトカムに対する効果と考えられるため、一般的には **DE** の指標として解釈される。

$$DE = \theta_1$$

3.1でこれらの推定値を $DE \cdot IE$ として因果的な解釈を可能にするための仮定について述べる。

もう一つの **Product Method** や **Product-of-Coefficients Method** と呼ばれる手法は、社会科学の分野でより頻繁に使用されてきたが、[Baron and Kenny \(1986\)](#) の論文以降広く用いられるようになった。**Product Method** においても同様に2つの回帰モデルが用いられる。1つ目の回帰モデルは、アウトカムを治療、媒介変数、共変量に対して回帰する。このモデルは、**Difference Method** における2つ目の回帰モデルと同様である。

$$E[Y|a, m, l] = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' l$$

2つ目の回帰モデルは、媒介変数を治療と共変量に対して回帰する。

$$E[M|a, l] = \beta_0 + \beta_1 a + \beta_2' l$$

DE は **Difference Method** と同様に θ_1 (媒介変数を含むアウトカム回帰モデルの治療の係数) で解釈される。 θ_1 は、固定した媒介変数の水準におけるアウトカムに対する効果とみることができる。

$$DE = \theta_1$$

一方、IE は媒介変数に対するモデルにおける治療の係数 β_1 とアウトカムに対するモデルにおける媒介変数の係数 θ_2 の積と考える。

$$IE = \beta_1\theta_2$$

媒介変数に対する治療の効果にアウトカムに対する媒介変数の効果を掛け合わせることで、媒介変数の水準を介したアウトカムに対する治療の効果という IE の指標として直感的に解釈できる。

この2つの方法について、どのように使い分ければよいのかという疑問が生じるのは自然であろう。幸いにも、連続量のアウトカム及び媒介変数に線形回帰モデルを当てはめて最小二乗法で推定する場合は、両方法は一致した推定値を与える。数値的には常に媒介効果として $\beta_1\theta_2 = \phi_1 - \theta_1$ となる。しかしながら、2値アウトカムのロジスティック回帰モデルの場合は、数値的に同一の結果とはならない。これらの方法についてロジスティック回帰モデルが妥当である場合について3.2.2で解説する。

媒介変数に対するこれらの古典的な方法（Difference Method 若しくは Product Method）と3章で解説する反事実ベースの方法は、いくつかの状況で一致する。線形モデル及び対数線形モデルでは、治療と媒介変数間の交互作用が存在しない場合に一致する。ロジスティックモデルでは、治療と媒介変数間の交互作用がなく、かつアウトカムがまれな場合に一致する（VanderWeele and Vansteelandt 2009, VanderWeele and Vansteelandt 2010）。したがって、解析者は、古典的な方法のいずれかを実行する場合は、(i) 治療とアウトカム間の交絡変数、媒介変数とアウトカム間の交絡変数及び治療と媒介変数間の交絡変数が制御されているかどうかを検討し、(ii) 治療と媒介変数間の交互作用があるかどうかを確認し、(iii) アウトカムが2値変数でロジスティック回帰モデルを用いる場合は、アウトカムがまれであるかどうかを確認しなくてはならない。測定されていない交絡変数（未測定交絡変数）がないという条件が満たされ、交互作用は認められず、ロジスティック回帰モデル使用時にはアウトカムがまれである（アウトカムが2値変数であるとき）場合に限り、古典的な方法を使用してもその推定値を因果的に解釈可能となる。その詳細や、曝露と媒介変数間の交互作用がある場合に使用可能な別の反事実ベースの手法については3章で解説する（Valeri and VanderWeele 2013）。

3. 反事実フレームワークを用いた因果媒介分析

3.1 効果の定義と識別条件

ここでは反事実フレームワークを用いた治療の TE の DE 及び IE への分解について触れ、各効果の定義及び識別条件について述べる。

反事実フレームワークを用いた TE は、研究対象集団全体が治療を受けた場合 ($A=1$) と受けなかった場合 ($A=0$) の差として以下のように定義される。

$$TE = E[Y^{a=1} - Y^{a=0}] = E[Y^{a=1, M^{a=1}}] - E[Y^{a=0, M^{a=0}}]$$

ここに、肩付き変数は潜在変数を表し、 $Y^{a=1}, Y^{a=0}$ はそれぞれ $A = 1$ (治療), $A = 0$ (未治療) のもとでの潜在アウトカム, $M^{a=1}, M^{a=0}$ はそれぞれ $A = 1, A = 0$ のもとでの潜在媒介変数を表している。つまり、 $Y^{a=1} = Y^{a=1, M^{a=1}}$, $Y^{a=0} = Y^{a=0, M^{a=0}}$ となる。TE は、自然な直接効果 (Natural Direct Effect ; NDE) と自然な間接効果 (Natural Indirect Effect ; NIE) に分解することができる (Robins and Greenland 1992, Pearl 2001)。

$$NDE = E[Y^{a=1, M^{a=0}} - Y^{a=0, M^{a=0}}] = E[Y^{a=1, M^{a=0}}] - E[Y^{a=0, M^{a=0}}]$$

$$NIE = E[Y^{a=1, M^{a=1}} - Y^{a=1, M^{a=0}}] = E[Y^{a=1, M^{a=1}}] - E[Y^{a=1, M^{a=0}}]$$

NDE は、 $A = 0$ (未治療) という「自然な」状態での媒介変数の値 $M^{a=0}$ のもとで治療効果を評価している。一方 NIE は、治療を $A = 1$ (治療) に固定したもとで、媒介変数が $M^{a=0}$ から $M^{a=1}$ に変化した場合の治療効果として定義される。NIE は、TE の中で、治療を受け媒介変数の値が変わることによるアウトカムへの影響を表すと解釈することができる (Hafeman and Schwartz 2009, Hafeman and VanderWeele 2011)。NDE と NIE の和は TE となり、TE が NDE と NIE へ分解されることが分かる。

$$\begin{aligned} NDE + NIE &= E[Y^{a=1, M^{a=0}}] - E[Y^{a=0, M^{a=0}}] + E[Y^{a=1, M^{a=1}}] - E[Y^{a=1, M^{a=0}}] \\ &= E[Y^{a=1, M^{a=1}}] - E[Y^{a=0, M^{a=0}}] \\ &= TE \end{aligned}$$

DE の別の定義として、制御された直接効果 (Controlled Direct Effect ; CDE) がある (Pearl 2001)。CDE は、媒介変数を $M = m$ で制御した場合の治療効果として以下のように表される。

$$CDE = E[Y^{a=1, M=m}] - E[Y^{a=0, M=m}]$$

一方で、治療がアウトカムに及ぼす直接的な影響を遮る形で、媒介変数を特定の値に固定することは不可能であり、媒介変数を制御する形で IE を定義することは困難である。とくに TE から CDE を引いたものを IE の指標として考えることは適切ではない (VanderWeele 2009)。これは、TE から CDE を引いたものは、IE の定義として必要な「治療が媒介変数に影響を及ぼさない (すべての対象者について $M^{a=1} = M^{a=0}$) 場合に、 $IE = 0$ 」という条件を必ずしも満足しないためである。

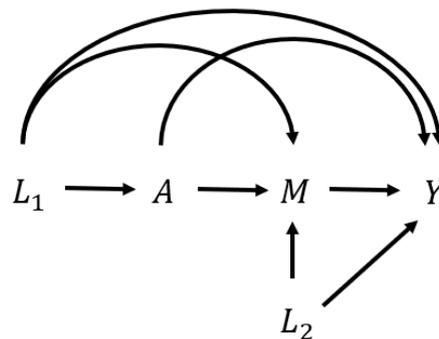
これに対して、NDE 及び NIE は、先に示した通り TE から NDE を引くことで NIE を得ることができる。しかし、NDE の定義に含まれる $E[Y^{a=1, M^{a=0}}]$ は、2つの治療 $A = 1$ 及び $A = 0$ のもとで与えられる、同一世界の同一個体では同時に観測し得ない反事実的な結果を含むため Cross-

World Quantity と呼ばれ (Hernán and Robins 2020), 現実に観察することは不可能という問題点がある (付録5.1).

次に, NDE 及び NIE を識別するための条件について述べる. NDE 及び NIE の推定値を因果効果として解釈するためには, かなり強い仮定を必要とする. まず, (A1) 治療とアウトカムの交絡の制御 (未測定交絡因子がないこと) が必要である. また, NDE 及び NIE を用いて媒介変数がアウトカムに及ぼす影響を評価するため, (A2) 媒介変数とアウトカムの交絡の制御も必要となる. さらに, 媒介分析では, 媒介変数を変化させる治療 (及びアウトカムに影響を及ぼす媒介変数の変化) に関する推論を行うため, (A3) 治療と媒介変数の交絡の制御が必要である. 最後に, (A4) 治療の影響を受ける媒介変数とアウトカムの交絡因子がない, という仮定も必要となる. これらの仮定を数式で表すと, 以下となる.

- A1: $(Y^{a,m} \perp A) \mid L, \forall a, m$ (治療とアウトカムの交絡が制御されている)
- A2: $(Y^{a,m} \perp M) \mid A, L, \forall a, m$ (媒介変数とアウトカムの交絡が制御されている)
- A3: $(M^a \perp A) \mid L, \forall a$ (治療と媒介変数の交絡が制御されている)
- A4: $(Y^{a,m} \perp M^{a=0}) \mid L, \forall a, m$ (治療の影響を受ける, 媒介変数とアウトカムの交絡因子がない)

図 3-1 に示す DAG を用いて上記の識別のための仮定を確認する.



[Valeri and VanderWeel (2013) Figure 2 を改変]

図 3-1 因果媒介分析における識別条件を検討するための DAG

DAG 中の変数は, 治療 (A), 媒介因子 (M), アウトカム (Y), 共変量 (L=(L1, L2)) であり, L1 は治療・アウトカム間および媒介変数・アウトカム間の交絡因子, L2 は媒介変数・アウトカム間の交絡因子である. 以下は, L1 が L2 に影響を及ぼす, もしくは L2 が L1 に影響を及ぼす場合も成立する. 研究者は L=(L1, L2) を慎重に検討しそのデータを収集する必要があるが, 治療の影響を受ける共変量がないことがわかっている場合は, これらすべての交絡因子を調整することが重要であり, 解析時 (モデル構築時) には治療・アウトカム間の交絡因子なのか, 媒介変数・アウトカム間の交絡因子なのかを区別する必要はないことに注意されたい.

仮定 A1 は, TE の推定を目的とする観察研究で一般に置かれる仮定である (例えば, JPMA DS 部会 2022a 2.2.2 等). 媒介分析の文脈で特異的に必要な仮定は, 仮定 A2~A4 である. これ

らの仮定は TE の推定には必要ではないが、仮定 A2は CDE を、仮定 A2～A4は NDE 及び NIE を因果的に解釈するために必要となる。注意すべき点は、CDE, NDE 及び NIE を推定するためには、治療がランダム化されていたとしてもこれらの仮定が必要ということである。これは、ランダム化比較試験では治療をランダム化することにより仮定 A1及び A3が満たされるが、一般に媒介変数はランダム化されることはないためである。ランダム化比較試験であったとしても、媒介変数とアウトカムの交絡が制御されない場合は、CDE, NDE 及び NIE の推定値は因果的解釈をもたず、大きなバイアスを伴う恐れがある。媒介分析が興味の対象である場合、媒介変数とアウトカムの交絡を引き起こし得る因子を慎重に検討してそのデータを収集し、交絡を制御しなければならない。未測定交絡因子の問題が懸念される場合には、3.3で紹介するような感度分析 (VanderWeele 2010, Imai et al. 2010) を実施する必要がある。なお、仮定 A1～A3には、時間的順序の仮定を前提としている点にも注意が必要である。治療・アウトカム間の未測定交絡因子がないという仮定は、治療がアウトカムに対して時間的に先行することを暗に仮定している。また、媒介変数・アウトカム間の未測定交絡因子がないという仮定は、媒介因子がアウトカムに対して時間的に先行することを暗に仮定している。同様に、治療・媒介因子間の未測定交絡因子がないという仮定は、治療が媒介因子に対して時間的に先行することを暗に仮定している。これらの時間的順序の仮定が満たされない場合は、結果を因果的に解釈することはできない。

仮定 A4は Cross-world conditional independence とも呼ばれ (Hernán and Robins 2020, 付録5.1), 数式を直接的に解釈することは難しいが、本質的には媒介変数・アウトカムの交絡因子が治療の影響を受けないことを意味する仮定であり、図 3-1の DAG では、治療 A から媒介変数・アウトカムの交絡因子 L2に向かう矢線がないことに対応する。治療が媒介変数・アウトカムの交絡因子 L2に影響を及ぼす場合、交絡因子 L2はそれ自体も治療 A がアウトカムに及ぼす影響の媒介変数となり、同時に関心の対象である媒介変数 M のアウトカム Y への効果に対する交絡因子にもなるため、NDE 及び NIE の推定において考慮する必要が生じ、推定問題は一層複雑となる。一般に、複数の媒介変数が考えられる場合は、媒介変数全体が治療のアウトカムに対する影響を媒介する程度を評価する回帰モデルに基づく方法を検討することができる。複数の媒介変数が考えられる場合の議論の詳細については VanderWeele and Vansteelandt (2014)を参照されたい。

仮定 A1～A4のもとで、NDE 及び NIE が識別可能であることを示すために、ここでは $E[Y^{a=1, M^{a=0}}]$ の識別性を確認する。なお、 $E[Y^{a=0, M^{a=1}}]$ の識別性も同様に確認することができる。まず一致性の仮定に基づいて以下を得る。

$$\begin{aligned} E[Y^{a=1, M^{a=0}}] &= \sum_l \sum_m E[Y^{a=1, M^{a=0}} | M^{a=0} = m, l] Pr[M^{a=0} = m | l] Pr(L = l) \\ &= \sum_l \sum_m E[Y^{a=1, m} | M^{a=0} = m, l] Pr[M^{a=0} = m | l] Pr(L = l) \end{aligned}$$

ここで上式の赤字部分と青字部分は、仮定 A1～A4を用いてそれぞれ以下のように表現することができる。

(赤字部分)

$$\begin{aligned}
 E[Y^{a=1,m}|M^{a=0} = m, l] &= E[Y^{a=1,m}|l] \quad (\because A4) \\
 &= E[Y^{a=1,m}|A = 1, l] \quad (\because A1) \\
 &= E[Y^{a=1,m}|A = 1, M = m, l] \quad (\because A2) \\
 &= E[Y|A = 1, M = m, l] \quad (\because \text{一致性})
 \end{aligned}$$

(青色部分)

$$\begin{aligned}
 Pr[M^{a=0} = m|l] &= Pr[M^{a=0} = m|A = 0, l] \quad (\because A3) \\
 &= Pr[M = m|A = 0, l] \quad (\because \text{一致性})
 \end{aligned}$$

以上から、 $E[Y^{a=1,M^{a=0}}] = \sum_l \sum_m E[Y|A = 1, M = m]Pr[M = m|A = 0, l]Pr(L = l)$ を得る。これは観察データから推定可能な量であり、[Hernan and Robins \(2020\)](#)では上式を媒介公式 (Mediation Formula) と呼んでいる。媒介公式を用いて、NDE 及び NIE を以下のように表現することができる。

$$\begin{aligned}
 \text{NDE} &= E[Y^{a=1,M^{a=0}}] - E[Y^{a=0,M^{a=0}}] \\
 &= \sum_l \sum_m \{E[Y|A = 1, M = m, l] - E[Y|A = 0, M = m, l]\}Pr[M = m|A = 0, l]Pr(L = l) \\
 \text{NIE} &= E[Y^{a=1,M^{a=1}}] - E[Y^{a=1,M^{a=0}}] \\
 &= \sum_l \sum_m E[Y|A = 1, M = m, l]\{Pr[M = m|A = 1, l] - Pr[M = m|A = 0, l]\}Pr(L = l)
 \end{aligned}$$

また、CDE についても以下のように表すことができる。

$$\text{CDE} = E[Y^{a=1,M=m}] - E[Y^{a=0,M=m}] = \sum_l \{E[Y|A = 1, M = m, l] - E[Y|A = 0, M = m, l]\}Pr(L = l)$$

なお、ここでは $\text{TE} = \{E[Y^{a=1,M^{a=0}}] - E[Y^{a=0,M^{a=0}}]\} + \{E[Y^{a=1,M^{a=1}}] - E[Y^{a=1,M^{a=0}}]\}$ という TE の分解を考えたが、 $\text{TE} = \{E[Y^{a=1,M^{a=1}}] - E[Y^{a=0,M^{a=1}}]\} + \{E[Y^{a=0,M^{a=1}}] - E[Y^{a=0,M^{a=0}}]\}$ という分解も可能である。[Robins and Greenland \(1992\)](#)では、 $\text{NDE} = E[Y^{a=1,M^a}] - E[Y^{a=0,M^a}]$, $\text{NIE} = E[Y^{a=1,M^a}] - E[Y^{a=1,M^a}]$ ($a = 0, 1$)のうち、 $a = 0$ の場合を純粋な直接効果 (Pure Direct Effect ; PDE) 及び純粋な間接効果 (Pure Indirect Effect ; PIE), $a = 1$ の場合を総合直接効果 (Total Direct Effect ; TDE) 及び総合間接効果 (Total Indirect Effect ; TIE) と呼んでいる。PDE 及び PIE が、それぞれ本節で紹介している NDE 及び NIE に相当し、TE は本節で紹介したように PDE + TIE と分解されるか、あるいは TDE + PIE へ分解される。PDE + TIE と TDE + PIE の違いは、治療と媒介変数の交互作用効果がどのように分解されるかの違いと考えることができる ([Robins and Greenland 1992](#), [Hafeman and Schwartz 2009](#))。具体的には、TDE および TIE は治療と媒介変数の交互作用効果を含む「総合的な」効果を、PDE および PIE は治療と媒介変数の交互作用効

果を含まない「純粋な」効果を意味しており，上述の分解の違いは，DE と IE のどちらに治療と媒介変数の交互作用効果を含めて解釈するか，という違いとなる（田栗 2014）．さらなる詳細は VanderWeele (2013a)を参照されたい．

3.2 推定方法

本節では，アウトカム変数の種類別に推定方法を紹介する．

3.2.1 連続値アウトカム

本節では，前節で説明した潜在変数に基づく識別条件が成立する下で連続値アウトカムにおける各効果が回帰モデルにより推定可能であることを示す．また，特定の条件下では2章で紹介した古典的な方法である Product Method による結果と一致することについても示す．

VanderWeele and Vansteelandt (2009)では，反事実フレームワークにおける DE と IE がどのように Product Method をアウトカム回帰モデルにおける治療と媒介変数間の交互作用が含まれる設定に拡張できるかを示している．アウトカム，媒介変数を連続量とし，媒介変数 M に対する回帰モデルとアウトカム Y に対する回帰モデルを以下とする．

$$E[M|a, l] = \beta_0 + \beta_1 a + \beta_2' l$$

$$E[Y|a, m, l] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' l$$

3.1で述べた4つの識別条件のうちA1とA2が成立し，モデル誤設定がないとする．このとき， $a \neq a^*$ として，CDE は以下で与えられる．

$$\begin{aligned} \text{CDE} &= E[Y^{A=a, M=m} - Y^{A=a^*, M=m} | l] \\ &= E[Y|A = a, M = m, l] - E[Y|A = a^*, M = m, l] \quad (\because A1, A2) \\ &= (\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' l) - (\theta_0 + \theta_1 a^* + \theta_2 m + \theta_3 a^* m + \theta_4' l) \\ &= (\theta_1 a + \theta_3 am - \theta_1 a^* - \theta_3 a^* m) \\ &= (\theta_1 + \theta_3 m)(a - a^*) \end{aligned}$$

また，4つの識別条件A1からA4すべてが成立し，モデル誤設定がないとすると，NDE と NIE は媒介公式を用いてそれぞれ以下で与えられる．

$$\begin{aligned} \text{NDE} &= E[Y^{A=a, M^{A=a^*}} - Y^{A=a^*, M^{A=a^*}} | l] \\ &= \sum_m \{E[Y|A = a, M = m, l] - E[Y|A = a^*, M = m, l]\} Pr[M = m | A = a^*, l] \quad (\because \text{媒介公式}) \\ &= \sum_m \{(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' l) - (\theta_0 + \theta_1 a^* + \theta_2 m + \theta_3 a^* m + \theta_4' l)\} Pr[M = m | A \\ &\quad = a^*, l] \\ &= \sum_m \{(\theta_1 a + \theta_2 m + \theta_3 am) - (\theta_1 a^* + \theta_2 m + \theta_3 a^* m)\} Pr[M = m | A = a^*, l] \end{aligned}$$

$$\begin{aligned}
 &= \{(\theta_1 a + \theta_2 E[M|A = a^*, l] + \theta_3 a E[M|A = a^*, l]) \\
 &\quad - (\theta_1 a^* + \theta_2 E[M|A = a^*, l] + \theta_3 a^* E[M|A = a^*, l])\} \\
 &= \{(\theta_1 a + \theta_2(\beta_0 + \beta_1 a^* + \beta_2' l) + \theta_3 a(\beta_0 + \beta_1 a^* + \beta_2' l)) \\
 &\quad - (\theta_1 a^* + \theta_2(\beta_0 + \beta_1 a^* + \beta_2' l) + \theta_3 a^*(\beta_0 + \beta_1 a^* + \beta_2' l))\} \\
 &= \{(\theta_1 a + \theta_3 a(\beta_0 + \beta_1 a^* + \beta_2' l)) - (\theta_1 a^* + \theta_3 a^*(\beta_0 + \beta_1 a^* + \beta_2' l))\} \\
 &= (\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2' l)(a - a^*),
 \end{aligned}$$

$$\begin{aligned}
 \text{NIE} &= E \left[Y^{A=a, M^{A=a}} - Y^{A=a, M^{A=a^*}} \mid l \right] \\
 &= \sum_m E[Y|A = a, M = m, l] Pr[M = m|A = a, l] \\
 &\quad - \sum_m E[Y|A = a, M = m, l] Pr[M = m|A = a^*, l] \quad (\because \text{媒介公式}) \\
 &= \sum_m (\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \theta_4' l) Pr[M = m|A = a, l] \\
 &\quad - \sum_m (\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \theta_4' l) Pr[M = m|A = a^*, l] \\
 &= (\theta_0 + \theta_1 a + \theta_2 E[M|A = a, l] + \theta_3 a E[M|A = a, l] + \theta_4' l) \\
 &\quad - (\theta_0 + \theta_1 a + \theta_2 E[M|A = a^*, l] + \theta_3 a E[M|A = a^*, l] + \theta_4' l) \\
 &= (\theta_0 + \theta_1 a + \theta_2(\beta_0 + \beta_1 a + \beta_2' l) + \theta_3 a(\beta_0 + \beta_1 a + \beta_2' l) + \theta_4' l) \\
 &\quad - (\theta_0 + \theta_1 a + \theta_2(\beta_0 + \beta_1 a^* + \beta_2' l) + \theta_3 a(\beta_0 + \beta_1 a^* + \beta_2' l) + \theta_4' l) \\
 &= (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*)
 \end{aligned}$$

これらの結果は Product Method を治療と媒介変数間の交互作用を含むモデルに一般化したものであり、交互作用がない場合、すなわち $\theta_3 = 0$ のとき CDE と NDE はともに $\theta_1(a - a^*)$ 、NIE は $\theta_2 \beta_1(a - a^*)$ となり Product Method による結果と一致する。

上記の回帰モデルにより推定された CDE, NDE, NIE について、標準誤差を計算することができる。媒介変数 M に対する回帰モデルとアウトカム Y に対する回帰モデルにおけるパラメータ $\beta \equiv (\beta_0, \beta_1, \beta_2')'$ と $\theta \equiv (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)'$ の推定値を $\hat{\beta}$ と $\hat{\theta}$ とし、それらの共分散行列を Σ_β と Σ_θ とすると $(\hat{\beta}', \hat{\theta}')$ の共分散行列は以下となる。

$$\Sigma \equiv \begin{pmatrix} \Sigma_\beta & \text{Cov}(\hat{\beta}, \hat{\theta}) \\ \text{Cov}(\hat{\beta}, \hat{\theta}) & \Sigma_\theta \end{pmatrix} = \begin{pmatrix} \Sigma_\beta & \mathbf{0} \\ \mathbf{0} & \Sigma_\theta \end{pmatrix}$$

ここで、 $\hat{\beta}$ と $\hat{\theta}$ の共分散は $\hat{\beta}$ が $M, A, 1$ のみの関数であることから

$$\begin{aligned}
 \text{Cov}(\hat{\beta}, \hat{\theta}) &= E[\text{Cov}(\hat{\beta}, \hat{\theta}|M, A, l)] + \text{Cov}(E[\hat{\beta}|M, A, l], E[\hat{\theta}|M, A, l]) \\
 &= \mathbf{0} + \text{Cov}(\hat{\beta}, \theta^*) \\
 &= \mathbf{0}
 \end{aligned}$$

を利用した。 θ^* は θ の真値である。 CDE, NDE, NIE の標準誤差はデルタ法を用いると以下で得られる。

$$\sqrt{\Gamma\Sigma\Gamma'}|a - a^*|$$

ここで CDE のとき $\Gamma \equiv (0, 0, 0', 0, 1, 0, m, 0')$, NDE のとき $\Gamma \equiv (\theta_3, \theta_3 a^*, \theta_3 l', 0, 1, 0, \beta_0 + \beta_1 a^* + \beta_2' l, 0')$, NIE のとき $\Gamma \equiv (0, \theta_2 + \theta_3 a, 0', 0, 0, \beta_1, \beta_1 a, 0')$ となる.

次にアウトカムが連続量であり, 媒介変数が2値の場合を考える. アウトカム Y に対する回帰モデルを以下とする.

$$E[Y|a, m, l] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' l$$

媒介変数 M については, 以下のロジスティック回帰モデルとする.

$$\text{logit}\{P(M = 1|a, l)\} = \beta_0 + \beta_1 a + \beta_2' l$$

媒介変数 M が連続量のときと同様に, 4つの識別条件 A1から A4のすべてが成立し, モデル誤設定がないとすると, CDE, NDE, NIE は媒介公式を用いてそれぞれ以下で与えられる.

$$\text{CDE} = (\theta_1 + \theta_3 m)(a - a^*),$$

$$\text{NDE} = \theta_1(a - a^*) + \{\theta_3(a - a^*)\} \frac{\exp(\beta_0 + \beta_1 a^* + \beta_2' l)}{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2' l)},$$

$$\text{NIE} = (\theta_2 + \theta_3 a) \left\{ \frac{\exp(\beta_0 + \beta_1 a + \beta_2' l)}{1 + \exp(\beta_0 + \beta_1 a + \beta_2' l)} - \frac{\exp(\beta_0 + \beta_1 a^* + \beta_2' l)}{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2' l)} \right\} (a - a^*)$$

3.2.2 2値アウトカム

2値アウトカム (例えば, 治療の成功失敗, 疾病の有無, 副作用の発生又は非発生等) に対する因果媒介分析として, [VanderWeele and Vansteelandt \(2010\)](#), [Valeri and VanderWeele \(2013\)](#), [VanderWeele \(2016\)](#)にて, ロジスティック回帰モデルを用いた提案がされている. 本節では, 2値アウトカムに対する推定方法について上記を基に推定方法を紹介する. ケースコントロール研究への拡張については [VanderWeele and Vansteelandt \(2010\)](#)を参照にされたい.

3.2.2.1 リスク差からオッズ比への拡張

Y を2値アウトカムとし, M は正規分布に従う連続量の媒介変数, L を治療の影響を受けないベースライン共変量とし, A を治療の2値 ($A = 1$ が治療あり, $A = 0$ が治療なし) と仮定する.

オッズ比 (Odds Ratio ; OR) で考えるために, リスク差からの拡張を行う. $L = l$ が与えられた下でのリスク差として, 治療 (A) の有無の2群の差の TE を $\text{TE} = E[Y^{A=1}|l] - E[Y^{A=0}|l]$ と定義する.

上記は条件 $L = l$ のもとでの平均因果効果を示している. OR での条件 $L = l$ のもとでの TE は以下のとおり定義でき (以降, OR の TE は「OR^{TE}」と記載し, NDE は「OR^{NDE}」, NIE は「OR^{NIE}」と記載する), オッズ (起こる確率と起こらない確率の比) の比を示すこととなる.

$$OR^{TE} = \frac{P(Y^{A=1} = 1|l) / \{1 - P(Y^{A=1} = 1|l)\}}{P(Y^{A=0} = 1|l) / \{1 - P(Y^{A=0} = 1|l)\}}$$

次に、媒介変数 (M) を考慮に入れ、リスク差におけるNDEを示すと $NDE = E[Y^{A=1, M^{a=0}} |l] - E[Y^{A=0, M^{a=0}} |l]$ であった。ORの OR^{NDE} も同様に以下のとおり定義する。

$$OR^{NDE} = \frac{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}}{P(Y^{A=0, M^{a=0}} = 1|l) / \{1 - P(Y^{A=0, M^{a=0}} = 1|l)\}}$$

NIEも同様に定義でき、リスク差は $NIE = E[Y^{A=1, M^{a=1}} |l] - E[Y^{A=1, M^{a=0}} |l]$ であることから、ORの OR^{NIE} は以下の定義となる。

$$OR^{NIE} = \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}}$$

OR^{NDE} は、治療 (A) が直接的にアウトカム (Y) に与える影響を媒介変数を $M^{a=0}$ と固定した状況で治療の有無のORで表したものであり、 OR^{NIE} は、治療 (A) が媒介変数 (M) を通じてアウトカム (Y) に与える影響を治療あり ($A = 1$) と固定した状況での媒介変数の変動をORで表わしたものとなる。

また、リスク差において、次式のとおり、TEをNDEとDIEに分解できる性質をもっていた。

$$\begin{aligned} TE &= E[Y^{A=1} |l] - E[Y^{A=0} |l] \\ &= E[Y^{A=1, M^{a=1}} |l] - E[Y^{A=0, M^{a=0}} |l] \\ &= E[Y^{A=1, M^{a=0}} |l] - E[Y^{A=0, M^{a=0}} |l] + E[Y^{A=1, M^{a=1}} |l] - E[Y^{A=1, M^{a=0}} |l] \end{aligned}$$

ORでは、 OR^{TE} は OR^{NDE} と OR^{NIE} の積に分解する。

$$\begin{aligned} OR^{TE} &= \frac{P(Y^{A=1} = 1|l) / \{1 - P(Y^{A=1} = 1|l)\}}{P(Y^{A=0} = 1|l) / \{1 - P(Y^{A=0} = 1|l)\}} \\ &= \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=0, M^{a=0}} = 1|l) / \{1 - P(Y^{A=0, M^{a=0}} = 1|l)\}} \\ &= \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}} \times \frac{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}}{P(Y^{A=0, M^{a=0}} = 1|l) / \{1 - P(Y^{A=0, M^{a=0}} = 1|l)\}} \end{aligned}$$

上記を対数スケールで表すと、 $\log(\text{OR}^{\text{TE}}) = \log(\text{OR}^{\text{NIE}}) + \log(\text{OR}^{\text{NDE}})$ となる。TE に対する NIE の寄与度を表す指標である Proportion Mediated (PM, 詳細は3.4参照) は、 $\log(\text{OR}^{\text{NIE}}) / \log(\text{OR}^{\text{TE}})$ となる。また、もしアウトカムがまれな場合、PM は、 $\text{OR}^{\text{NDE}} \times \{\text{OR}^{\text{NIE}} - 1\} / \{\text{OR}^{\text{NDE}} \times \text{OR}^{\text{NIE}} - 1\}$ からも得ることができる。

3.2.2.2 アウトカムの発生頻度及び媒介変数の種類別の導出

以降、詳細な導出は付録5.2を参照されたい。

3.2.2.2.1 アウトカム (Y) がまれであり、媒介変数 (M) が連続量の場合

アウトカム (Y) が治療 (A) と独立なら ($Y^a \perp A | L$) なら、 OR^{TE} を識別することができ、次式の左辺の OR^{TE} を右辺のデータから推定できる。

$$\frac{P(Y^{A=1} = 1|l) / \{1 - P(Y^{A=1} = 1|l)\}}{P(Y^{A=0} = 1|l) / \{1 - P(Y^{A=0} = 1|l)\}} = \frac{P(Y = 1|A = 1, l) / \{1 - P(Y = 1|A = 1, l)\}}{P(Y = 1|A = 0, l) / \{1 - P(Y = 1|A = 0, l)\}}$$

OR^{NDE} , OR^{NIE} は、仮定 A1~A4, 線形回帰モデル及びロジスティック回帰モデルを組み合わせ、シンプルな回帰モデルベースの推定を行う。OR をリスク比に近似することから、前提として、アウトカム (Y) がまれである必要があり、カットオフ値として10%未満を想定されている (VanderWeele and Vansteelandt 2010)。

以下の2つのモデルを定義する。1つ目は、アウトカム (Y) のロジスティック回帰モデル (交互作用項なし) を考える。

$$\text{logit}\{P(Y = 1|A = a, M = m, L = l)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_4 l$$

次に媒介変数 (M) に線形回帰モデルを考える。この線形回帰モデルの誤差項は、平均0・分散 σ^2 の正規分布に従うと仮定する。

$$E[M|A = a, L = l] = \beta_0 + \beta_1 a + \beta_2 l$$

これらを満たすと、 $\text{OR}^{\text{CDE}}(m)$ は次のようになる。

$$\text{OR}^{\text{CDE}}(m) \approx \exp[\theta_1 a - \theta_1 a^*]$$

OR^{NDE} は以下となる。

$$\text{OR}^{\text{NDE}} \approx \exp[\theta_1 (a - a^*) + 0.5\sigma^2 (a^2 - a^{*2})]$$

OR^{NIE} も同様に以下になる。

$$\text{OR}^{\text{NIE}} \approx \exp[\theta_2 \beta_1 (a - a^*)]$$

続いて、 OR^{TE} は以下となる。

$$\text{OR}^{\text{TE}} \approx \exp[(\theta_1 + \theta_2 \beta_1)(a - a^*) + 0.5\sigma^2 (a^2 - a^{*2})]$$

PMについては、以下となる。

$$PM = \frac{\theta_2 \beta_1 (a - a^*)}{(\theta_1 + \theta_2 \beta_1)(a - a^*) + 0.5\sigma^2(a^2 - a^{*2})}$$

3.2.2.2.2 アウトカム (Y) がまれでなく、媒介変数 (M) が連続量の場合

アウトカム (Y) がまれではなく、アウトカムに対してもロジスティック回帰モデルではなく、対数線形回帰モデルを適用する場合でも、先の式は適用可能であり、アウトカム (Y) がカウント変数の場合にも適用することができる。この場合は、OR ではなくリスク比での解釈になる。

アウトカム (Y) が以下に従う場合を想定する。

$$Y \sim Poi(\lambda) \text{ for } \lambda = \exp(\theta_0 + \theta_1 a + \theta_2 m + \theta_4'c)$$

このとき、回帰モデルは以下のとおり定義できる。

$$\log\{E(Y|A = a, M = m, L = l)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_4'c$$

2値アウトカムに対するリスク比 (Risk Ratio ; RR) のNDEは、ポアソンのカウントアウトカムのNDEと一致し、これはNIEでも同様である (以降、RRのTEは「RR^{TE}」と記載し、NDEは「RR^{NDE}」、NIEは「RR^{NIE}」と記載する)。

$$RR^{NDE} = \exp \left[\log \left\{ \frac{E(Y^{a=1, M^{a=0}} | l)}{E(Y^{a=0, M^{a=0}} | l)} \right\} \right]$$

また、負の二項分布が過分散ポアソン分布として、平均が一致することから、カウントアウトカムが負の二項分布でモデル化される場合にも拡張できるとしている。

3.2.2.2.3 アウトカム (Y) がまれであり、媒介変数 (M) が2値の場合

ここでは、アウトカム (Y) 及び媒介変数 (M) がともに2値である場合を想定し、次のモデルを考える。また、アウトカム (Y) はまれであるという仮定を置く。

$$\text{logit}\{P(M = 1|A = a, L = l)\} = \beta_0 + \beta_1 a + \beta_2' l$$

$$\text{logit}\{P(Y = 1|A = a, M = m, L = l)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' l$$

OR^{CDE}については、媒介変数 (M) が連続量とした場合と同様に、M = mと固定した場合であることから計算可能である。OR^{NDE}は以下のとおりとなる。

$$OR^{NDE} = \left\{ \frac{\exp[\theta_1 a] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2' c])}{\exp[\theta_1 a^*] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2' c])} \right\}$$

OR^{NIE}も同様に以下のとおりとなる。

$$OR^{NDE} = \left\{ \frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2' c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a + \beta_2' c]\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta_2' c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2' c]\}} \right\}$$

以上より、OR^{TE}は以下となる。

$$OR^{TE} = \left\{ \frac{\exp[\theta_1 a] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2'c])}{\exp[\theta_1 a^*] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2'c])} \right\} \\ \times \left\{ \frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2'c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a + \beta_2'c]\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta_2'c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2'c]\}} \right\}$$

PMについても、同様に示す。

$$PM = \log \left[\frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2'c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a + \beta_2'c]\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta_2'c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2'c]\}} \right] \\ / \log \left[\frac{\exp[\theta_1 a] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2'c])}{\exp[\theta_1 a^*] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2'c])} \right] \\ \times \left[\frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2'c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a + \beta_2'c]\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta_2'c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2'c]\}} \right]$$

3.2.2.2.4 アウトカム (Y) がまれでなく、媒介変数 (M) が2値の場合

前述したアウトカム (Y) がまれでなく、媒介変数 (M) が連続量の場合と同様である。

3.2.2.3 交互作用について

アウトカム (Y) がまれであり、媒介変数 (M) が連続量の場合を想定し、仮定 A1～A4を満たすとき、以下の2つのモデルを想定した。

$$\text{logit}\{P(Y = 1|A = a, M = m, L = l)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_4'c$$

$$E[M|A = a, C = c] = \beta_0 + \beta_1 a + \beta_2'c$$

また、上記の場合、以下で近似できる。

$$OR^{NDE} \approx OR^{CDE} = \exp\{\theta_1 (a - a^*)\}$$

$$OR^{NIE} \approx \exp\{\theta_2 \beta_1 (a - a^*)\}$$

OR^{NDE}に θ_1 を用い、OR^{NIE}に $\theta_2 \beta_1$ を用いている。アウトカム (Y) について A, L のみを用いた回帰モデルを用いた方法も存在する。

$$\text{logit}\{P(Y = 1|A = a, L = l)\} = \phi_0 + \phi_1 a + \phi_4'c$$

このとき、A の係数の差である $\phi_1 - \theta_1$ が間接効果として解釈されることがある。これは A の変化が媒介変数 (M) を通じてアウトカム (Y) にどの程度提供を与えるかを示している。治療 (A) と媒介変数 (M) の交互作用がない場合、以下のとおりほぼ同等となるが、交互作用がある場合、以下は成立しない。

$$\phi_1 - \theta_1 \approx \theta_2 \beta_1$$

3.2.2.3.1 アウトカム (Y) がまれであり、媒介変数 (M) が連続量の場合

ロジスティック回帰モデルに交互作用を追加し、交互作用がある場合に拡張する。以下のロ

ジスティック回帰モデルを考える。

$$\text{logit}\{P(Y = 1|A = a, M = m, L = l)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4 l$$

上記に変更した場合，OR^{CDE}とOR^{NIE}は次の式になる。

$$\text{OR}^{\text{CDE}} = \exp\{(\theta_1 + \theta_3 m)(a - a^*)\}$$

$$\text{OR}^{\text{NIE}} \approx \exp\{(\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*)\}$$

また，OR^{NDE}は以下となる。

$$\text{OR}^{\text{NIE}} \approx \exp\{[\theta_1 + \theta_3 (\beta_0 + \beta_1 a^* + \beta_2' l + \theta_2 + \sigma^2)](a - a^*) + 0.5\theta_3^2 \sigma^2 (a^2 - a^{*2})\}$$

3.2.2.3.2 アウトカム (Y) がまれであり，媒介変数 (M) が2値の場合

OR^{NIE}とOR^{NDE}は次の式になる。

$$\text{OR}^{\text{NDE}} \approx \left\{ \frac{\exp[\theta_1 a] (1 + \exp[\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta_2' l])}{\exp[\theta_1 a^*] (1 + \exp[\theta_2 + \theta_3 a^* + \beta_0 + \beta_1 a^* + \beta_2' l])} \right\}$$

$$\text{OR}^{\text{NIE}} \approx \left\{ \frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2' c)\} \{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta_2' c)\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta_2' c)\} \{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta_2' c)\}} \right\}$$

3.2.2.4 因果媒介分析におけるロジスティック回帰モデルの特性

先に述べたとおり，アウトカム (Y) がまれであり，交互作用を考慮しない場合，以下の数式が得られ Difference Method と Product Method において，ほぼ同様の結果を得ることができる。

$$\phi_1 - \theta_1 \approx \theta_2 \beta_1$$

しかし，一般的にはロジスティック回帰モデルを用いた場合，Difference Method と Product Method の結果が同様の値になることはない。

アウトカム (Y) がまれでなく，交互作用を考慮しない場合で，対数二項モデルを用いるとき，上記の式はリスク比に対する DE 及び IE に対して有効であり，Product Method と Difference Method は一致する。

また，TE = NDE + NIE という関係も成り立たない。いままで見てきたように，OR^{TE}はOR^{NDE}とOR^{NIE}の積であり，OR^{NDE}とOR^{NIE}の合算がOR^{TE}にならないことに注意が必要である。

3.2.2.4.1 非線形性及び交互作用

Difference Method と Product Method の結果が同様の値にならない要因としては大きく2つ考えられる。一つはロジスティック回帰モデルが非線形性であること，もう一つは交互作用である。ロジスティック回帰モデルでは，説明変数の線形結合をロジスティック関数に通すことで，目的変数が特定のクラスに属する確率を計算する。ロジスティック関数は S 字型の曲線を描き，説明変数と目的変数との関係が非線形となる。その出力は0から1の間の値を取ることから，確率としての解釈は容易である。この非線形性により，線形回帰モデルで可能な加法性が必ずし

も成り立たず、一致しない原因となる。

また、交互作用を考慮した場合、ある説明変数が他の説明変数の値によって変わるため、各説明変数が独立に目的変数に影響しているという加法性を満たせず、必ずしも一致しない原因となる。これは線形モデルでも同様である。

3.2.2.4.2 Non-collapsibility

ロジスティック回帰モデルが OR を用いるために生じる non-collapsibility についても注意が必要である。

Non-collapsibility とは、ここでは、周辺オッズ比と共有オッズ比（条件付きオッズ比の重み付き平均）と一致しないということである。

線形回帰モデルでは、目的変数に影響を与えない説明変数をモデルに追加した場合、その係数はゼロに近くなる場合がある。その結果、説明変数が大きく変動しても目的変数には影響を与えず、その説明変数をモデルから除いても、他の説明変数の係数やモデル全体の適合度にはほとんど影響しない。

一方、ロジスティック回帰モデルの場合、同じく目的変数に影響を与えない説明変数をモデルに追加した場合、その係数はゼロに近い値だったとしても、その変数をモデルから除いたときの周辺オッズ比と共有オッズ比が異なってしまう可能性があり、これを non-collapsibility という。ロジスティック回帰モデルに共変量を追加すると（交絡因子ではない場合でも）、係数の大きさが増加する傾向があるために発生する（Greenland 2021a, Greenland 2021b）。

この性質は、まれでない場合にオッズ比はリスク比とは異なる値を示すため、まれでない場合に問題となる。リスク比は「collapsibility」を持つとされており、周辺リスク比と共有リスク比は直接比較可能である。（Hernán and Robins 2020, 田中 2019）

オッズ比の non-collapsibility により、A1～A4の仮定が成り立つとし、Difference Method を使用した場合、一般に、ロジスティック回帰モデルは NIE を過小評価することが知られており、保守的になる。

3.2.2.4.3 媒介変数と説明変数の影響

治療によるアウトカム（Y）への影響は媒介変数（M）によって低下する可能性があるが、モデル内の説明変数によって上昇する可能性もある。その結果、媒介効果があるにもかかわらず、治療の影響が変化していないように見える可能性がある。先ほど記載したとおり、ロジスティック回帰モデルによる Difference Method の NIE がゼロを示したとしても、Difference Method が保守的であることから、媒介効果が存在しないとは限らず、結論を出すことが難しい。

アウトカムがまれな場合は、オッズ比はリスク比に近似し、これらの問題は解消される可能性がある。また、まれなアウトカムでない場合、ロジスティック回帰モデルではなく対数二項モデルを適用することで、これらの問題を回避できる。（VanderWeele 2016）

3.2.2.4.4 注意点と課題

2値アウトカムに対する研究は現在でも盛んに行われており、先に述べてきたアウトカム（Y）

がまれな場合という仮定は、因果関係のある媒介変数に対しては不適切であるという意見もある (Samoilenko and Lefebvre 2021)。また、アウトカムがまれではないという仮定の下で、NDE 及び NIE についてのオッズを取得する手法も提案されている (Muthén et al. 2016)。この場合、まれではない結果に関連するオッズ比はリスク比で近似できないため、解釈を慎重に行う必要がある。さらに、Gaynor et al. (2019)は、ロジット関数のプロビット近似を使用し、アウトカムが共通である場合に使用できるオッズでの NDE と NIE の推定量を提供した。デルタ法等を用いた NDE と NIE に対する正確な推定等も提案されている (Samoilenko and Lefebvre 2021)。

これまで述べてきたことから、ロジスティック回帰モデルを用いた因果媒介分析では、非線形性等を考慮した解釈が重要である。

3.2.3 Time-to-event アウトカム

本節では、アウトカムが生存時間データの場合を考える。これまで通り治療を A 、媒介変数を M 、共変量を L とし、新たに生存時間アウトカムを T とする。また、時点 t における生存関数を $S_T(t) = P(T > t)$ とし、時点 t におけるハザード関数を $\lambda_T(t)$ とする。同様に、共変量 $L = l$ で条件づけた時点 t の生存関数を $S_T(t|l) = P(T > t|l)$ とし、共変量 $L = l$ で条件づけた時点 t のハザード関数を $\lambda_T(t|l)$ とする。さらに、前節までと同様に、仮定 A1～A4が成立している状況を考える。打ち切りがある場合には、それらは情報のない打ち切りであるとする (VanderWeele 2015)。

3.2.1にて、アウトカム (Y) 及び媒介変数 (M) が連続値の場合、治療 (A) と媒介変数 (M) の交互作用がないときには、Difference Method, Product Method によって推定される因果媒介効果は一致することを述べた。また、3.2.2にて、アウトカム (Y) が2値の場合、治療 (A) と媒介変数 (M) の交互作用がなく、媒介変数 (M) が正規分布に従い、さらにアウトカムがまれであるとき、Difference Method, Product Method によって推定される因果媒介効果は一致することを述べた。

本節では、生存時間アウトカムに対して加速 (Accelerated Failure Time ; AFT) モデルもしくは比例ハザード (Proportional Hazard ; PH) モデルを当てはめた場合に、NDE, NIE をどのように推定できるかについて考える。その際、どのような場合に Difference Method, Product Method による結果が一致するのかについても触れる。

3.2.3.1 生存時間アウトカムにおける効果指標

アウトカムが連続値の場合には、因果効果である TE の効果指標には平均の差を用いることが多い。また、アウトカムが2値の場合には、因果効果である TE の効果指標にはリスク差, リスク比, オッズ比等を用いることが多い。一方、アウトカムが生存時間の場合には、因果効果である TE の効果指標には、平均生存時間の差, 生存関数の差, ハザード比などの指標を用いることが多いだろう。TE の効果指標がいずれであっても、NDE と NIE に分解することが可能であり、そして CDE を定義することが可能である。以下に代表的な例として、効果指標が生存関数の差, ハザード比の場合を示す (VanderWeele 2011)。

$$\text{生存関数の差 } (S_{T^a}(t) - S_{T^{a^*}}(t)) :$$

$$\begin{aligned}
 \text{CDE} &= S_{T^{am}}(t) - S_{T^{a^*m}}(t) \\
 \text{NIE} &= S_{T^{aM^a}}(t) - S_{T^{aM^{a^*}}}(t) \\
 \text{NDE} &= S_{T^{aM^{a^*}}}(t) - S_{T^{a^*M^{a^*}}}(t) \\
 \text{ハザード比 } (\lambda_{T^a}(t)/\lambda_{T^{a^*}}(t)) : \\
 \text{CDE}^{\text{PH}} &= \lambda_{T^{am}}(t)/\lambda_{T^{a^*m}}(t) \\
 \text{NIE}^{\text{PH}} &= \lambda_{T^{aM^a}}(t)/\lambda_{T^{aM^{a^*}}}(t) \\
 \text{NDE}^{\text{PH}} &= \lambda_{T^{aM^{a^*}}}(t)/\lambda_{T^{a^*M^{a^*}}}(t)
 \end{aligned}$$

効果指標は、解析手法と関連する。解析手法に後述する AFT モデル、PH モデルを用いると、それぞれ効果指標に対数平均生存時間の差、ハザード比を選択することとなる。

3.2.3.2 AFT モデルを用いた因果媒介分析

生存時間アウトカム (T) に、次のような AFT モデルを当てはめることを考える。

$$\log(T|A = a, M = m, L = l) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta'_4 l + v\varepsilon$$

ここで、 ε は極値分布に従う誤差、 v は尺度パラメータとする。このとき、生存時間 T はワイブル分布に従う。

媒介変数 (M) が連続値である場合、媒介変数 (M) に次のような線形モデルを当てはめることができる。

$$E(M|A = a, L = l) = \beta_0 + \beta_1 a + \beta'_2 l$$

ここで、媒介変数 (M) は、治療 (A) 及び共変量 (L) によって条件づけた下で、分散 σ^2 の正規分布に従うとする。このとき、モデルの誤特定がなく、仮定 A1, A2が成立している場合には、CDEは次式で得られる。

$$\text{CDE}^{\text{AFT}} = \log\{E(T^{am}|L = l)\} - \log\{E(T^{a^*m}|L = l)\} = (\theta_1 + \theta_3 m)(a - a^*)$$

また、仮定 A1～A4が成立している場合には、NIE, NDE は次式で得られる。証明は、付録 5.3を参照されたい。(VanderWeele 2011, VanderWeele 2015)

$$\begin{aligned}
 \text{NIE}^{\text{AFT}} &= \log\{E(T^{aM^a}|L = l)\} - \log\{E(T^{aM^{a^*}}|L = l)\} \\
 &= (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*)
 \end{aligned}$$

$$\begin{aligned}
 \text{NDE}^{\text{AFT}} &= \log\{E(T^{aM^{a^*}}|L = l)\} - \log\{E(T^{a^*M^{a^*}}|L = l)\} \\
 &= \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta'_2 l + \theta_2 \sigma^2)\}(a - a^*) + 0.5\theta_3^2 \sigma^2 (a^2 - a^{*2})
 \end{aligned}$$

上式は、アウトカムが2値で、媒介変数が連続値であった場合に得られるオッズ比の尺度に関するNIE, NDEと類似したものであることが確認できる (3.2.2)。なお、治療 (A) と媒介変数 (M) に交互作用がない場合には、Difference Method, Product Method は一致する。

以上、媒介変数 (M) が連続値である場合を考えたが、2値である場合も同様である。以下では、NIE, NDE についてのみ紹介する。媒介変数 (M) が2値である場合、媒介変数 (M) に次のようなロジスティック回帰モデルを当てはめることができる。

$$\text{logit}\{P(M = 1|A = a, L = l)\} = \beta_0 + \beta_1 a + \beta'_2 l$$

このとき、モデルの誤特定がなく、仮定 A1～A4が成立している場合には、NIE, NDE は次式で得られる (VanderWeele 2015)。

$$\begin{aligned} \text{NIE}^{\text{AFT}} &= \frac{E(T^{aM^a} | A = a, L = l)}{E(T^{aM^{a^*}} | A = a, L = l)} \\ &= \frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta'_2 l)\} \{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta'_2 l)\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta'_2 l)\} \{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta'_2 l)\}} \\ \text{NDE}^{\text{AFT}} &= \frac{E(T^{aM^{a^*}} | A = a, L = l)}{E(T^{a^*M^{a^*}} | A = a, L = l)} \\ &= \frac{\exp(\theta_1 a) \{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta'_2 l)\}}{\exp(\theta_1 a^*) \{1 + \exp(\theta_2 + \theta_3 a^* + \beta_0 + \beta_1 a^* + \beta'_2 l)\}} \end{aligned}$$

上式は、アウトカムが2値で、媒介変数が2値であった場合に得られる NIE, NDE と類似したものであることが確認できる (3.2.2)。

3.2.3.3 PH モデルを用いた因果媒介分析

次に、生存時間 (T) に、次のような PH モデルを当てはめることを考える。

$$\lambda_T(t|A = a, M = m, L = l) = \lambda_T(t|A = 0, M = 0, L = 0) \exp(\gamma_1 a + \gamma_2 m + \gamma_3 am + \gamma'_4 l)$$

媒介変数 (M) が連続値である場合、媒介変数 (M) に、次のような線形モデルを当てはめることができる。

$$E(M|A = a, L = l) = \beta_0 + \beta_1 a + \beta'_2 l$$

ここで、媒介変数 (M) は、治療 (A) 及び共変量 (L) によって条件づけた下で、分散 σ^2 の正規分布に従うとする。このとき、モデルの誤特定がなく、仮定 A1, A2が成立している場合には、CDE は次式で得られる。

$$\text{CDE}^{\text{PH}} = \log\{\lambda_{T^{am}}(t|L = l)\} - \log\{\lambda_{T^{a^*m}}(t|L = l)\} = (\gamma_1 + \gamma_3 m)(a - a^*)$$

また、アウトカムがまれであり、仮定 A1～A4が成立している場合には、NIE, NDE は次式のように近似される。証明は、付録5.4を参照されたい。(VanderWeele 2011, VanderWeele 2015)

$$\begin{aligned} \text{NIE}^{\text{PH}} &= \log\{\lambda_{T^{am^a}}(t|L = l)\} - \log\{\lambda_{T^{am^{a^*}}}(t|L = l)\} \\ &\approx (\gamma_2 \beta_1 + \gamma_3 \beta_1 a)(a - a^*) \end{aligned}$$

$$\text{NDE}^{\text{PH}} = \log\{\lambda_{T^{aM^{a^*}}}(t|L = l)\} - \log\{\lambda_{T^{a^*M^{a^*}}}(t|L = l)\}$$

$$\approx \{\gamma_1 + \gamma_3(\beta_0 + \beta_1 a^* + \beta'_2 l + \gamma_2 \sigma^2)\}(a - a^*) + 0.5\gamma_3^2 \sigma^2 (a^2 - a^{*2})$$

なお、治療 (A) と媒介変数 (M) に交互作用がない場合には、Difference Method, Product Method は近似的に一致する。

以上、媒介変数 (M) が連続値である場合を考えたが、2値である場合も同様である。以下では、NIE, NDE についてのみ紹介する。媒介変数 (M) が2値である場合、媒介変数 (M) に次のようなロジスティック回帰モデルを当てはめることができる。

$$\text{logit}\{P(M = 1|A = a, L = l)\} = \beta_0 + \beta_1 a + \beta'_2 l$$

このとき、モデルの誤特定がなく、アウトカムがまれであり、仮定 A1～A4が成立している場合には、NIE, NDE は次式で与えられる (VanderWeele 2015)。

$$\begin{aligned} \text{NIE}^{\text{PH}} &= \frac{\lambda_{TAMa}(t|A = a, L = l)}{\lambda_{TAMa^*}(t|A = a, L = l)} \\ &\approx \frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta'_2 l)\}\{1 + \exp(\gamma_2 + \gamma_3 a + \beta_0 + \beta_1 a + \beta'_2 l)\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta'_2 l)\}\{1 + \exp(\gamma_2 + \gamma_3 a + \beta_0 + \beta_1 a^* + \beta'_2 l)\}} \\ \text{NDE}^{\text{PH}} &= \frac{\lambda_{TAMa^*}(t|A = a, L = l)}{\lambda_{TAMa}(t|A = a, L = l)} \\ &\approx \frac{\exp(\gamma_1 a)\{1 + \exp(\gamma_2 + \gamma_3 a + \beta_0 + \beta_1 a^* + \beta'_2 l)\}}{\exp(\gamma_1 a^*)\{1 + \exp(\gamma_2 + \gamma_3 a^* + \beta_0 + \beta_1 a^* + \beta'_2 l)\}} \end{aligned}$$

なお、アウトカムがまれでない場合、Difference Method も Product Method も因果的な解釈ができない。しかしながら、アウトカムがまれな場合であっても、Product Method は、モデルが正しく特定されており、A1～A4の仮定を満たせば、因果媒介効果の有無に対する妥当な検定を与える (VanderWeele 2011)。なお、アウトカムがまれでない場合に、PH モデルを用いて NIE, NDE を推定する方法については、Lange et al. (2012)や Vansteelandt et al. (2012)を参照されたい。

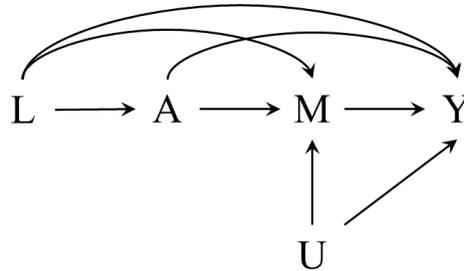
3.3 感度分析

DE 及び IE に関する結論を導くために必要な仮定 (A1～A4) は非常に強い仮定であり、応用する場合において満たされない可能性がある。したがって、DE 及び IE に関する結論が、仮定が満たされない場合にどれだけ頑健性を持っているかを評価することが重要である。感度分析は、例えば、未測定の変数因子が媒介変数とアウトカムの両方にどれだけ強く関連しているかを評価し、DE 及び IE に関する結論が大きく変わるかどうかを評価するのに役立つ。

本節では、VanderWeele (2016)及び Ding and VanderWeele (2016)を参考に、広範な適用可能性を持ち、比較的少数の仮定を置く、未測定変数因子の感度分析について紹介する。この解析では、2つの感度分析パラメータを使用し、治療変数-アウトカムの交絡 (仮定 A1) 及び治療変数-媒介変数の交絡 (仮定 A3) が制御されていると仮定する。

媒介変数-アウトカムの未測定変数因子が存在する可能性がある (つまり、仮定 A2が満たされない) が、これらのいずれも治療変数に影響を受けないと仮定する (つまり、仮定 A4が満

たされている). そのため, 図 3-2のような因果構造を考慮する. ここでは, L: 共変量, A: 治療変数, M: 媒介変数, Y: アウトカムとし, U: 治療変数-アウトカムの未測定交絡因子とする.



[VanderWeele (2016) Figure 3を改変]

図 3-2 媒介変数-アウトカム間の未測定交絡因子 U

3.3.1 感度分析パラメータ γ

本節では, 2値のアウトカム Y を考える. (A=1, M, L=l)が与えられた条件下での U と Y の間の条件付き関連性尺度を導入し, 最初の感度分析パラメータ γ を以下のように定義する.

$$\gamma = RR_{UY|(A=1,M,l)} = \max_m RR_{UY|(A=1,m,l)} = \max_m \frac{\max_u P(Y = 1|A = 1, m, l, u)}{\min_u P(Y = 1|A = 1, m, l, u)}$$

ここで $RR_{UY|(A=1,M,l)}$ は, u を通じての確率 $P(Y = 1|A = 1, m, l, u)$ の最大値を最小値で割ったものである. U が2値の場合, $RR_{UY|(A=1,M,l)}$ は通常の場合の条件付きリスク比であり, $\gamma = RR_{UY|(A=1,M,l)}$ は, M の層別における治療群の被験者において, 共変量の条件付きで, U と Y との関連性の最大リスク比を示している. すなわち, U が M を経由せずに Y の確率を増加させる最大の比率である. U と Y が(A, M, L)が与えられた条件下で条件付き独立である場合, $\gamma = RR_{UY|(A=1,M,l)} = 1$ となる.

3.3.2 感度分析パラメータ λ

M が与えられた条件下での A と U の条件付き関連性を導入する.

図 3-2に示されているように, もし A と U が M に影響を与える場合, A と U の共通の子孫 M に条件付けられた場合, どちらも他方に影響を与えていなくても ($A \perp U | L$), A と U の間に関連性が生じる. これは collider bias とも呼ばれる. 第二の感度分析パラメータは, collider bias によって生成されるこの関連性の大きさを評価する.

第二の感度分析パラメータ λ を以下のように定義する.

$$\lambda = RR_{AU|(M,l)} = \max_m RR_{AU|(m,l)} = \max_{u,m} \frac{P(u|A = 1, m, l)}{P(u|A = 0, m, l)}$$

ここで, $RR_{AU|(M,l)}$ は, $M = m$ 及び $L = l$ が与えられたもとで, A と U のリスク比の最大値を取

る。U が2値の場合、 $RR_{AU|(M,l)}$ は通常の状態付きリスク比である。すなわち、第二の感度分析パラメータ λ は、層 $M = m$ における条件付けを通じた collider bias 比の最大値である。

$RR_{AU|(M,l)}$ の代替形を以下のように示すことができる。

$$RR_{AU|(m,l)} = \max_u \frac{P(m|A = 1, l, u) / P(m|A = 1, l)}{P(m|A = 0, l, u) / P(m|A = 0, l)}$$

これは、 $U = u$ の層内で A が $M = m$ に対して持つ最大の条件付き相対リスクを、A が $M = m$ に対して持つ無条件の相対リスクで割ったものである。U に対する無条件の相対リスクは観測データから特定可能であり、したがって、第二の感度分析パラメータ λ はUに条件づけられた相対リスクに大きく依存する。

さらに、この第二の感度分析パラメータ λ の指定を支援するために、Greenland (2003)は、関連性の大きさに応じて、多くのがほとんどすべての状況で、Mによる条件付けによって導入されるAとUの関連性を示す比の大きさが、AとMの関連性及びUとMの関連性と比較して小さいことを示した。したがって、これらの比のうち小さい方は、第二の感度分析パラメータ λ の指定に役立つことがある。特に、AとMの関連が弱い場合、collider bias は小さくなる。もし、 $A \perp M|C$ であるならば、collider bias はゼロであり、すなわち $RR_{AU|(M,l)} = 1$ となる。

3.3.3 Bounding factor B

図 3-2が正しい因果構造を表している場合、未測定の交絡因子が DE を減少させるか、もしくは IE を増大させる可能性のある最大の比率は次のように示すことができる。

$$B = \frac{\gamma\lambda}{\gamma + \lambda - 1}$$

これは、 γ 及び λ の両方に対して対称で単調な性質を持ち、いずれの感度分析パラメータよりも大きくなる。感度分析パラメータのうちの一つが1の場合、Bounding factor B も1に等しくなる。Bounding factor は、未測定の媒介変数とアウトカムの交絡の強さを測る指標であり、Ding and VanderWeele (2016)に示される定理より、調整された推定値（つまり、未測定の交絡因子が DE 及び IE をどれだけ変える可能性があるかを理解するための調整された値）は以下の方法で得られる。

- DE の推定値とその信頼区間の両側限界をバイアス因子 B で除算
- IE の推定値とその信頼区間の両側限界をバイアス因子 B で乗算

この方法は、図 3-2が正しい因果構造であるという仮定（つまり、媒介変数-アウトカムの交絡因子のみ未測定、その他の交絡因子の仮定 (A1, A3, 及び A4) も成り立っていると仮定)のもとで成り立つ。

3.3.4 感度分析の報告

感度分析パラメータ γ と λ の実際の値はわからないが、値を変化させて、推定値がどのよう

に変化するか、または例えばゼロに減少するかを確認することができる。実際には、感度分析パラメータのさまざまな値にわたる調整された DE 及び IE の推定値の表、例えば、1つの感度分析パラメータ λ を行に、もう1つの感度分析パラメータ γ を列にして、これらの設定ごとに調整された推定値と信頼区間を報告することが考えられる。これにより、仮定が満たされない場合に推定値の感度がどれだけあるかについての詳細な情報が得られる。少なくとも、未測定の変数因子が DE の推定値をゼロにするためにどれだけの影響を持つ必要があり、また信頼区間がゼロを含むようにするためにどれだけの交絡が必要かを報告するのは適切で、また、比較的簡単な方法である。

3.3.5 2値アウトカム以外の場合

ここで説明した方法は、2値アウトカムに関してのものであるが、連続量、カウントデータ、イベント発生までの時間に関する同様の方法が使用され、比率の代わりに差でも適用できる。さらなる議論については [Ding and VanderWeele \(2016\)](#) を参照されたい。

3.4 媒介変数への介入効果指標

本節では、媒介変数の介入効果の程度を表す指標として、Proportion Mediated (PM)、及び Proportion Eliminated (PE) について触れたい。

3.1では、反事実フレームワークを用いた治療の TE の DE 及び IE への分解を行った。

$$\begin{aligned} TE &= E[Y^{a=1} - Y^{a=0}] = E[Y^{a=1, M^{a=1}}] - E[Y^{a=0, M^{a=0}}] \\ &= (E[Y^{a=1, M^{a=0}}] - E[Y^{a=0, M^{a=0}}]) + (E[Y^{a=1, M^{a=1}}] - E[Y^{a=1, M^{a=0}}]) \\ &= NDE + NIE \end{aligned}$$

PM は、NIE が TE に対してどの程度影響を与えているかを測る指標として以下のように表される。PM の算出に必要な NIE (または NDE) の値は、3.1で言及した通り、A1～A4の仮定が成り立つ下では媒介公式により算出が可能である。

$$PM = NIE / TE = (TE - NDE) / TE$$

また、3.1では、DE の別の定義として、CDE を紹介している。例えば、媒介変数について $M = m$ という介入を行う場合、CDE は

$$CDE(m) = E[Y^{a=1, M=m}] - E[Y^{a=0, M=m}]$$

と表わされる。このように媒介変数への介入を行った時 (M をある値に制御した時) に TE のうちの程度除去できるのかを表す指標として、以下のように表される PE が提唱されている ([VanderWeele 2013b](#))。PE の算出に必要な CDE(m) の値は、A1及び A2の仮定が成り立つ下では算出が可能である ([田中 2022](#))。

$$PE(m) = (TE - CDE(m)) / TE$$

PM と PE は異なるものを捉える指標と考えられる。反事実フレームワークの下で TE を DE, IE への分解することにより、媒介変数を通して治療がどのような Pathway でアウトカムに影響しているかを理解することができるが、PM はその影響の程度を示すことが期待される。一方で、媒介変数への介入によって TE をどの制御できるかを見る指標ととらえることのできる PE のほうが、医療政策や介入決定の検討に用いやすい、より Policy relevant な指標と言えるかもしれない。ただし、PM と PE のいずれも0から1の間を取る場合は上記のような解釈を考えられるが (田栗 2014, 黒木 2017)、実際には算出される値が必ず0から1の間を取るとは限らない点にも留意されたい。

また、NDE は本質的には様々な CDE の平均化したものであり、PM と PE が同じような値をとるとは限らない点にも留意されたい。例えば、治療 A 自体は媒介 M に変化を及ぼさない ($M^{a=0} = M^{a=1}$) が、A と M に交互作用がある場合を考える。前者の条件のため、NIE 及び PM も0となる。一方で、後者の条件により、CDE(m)の値は媒介変数の値 m によっても変わってくる。例えば、CDE($m = 1$) が比較的大きい場合はPE($m = 1$) = 0、逆にCDE($m = 0$) が比較的小さい場合はPE($m = 0$) = 1 のようになると考えられる。このような交互作用がない場合は、PM と PE は一致すると考えられる。

PM と PE が乖離する事例としては、15q25.1遺伝子の変異と肺癌についての関連性について考察した研究が挙げられる (Lash et al. 2021, VanderWeele 2011)。ここでは治療は遺伝子の変異の有無、アウトカムは肺癌の発症、媒介変数としては1日当たりの喫煙本数としている。解析の結果、Odds 比として算出された NDE の大きさは1.26、NIE の大きさは1.01であり、PM は約3%となった。即ち、喫煙によって媒介される効果の大きさは極めて限定的で、それ以外の Pathway によって遺伝子の変異の有無と肺癌の発症は関連付けられることが示唆された。遺伝子の変異の喫煙に対する影響は喫煙本数として1本程度でしかないことが示唆されている。一方で、遺伝子の変異と1日当たりの喫煙本数には交互作用があり、非喫煙者には遺伝子の変異は肺癌に影響を与えないことが考えられる。この場合は CDE($m = 0$) = 0、即ち、PE($m = 0$) = (TE - CDE($m = 0$))/TE = 1 となり、遺伝子の変異は喫煙がない場合には肺癌には影響を与えない (禁煙を勧めることで肺癌のリスクを排除できる割合が大きい) 可能性がある。

4. 解析コード

4.1 SAS

4.1.1では本稿で紹介した因果媒介分析に対応した公開されている SAS マクロを, 4.1.2ではその公開マクロをもとに作成された SAS プロシージャを紹介する.

4.1.1 公開マクロ

ここまで紹介してきた因果媒介分析に対応した SAS マクロが VanderWeele のウェブサイトにて公開 (VanderWeele 2013c) されており, Valeri and VanderWeele (2013)及び Valeri and VanderWeele (2015)で紹介がされている. 本マクロは, 異なる種類のアウトカムや媒介変数, 回帰モデルに対応できるだけでなく, 媒介変数への交互作用が存在する場合にも適用可能である. 本マクロを用いることで, 因果媒介分析を実行することができ, モデルの仮定が正しく, 前項で考察した識別可能性の仮定 (A1~A4) が成立すると仮定したもとの, DE 及び IE の推定値及び信頼区間を計算できる. ケースコントロール研究 (VanderWeele and Vansteelandt 2010) でも適用可能なことが本マクロのもう一つの大きな特徴である. なお, 欠測データがアウトカム, 媒介変数及び共変量に含まれている場合は解析から除外されること, 及び共変量にカテゴリ変数を入れたい場合はダミー変数化が必要なことには留意いただきたい. Valeri and VanderWeele (2013)で述べられたマクロは, 生存時間型のアウトカムは対応していなかったが, 現在公開されているマクロは Valeri and VanderWeele (2015)に基づき生存時間型のアウトカムにも対応しており, PHモデル及び AFTモデルによる解析が選択可能である. 基本的なコードは同じとなっており, 打ち切りの変数が追加になっている.

表 4-1 SAS マクロの引数とその説明

引数	説明
data	データセット名
yvar	アウトカムの変数名
avar	治療の変数名
mvar	媒介変数の変数名
cvar	その他の共変量の変数名
a0	対照とする治療レベル
a1	比較する治療レベル
m	controlled direct effect を推定する媒介変数のレベル
nc	共変量の数
yreg	アウトカムに対して用いる回帰モデル linear, logistic, loglinear, poisson, negbin (負の二項分布), survCox, survAFT_exp, survAFT_weibull のいずれかを入力
mreg	媒介変数に対して用いる回帰モデル linear か logistic のいずれかを入力
interaction	治療と媒介変数の交互作用の存在有無 true か false を入力
output	[optional] アウトプットの制御を行う. デフォルト: Controlled direct effect, natural indirect effect, total effect のみ出力 full: 条件付き効果及び平均共変量レベルで評価した効果の両方も出力
c	[optional] output=full の際に条件付き効果を計算する共変量の固定値. ここを指定しないと共変量の平均レベルでの治療効果が推定される.

casecontrol	[optional] 試験デザインの指定 デフォルト:ランダム化比較試験やコホート研究の前向き比較試験 true: ケースコントロール試験での解析も可能になる
boot	[optional] 標準誤差の算出方法の指定. デフォルト: デルタ法に基づく標準誤差を算出 true: ブートストラップ法に基づく標準誤差を算出. デフォルトでは1000回のブートストラップサンプリングから算出. 回数を変更したい場合は boot=5000のように指定する
cens	[optional] Time-to-event アウトカムの場合 (yreg = survCox, survAFT exp, または survAFT weibull) に, 打ち切り変数として指定が必要. 1がイベント, 0が打ち切りを示す.

4.1.2 プロシージャ

本節では SAS の CAUSALMED プロシージャ (SAS Institute Inc. 2017) の概要について説明する. 関心のある変数として以下の4変数を解析に用いる. また, 本プロシージャが適用可能な変数のタイプは以下である.

- アウトカム Y : 2 値, 連続量, カウント変数
- 治療変数 A : 2 値, 連続量
- 媒介変数 M : 2 値, 連続量
- 交絡変数 C : カテゴリカル, 連続量

本プロシージャでは一般化線形モデルを用いて解析を行うため, 結果に因果的な解釈を持たせるためには3章で述べた識別性の仮定に加えてモデルの特定にも注意が必要である.

以下の3つのステートメントで回帰モデルを指定する.

- MODEL ステートメント;アウトカム変数に対する一般化線形モデルを指定する. 指定可能な分布は 2 値, 負の二項分布, ポアソン分布, 及び正規分布
- MEDIATOR ステートメント; 媒介変数に対する一般化線形モデルを指定する. 指定可能な分布は 2 値, 及び正規分布
- COVAR ステートメント;上記の回帰モデルで用いる共変量を指定する

GLM プロシージャ等とは異なり回帰モデル中の共変量は別ステートメント (COVAR ステートメント) で指定する点に注意が必要である.

本プロシージャでは以下の因果媒介効果を計算可能である.

- 総合効果 (TE)
- 制御された直接効果 (CDE)
- 自然な直接効果 (NDE)
- 自然な間接効果 (NIE)

アウトカムが連続量の場合には上記の効果は元々のスケールで計算され, 2値の場合にはオッズ比及び超過相対リスクのスケールで計算される. また, 上記の効果の TE に対する割合も

計算される.

- Proportion Mediated (PM)
- Proportion due to interaction (PI, [VanderWeele and Tchetgen 2014](#))
- Proportion Eliminated (PE)

また、本プロシージャでは NDE と NIE への2つの分解以外にも3つ、4つの効果の分解が可能であるが本報告書の内容を超えるためここでは割愛する. 詳しくは [VanderWeele \(2014\)](#)等を参照されたい. 各効果の標準誤差及び信頼区間の算出も可能である. デフォルトは最尤推定量による漸近分散に基づいて算出されるが、後述する BOOTSTRAP ステートメントを用いてブートストラップ法によるリサンプリングに基づく方法も計算可能である.

各ステートメントについて簡単に紹介する.

- PROC CAUSALMED ステートメント :
CAUSALMED プロシージャを使用するためのステートメント. CAUSALMED プロシージャ特有のいくつかのオプションについて紹介する.
 - CASECONTROL オプション : ケースコントロール研究の解析で用いる. このオプションを指定することで MEDIATOR モデルにおいて対照群のデータのみが用いられる.
 - DECOMP オプション : TE の分解について指定することができる. 分解する効果の数として 2, 3, 4 を指定することができる.
- BOOTSTRAP ステートメント :
各効果及びそれらの TE に対する割合についての標準誤差及び信頼区間のブートストラップ法による推定が可能. 以下に代表的なオプションを紹介する. その他のオプションについては CAUSALMED のユーザーガイドを参照されたい.
 - BOOTCI オプション : 各効果及びそれらの TE に対する割合の信頼区間を算出
 - NBOOT オプション : ブートストラップサンプルのデータセット数を指定
 - SEED オプション : ランダムナンバーの seed を指定
- COVAR ステートメント :
因果媒介モデルにおける共変量を指定する. 指定する共変量が治療-アウトカム間, 治療-媒介変数間, 媒介変数-アウトカム間のどの交絡因子であるか区別する必要はなく, すべての交絡因子を指定すればよい.
- EVALUATE ステートメント :
各因果媒介効果の水準を指定する. 以下の変数の水準について指定が可能である.
 - 全ての効果に対する治療と対照の水準
 - CDE の評価における媒介変数の水準
 - 様々な条件付き効果の評価における共変量の水準

交互作用効果及び非線形のモデルにおいては, 計算された因果媒介効果が治療, 対照, 媒介変数, 共変量の水準に依存するため本ステートメントにおいて特定の水準を指定することが可

能.

- **FREQ** ステートメント：
データセット中の変数が他の変数の頻度を表す場合、本ステートメントで指定が可能.
- **MEDIATOR** ステートメント：
媒介変数のモデルを指定する。等式の左辺に媒介変数を指定し、右辺に治療変数を指定する。本ステートメントで指定する媒介変数と治療変数は **MODEL** ステートメントで指定する変数と一致していることが必要である。前述した通り、モデルに含める共変量は本ステートメントではなく **COVAR** ステートメントで指定する。媒介変数及び治療変数は 2 値または連続量の変数が指定可能であり、2 値の場合には **CLASS** ステートメントで指定することが必要である。指定しない場合は連続量とみなされる。連続量の媒介変数に関しては正規分布で **identity** のリンク関数が仮定され、2 値の媒介変数に関しては二項分布及び **logit** リンク関数が仮定される.
- **MODEL** ステートメント：
アウトカム変数のモデルを指定する。等式の左辺にアウトカム変数を指定し、右辺に媒介変数と治療変数を指定する。本ステートメントで指定する媒介変数及び治療変数は **MEDIATOR** ステートメントと一致している必要がある。**MEDIATOR** ステートメントと同様にモデルに含める共変量はここでは指定せずに **COVAR** ステートメントで指定する。アウトカム変数には 2 値、連続量、カウント変数が指定することが可能であるが、2 水準を超える水準数を持つ名義変数、順序変数 ([SAS Institute Inc. 2017](#)) を指定することはできない。
 - **DISTRIBUTION/DIST** オプション：二項分布、負の二項分布、正規分布、Poisson 分布の指定が可能。デフォルトのリンク関数は、二項分布=Logit、負の二項分布=log、正規分布=identity、ポアソン分布=Log となる。**DISTRIBUTION/DIST** 及び **LINK** オプションのどちらも指定しない場合、デフォルトはアウトカム変数が **CLASS** ステートメントに指定があれば **logit** リンクを用いた二項分布となり、**CLASS** ステートメントに指定がなければ **identity** リンクを用いた正規分布となる。
 - **LINK** オプション：モデル中のリンク関数を指定する。**identity**、**log**、**logit** の 3 種類が指定可能。

本プロシージャの適用事例については[4.1.3.1.4](#)を参照されたい。

4.1.3 適用事例

本節は、実際にプログラムを実行することで、[2](#)章及び[3](#)章で解説した各手法の具体的な適用方法や手法そのものの理解を促進することを目的とする。[4.1.3.1](#)では、シミュレーションにより、各個人の複数の潜在変数を全て知っている仮想的状況を考え、本書で紹介した各手法の **SAS** での適用方法を具体的に示す。[4.1.3.1](#)は以下の複数の節で構成されている。まず、[4.1.3.1.1](#)にて、シミュレーションデータを作成し、[4.1.3.1.2](#)にて、[2](#)章で紹介した **traditional** な手法の適用方法を示し、また **traditional** な手法の限界も示す。その後、[3.2.1](#)で解説した手法の適用方法

と結果について、引き続き4.1.3.1.3で%medication, 4.1.3.1.4で SAS/STAT の PROC CAUSALMED を適用する。また4.1.3.2では、CAUSALMED プロシージャのマニュアル (SAS Institute Inc. 2017) で用いている Cognitive データセットに対する解析事例を紹介し、因果媒介分析を適用した際の結果の解釈の仕方についても理解を深める。

なお、4.1.3.1.1で作成するシミュレーションデータは、4.2の R を用いたシミュレーションでも利用される。

4.1.3.1 シミュレーションデータでの媒介分析の適用

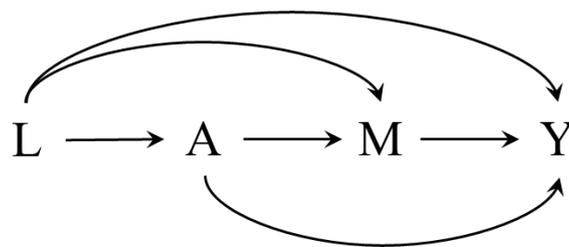
4.1.3.1.1 シミュレーションデータの生成

本節では、まずシミュレーションデータの設定を説明し、生成プログラムを紹介する。生成されたデータセット内には、現実では反事実であり観測され得ない、各個人の全ての潜在変数が含まれている。この全ての潜在変数を用いて、シミュレーションデータでの因果媒介効果の真値を算出する。

なお、シミュレーションデータは、手法の理解しやすさを重視し、単純な数値を各パラメータの真値として指定した。特定の現実の状況を想定したものではないため、結果の解釈を含めた実際の適用例については、4.1.3.2を参照いただきたい。

【シミュレーションデータの設定】

DAGは図4-1を想定する。アウトカム Y と媒介変数 M は連続値とする。なお、媒介変数 M は 1 変数とする。ベースライン共変量 L は 1 変数で 2 値、治療変数 A も 2 値とする。また、アウトカム Y の発生メカニズムでは、治療変数 A と媒介変数 M の交互作用があるものとする。さらにベースライン共変量 L は前述 (3.1) の 4 つの仮定を満たすものとする。



[VanderWeele (2016) Figure 1 を改変]

図 4-1 シミュレーションデータで想定する DAG

十分に大きな集団を対象としているものとし、患者数は10000人とする。2値変数であるベースライン共変量 L は全患者を1:1で2群に分割する。治療変数 A は、L = 1 のときの生起確率70%、L = 0 のときの生起確率は30%の Bernoulli 分布に従う。

ベースライン共変量 L 及び治療変数 A が与えられた下での、各個人の潜在変数発生モデルは以下の通り。

- $M^a = \beta_0 + \beta_1 a + \beta_2 l + \varepsilon_1$ $\varepsilon_1 \sim N(0,1)$

- $Y^{a,m} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4 l + \varepsilon_2 \quad \varepsilon_2 \sim N(0,1)$

ここで、 $\beta = (\beta_0, \beta_1, \beta_2) = (1.0, 2.0, 3.0)$, $\theta = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4) = (0.5, 1.5, 2.5, 3.5, 4.5)$ とする。つまり、以下の式によりデータを発生させる。

- $M^a = 1.0 + 2.0a + 3.0l + \varepsilon_1 \quad \varepsilon_1 \sim N(0,1)$

- $Y^{a,m} = 0.5 + 1.5a + 2.5m + 3.5am + 4.5l + \varepsilon_2 \quad \varepsilon_2 \sim N(0,1)$

本シミュレーション設定において、各因果媒介効果の真値は以下のように求められる。ただし、1つ目の式のCDE($m = 3.48$)で与えた3.48という値は、本シミュレーション設定でのデータ発生の結果得られる m の観測値の算術平均値である (Output 4-1で後述)。 $m = 0, 1$ は M が2値データの場合に利用されるものであるが、手法の理解しやすさのために補足的に出力した。

- $CDE(m = 3.48) = \{(\theta_1 + \theta_3 m)(a - a^*) | m = 3.48\} = (1.5 + 3.5 * 3.48)(1 - 0) = 13.68$
- $CDE(m = 0) = \{(\theta_1 + \theta_3 m)(a - a^*) | m = 0\} = (1.5 + 3.5 * 0)(1 - 0) = 1.5$
- $CDE(m = 1) = \{(\theta_1 + \theta_3 m)(a - a^*) | m = 1\} = (1.5 + 3.5 * 1)(1 - 0) = 5.0$
- $NDE = (\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2 \bar{l}) * (a - a^*) = (1.5 + 3.5 * 1.0 + 3.5 * 2.0 * 0 + 3.5 * 3.0 * 0.5)(1 - 0) = 10.25$
- $NIE = (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*) = (2.5 * 2.0 + 3.5 * 2.0 * 1)(1 - 0) = 12$
- $TE = NDE + NIE = 22.25$

なお、3.1の最終段落で解説したように、TEの分解にはいくつかの異なる方法が存在する。上記のNDE及びNIEは、本稿を通じた分解方法に倣い、それぞれPDE及びTIEを指している。また、3.1の最終段落で紹介した分解 $TE=TDE+PIE$ についても以下に真値を示しておく。ただし、以下の式では、TDEとPIEをそれぞれNDE(TDE)及びNIE(PIE)と表記した。本稿を通じた分解方法に倣い、4.1.3.1でも、特に断りのない限り、NDEとNIEはそれぞれPDEとTIEを指しているが、2つの分解方法を対比する場合には、必要に応じて、それぞれ $TE=NDE(PDE)+NIE(TIE)$, $TE=NDE(TDE)+NIE(PIE)$ のように、NDE及びNIEがどちらの分解に従っているかを括弧書きで明確にすることとした。なお、本稿ではNDE(TDE)及びNIE(PIE)の推定量を導出していないため、興味のある読者は3.2.1のNDE(PDE)及びNIE(TIE)の推定量の導出過程を参考に導出を試みて欲しい。

- $NDE(TDE) = (\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a + \theta_3 \beta_2 \bar{l}) * (a - a^*) = (1.5 + 3.5 * 1.0 + 3.5 * 2.0 * 1 + 3.5 * 3.0 * 0.5)(1 - 0) = 17.25$
- $NIE(PIE) = (\theta_2 \beta_1 + \theta_3 \beta_1 a^*)(a - a^*) = (2.5 * 2.0 + 3.5 * 2.0 * 0)(1 - 0) = 5$

【シミュレーションデータの作成】

次に、シミュレーションデータを生成する。Program 4-1にSASデータステップによるプログラム例を示した。本プログラムは、シミュレーション設定での変数のタイプ・数に従って作成されており、最初の数行のマクロ変数でパラメータの値が与えられている。なお、変数のタイ

プ・数をシミュレーション設定とは異なるものに変えたい場合には、`data` ステップ内のプログラムを適切に変更する必要がある。例えば、アウトカムを2値に変更したい場合には、3.2.2でモデルを確認し、変数生成モデルを書き下す必要がある。

プログラムの流れは以下のとおりである。まず、各パラメータ用のマクロ変数に、パラメータ値を設定する。Data ステップにおいては、被験者毎に、ベースライン共変量 L 、治療変数 A 、媒介変数 M 、アウトカム Y の順でデータが生成される。媒介変数 M 、アウトカム Y については、各被験者の潜在変数 $M^1, M^0, Y^{1,M^1}, Y^{1,M^0}, Y^{0,M^1}, Y^{0,M^0}, Y^{1,1}, Y^{1,0}, Y^{0,1}, Y^{0,0}$ と各被験者に1つのみ観測される治療 $A = a$ 及び媒介変数 $M = m$ の下での観測値も生成される。なお、各被験者で1つの変数に対して複数の潜在変数を作成する場合には、個人内で1度発生させた乱数を繰り返し共通で利用する必要があることに留意が必要である（プログラム内では、`RAND_M`、`RAND_Y` が相当する。）。

Program 4-1 simulation データの作成

```
%let N=10000; * sample size ;
%let pr_l=0.5; /*Lは2値*/
%let pr_a_l1=0.7; %let pr_a_l0=0.3; /*Aは2値*/
%let beta0=1.0; %let beta1=2.0; %let beta2=3.0;
%let theta0=0.5; %let theta1=1.5; %let theta2=2.5; %let theta3=3.5; %let theta4=4.5; /* 交互作用なしにするならtheta3=0*/

data dat;
  call streaminit(1234);
  N=&N;
  pr_l=&pr_l;
  pr_a_l1=&pr_a_l1; pr_a_l0=&pr_a_l0;
  beta0=&beta0; beta1=&beta1; beta2=&beta2;
  theta0=&theta0; theta1=&theta1; theta2=&theta2; theta3=&theta3; theta4=&theta4;
  do l=0 to 1;
    do id=1 to round(N*pr_l,1);
      * treatment a ;
      if l=1 then a=rand('BERNOULLI',pr_a_l1);
      else a=rand('BERNOULLI',pr_a_l0);
      * mediator m ;
      * potential variable ;
      RAND_M=rand("NORMAL"); /*個人内の複数の潜在変数で使用する共通の乱数*/
      m_l=beta0+beta1+beta2*1+RAND_M;
      m_0=beta0 +beta2*1+RAND_M;
      * observation ;
      if a=1 then m=m_l; else m=m_0;
      * outcome y ;
      * potential variable ;
      RAND_Y=rand("NORMAL"); /*個人内の複数の潜在変数で使用する共通の乱数*/
      /*Y a M m=theta0+theta1*a+THETA2*M m+THETA3*M m*A+theta4*1+RAND Y;*/
      y_l_m_l=theta0+theta1 +theta2*m_l+theta3*m_l +theta4*1+RAND_Y;
      y_l_m_0=theta0+theta1 +theta2*m_0+theta3*m_0 +theta4*1+RAND_Y;
      y_0_m_l=theta0 +theta2*m_l +theta4*1+RAND_Y;
      y_0_m_0=theta0 +theta2*m_0 +theta4*1+RAND_Y;
      /*CDE用. m=1,0に固定*/
      /*Y a m =theta0+theta1*a+THETA2*M m+THETA3*M m*A+theta4*1+RAND Y;*/
      y_l_1 =theta0+theta1 +theta2*1 +theta3*1 +theta4*1+RAND_Y;
      y_l_0 =theta0+theta1 +theta4*1+RAND_Y;
      y_0_1 =theta0 +theta2*1 +theta4*1+RAND_Y;
      y_0_0 =theta0 +theta4*1+RAND_Y;
```

```

        * observation ;
        if a=1 then y=y_1_m_1; else y=y_0_m_0;
    output;
end;
end;
run;

```

【因果効果の真値の計算】

次に、シミュレーションデータでの各因果媒介効果の真値を計算する。Program 4-2にプログラムを示した。本シミュレーションデータでは、各被験者について現実には観測されない反事実を含む全ての潜在変数を持っているため、各潜在変数の単純な算術平均にて因果媒介効果の真値がわかる。

Program 4-2 シミュレーションデータでの因果媒介効果の真値の確認

```

ods select none;
ods output Summary=MM;
proc means data=dat mean;
    var _all_;
run;

/*因果効果の真値の計算. 潜在変数を直接利用している*/
data result;
    set MM;
    r TE=y 1 m 1 mean-y 0 m 0 mean;
    r CDE m0=y 1 0 mean-y 0 0 mean;
    r_CDE_m1=y_1_1_mean-y_0_1_mean;
    r PDE=y 1 m 0 mean-y 0 m 0 mean; /*=NDE*/
    r TIE=y 1 m 1 mean-y_1_m_0_mean; /*=NIE*/
    keep r_ : m_Mean;
run;

ods select all;
proc print data=result noobs; format _all_ 8.2; run;

```

本プログラムの出力結果を Output 4-1に示した。CDE($m = 0$) = 1.5, CDE($m = 1$) = 5.0, PDE = NDE = 10.19 \approx 10.25, TIE = 12.00となり、シミュレーション設定のパラメータの真値とのずれは、被験者数10000であるため、わずかであった。なお、追加で出力している m_Mean は媒介変数 M の観測値の算術平均値であり、4.1.3.1.3で利用するために出力している。

Output 4-1 シミュレーションデータでの因果媒介効果の真値

m_Mean	r_TE	r_CDE_m0	r_CDE_m1	r_PDE	r_TIE
3.48	22.19	1.50	5.00	10.19	12.00

4.1.3.1.2 回帰モデルを用いた古典的な媒介分析の適用

本節では、シミュレーションデータに対して、2章で紹介した古典的な手法である Difference method と Product method を SAS で適用する。なお、これらの方法は一部の例外を除いて、交互作用項がある場合の DE・IE の正確な定式化がなされていない。そのため、本節では、交互作

用項を含まないモデル誤特定下での適用方法を紹介し、結果を考察することとする。

【Difference method (交互作用項無し)】

Difference method では、媒介変数 M を考慮しない回帰及び考慮する回帰を実施する。媒介変数 M を考慮しない回帰モデルから得られる治療変数 A のパラメータ値 ϕ_1 を TE, 考慮する回帰モデルから得られる治療変数 A のパラメータ値 θ_1 を、媒介変数 M で説明される効果を DE と考える。また、IE は、 $TE = \phi_1$ から $DE = \theta_1$ を差し引くことで、 $\phi_1 - \theta_1$ で与えられる。この IE は、それぞれの回帰モデルでの治療変数 A のパラメータ値 ϕ_1 及び θ_1 からの、媒介因子 M 調整による減少分 $\phi_1 - \theta_1$ と解釈することが出来る。上記の2つの回帰モデルを Program 4-3に示した。

Program 4-3 Difference method によるシミュレーションデータの解析

```

title 'Difference method 1: A⇒Y : Aの推定値(TE) [  $\phi_1$  ]';
proc mixed data=dat;
  class A L/ref=first;
  model Y=A L/s;
  lsmeans A/pdiff;
run;
title 'Difference method 2: A⇒Y : Aの推定値 [  $\theta_1$  ] (Mで説明される効果を除いたもの=DE). [  $\phi_1 - \theta_1 = IE$  ]';
proc mixed data=dat;
  class A L/ref=first;
  model Y=A M L/s;
  lsmeans A/pdiff;
run;
title;

```

Program 4-3の2つの回帰モデルに対する2つの出力を以下に示す。Output 4-2には TE が出力されており、Output 4-3には、DE が出力されている。それぞれの出力結果から治療変数 A の推定値を確認することにより、TE は22.3028, DE は13.6237, IE は $22.3038 - 13.6237 = 8.6801$ となる。ただし、真のアウトカムモデルは治療変数 A と媒介変数 M の交互作用項を含むため、誤特定下での推定結果であることに留意されたい。なお、シミュレーションデータの真値は Output 4-1に示したように、TE は22.19, DE は10.19, IE は12.00であった。

Output 4-2 TE の出力結果

Differences of Least Squares Means							
Effect	A	_a	Estimate	Standard Error	DF	t Value	Pr > t
a	1	0	22.3038	0.1150	9997	193.97	<.0001

Output 4-3 DE の出力結果

Differences of Least Squares Means							
Effect	a	_a	Estimate	Standard Error	DF	t Value	Pr > t
a	1	0	13.6237	0.09359	9996	145.57	<.0001

【Product method (交互作用項無し)】

Product method では、「治療 A が媒介変数 M をどの程度変化させるか」「媒介変数 M は結果変数 Y をどの程度変化させるか」の積を IE と考える。上記の2つの回帰モデルを Program 4-4に示した。

Program 4-4 Product method によるシミュレーションデータの解析

```

title 'product method 1: A⇒M: Aの推定値 β1';
proc mixed data=dat;
  class A L/ref=first;
  model M=A L/s;
  lsmeans A/pdiff;
run;
title 'product method 2: M⇒Y: Mの推定値 θ2, Aの推定値 θ1 (=DE)';
proc mixed data=dat;
  class A L/ref=first;
  model Y=A M L/s;
run;
title;
    
```

Program 4-4の2つの回帰モデルに対する2つの出力を以下に示す。Output 4-4には治療変数 A の媒介変数 M への推定値が出力されており、Output 4-5には、媒介変数 M のアウトカム Y への推定値が出力されている。それぞれの出力結果から DE は13.6237、IE は2.0307*4.2745=8.6782 となった。DE は Differential Method の結果と完全に一致し、IE は Differential Method の結果 8.6801とほぼ一致した。ただし、真のアウトカムモデルは治療変数 A と媒介変数 M の交互作用項を含むため、本結果はそもそもモデル誤特定下での推定結果であることに留意されたい。

Output 4-4 媒介変数回帰モデルの推定結果

Differences of Least Squares Means						
Effect	a	_a	Estimate	Standard Error	DF	t Value Pr > t
a	1	0	2.0307	0.02165	9997	93.80 <.0001

Output 4-5 アウトカム回帰モデルの推定結果

Solution for Fixed Effects						
Effect	a	c	Estimate	Standard Error	DF	t Value Pr > t
Intercept			-2.8166	0.05778	9996	-48.75 <.0001
a	1		13.6237	0.09359	9996	145.57 <.0001
a	0		0	.	.	.
m			4.2745	0.03153	9996	135.56 <.0001
l		1	4.4268	0.1170	9996	37.82 <.0001
l		0	0	.	.	.

4.1.3.1.3 公開マクロ%mediationの適用

本節では、4.1.1で紹介した公開マクロ%mediationを実際にシミュレーションデータへ適用し、出力結果を解説する。マクロが背後で生成しているプログラムも紹介することで、実際に適用されているモデルの確認や、出力されたモデルパラメータ推定値からの因果媒介効果の導出を通して、因果媒介分析の理解を深める。本節では治療変数 A と媒介変数 M の交互作用の有無により、2つの解析を実施する。まず、交互作用項を含む生成データの正しいモデル化により、反事実フレームワークを用いた因果媒介分析が一致推定を導いていることを示し、その後、交互作用項を含まずモデルが誤特定されている状況での解析結果も紹介し、解析結果への影響を確認する。

【%mediation マクロの実行プログラムと生成されたプログラム（交互作用項有り）】

Program 4-5にて、4.1.3.1.1で作成したシミュレーションデータへの%mediationマクロの適用方法を示した。Yvar=, avar=, mvar=, cvar=では、解析に用いるアウトカム Y, 治療変数 A, 媒介変数 M, 共変量 L のデータセット内の変数名を指定する。シミュレーションデータでは簡単のため、それぞれ、y, a, m, l という変数名を使用した。また、m=0として、CDE で制御される媒介変数 M の水準を指定している。Yreg=及び mreg=はそれぞれアウトカム回帰モデル及び媒介変数回帰モデルを特定するもので、シミュレーション設定に基づき、両方に線形回帰モデルとして linear を指定した。Interaction=true として治療変数 A と媒介変数 M の交互作用項をアウトカム回帰モデルで考慮することを指定した。また、標準誤差及び95%信頼区間の算出の方法の指定方法としては、計算負荷の都合により、boot=false としてデルタ法により実施する。Boot=true とすることでbootstrap法が利用可能である。%mediationマクロの詳細については4.1.1を参照されたい。

なお、マクロ生成プログラムをログで確認するために、options ステートメントで、2つのオプションを指定している。Mprint オプションは、実際にマクロで生成されたステートメントをログに出力するための指定である。また、mprintnest オプションは、マクロ内でさらにマクロが実行された場合に、ログに関連情報を出力するための指定である。公開マクロのように他者により作成されたSASマクロプログラムを利用する際には、マクロプログラムそのものの確認だけではなく、ログで実際に出力されたプログラムも確認することが、QCの観点からも重要な手順であると考えられる。Options noprint nomprintnest;にて、マクロ出力を抑制するデフォルト設定に戻すことが出来る。

Program 4-5 %mediation マクロの実行（交互作用項有り）

```
options mprint mprintnest;
%mediation(data=dat, yvar=y, avar=a, mvar=m, cvar=l, a0=0, a1=1, m=0, nc=1, yreg=linear, mreg=linear, interaction=true, casecontrol=, output=full, c=, boot=false);
options noprint nomprintnest ;
```

Program 4-6は、ログ出力結果から一部を抽出したものである。Interaction=true の指定により avar=で指定された変数 a と、mvar=で指定された変数 m の交互作用が int=a*m として作成され

ている。その次に2つの `reg` プロシージャにて、アウトカム Y に対する回帰モデルと、媒介変数 M に対する回帰モデルが実行されている。

また、`iml` プロシージャにて、出力されたパラメータを回帰推定量に代入することにより、種々の DE, IE が算出されていることがわかる。

Program 4-6 %mediation マクロで生成されたプログラム (ログからの抽出)

```

data datal;set datal;
int=a*m;
run;

proc reg data=datall covout outest=out1(drop=_model_/*省略*/);
model y=a m int l;
proc reg data=datall covout outest=out2(drop=_model_/*省略*/);
model m=a l;

proc iml;
/*各出力データout1,out2からtheta1-3, beta0-2を作成している箇所は省略*/

/*MARGINAL CDE*/;
effect[,1]=(theta1+theta3*0)*(1-0);
/*MARGINAL NDE*/;
effect[,2]=(theta1+theta3*beta0+theta3*beta1*0+(theta3*beta2*t(cmean)))*(1-0);
/*MARGINAL NIE*/;
effect[,3]=(theta2*beta1+theta3*beta1*0)*(1-0);
/* MARGINAL TNDE*/;
effect[,4]=(theta1+theta3*beta0+theta3*beta1*1+(theta3*beta2*t(cmean)))*(1-0);
/* MARGINAL TNIE*/;
effect[,5]=(theta2*beta1+theta3*beta1*1)*(1-0);
/*te marginal*/;
effect[,6]=(theta1+theta3*beta0+theta3*beta1*0+(theta3*beta2*t(cmean))+theta2*beta1+theta3*beta1*1)*(1-0);
    
```

【%mediation マクロプログラムの出力結果(交互作用項有り)】

Program 4-6の2つの回帰モデルに対する2つの出力を以下に示す。Output 4-6には媒介変数 M のアウトカム Y への推定値が出力されており、Output 4-7には、治療変数 A の媒介変数 M への推定値が出力されている。これらの推定値は、4.1.3.1.1で示したシミュレーションデータの真値を精度高く推定できている。

Output 4-6 &mediation による出力 (媒介変数 M のアウトカム Y への推定値)

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept(= θ_0)	1	0.49296	0.02156	22.86	<.0001
a(= θ_1)	1	1.51550	0.05059	29.95	<.0001
m(= θ_2)	1	2.51256	0.01169	214.88	<.0001
int(= θ_3)	1	3.48472	0.01175	296.63	<.0001
l(= θ_4)	1	4.49549	0.03738	120.26	<.0001

Output 4-7 &mediation による出力（治療変数 A の媒介変数 M への推定値）

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept(= β 0)	1	0.95990	0.01561	61.50	<.0001
a(= β 1)	1	2.03069	0.02165	93.80	<.0001
l(= β 2)	1	3.01504	0.02165	139.27	<.0001

それぞれの出力結果から各因果媒介効果の推定値は以下のように求められる。

- $CDE(m = 0) = \{(\hat{\theta}_1 + \hat{\theta}_3 m)(a - a^*) | m = 0\} = (1.51550 + 3.48472 * 0)(1 - 0) = 1.51550$
- $CDE(m = 1) = \{(\hat{\theta}_1 + \hat{\theta}_3 m)(a - a^*) | m = 1\} = (1.51550 + 3.48472 * 1)(1 - 0) = 5.00022$
- $NDE = (\hat{\theta}_1 + \hat{\theta}_3 \hat{\beta}_0 + \hat{\theta}_3 \hat{\beta}_1 a^* + \hat{\theta}_3 \hat{\beta}_2 \bar{l}) * (a - a^*)$
- $= (1.51550 + 3.48472 * 0.95990 + 3.48472 * 2.03069 * 0 + 3.48472 * 3.01504 * 0.5)(1 - 0) = 10.11377$
- $NIE = (\hat{\theta}_2 \hat{\beta}_1 + \hat{\theta}_3 \hat{\beta}_1 a)(a - a^*)$
- $= (2.51256 * 2.03069 + 3.48472 * 2.03069 * 1)(1 - 0) = 12.17862$
- $TE = NDE + NIE = 10.11377 + 12.17862 = 22.29239$

Program 4-6で iml で算出された各因果媒介効果の推定値の出力を Output 4-8に示す。アウトプット内の effect 列での「Pnde」と「tnie」がそれぞれ NDE(PDE)と NIE(TIE)に該当する。マクロで m=0を指定したため、CDE(M=0)が出力されている。これは θ1そのものであり、Output4.1.3.1.3-1の出力からもわかるように、CDE(m=0)=1.5155と一致している。また、NDE、NIE 及び TE は、10.1138, 12.1786, 22.2924となり、各回帰モデルからの推定値を因果媒介効果の推定量に代入して得られた値と一致していることもわかる。（なお、アウトプット内の effect 列での「tnde」と「pnie」がそれぞれ NDE(TDE)と NIE(PIE)に該当する）

Output 4-8 &mediation による出力（各因果媒介効果の推定値）

Obs	effect	Estimate	s_e	p_value	_95_CI_lower	_95_CI_upper
1	cde	1.5155	0.5059	0	1.4163	1.6147
2	pnde	10.1138	0.06049	0	9.9952	10.2323
3	pnie	5.1022	0.05935	0	4.9859	5.2186
4	tnde	17.1901	0.06070	0	17.0712	17.3091
5	tnie	12.1786	0.13197	0	11.9200	12.4373
6	total effect	22.2924	0.10099	0	22.0944	22.4903

【%mediation マクロプログラムの実行と生成プログラム（交互作用項無し）】

交互作用項を含まないモデルにて実行するプログラムを Program 4-7に示した。Program 4-5との違いは、Interaction=falseの変更だけである。この指定により、治療変数 A と媒介変数 M の交互作用項はアウトカム回帰モデルで考慮されないこととなる。

Program 4-7 %mediation マクロの実行（交互作用項無し）

```
options mprint mprintnest;
%mediation (data=dat, yvar=y, avar=a, mvar=m, cvar=1, a0=0, a1=1, m=0, nc=1, yreg=linear, mreg=linear, i
nteraction=false, casecontrol=, output=full, c=, boot=false);
options nomprint nomprintnest ;
```

Program 4-8は、ログ出力結果から一部を抽出したものである。Interaction=false の指定のため、Program 4-6とは異なり、avar=で指定された変数 a と、mvar=で指定された変数 m の交互作用が $int=a*m$ として作成されるプログラムは生成されていない。その次に2つの reg プロシージャにて、アウトカム Y に対する回帰モデルと、媒介変数 M に対する回帰モデルが実行されている。アウトカム Y に対する回帰モデルは主効果だけで構成されている。

また、iml プロシージャにて、出力されたパラメータを回帰推定量に代入することにより、種々の DE, IE が算出されているが、交互作用が無いため、Program 4-6とは異なり、回帰推定量は単純な式で与えられている。

Program 4-8 %mediation マクロで生成されたプログラム（ログからの抽出）

```
proc reg data=dat1 covout outest=out1(drop=_model_/*省略*/);
model y=a m 1;
proc reg data=dat1 covout outest=out2(drop=_model_/*省略*/);
model m=a 1;

proc iml;
/*各出力データout1,out2からtheta1-3, beta0-2を作成している箇所は省略*/

effect[,1]=(theta1)*(1-0);
effect[,2]=(theta2*beta1)*(1-0);
effect[,3]=(theta1+theta2*beta1)*(1-0);
```

【%mediation マクロプログラムの出力結果（交互作用項無し）】

Program 4-8の2つの回帰モデルに対する2つの出力を以下に示す。Output 4-9には媒介変数 M のアウトカム Y への推定値が出力されている。アウトカムモデルを誤特定しているため、4.1.3.1.1で与えられたシミュレーション設定の真のパラメータとは異なる値が推定されている。Output 4-10には、治療変数 A の媒介変数 M への推定値が出力されている。なお、Output 4-10は交互作用項の有無によらないため Output 4-7と完全に一致している。

Output 4-9 &mediation による出力（媒介変数 M のアウトカム Y への推定値）

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept(= θ_0)	1	-2.81659	0.05778	-48.75	<.0001
a(= θ_1)	1	13.62368	0.09359	145.57	<.0001
m(= θ_2)	1	4.27447	0.03153	135.56	<.0001
l(= θ_4)	1	4.42681	0.11704	37.82	<.0001

Output 4-10 &mediation による出力（治療変数 A の媒介変数 M への推定値）

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept(= β_0)	1	0.95990	0.01561	61.50	<.0001
a(= β_1)	1	2.03069	0.02165	93.80	<.0001
l(= β_2)	1	3.01504	0.02165	139.27	<.0001

それぞれの出力結果から各因果媒介効果の推定値は以下のように求められる。

- $CDE = \hat{\theta}_1(a - a^*) = 13.62368 * (1 - 0) = 13.62368$
- $NDE = \hat{\theta}_1(a - a^*) * (a - a^*) = 13.62368 * (1 - 0) = 13.62368$
- $NIE = (\hat{\theta}_2\hat{\beta}_1)(a - a^*) = (4.27447 * 2.03069)(1 - 0) = 8.68012$
- $TE = NDE + NIE = 13.62368 + 8.68012 = 22.3038$

Program 4-8で iml で算出された各因果媒介効果の推定値の出力を Output 4-11に示す。交互作用項が無いため、CDE と NDE が一致している。NDE(=CDE)、NIE 及び TE は、13.6237, 8.6801, 22.3028 となり、各回帰モデルからの推定値を因果媒介効果の推定量に代入して得られた値と一致していることもわかる。ただし、アウトカムモデルを誤特定しているため、因果媒介効果は正しくは推定されていない。なお、この結果は、4.1.3.1.2にて同様のモデル誤特定下で古典的な手法を適用した結果と一致していることに触れておく。

Output 4-11 &mediation による出力（各因果媒介効果の推定値）

Obs	Effect	Estimate	s_e	p_value	_95_CI_lower	_95_CI_upper
1	cde=nde	13.6237	0.09359	0	13.4402	13.8071
2	nie	8.6801	0.11253	0	8.4595	8.9007
3	total effect	22.3038	0.11499	0	22.0784	22.5292

4.1.3.1.4 PROC CAUSALMED の適用

本節では、4.1.2で紹介した PROC CAUSALMED を実際にシミュレーションデータへ適用し、出力結果を解説する。本節では治療変数 A と媒介変数 M の交互作用の有無により、2つの解析を実施する。まず、交互作用項を含む生成データの正しいモデル化により、反事実フレームワークを用いた因果媒介分析が一致推定を導いていることを示し、その後、交互作用項を含まずモデルが誤特定されている状況での解析結果も紹介し、解析結果への影響を確認する。

【PROC CAUSALMED の実行プログラム（交互作用項有り）】

Program 4-9にて、4.1.3.1.1で作成したシミュレーションデータへの PROC CAUSALMED の適用方法を示した。Model ステートメントでアウトカム回帰モデルを指定している。治療変数 A と媒介変数 M の交互作用項を A|M により、指定している(A M A*M という指定と等価)。

mediator ステートメントで媒介変数回帰モデルを特定する。シミュレーション設定に基づき、両方に線形回帰モデルを用いるため、特別なオプションは設定していない。また、covar ステートメントで共変量 L を指定しており、アウトカム回帰モデルと媒介変数回帰モデルの両方に利用される。標準誤差は bootstrap 法を用いて算出される。また、evaluate ステートメントでは、m=0及び1として、CDE で制御される媒介変数 M の水準を指定している。PROC CAUSALMED では、CDE はデフォルトで m の観測平均値を用いて算出されるため、確認のために evaluate 文を追加し、明示的に CDE に m の観測平均値を代入してみる。PROC CAUSALMED の詳細については4.1.2節を参照されたい。

PROC CAUSALMED プロシージャのオプション poutcomemod 及び pmedmod により、アウトカム及び媒介変数それぞれの回帰モデルのパラメータ推定値が得られる。まずはそれぞれの出力結果を確認し、さらに各種因果媒介効果推定量に代入することで CDE・NDE・NIE を確認してみよう。

Program 4-9 PROC CAUSALMED の実行（交互作用項有り）

```
proc causalmed data=dat /*decomp*/
                                poutcomemod /*Yへの回帰でのパラメータを出力*/
                                pmedmod      /*Mへの回帰でのパラメータを出力*/
                                ;
model   Y = A |M; /*AMの交互作用項がある場合(モデルが正しく特定されている)*/
mediator M = A;
covar L;
bootstrap bootci(bc) nsamples=1000 seed=4989;
evaluate 'm=0' M=0;
evaluate 'm=1' M=1;
evaluate 'm=E(M)' M=3.4789042661; /*CDEのデフォルト*/
evaluate 'high m(m=E(M)+1SD)' M=1(SD); /*CDEでmが高めの人(+1SD) */
run;
```

【因果媒介効果の出力結果（交互作用項有り）】

Program 4-9の2つの回帰モデルに対する2つの出力を以下に示す。Output 4-12には媒介変数 M のアウトカム Y への推定値が出力されており、Output 4-13には、治療変数 A の媒介変数 M への推定値が出力されている。これらの推定値は、4.1.3.1.1で示したシミュレーションデータの真値を精度高く推定できている。

Output 4-12 CAUSALMED による出力（媒介変数 M のアウトカム Y への推定値）

Outcome Model Estimates						
Parameter	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept(= θ_0)	0.4930	0.02156	0.4507	0.5352	522.8469	<.0001
l(= θ_4)	4.4955	0.03737	4.4222	4.5687	14468.9250	<.0001
a(= θ_1)	1.5155	0.05058	1.4164	1.6146	897.6880	<.0001
m(= θ_2)	2.5126	0.01169	2.4897	2.5355	46198.0816	<.0001
a*m(= θ_3)	3.4847	0.01174	3.4617	3.5077	88032.6176	<.0001
Scale	1.0061	0.007114	0.9923	1.0202		

Output 4-13 CAUSALMED による出力（治療変数 A の媒介変数 M への推定値）

Mediator Model Estimates						
Parameter	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept(= β 0)	0.9599	0.01561	0.9293	0.9905	3783.0189	<.0001
l(= β 2)	3.0150	0.02165	2.9726	3.0575	19401.4590	<.0001
a(= β 1)	2.0307	0.02165	1.9883	2.0731	8800.9318	<.0001
Scale	0.9992	0.007066	0.9855	1.0132		

それぞれの出力結果から各因果媒介効果の推定値は以下のように求められる。

- $CDE(m = 0) = \{(\hat{\theta}_1 + \hat{\theta}_3 m)(a - a^*) | m = 0\} = (1.5155 + 3.4847 * 0)(1 - 0) = 1.5155$
- $CDE(m = 1) = \{(\hat{\theta}_1 + \hat{\theta}_3 m)(a - a^*) | m = 1\} = (1.5155 + 3.4847 * 1)(1 - 0) = 5.0002$
- $CDE(m = 3.4789042661) = \{(\hat{\theta}_1 + \hat{\theta}_3 m)(a - a^*) | m = 3.4789042661\} = (1.5155 + 3.4847 * 3.4789042661)(1 - 0) = 13.6384$
- $NDE = (\hat{\theta}_1 + \hat{\theta}_3 \hat{\beta}_0 + \hat{\theta}_3 \hat{\beta}_1 a^* + \hat{\theta}_3 \hat{\beta}_2 \bar{l}) * (a - a^*)$
- $= (1.5155 + 3.4847 * 0.9599 + 3.4847 * 2.0307 * 0 + 3.4847 * 3.0150 * 0.5)(1 - 0) = 10.11365$
- $NIE = (\hat{\theta}_2 \hat{\beta}_1 + \hat{\theta}_3 \hat{\beta}_1 a)(a - a^*)$
- $= (2.5126 * 2.0307 + 3.4847 * 2.0307 * 1)(1 - 0) = 12.17872$
- $TE = NDE + NIE = 10.11365 + 12.17872 = 22.29237$

続いて、Program 4-9から得られる因果媒介効果の推定値を確認する。Output 4-14では、NDE 及び NIE はそれぞれのシミュレーションデータでの真値10.19及び12.00を一致推定できていると推察される。なお、BOOTSTRAP ステートメントにて bootci(bc)を指定したため、ブートストラップに基づく標準誤差及び95%信頼区間が出力されている。

Output 4-14 CAUSALMED による出力（各種因果媒介効果の推定値）

Summary of Effects									
	Estimate	Standard Error	Bootstrap Standard Error	Wald 95% Confidence Limits		Bootstrap Bias Corrected 95% Confidence Limits		Z	Pr > Z
Total Effect	22.2655	0.1007	0.09947	22.0681	22.4629	22.0647	22.4519	221.07	<.0001
Controlled Direct Effect (CDE)	13.6385	0.02989	0.03020	13.5799	13.6971	13.5824	13.6995	456.36	<.0001
Natural Direct Effect (NDE)	10.1003	0.06057	0.06143	9.9816	10.2190	9.9890	10.2280	166.75	<.0001
Natural Indirect Effect (NIE)	12.1652	0.1318	0.1317	11.9068	12.4235	11.9022	12.4071	92.30	<.0001
Percentage Mediated	54.6369	0.3856	0.3881	53.8811	55.3926	53.8597	55.3564	141.70	<.0001
Percentage Due to Interaction	15.8909	0.1766	0.1758	15.5448	16.2370	15.5298	16.2098	89.98	<.0001
Percentage Eliminated	38.7459	0.2882	0.2838	38.1811	39.3108	38.1803	39.2806	134.45	<.0001

次に、evaluate ステートメントで、媒介変数 M を1に制御した場合の出力結果が Output 4-15で

ある。タイトルに制御情報が表示されている。CDE 行の estimate を確認すると、シミュレーションデータでの真値5.00を推定できていることがわかる。Percent eliminated=(22.2655-5.0002)/22.2655=0.775428であることも確認できる。なお、NDE 及び NIE は、Output 4-8で出力されている pnde 及び tnie と同様の結果が得られている。

Output 4-15 CAUSALMED による出力 (CDE(m=1)の推定値)

Summary of Effects: m=1									
	Estimate	Standard Error	Bootstrap Standard Error	Wald 95% Confidence Limits		Bootstrap Bias Corrected 95% Confidence Limits		Z	Pr > Z
Total Effect	22.2655	0.1007	0.09947	22.0681	22.4629	22.0647	22.4519	221.07	<.0001
Controlled Direct Effect (CDE)	5.0002	0.04169	0.04213	4.9185	5.0819	4.9226	5.0856	119.94	<.0001
Natural Direct Effect (NDE)	10.1003	0.06057	0.06143	9.9816	10.2190	9.9890	10.2280	166.75	<.0001
Natural Indirect Effect (NIE)	12.1652	0.1318	0.1317	11.9068	12.4235	11.9022	12.4071	92.30	<.0001
Percentage Mediated	54.6369	0.3856	0.3881	53.8811	55.3926	53.8597	55.3564	141.70	<.0001
Percentage Due to Interaction	54.6877	0.2125	0.2109	54.2711	55.1042	54.2908	55.1198	257.32	<.0001
Percentage Eliminated	77.5427	0.2026	0.2051	77.1457	77.9398	77.1425	77.9346	382.78	<.0001

【PROC CAUSALMED プログラム (交互作用項無し)】

交互作用項を含まないモデルにて実行するプログラムを Program 4-10に示した。Program 4-9との違いは、model 文から A と M の交互作用項を除いたことである。また、交互作用項を含まない場合、CDEはNDEと等しくなるため、evaluate ステートメントは指定していない。

Program 4-10 PROC CAUSALMED の実行 (交互作用項無し)

```
proc causalmed data=dat /*decomp*/
                                poutcomemod /*Yへの回帰でのパラメータを出力*/
                                pmedmod      /*Mへの回帰でのパラメータを出力*/
                                ;
model Y = A M; /*交互作用項が無い場合(モデルが正しく特定されていない)*/
mediator M = A;
covar L;
bootstrap bootci(bc) nsamples=1000 seed=4989;
run;
```

【因果媒介効果の出力結果 (交互作用項無し)】

Program 4-10の2つの回帰モデルに対する2つの出力を以下に示す。Output 4-16には媒介変数 M のアウトカム Y への推定値が出力されている。アウトカムモデルを誤特定しているため、4.1.3.1.1で与えられたシミュレーション設定の真のパラメータとは異なる値が推定されている。同様に誤特定している、古典的な手法を利用した結果 Output 4-5及び%mediation を利用した結果 Output 4-15とほぼ同様の結果が出力されている。Output 4-17には、治療変数 A の媒介変数 M への推定値が出力されている。なお、Output 4-17は交互作用項の有無によらないため Output 4-13と完全に一致している。

Output 4-16 CAUSALMED による出力（媒介変数 M のアウトカム Y への推定値）（交互作用項無し）

Outcome Model Estimates						
Parameter	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
			Intercept(= θ_0)	-2.8166		
l(= θ_4)	4.4268	0.1170	4.1975	4.6562	1431.2298	<.0001
a(= θ_1)	13.6237	0.09357	13.4403	13.8071	21198.2479	<.0001
m(= θ_2)	4.2745	0.03153	4.2127	4.3363	18382.8267	<.0001
Scale	3.1502	0.02228	3.1069	3.1942		

Output 4-17 CAUSALMED による出力（治療変数 A の媒介変数 M への推定値）（交互作用項無し）

Mediator Model Estimates						
Parameter	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
			Intercept(= β_0)	0.9599		
l(= β_2)	3.0150	0.02165	2.9726	3.0575	19401.4590	<.0001
a(= β_1)	2.0307	0.02165	1.9883	2.0731	8800.9318	<.0001
Scale	0.9992	0.007066	0.9855	1.0132		

それぞれの出力結果から各因果媒介効果の推定値は以下のように求められる。

- $CDE = \hat{\theta}_1(a - a^*) = 13.6237 * (1 - 0) = 13.6237$
- $NDE = \hat{\theta}_1(a - a^*) * (a - a^*) = 13.6237 * (1 - 0) = 13.6237$
- $NIE = (\hat{\theta}_2 \hat{\beta}_1)(a - a^*) = (4.2745 * 2.0307)(1 - 0) = 8.6802$
- $TE = NDE + NIE = 13.6237 + 8.6802 = 22.3039$

続いて、Program 4-10から得られる因果媒介効果の推定値の出力を Output 4-18に示す。交互作用項が無いため、CDE と NDE が一致している。NDE(=CDE)、NIE 及び TE は、13.6237, 8.6801, 22.3038となり、各回帰モデルからの推定値を因果媒介効果の推定量に代入して得られた値とほぼ一致していることもわかる。ただし、アウトカムモデルを誤特定しているため、因果媒介効果は正しくは推定されていない。なお、NDE 及び NIE は、Output 4-11で出力されている pnde 及び tnie と同様の結果が得られている。

また、交互作用項を指定していないため、Percentage Due to Interaction の行に0が出力されている。

なお、NDE 及び NIE の推定値は、4.1.3.1.1の Difference Method による DE・IE の値と完全一致し、Product method による推定値とも類似している。モデルが正しく指定されていないためこの結果の解釈は難しいものの、交互作用項を考慮しない場合には、Difference Method の DE 及び IE は、因果媒介分析での CDE/NDE 及び NIE と同一の値を与えることがわかった。ただし、既に PROC CAUSALMED がパッケージ化されている状況下にて Difference Method や Product

Method を用いる利点は無いと思われる。

Output 4-18 CAUSALMED による出力（各種因果媒介効果の推定値）（交互作用項無し）

Summary of Effects									
	Estimate	Standard Error	Bootstrap Standard Error	Wald 95% Confidence Limits		Bootstrap Bias Corrected 95% Confidence Limits		Z	Pr > Z
Total Effect	22.3038	0.1150	0.1223	22.0785	22.5291	22.0476	22.5236	194.00	<.0001
Controlled Direct Effect (CDE)	13.6237	0.09357	0.1274	13.4403	13.8071	13.3795	13.8856	145.60	<.0001
Natural Direct Effect (NDE)	13.6237	0.09357	0.1274	13.4403	13.8071	13.3795	13.8856	145.60	<.0001
Natural Indirect Effect (NIE)	8.6801	0.1125	0.1236	8.4596	8.9006	8.4463	8.9457	77.15	<.0001
Percentage Mediated	38.9176	0.4010	0.4927	38.1318	39.7035	37.9868	39.9232	97.06	<.0001
Percentage Due to Interaction	0
Percentage Eliminated	38.9176	0.4010	0.4927	38.1318	39.7035	37.9868	39.9232	97.06	<.0001

4.1.3.2 SAS help 事例紹介

本節では SAS/STAT の CAUSALMED プロシージャのマニュアル (SAS Institute Inc. 2017) で用いている Cognitive データセットに対する適用事例を紹介する。Cognitive データセットに対する CAUSALMED プロシージャの適用事例については、上記の SAS マニュアル、矢田・魚住 (2019)、及び矢田ら (2020) に結果の解釈も含めて詳細に記載されているため、適宜そちらを参照されたい。そのため、ここでは簡略的な記載に書き留めるが、実データに対して容易に実行が可能であり、因果媒介分析を実感するには適した事例であると考えられるため、是非一度実行してみることをお勧めする。

Cognitive データセットは、Marjoribanks (1974) で議論された教育モデルに基づき作成されており、シミュレートされたデータセットである。以下に示す変数を含む300名からなるデータセットである。

- SubjectID : 患者 ID
- FamSize : 家族の人数
- SocStatus : 子供の社会的ステータス
- Encourage : 親の子供への励ましスコア (治療変数)
- Motivation : 子供の学習意欲スコア (媒介変数)
- CogPerform : 子供の認知テストのスコア (アウトカム)

https://go.documentation.sas.com/doc/en/pgmsascdc/9.4_3.3/statug/statug_code_cmedex1.htm

「子供の社会的ステータス」とは世帯収入、両親の職業、両親の教育レベルからスコア化された値であり、「親の子供への励ましスコア」は質問票への回答結果から得られたスコアであり、「子供の学習意欲スコア」は子供、教師等が評価した子供の学習意欲レベルの合計スコアである。

このデータの解析目的は、親の子供への励ましスコア (Encourage) が、子供の認知テストの

スコア (CogPerform) に影響を与えているかどうかを検討することと、子供の認知テストのスコアにおける子供の学習意欲スコア (Motivation) の媒介による効果を検討することである。連続値アウトカムであることから、回帰モデルによる効果の推定を考える。このとき、アウトカムに対する回帰モデル、及び媒介変数に対する回帰モデルは以下のように定義できる。Aは治療変数 (Encourage), Mは媒介変数 (Motivation), Yはアウトカム (CogPerform), Lは共変量ベクトル (FamSize, SocStatus) とし, $\beta_0 \sim \beta_2$ や $\theta_0 \sim \theta_4$ は対応する各パラメータを意味する。

$$E[M|A = a, L = l] = \beta_0 + \beta_1 a + \beta_2^T l$$

$$E[Y|A = a, M = m, L = l] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4^T l$$

このもとで CAUSALMED プロシージャの実行例を Program 4-11 に示す。各ステートメントの意味合いやオプションについては 4.1.2 参照。

Program 4-11 PROC CAUSALMED の実行 (交互作用項有り)

```
proc causalmed data= Cognitive decomp poutcomemod pmedmod;
  model CogPerform = Encourage | Motivation;
  mediator Motivation = Encourage;
  covar FamSize SocStatus;
run;
```

プロシージャの実行結果の一部抜粋を表 4-2～表 4-4 に示す。表 4-2 は各因果効果の推定値、表 4-3 は子供の認知テストのスコアに対する回帰モデルの推定値、表 4-4 は子供の学習意欲スコアに対する回帰モデルの推定値を示している。表 4-2 より、各効果の推定値はすべて有意であった。TE の推定値及び95%信頼区間は6.8421 (6.5618, 7.1224) であり、CDE, NDE, NIE の推定値及び95%信頼区間はそれぞれ4.1797 (4.0876, 4.2717), 4.1509 (4.0587, 4.2432), 2.6912 (2.4065, 2.9759) であった。TE は NDE (4.1509) と NIE (2.6912) に分解される。媒介された割合は $PM = NIE / TE = (TE - NDE) / TE$ であり、約40%であった。子供の認知テストのスコアに対する親の子供への励まし効果の4割が、子供の学習意欲の強化に起因した IE であったことを意味する。

表 4-3 及び表 4-4 の回帰モデルの推定値から、表 4-2 の各因果効果の確認が可能である。CAUSALMED プロシージャでは特に指定しない限り、治療変数 a, a^* にはそれぞれ $\bar{a} + 0.5, \bar{a}^* - 0.5$ が、その他変数に対しては平均値が用いられている。Cognitive データセットにおいて、FamSize, SocStatus, Encourage, Motivation の平均値はそれぞれ3.8033, 25.0733, 34.4167, 38.4700である。このとき、以下のように CDE, NDE, NIE は回帰パラメータの値と一致することが確認できる。

- $CDE = (\hat{\theta}_1 + \hat{\theta}_3 \bar{m})(\bar{a} + 0.5) - (\bar{a}^* - 0.5)$
 $= (0.9680 + 0.08349 * 38.4700)((34.4167 + 0.5) - (34.4167 - 0.5)) = 4.1797$
- $NDE = (\hat{\theta}_1 + \hat{\theta}_3(\hat{\beta}_0 + \hat{\beta}_1(\bar{a}^* - 0.5) + \hat{\beta}_2^T l))(\bar{a} + 0.5) - (\bar{a}^* - 0.5)$
 $= (0.9680 + 0.08349(11.2993 + 0.6879(34.4167 - 0.5) + (-0.1104 * 3.8033 + 0.1562 * 25.0733)))$

- * $((34.4167 + 0.5) - (34.4167 - 0.5)) = 4.1509$
- $$\begin{aligned} \text{NIE} &= (\hat{\beta}_1 \hat{\theta}_2 + \hat{\beta}_1 \hat{\theta}_3 (\bar{a} + 0.5))((\bar{a} + 0.5) - (\bar{a}^* - 0.5)) \\ &= (0.6879 * 0.9974 + 0.6879 * 0.08349(34.4167 + 0.5))((34.4167 + 0.5) - (34.4167 - 0.5)) \\ &= 2.6912 \end{aligned}$$

表 4-2 Cognitive データセットにおける TE, DE, IE

	推定値	標準誤差	95%信頼区間		Z	Pr > Z
総合効果 (TE)	6.8421	0.1430	6.5618	7.1224	47.84	<.0001
制御された直接効果 (CDE)	4.1797	0.04696	4.0876	4.2717	89.00	<.0001
自然な直接効果 (NDE)	4.1509	0.04706	4.0587	4.2432	88.21	<.0001
自然な間接効果 (NIE)	2.6912	0.1453	2.4065	2.9759	18.53	<.0001
媒介された割合 (PM)	39.3325	1.3704	36.6465	42.0184	28.70	<.0001

表 4-3 子供の認知テストのスコアに対する回帰モデルの推定値

パラメータ	推定値	標準誤差	95%信頼区間		Wald カイ 2 乗値	Pr>ChiSq
Intercept	-91.9579	3.0824	-97.9994	-85.9165	890.0082	<.0001
FamSize	-0.02471	0.01262	-0.04944	0.000023	3.8345	0.0502
SocStatus	0.01212	0.01763	-0.02243	0.04667	0.4725	0.4918
Encourage	0.9680	0.1021	0.7679	1.1681	89.8917	<.0001
Motivation	0.9974	0.08935	0.8223	1.1725	124.6085	<.0001
Encourage*Motivation	0.08349	0.002275	0.07903	0.08794	1346.3079	<.0001

表 4-4 子供の学習意欲スコアに対する回帰モデルの推定値

パラメータ	推定値	標準誤差	95%信頼区間		Wald カイ 2 乗値	Pr>ChiSq
Intercept	11.2993	0.8412	9.6506	12.9480	180.4392	<.0001
FamSize	-0.1104	0.01294	-0.1357	-0.08500	72.7097	<.0001
SocStatus	0.1562	0.01805	0.1208	0.1916	74.9112	<.0001
Encourage	0.6879	0.03604	0.6172	0.7585	364.3009	<.0001

4.2 R

本節では、4.1.3.1と同様のシミュレーションデータを用いて、R パッケージ “regmedint” (Yoshida et al. 2022) による本稿で紹介した因果媒介分析の実装方法を紹介する。R パッケージ “regmedint” は次を実行することでインストールできる。

```
install.packages("regmedint")
library(regmedint)
```

本パッケージは、4.1で紹介した SAS の公開マクロやプロシージャと概ね同様のものとなっている。異なる点としては、本パッケージは媒介因子回帰モデル又はアウトカム回帰モデルに共変量との交互作用項を追加したモデルにも対応している。本パッケージでは、関数 regmedint を用いて回帰モデルに基づく因果媒介分析を実行できる。関数 regmedint の引数の詳細は以下の通りであり、引数に与えている値はデフォルト値である。

```

regmedint(
  data, # 解析に用いるデータフレームを指定
  yvar, # アウトカムの変数名を指定 (スカラー値)
  avar, # 治療の変数名を指定 (スカラー値)
  mvar, # 媒介因子の変数名を指定 (スカラー値)
  cvar, # 共変量の変数名を指定 (ベクトル値)
  emm_ac_mreg = NULL, # 媒介因子回帰モデルにおける治療と交互作用項をもつ
                        # 共変量の変数名を指定 (ベクトル値)
  emm_ac_yreg = NULL, # アウトカム回帰モデルにおける治療と交互作用項をもつ
                        # 共変量の変数名を指定 (ベクトル値)
  emm_mc_yreg = NULL, # アウトカム回帰モデルにおける媒介因子と交互作用項をもつ
                        # 共変量の変数名を指定 (ベクトル値)
  eventvar = NULL, # 生存時間アウトカムの場合のみイベントの指示変数を指定 (スカラー値)
                   # survival package 利用時と同様にイベントで1, 打ち切りで0をもつ
  a0, # 対照治療を示す変数値を指定
  a1, # 被験治療を示す変数値を指定
  m_cde, # CDE 推定時の媒介因子の変数値を指定
  c_cond, # 潜在アウトカムの条件付き期待値算出時に用いる共変量の変数値を指定 (ベクトル値)
           # ベクトル値の長さは cvar と同一
           # 共変量が連続変数の場合は共変量の平均値を指定することで周辺効果を推定可能
  mreg, # 媒介因子回帰モデルタイプを指定; "linear" or "logistic"
  yreg, # アウトカム回帰モデルタイプを指定: "linear", "logistic", "loglinear",
         # "poisson", "negbin", "survCox", "survAFT_exp", or "survAFT_weibull"
  interaction = TRUE, # アウトカム回帰モデルにおける治療と媒介因子の交互作用有無を指定
  casecontrol = FALSE, # ケースコントロール研究の場合は TRUE を指定
  na_omit = FALSE # NA が格納された変数をもつレコードを削除する場合は TRUE を指定
)

```

4.1.3.1のシミュレーションデータに対して上記関数を適用するには、以下のコードを実行すれば良い。まずは周辺効果を推定するために、`c_cond`には共変量変数1の平均値を指定した。

```

regmedint_result <- regmedint(data=dat,
                              yvar="y",
                              avar="a",
                              mvar="m",
                              cvar=c("1"),
                              emm_ac_mreg = NULL,
                              emm_ac_yreg = NULL,
                              emm_mc_yreg = NULL,
                              a0=0,
                              a1=1,
                              m_cde=1,
                              # c_cond は周辺効果を推定するために平均値を指定
                              c_cond=mean(dat$1),
                              mreg="linear",
                              yreg="linear",
                              interaction = TRUE,
                              casecontrol = FALSE,
                              na_omit = FALSE)
summary(regmedint_result)

```

上記コードの実行結果は以下の通りである。標準誤差 (se) はデルタ法により算出されており、特に指定がなければ95%信頼区間が出力される。アウトカム回帰モデルにおける治療と媒介因子の交互作用項を設定している場合は、`m_cde`と`c_cond`で指定する値に応じて結果が変わ

るため注意が必要である。なお、cde は CDE(m=1), pnde は NDE, tnie は NIE, tnde は TDE, pnle は PIE を指す。

```
### Mediation analysis
      est      se      z p      lower      upper
cde  5.0002250 0.041698196 119.91466 0  4.9184981  5.0819520
pnde 10.1137708 0.060485405 167.21010 0  9.9952216 10.2323200
tnie 12.1786073 0.131966076  92.28589 0 11.9199585 12.4372561
tnde 17.1901486 0.060701402 283.19195 0 17.0711761 17.3091212
pnle  5.1022295 0.059351712  85.96600 0  4.9859023  5.2185567
te   22.2923781 0.100994609 220.72840 0 22.0944323 22.4903239
pm   0.5463126 0.003851442 141.84625 0  0.5387639  0.5538613

Evaluated at:
avar: a
  a1 (intervened value of avar) = 1
  a0 (reference value of avar) = 0
mvar: m
  m_cde (intervend value of mvar for cde) = 1
cvar: l
  c_cond (covariate vector value) = 0.5
```

信頼区間幅を変更したい場合や、CDE 算出時の媒介因子の変数値を変更したい場合、潜在アウトカムの条件付き期待値算出時の共変量値を変更したい場合等は、以下のように関数 `summary` を用いて簡単に出力することができる。

```
# 信頼区間幅を変更したい場合
summary(regmedint_result, level=0.90)

# CDE 算出時の媒介因子の変数値を変更したい場合
summary(regmedint_result, c_cond=1)

# 潜在アウトカムの条件付き期待値算出時の共変量値を変更したい場合
summary(regmedint_result, m_cde=0)
```

関数 `regmedint` ではどのように点推定値が算出されているかを確認するために、関数 `regmedint` を用いずに同一の解析結果を導くことを試みる。媒介因子回帰モデルとアウトカム回帰モデルを別々に実行し、それぞれの回帰係数の推定値と3.2.1で示した潜在アウトカムの期待値を求める式を用いて因果効果を分解した。その R コード実行後に `Console` に出力されたものを解析結果とともに以下に示す。今回の事例のような簡単な状況であれば、関数 `regmedint` を用いずとも簡単に点推定を求めることができる。ただし、推定値の標準誤差を算出するのは複雑であるため、関数 `regmedint` を用いるのが望ましいだろう。

```
> # mediator
> model_med <- summary(lm(formula = m ~ a + l, data = dat))
>
> # outcome
> model_out <- summary(lm(formula = y ~ a + m + a:m + l, data = dat))
>
> beta_0 <- model_med$coefficients[1,1]
> beta_1 <- model_med$coefficients[2,1]
> beta_2 <- model_med$coefficients[3,1]
```

```

>
> theta_0 <- model_out$coefficients[1,1]
> theta_1 <- model_out$coefficients[2,1]
> theta_2 <- model_out$coefficients[3,1]
> theta_3 <- model_out$coefficients[5,1]
> theta_4 <- model_out$coefficients[4,1]
>
> # CDE(m=0)
> (CDEm0 <- theta_1)
[1] 1.515504
> # CDE(m=1)
> (CDEm1 <- theta_1+theta_3)
[1] 5.000225
> # NDE
> (PNDE <- theta_1+theta_3*(beta_0+beta_2*mean(dat$1)))
[1] 10.11377
> # NIE
> (TNIE <- (theta_2+theta_3)*beta_1)
[1] 12.17861
> # TDE
> (TNDE <- theta_1+theta_3*(beta_0+beta_1+beta_2*mean(dat$1)))
[1] 17.19015
> # PIE
> (PNIE <- theta_2*beta_1)
[1] 5.10223
> # TE
> (TE <- ePDE+eTIE)
[1] 22.29238
> # PM
> (PM <- TNIE/(PNDE+TNIE))
[1] 0.5463126

```

また、ブートストラップ法に基づく信頼区間の算出も可能である。実行例を以下に示す。この例では、Rのパッケージとして“tidyverse”，“modelr”，“future”，及び“furr”を用いている。

```

install.packages("modelr")
library(modelr)
install.packages("future")
library(future)
future::plan(sequential)
install.packages("furr")
library(furr)

## Bootstrapping
tib_obj <- dat %>%
  modelr::bootstrap(n = 1000) %>%
  ## Resampled data objects are in the list column named strap.
  mutate(boot_fit = future_map(strap, function(strap) {
    ## Note the change in the data argument.
    regmedint_result <- regmedint(data = as_tibble(strap),
                                   yvar="y",
                                   avar="a",
                                   mvar="m",
                                   cvar=c("1"),
                                   a0=0,
                                   a1=1,

```

```

        m_cde=1,
        c_cond=mean(dat$1),
        mreg="linear",
        yreg="linear",
        interaction = TRUE)

## Trick to return a row tibble
mat <- t(matrix(coef(regmedint_result)))
colnames(mat) <- names(coef(regmedint_result))
as_tibble(mat)
})) %>%
select(-strap) %>%
unnest(cols = c(boot_fit))

tib_obj %>%
pivot_longer(-.id) %>%
mutate(name = factor(name, levels = names(coef(regmedint_result)))) %>%
group_by(name) %>%
summarize(lower_boot = quantile(value, probs = 0.025),
           upper_boot = quantile(value, probs = 0.975)) %>%
mutate(lower_delta = confint(regmedint_result)[,"lower"],
       upper_delta = confint(regmedint_result)[,"upper"])

```

本節のシミュレーションデータは N=10,000 と大標本であるため、以下の通りデルタ法とブートストラップ法の計算結果は一致した結果となった。なお、cde は CDE(m=1)、pnde は NDE、tnie は NIE、tnde は TDE、pnie は PIE を指す。

```

# A tibble: 7 x 5
  name lower_boot upper_boot lower_delta upper_delta
<fct> <dbl> <dbl> <dbl> <dbl>
1 cde 4.92 5.08 4.92 5.08
2 pnde 10.0 10.2 10.0 10.2
3 tnie 11.9 12.4 11.9 12.4
4 tnde 17.1 17.3 17.1 17.3
5 pnie 4.99 5.22 4.99 5.22
6 te 22.1 22.5 22.1 22.5
7 pm 0.539 0.554 0.539 0.554

```

最後にモデル誤特定下では適切な推定ができないことを確認しておく。4.1.3.1.1で示したとおり、本節で利用したシミュレーションデータの真の構造は、治療と媒介因子の交互作用項をもつ。このとき、その交互作用項を設定せずに推定したときの R コードと結果を以下に示す。推定結果をみると、モデル誤特定下では因果効果の分解結果が大きく変わることが分かる一方、TE への影響は軽微であった。

```

> regmedint_result_modelmiss <- regmedint(data=dat,
+                                       yvar="y",
+                                       avar="a",
+                                       mvar="m",
+                                       cvar=c("1"),
+                                       emm_ac_mreg = NULL,
+                                       emm_ac_yreg = NULL,
+                                       emm_mc_yreg = NULL,
+                                       a0=0,
+                                       a1=1,

```

```

+           m_cde=1,
+           c_cond=mean(dat$l),
+           mreg="linear",
+           yreg="linear",
+           interaction = FALSE,
+           casecontrol = FALSE,
+           na_omit = FALSE
+ )
> summary(regmedint_result_modelmiss)
## (中略)
### Mediation analysis
      est      se      Z p      lower      upper
cde  13.6236821 0.09359042 145.56706 0 13.4402483 13.8071160
pnde 13.6236821 0.09359042 145.56706 0 13.4402483 13.8071160
tnie  8.6801115 0.11253349  77.13359 0  8.4595499  8.9006731
tnde 13.6236821 0.09359042 145.56706 0 13.4402483 13.8071160
pnie  8.6801115 0.11253349  77.13359 0  8.4595499  8.9006731
te   22.3037936 0.11498878 193.96496 0 22.0784198 22.5291675
pm    0.3891765 0.00401045  97.04059 0  0.3813161  0.3970368

Evaluated at:
avar: a
  a1 (intervened value of avar) = 1
  a0 (reference value of avar)  = 0
mvar: m
  m_cde (intervend value of mvar for cde) = 1
cvar: l
  c_cond (covariate vector value) = 0.5

Note that effect estimates do not vary over m_cde and c_cond values when interaction = FALSE.

```

5. 付録

5.1 Cross-world conditional independence

NPSEM (Non-Parametric Structural Equation Model) は因果 DAG が表す反事実モデルであり, [Robins \(1986\)](#)によって提案された (Robins は FCISTG (Finest Causally Interpreted Structural Tree Graph) と呼んでいる). NPSEM で表される因果 DAG を実際に得られるデータと結びつけるためには因果マルコフ仮定 (当該変数の親を条件として, 変数とその非子孫と独立) を満たす必要があり, 追加の統計的独立性の仮定が必要である. [Pearl \(2000\)](#)はすべての反事実を含む誤差項が独立であることを仮定した NPSEM-IE (NPSEM with Independent Errors) を提案した. 一方 [Robins \(1986\)](#)は, 当該変数の親の反事実を含む値 (one-step ahead counterfactuals) を条件としたときの変数の同時独立のみを仮定し, これを FFRCISTG (Finest Fully Randomized Causally Interpreted Structured Tree Graph) と呼んだ. NPSEM-IE は FFRCISTG よりも多くの独立性の仮定を置いており, NPSEM-IE であれば FFRCISTG であるがその逆は成り立たない ([Robins and Richardson 2010](#)). 仮定 A4は Cross-world conditional independence と呼ばれ, NPSEM-IE に基づく因果 DAG ではその仮定が含まれるが, FFRCISTG の因果 DAG では仮定されない. 具体的に, 仮定 A4 : $(Y^{a,m} \perp M^{a=0}) \mid L, \forall a, m$ は, 仮にバックドア基準を満たさない共変量 L があつたとしても, 完全交換可能性 ($(Y^{a=0}, Y^{a=1}) \perp A$) を想定する NPSEM-IE の下であれば成立すると考えられるが, FFRCISTG では $a = 0$ および $a = 1$ に対する周辺交換可能性のみを想定するため, バックドア基準を満たさない共変量 L に対して $(Y^{a=1,m} \perp M^{a=0}) \mid L$ は成立しない仮定であると解釈される. 共変量のバックドア基準は, (1) 治療とアウトカムのバックドア経路が当該共変量で条件づけることによりブロックされる, (2) かつ, 当該共変量が治療の子孫ではない, であり, 共変量がバックドア基準を満たすとき, 当該共変量で条件付けたもとで条件付交換可能性が成り立つ. FFRCISTG のもとでは, Faithful (治療からアウトカムへの矢線がない場合, すべての個人について治療の因果効果が存在しない) の場合に, ある共変量で条件付けたもとで条件付交換可能性が満たされるとき, 当該共変量はバックドア基準を満たすことが示されている. [Pearl \(2001\)](#)では, 仮定 A2が満たされ, かつ治療の影響を受ける媒介変数・アウトカム間の交絡因子が存在しない場合に, 一般に仮定 A4が成り立つとしているが, 一方で [Nguyen et al. \(2023\)](#)では, 治療後の交絡因子がバックドア基準を満たさない場合があることが指摘されている. 仮定 A4に対する議論の詳細については, [VanderWeele and Vansteelandt \(2009\)](#)も参照されたい.

5.2 2値アウトカムに対する推定方法の詳細

5.2.1 アウトカム (Y) がまれであり, 媒介変数 (M) が連続量の場合

アウトカム (Y) が治療 (A) と独立なら $(Y^a \perp A \mid L)$ なら, OR^{TE} を識別することができ, 次式の左辺の OR^{TE} を右辺のデータから推定できる.

$$\frac{P(Y^{A=1} = 1 \mid l) / \{1 - P(Y^{A=1} = 1 \mid l)\}}{P(Y^{A=0} = 1 \mid l) / \{1 - P(Y^{A=0} = 1 \mid l)\}} = \frac{P(Y = 1 \mid A = 1, l) / \{1 - P(Y = 1 \mid A = 1, l)\}}{P(Y = 1 \mid A = 0, l) / \{1 - P(Y = 1 \mid A = 0, l)\}}$$

OR^{NDE}, OR^{NIE}は, 仮定 A1～A4, 線形回帰モデル及びロジスティック回帰モデルを組み合わせ、シンプルな回帰モデルベースの推定を行う。オッズ比をリスク比に近似することから、前提として、アウトカム (Y) がまれである必要があり、カットオフ値として10%未満を想定されてる (VanderWeele and Vansteelandt 2010)。

媒介変数 (M) を m に固定したオッズ比における CDE (以降, 「OR^{CDE}」と記載する) について、以下のとおり定義する。なお, OR^{CDE}では, アウトカム (Y) がまれである仮定は要らず, また, 治療 (A) と媒介変数 (M) に交互作用する場合, m の値によってOR^{CDE}は変化する。

$$OR^{CDE}(m) = \frac{P(Y^{A=1, M^m} = 1|l) / \{1 - P(Y^{A=1, M^m} = 1|l)\}}{P(Y^{A=0, M^m} = 1|l) / \{1 - P(Y^{A=0, M^m} = 1|l)\}}$$

さらに, 以下の2つのモデルを前提とする。1つ目は, アウトカム (Y) のロジスティック回帰モデル (交互作用項なし) を考える。

$$\text{logit}\{P(Y = 1|A = a, M = m, L = l)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' c$$

次に, 媒介変数 (M) に線形回帰モデルを考える。この線形回帰モデルの誤差項は, 平均0の分散 σ^2 の正規分布に従うと仮定する。

$$E[M|A = a, C = c] = \beta_0 + \beta_1 a + \beta_2' c$$

これらを満たすと, OR^{CDE}(m)は次のようになる。記載都合上, 展開途中で $A = 1$ を a , $A = 0$ を a^* に記載を変更している。以降も同様に記載の変更を行っている箇所がある。

$$\begin{aligned} OR^{CDE}(m) &= \frac{P(Y^{A=1, M^m} = 1|l) / \{1 - P(Y^{A=1, M^m} = 1|l)\}}{P(Y^{A=0, M^m} = 1|l) / \{1 - P(Y^{A=0, M^m} = 1|l)\}} \\ &= \frac{P(Y = 1|A = 1, m, l) / \{1 - P(Y = 1|A = 1, m, l)\}}{P(Y = 1|A = 0, m, l) / \{1 - P(Y = 1|A = 0, m, l)\}} \\ &= \frac{\exp[\theta_0 + \theta_1 a + \theta_2 m + \theta_4' c]}{\exp[\theta_0 + \theta_1 a^* + \theta_2 m + \theta_4' c]} \\ &= \exp[\theta_1 a - \theta_1 a^*] \end{aligned}$$

以上から, OR^{NDE}は以下になる。

$$\begin{aligned} OR^{NDE} &= \exp \left[\log \left\{ \frac{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}}{P(Y^{A=0, M^{a=0}} = 1|l) / \{1 - P(Y^{A=0, M^{a=0}} = 1|l)\}} \right\} \right] \\ &= \exp[\text{logit}\{P(Y^{A=1, M^{a=0}} = 1|l)\} - \text{logit}\{P(Y^{A=0, M^{a=0}} = 1|l)\}] \end{aligned}$$

$$\begin{aligned} & \sim \exp \left[\theta_0 + \theta_1 a + \theta_4' c + \theta_2 (\beta_0 + \beta_1 a^* + \beta_2' c) + \frac{1}{2} (\theta_2)^2 \sigma^2 \right. \\ & \quad \left. - \left\{ \theta_0 + \theta_1 a^* + \theta_4' c + \theta_2 (\beta_0 + \beta_1 a^* + \beta_2' c) + \frac{1}{2} (\theta_2)^2 \sigma^2 \right\} \right] \\ & = \exp [\theta_1 (a - a^*) + 0.5 \sigma^2 (a^2 - a^{*2})] \end{aligned}$$

OR^{NIE}も同様に以下になる.

$$\begin{aligned} \text{OR}^{\text{NIE}} & = \exp \left[\log \left\{ \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}} \right\} \right] \\ & = \exp [\text{logit}\{P(Y^{A=1, M^{a=1}} = 1|l)\} - \text{logit}\{P(Y^{A=1, M^{a=0}} = 1|l)\}] \\ & \sim \exp \left[\theta_0 + \theta_1 a + \theta_4' c + \theta_2 (\beta_0 + \beta_1 a + \beta_2' c) + \frac{1}{2} (\theta_2)^2 \sigma^2 \right. \\ & \quad \left. - \left\{ \theta_0 + \theta_1 a + \theta_4' c + \theta_2 (\beta_0 + \beta_1 a^* + \beta_2' c) + \frac{1}{2} (\theta_2)^2 \sigma^2 \right\} \right] \\ & = \exp [\theta_2 \beta_1 (a - a^*)] \end{aligned}$$

続いて, OR^{TE}は以下となる.

$$\begin{aligned} \text{OR}^{\text{TE}} & = \exp \left[\log \left\{ \frac{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}}{P(Y^{A=0, M^{a=0}} = 1|l) / \{1 - P(Y^{A=0, M^{a=0}} = 1|l)\}} \right\} \right] \\ & \quad \times \exp \left[\log \left\{ \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}} \right\} \right] \\ & = E[Y^{A=1, M^{a=0}} - Y^{A=0, M^{a=0}} | l] \times E[Y^{A=1, M^{a=1}} - Y^{A=1, M^{a=0}} | l] \\ & = \exp [(\theta_1 + \theta_2 \beta_1)(a - a^*) + 0.5 \sigma^2 (a^2 - a^{*2})] \end{aligned}$$

また, PMについては, 以下となる.

$$\begin{aligned}
 \text{PM} &= \log \left\{ \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}} \right\} \\
 &\quad / \left[\log \left\{ \frac{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}}{P(Y^{A=0, M^{a=0}} = 1|l) / \{1 - P(Y^{A=0, M^{a=0}} = 1|l)\}} \right\} \right. \\
 &\quad \left. + \log \left\{ \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}} \right\} \right] \\
 &= \frac{\theta_2 \beta_1 (a - a^*)}{(\theta_1 + \theta_2 \beta_1)(a - a^*) + 0.5\sigma^2(a^2 - a^{*2})}
 \end{aligned}$$

5.2.2 アウトカム (Y) がまれでなく、媒介変数 (M) が連続量の場合

アウトカム (Y) がまれではなく、アウトカムに対してもロジスティック回帰モデルではなく、対数線形回帰モデルを適用する場合でも、先の式は適用可能であり、アウトカム (Y) がカウント変数の場合にも適用することができる。この場合は、オッズ比ではなくリスク比での解釈になる。

アウトカム (Y) が以下に従う場合を想定する。

$$Y \sim \text{Poi}(\lambda) \text{ for } \lambda = \exp(\theta_0 + \theta_1 a + \theta_2 m + \theta_4' c)$$

このとき、回帰モデルは以下のとおり定義できる。

$$\log\{E(Y|A = a, M = m, L = l)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' c$$

リスク比の2値アウトカムに対する NDE は、ポアソンのカウントアウトカムの NDE と一致し、これは NIE でも同様である。

$$\text{RR}^{\text{NDE}} = \exp \left[\log \left\{ \frac{E(Y^{a=1, M^{a=0}} | l)}{E(Y^{a=0, M^{a=0}} | l)} \right\} \right]$$

また、負の二項分布が過分散ポアソン分布として、平均が一致することから、カウントアウトカムが負の二項分布でモデル化される場合にも拡張できるとしている。

5.2.3 アウトカム (Y) がまれであり、媒介変数 (M) が2値の場合

ここでは、アウトカム (Y) 及び媒介変数 (M) がともに2値である場合を想定し、次のモデルを考える。また、アウトカム (Y) はまれであるという仮定を置く。

$$\text{logit}\{P(M = 1|A = a, L = l)\} = \beta_0 + \beta_1 a + \beta'_2 c$$

$$\text{logit}\{P(Y = 1|A = a, M = m, L = l)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta'_4 c$$

OR^{CDE}については、媒介変数 (M) が連続量とした場合と同様に、 $M = m$ と固定した場合であることから計算可能である。OR^{NDE}は以下のとおりとなる。

$$\begin{aligned} \text{OR}^{\text{NDE}} &= \exp \left[\log \left\{ \frac{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}}{P(Y^{A=0, M^{a=0}} = 1|l) / \{1 - P(Y^{A=0, M^{a=0}} = 1|l)\}} \right\} \right] \\ &= \exp[\text{logit}\{P(Y^{A=1, M^{a=0}} = 1|l)\} - \text{logit}\{P(Y^{A=0, M^{a=0}} = 1|l)\}] \\ &\sim \exp \left[\log \left\{ \frac{\exp(\theta_0 + \theta_1 a + \theta'_4 c) + \exp(\theta_0 + \theta_1 a + \theta'_4 c + \theta_2 + \beta_0 + \beta_1 a^* + \beta'_2 c)}{1 + \exp[\beta_0 + \beta_1 a^* + \beta'_2 c]} \right\} \right. \\ &\quad \left. - \log \left\{ \frac{\exp(\theta_0 + \theta_1 a^* + \theta'_4 c) + \exp(\theta_0 + \theta_1 a^* + \theta'_4 c + \theta_2 + \beta_0 + \beta_1 a^* + \beta'_2 c)}{1 + \exp[\beta_0 + \beta_1 a^* + \beta'_2 c]} \right\} \right] \\ &= \left\{ \frac{\exp[\theta_0 + \theta_1 a + \theta'_4 c] + \exp[\theta_0 + \theta_1 a + \theta'_4 c + \theta_2 + \beta_0 + \beta_1 a^* + \beta'_2 c]}{\exp[\theta_0 + \theta_1 a^* + \theta'_4 c] + \exp[\theta_0 + \theta_1 a^* + \theta'_4 c + \theta_2 + \beta_0 + \beta_1 a^* + \beta'_2 c]} \right\} \\ &= \left\{ \frac{\exp[\theta_1 a] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta'_2 c])}{\exp[\theta_1 a^*] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta'_2 c])} \right\} \end{aligned}$$

OR^{NIE}も同様に以下のとおりとなる。

$$\begin{aligned} \text{OR}^{\text{NDE}} &= \exp \left[\log \left\{ \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}} \right\} \right] \\ &= \exp[\text{logit}\{P(Y^{A=1, M^{a=1}} = 1|l)\} - \text{logit}\{P(Y^{A=1, M^{a=0}} = 1|l)\}] \\ &\sim \exp \left[\log \left\{ \frac{\exp(\theta_0 + \theta_1 a + \theta'_4 c) + \exp(\theta_0 + \theta_1 a + \theta'_4 c + \theta_2 + \beta_0 + \beta_1 a + \beta'_2 c)}{1 + \exp[\beta_0 + \beta_1 a + \beta'_2 c]} \right\} \right. \\ &\quad \left. - \log \left\{ \frac{\exp(\theta_0 + \theta_1 a + \theta'_4 c) + \exp(\theta_0 + \theta_1 a + \theta'_4 c + \theta_2 + \beta_0 + \beta_1 a^* + \beta'_2 c)}{1 + \exp[\beta_0 + \beta_1 a^* + \beta'_2 c]} \right\} \right] \\ &= \left\{ \frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta'_2 c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a + \beta'_2 c]\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta'_2 c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta'_2 c]\}} \right\} \end{aligned}$$

以上より、OR^{TE}は以下となる。

$$\begin{aligned}
 \text{OR}^{\text{TE}} &= \exp \left[\log \left\{ \frac{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}}{P(Y^{A=0, M^{a=0}} = 1|l) / \{1 - P(Y^{A=0, M^{a=0}} = 1|l)\}} \right\} \right] \\
 &\quad \times \exp \left[\log \left\{ \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}} \right\} \right] \\
 &= \frac{\left\{ \exp[\theta_1 a] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2' c]) \right\}}{\left\{ \exp[\theta_1 a^*] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2' c]) \right\}} \\
 &\quad \times \frac{\left\{ \{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2' c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a + \beta_2' c]\} \right\}}{\left\{ \{1 + \exp(\beta_0 + \beta_1 a + \beta_2' c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2' c]\} \right\}}
 \end{aligned}$$

PM についても、同様に示す.

$$\begin{aligned}
 \text{PM} &= \log \left\{ \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}} \right\} \\
 &\quad / \left[\log \left\{ \frac{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}}{P(Y^{A=0, M^{a=0}} = 1|l) / \{1 - P(Y^{A=0, M^{a=0}} = 1|l)\}} \right\} \right] \\
 &\quad + \log \left\{ \frac{P(Y^{A=1, M^{a=1}} = 1|l) / \{1 - P(Y^{A=1, M^{a=1}} = 1|l)\}}{P(Y^{A=1, M^{a=0}} = 1|l) / \{1 - P(Y^{A=1, M^{a=0}} = 1|l)\}} \right\} \\
 &= \log \frac{\left\{ \{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2' c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a + \beta_2' c]\} \right\}}{\left\{ \{1 + \exp(\beta_0 + \beta_1 a + \beta_2' c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2' c]\} \right\}} \\
 &\quad / \left[\log \left[\frac{\left\{ \exp[\theta_1 a] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2' c]) \right\}}{\left\{ \exp[\theta_1 a^*] (1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2' c]) \right\}} \right] \right] \\
 &\quad \times \frac{\left\{ \{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2' c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a + \beta_2' c]\} \right\}}{\left\{ \{1 + \exp(\beta_0 + \beta_1 a + \beta_2' c)\} \{1 + \exp[\theta_2 + \beta_0 + \beta_1 a^* + \beta_2' c]\} \right\}}
 \end{aligned}$$

5.2.4 アウトカム (Y) がまれな場合の Difference Method と Product Method の近似

アウトカム (Y) がまれな場合は以下になる必要がある.

$$\phi_0 + \phi_1 a + \phi_4 c = \text{logit}\{P(Y = 1|A = a, L = l)\} \approx \log \{P(Y = 1|A = a, L = l)\}$$

そのため、以下のようになる.

$$\begin{aligned}
 \exp\{\phi_0 + \phi_1 a + \phi_4' c\} &\approx P(Y = 1|A = a, L = l) \\
 &= E[P(Y = 1|A = a, M = m, L = l)|A = a, L = l] \\
 &\approx E[\exp\{\theta_0 + \theta_1 a + \theta_2 M + \theta_4' l\}|A = a, L = l] \\
 &= \exp(\theta_0 + \theta_1 a + \theta_4' l) E[\exp(\theta_2 M) |A = a, L = l] \\
 &= \exp(\theta_0 + \theta_1 a + \theta_4' l) \exp\{\theta_2 (\beta_0 + \beta_1 a + \beta_2' l)\} + \frac{1}{2} \theta_2^2 \sigma^2 \\
 &= \exp\left(\theta_0 + \frac{1}{2} \theta_2^2 \sigma^2 + \beta_0 \theta_2\right) + (\theta_1 + \beta_1 \theta_2) a + (\theta_4 + \beta_2 \theta_2)' l
 \end{aligned}$$

すべての $A = a$ に以下は適用される.

$$\phi_1 \approx (\theta_1 + \theta_2 \beta_1)$$

そのため, 以下となる.

$$\phi_1 - \theta_1 \approx \theta_2 \beta_1$$

しかし, アウトカムがまれでない場合は, 媒介変数 (M) に対する回帰モデルの誤差項が等分散ではない, 又は正規分布していない場合, この近似は成り立たない. また, その場合は, $\beta_1 \theta_2$ 及び $\phi_1 - \theta_1$ のどちらも NIE としての解釈が保証されるものではない.

5.3 AFT モデルにおける NIE と NDE

生存時間 T には, 次のような AFT モデルを仮定する.

$$\log(T|A = a, M = m, L = l) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \theta_4' l + v\varepsilon$$

また, 分散 σ^2 の正規分布に従う媒介変数 M (連続値) には, 次のような線形モデルを仮定する.

$$E(M|A = a, L = l) = \beta_0 + \beta_1 a + \beta_2' l$$

仮定 A1~A4 の成立下において, NIE, NDE が次式で得られることを示す.

$$\begin{aligned}
 \text{NIE}^{\text{AFT}} &= \log\{E(T^{aM^a} | L = l)\} - \log\{E(T^{aM^{a^*}} | L = l)\} \\
 &= (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*) \\
 \text{NDE}^{\text{AFT}} &= \log\{E(T^{aM^{a^*}} | L = l)\} - \log\{E(T^{a^* M^{a^*}} | L = l)\} \\
 &= \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' l + \theta_2 \sigma^2)\}(a - a^*) + 0.5 \theta_3^2 \sigma^2 (a^2 - a^{*2})
 \end{aligned}$$

証明:

$$\begin{aligned}
 &E[T^{aM^{a^*}} | L = l] \\
 &= \int E[T^{am} | L = l, M^{a^*} = m] dP_{M^{a^*}}(M^{a^*} = m | L = l) \quad (\because \text{期待値の繰り返しの公式})
 \end{aligned}$$

$$\begin{aligned}
 &= \int E[T^{am}|L=l]dP_{M^{a^*}}(M^{a^*}=m|L=l) \quad (\because A4) \\
 &= \int E[T|A=a, M=m, L=l]dP_M(M=m|A=a^*, L=l) \quad (\because A1-A3) \\
 &= \int E\left[\exp(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' l + v\varepsilon)\right]dP_M(M=m|A=a^*, L=l) \quad (\because \text{AFT モデル}) \\
 &= \exp(\theta_0 + \theta_1 a + \theta_4' l)E[\exp(v\varepsilon)]E[\exp(\theta_2 M + \theta_3 aM)] \\
 &= \exp(\theta_0 + \theta_1 a + \theta_4' l)E[\exp(v\varepsilon)]\exp((\theta_2 + \theta_3 a)(\beta_0 + \beta_1 a^* + \beta_2' l) \\
 &\quad + 0.5(\theta_2 + \theta_3 a)^2 \sigma^2) \quad (\because M \text{ は分散 } \sigma^2 \text{ の正規分布})
 \end{aligned}$$

したがって、

$$\begin{aligned}
 &\log\{E[T^{aM^{a^*}}|L=l]\} \\
 &= \log\{E[e^{v\varepsilon}]\} + \theta_0 + \theta_1 a + \theta_4' l + (\theta_2 + \theta_3 a)(\beta_0 + \beta_1 a^* + \beta_2' l) + 0.5(\theta_2 + \theta_3 a)^2 \sigma^2
 \end{aligned}$$

よって、次式を得る。

$$\begin{aligned}
 \text{NIE}^{\text{AFT}} &= \log\{E(T^{aM^a}|L=l)\} - \log\{E(T^{aM^{a^*}}|L=l)\} \\
 &= (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*) \\
 \text{NDE}^{\text{AFT}} &= \log\{E(T^{aM^{a^*}}|L=l)\} - \log\{E(T^{a^*M^{a^*}}|L=l)\} \\
 &= \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' l + \theta_2 \sigma^2)\}(a - a^*) + 0.5\theta_3^2 \sigma^2 (a^2 - a^{*2})
 \end{aligned}$$

5.4 PHモデルにおけるNIEとNDE

生存時間 T には、次のような PH モデルを仮定する。

$$\lambda_T(t|A=a, M=m, L=l) = \lambda_T(t|A=0, M=0, L=0)e^{\gamma_1 a + \gamma_2 m + \gamma_3 am + \gamma_4' l}$$

また、分散 σ^2 の正規分布に従う媒介変数 M (連続値) には、次のような線形モデルを仮定する。

$$E(M|A=a, L=l) = \beta_0 + \beta_1 a + \beta_2' l$$

仮定 A1～A4の成立下において、アウトカムが稀な場合、NIE, NDE が次式で得られることを示す。

$$\begin{aligned}
 \text{NIE}^{\text{PH}} &= \log\{\lambda_{T^{aM^a}}(t|L=l)\} - \log\{\lambda_{T^{aM^{a^*}}}(t|L=l)\} \\
 &\approx (\gamma_2 \beta_1 + \gamma_3 \beta_1 a)(a - a^*) \\
 \text{NDE}^{\text{PH}} &= \log\{\lambda_{T^{aM^{a^*}}}(t|L=l)\} - \log\{\lambda_{T^{a^*M^{a^*}}}(t|L=l)\}
 \end{aligned}$$

$$\approx \{\gamma_1 + \gamma_3(\beta_0 + \beta_1 a^* + \beta'_2 l + \gamma_2 \sigma^2)\}(a - a^*) + 0.5\gamma_3^2 \sigma^2 (a^2 - a^{*2})$$

証明：

PH モデルの下では、次式がなりたつ。

$$\lambda_{T^{aM}a^*}(t|L=l) = \frac{f_{T^{aM}a^*}(t|L=l)}{S_{T^{aM}a^*}(t|L=l)}$$

ここで、 $f_{T^{aM}a^*}(t|L=l)$ は $T^{aM}a^*$ に関する条件付密度関数、 $S_{T^{aM}a^*}(t|L=l)$ は $T^{aM}a^*$ に関する条件付生存関数である。それぞれ、次式のようになる。

$$\begin{aligned} & f_{T^{aM}a^*}(t|L=l) \\ &= \int f_{T^{am}}(t|L=l, M^{a^*}=m) dP_{M^{a^*}}(M^{a^*}=m|L=l) \\ &= \int f_{T^{am}}(t|L=l) dP_{M^{a^*}}(M^{a^*}=m|L=l) \quad (\because A4) \\ &= \int f_{T^{am}}(t|A=a, M=m, L=l) dP_M(M=m|A=a^*, L=l) \quad (\because A1-A3) \\ &= \int \lambda_T(t|A=0, M=0, L=0) \exp(\gamma_1 a + \gamma_2 m + \gamma_3 am + \gamma'_4 l) \\ &\quad \times \exp(-\Lambda_T(t|A=0, M=0, L=0) e^{\gamma_1 a + \gamma_2 m + \gamma_3 am + \gamma'_4 l}) dP_M(M=m|A=a^*, L=l) \end{aligned}$$

ここで、 $\Lambda_T(t|A=0, M=0, L=0) = \int_0^t \lambda_T(t|A=0, M=0, L=0) dt$ である。同様に、

$$\begin{aligned} & S_{T^{aM}a^*}(t|L=l) \\ &= \int \exp(-\Lambda_T(t|A=0, M=0, L=0) e^{\gamma_1 a + \gamma_2 m + \gamma_3 am + \gamma'_4 l}) dP_M(M=m|A=a^*, L=l) \end{aligned}$$

したがって、 $\lambda_{T^{aM}a^*}(t|L=l)$ は次式となる。

$$\lambda_{T^{aM}a^*}(t|L=l) = \lambda_T(t|A=0, M=0, L=0) \exp(\gamma_1 a + \gamma'_4 l) r(t; a, a^*, l)$$

ここで、 $r(t; a, a^*, l)$ は、次式の通りである。

$$\begin{aligned} & r(t; a, a^*, l) \\ &= \frac{\int \exp(\gamma_2 m + \gamma_3 am) \exp(-\Lambda_T(t|A=0, M=0, L=0) e^{\gamma_1 a + \gamma_2 m + \gamma_3 am + \gamma'_4 l}) dP_M(M=m|A=a^*, L=l)}{\int \exp(-\Lambda_T(t|A=0, M=0, L=0) e^{\gamma_1 a + \gamma_2 m + \gamma_3 am + \gamma'_4 l}) dP_M(M=m|A=a^*, L=l)} \end{aligned}$$

$$\begin{aligned}
 &= \exp\left((\gamma_2 + \gamma_3 a)\left(\beta_0 + \beta_1 a^* + \beta'_2 l\right) + 0.5(\gamma_2 + \gamma_3 a)^2 \sigma^2\right) \\
 &\times \frac{\int \exp\left(-\Lambda_T(t|A=0, M=0, L=0) e^{(\gamma_2 + \gamma_3 a)^2 + \gamma_1 a + \gamma_2 m + \gamma_3 a m + \gamma'_4 l}\right) \exp\left(-\frac{(m - ((\beta_0 + \beta_1 a^* + \beta'_2 l))^2)}{2}\right) dm}{\int \exp\left(-\Lambda_T(t|A=0, M=0, L=0) e^{\gamma_1 a + \gamma_2 m + \gamma_3 a m + \gamma'_4 l}\right) \exp\left(-\frac{(m - ((\beta_0 + \beta_1 a^* + \beta'_2 l))^2)}{2}\right) dm}
 \end{aligned}$$

アウトカムがまれなとき、 $\Lambda_T(t|A=0, M=0, L=0)$ は十分小さくなるため、 $r(t; a, a^*, l)$ は、およそ $\exp((\gamma_2 + \gamma_3 a)(\beta_0 + \beta_1 a^* + \beta'_2 l) + 0.5(\gamma_2 + \gamma_3 a)^2 \sigma^2)$ に近似できる。

したがって、 $\lambda_{T a M a^*}(t|L=l)$ は、次式となる。

$$\begin{aligned}
 \lambda_{T a M a^*}(t|L=l) &\approx \lambda_T(t|A=0, M=0, L=0) \exp(\gamma_1 a + \gamma'_4 l) \exp\left((\gamma_2 + \gamma_3 a)\left(\beta_0 + \beta_1 a^* + \beta'_2 l\right) \right. \\
 &\quad \left. + 0.5(\gamma_2 + \gamma_3 a)^2 \sigma^2\right)
 \end{aligned}$$

よって、

$$\begin{aligned}
 \log\{\lambda_{T a M a^*}(t|L=l)\} &\approx \log(\lambda_T(t|A=0, M=0, L=0)) + \gamma_1 a + \gamma'_4 l + (\gamma_2 + \gamma_3 a)(\beta_0 + \beta_1 a^* + \beta'_2 l) \\
 &\quad + 0.5(\gamma_2 + \gamma_3 a)^2 \sigma^2
 \end{aligned}$$

となり、次式を得る。

$$\begin{aligned}
 \text{NIE}^{\text{PH}} &= \log\{\lambda_{T a M a^*}(t|L=l)\} - \log\{\lambda_{T a^* M a^*}(t|L=l)\} \\
 &\approx (\gamma_2 \beta_1 + \gamma_3 \beta_1 a)(a - a^*)
 \end{aligned}$$

$$\begin{aligned}
 \text{NDE}^{\text{PH}} &= \log\{\lambda_{T a M a^*}(t|L=l)\} - \log\{\lambda_{T a^* M a^*}(t|L=l)\} \\
 &\approx \{\gamma_1 + \gamma_3(\beta_0 + \beta_1 a^* + \beta'_2 l + \gamma_2 \sigma^2)\}(a - a^*) + 0.5\gamma_3^2 \sigma^2 (a^2 - a^{*2})
 \end{aligned}$$

6. 参考文献

- Avin, C., Shpister, I., and Pearl, J. (2005). Identifiability of path-specific effects. *Proceedings of the International Joint Conferences on Artificial intelligence*, 357-363.
- Baron, R. M., and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, **51**, 1173-1182.
- Ding, P., and Vanderweele, T. J. (2016). Sharp sensitivity bounds for mediation under unmeasured mediator-outcome confounding. *Biometrika*, **103**, 483-490.
- Gaynor, S. M., Schwartz, J., and Lin, X. (2019). Mediation analysis for common binary outcomes. *Statistics in medicine*, **38**, 512-529.
- Greenland, S. (2021a). Noncollapsibility, confounding, and sparse-data bias. Part 1: The oddities of odds. *Journal of clinical epidemiology*, **138**, 178-181.
- Greenland, S. (2021b). Noncollapsibility, confounding, and sparse-data bias. Part 2: What should researchers make of persistent controversies about the odds ratio?. *Journal of clinical epidemiology*, **139**, 264-268.
- Greenland, S. (2003). Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*, **14**, 300-306.
- Hafeman, D. M. and Schwartz, S. (2009). Opening the black box: A motivation for the assessment of mediation. *International Journal of Epidemiology*, **38**, 838–845.
- Hafeman, D. M. and VanderWeele, T. J. (2011). Alternative assumptions for the identification of direct and indirect effects, *Epidemiology*, **22**, 753–764.
- Hernán, M. A. and Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman and Hall/CRC Press.
- Imai K, Keele L, and Tingley D. 2010. A general approach to causal mediation analysis. *Psychol. Methods*, **15**:309–34
- Lange, T., Vansteelandt, S., and Bekaert, M. (2012). A simple unified approach for estimating natural direct and indirect effects. *American Journal of Epidemiology*, **176**, 190-195.
- Lash, T. L., VanderWeele, T. J., Haneuse, S., and Rothman, K. L. (2021). *Modern Epidemiology, 4th edition*. Wolters Kluwer.
- Marjoribanks, K. (1974). *Environments for Learning*. London: National Foundation for Educational Research Publications.
- Samoilenko, M., & Lefebvre, G. (2021). Parametric-regression-based causal mediation analysis of binary outcomes and binary mediators: moving beyond the rareness or commonness of the outcome. *American journal of epidemiology*, **190**, 1846-1858.
- Muthén, B. O., Muthén, L. K., and Asparouhov, T. (2017). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén.
- Nguyen, T. Q., Ogburn, E. L., Schmid, I., Sarker, E. B., Greifer, N., Koning, I. M., and Stuart, E. A. (2023). Causal mediation analysis: From simple to more robust strategies for estimation of marginal natural (in) direct effects. *Statistic Surveys*, **17**, 1-41.

- Pearl, J. (2001). Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 401-420.
- Robins JM (1986). A new approach to causal Inference in mortality studies with sustained exposure periods -Application to control of the healthy worker survivor effect. *Mathematical Modelling*, **7**, 1393-1512.
- Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, **3**, 143-155.
- Robins, J. M., and Richardson, T. S. (2010). Alternative graphical causal models and the identification of direct effects. *Causality and psychopathology: Finding the determinants of disorders and their cures*, **84**, 103-58.
- SAS Institute Inc. (2017). User's Guide The CAUSALMED Procedure. In *SAS/STAT® 14.3 User's Guide*. <https://support.sas.com/documentation/onlinedoc/stat/143/causalmed.pdf> (accessed on 15JAN2024).
- Valeri, L., and VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological methods*, **18**, 137.
- Valeri, L., and VanderWeele, T. J. (2015). SAS macro for causal mediation analysis with survival data. *Epidemiology*, **26**, e23-e24.
- VanderWeele, T. J. (2009). Concerning the consistency assumption in causal inference, *Epidemiology*, **20**, 880–883.
- VanderWeele TJ. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, **21**:540–51
- VanderWeele, T.J. (2011). Causal mediation analysis with survival data. *Epidemiology*, **22**:575-581.
- VanderWeele, T. J. (2013a). A three way decomposition of a total effect into direct, indirect, and interactive effects, *Epidemiology*, **24**, 224–232.
- VanderWeele, T. J. (2013b). Policy-relevant proportions for direct effects. *Epidemiology*, **24**, 175.
- VanderWeele, T. J. (2013c). “Tyler VanderWeele's Faculty Website”. MEDIATION TOOLS AND TUTORIALS: A tutorial on mediation with SAS, Stata, SPSS, and R macros. <https://www.hsph.harvard.edu/tyler-vanderweele/tools-and-tutorials/> (accessed on 01MAR2024)
- VanderWeele, T. J. (2014). A Unification of Mediation and Interaction: A 4-Way Decomposition. *Epidemiology*, **25**, 749–761.
- VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.
- VanderWeele, T. J. (2016). Mediation analysis: a practitioner's guide. *Annual review of public health*, **37**, 17-32.
- VanderWeele, T. J., Asomaning, K., Tchetgen Tchetgen, E. J., Han, Y., Spitz, M. R., Shete, S. et al. (2012). Genetic variants on 15q25. 1, smoking, and lung cancer: an assessment of mediation and interaction. *American journal of epidemiology*, **175**, 1013-1020.
- Vansteelandt, S., Bekaert, M., and Lange, T. (2012). Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiologic Methods*, **1**,131-158.

- Vansteelandt, S., and Daniel, R. M. (2017). Interventional effects for mediation analysis with multiple mediators. *Epidemiology*, **28**, 258.
- VanderWeele, T. J., and Tchetgen, E. J. T. (2014). Attributing effects to interactions. *Epidemiology*, **25**, 711-722.
- VanderWeele, T. J., and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*, **2**, 457-468.
- VanderWeele, T. J., and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American journal of epidemiology*, **172**, 1339-1348.
- VanderWeele, T. J. and Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiologic methods*, **2**, 95-115.
- Yoshida, K., Li, Y., and Mathur, M. B. (2022). regmedint: Regression-Based Causal Mediation Analysis with an Interaction Term Regression-Based Causal Mediation Analysis with an Interaction Term. R package version 1.0. 0.
- 黒木学 (2017). 『構造的因果モデルの基礎』, 共立出版.
- 田栗正隆 (2014). 直接効果・間接効果の推定及び未測定の交絡に対する感度解析. *統計数理*, **62**, 59-75
- 田中司朗 (2019). 医学のための因果推論の基礎概念. *計量生物学*, **40**, 35-62.
- 田中司朗 (2022). 『医学のための因果推論 II』, 朝倉書店.
- 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 (2022a). “ICH E9(R1)の理解に役立つ因果推論”, 医薬品評価委員会の成果物一覧. https://www.jpma.or.jp/information/evaluation/results/allotment/ds_202207_causal.html (参照 2024-02-26)
- 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 2022 年度継続タスクフォース 4 因果推論による時間依存性治療の推定サブチーム (2022b). “ICH E9(R1)の理解に役立つ因果推論～時間依存性治療編～”, 医薬品評価委員会の成果物一覧. https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202209_causal-tv.html (参照 2024-02-26)
- 矢田真城・魚住龍史 (2019). CAUSALGRAPH プロシジャによる因果効果の識別. *SAS ユーザー会論文集*, **2019**, 51-72.
- 矢田真城・魚住龍史・田栗正隆 (2020). 反事実モデルに基づく直接効果と間接効果の推定. *計量生物学*, **40**, 81-116.

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

2023年度タスクフォース3-1 STAT 新時代

【チームメンバー（本稿執筆者，50音順）】

大江 基貴	ファイザーR&D 合同会社	担当：2章, 3.1, 5.1
尾崎 凌斗	中外製薬株式会社	担当：1章, 4.2
小谷 基	小野薬品工業株式会社	担当：2章
作井 将	武田薬品工業株式会社	担当：4.1.1, 4.1.2
佐野 文哉	住友ファーマ株式会社	担当：3.2.1
重乃 広宇	千寿製薬株式会社	担当：3.2.2, 5.2
島内 順一郎	アムジェン株式会社	担当：3.4
竹ノ内 一雅	日本イーライリリー株式会社	担当：3.3
土川 克	ゼリア新薬工業株式会社	担当：4.1.3
豊泉 樹一郎	ヤンセンファーマ株式会社	担当：4.1.1, 4.1.2
藤原 正和	アストラゼネカ株式会社	担当：4.1.3
降旗 啓	第一三共株式会社	担当：3.2.3, 5.3, 5.4

【タスクフォースリーダー，推進委員】

尾崎 凌斗	中外製薬株式会社
竹ノ内 一雅	日本イーライリリー株式会社
土川 克	ゼリア新薬工業株式会社
豊泉 樹一郎	ヤンセンファーマ株式会社

【担当副部長】

菅波 秀規	興和株式会社
-------	--------